UNIVERSIDAD COMPLUTENSE DE MADRID
FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

TESIS DOCTORAL

Essays on information and prediction
Ensayos sobre información y predicción

MEMORIA PARA OPTAR AL GRADO DE DOCTOR

PRESENTADA POR

Víctor Hernández García

DIRECTORES

J. Ignacio Conde-Ruiz
Marco Celentani

Madrid, 2018

UNIVERSIDAD COMPLUTENSE DE MADRID

DOCTORAL THESIS

Essays on Information and Prediction
Ensayos sobre Información y Predicción

Author:
Víctor HERNÁNDEZ GARCÍA

Supervisor:
J. Ignacio CONDE-RUIZ,
Marco CELENTANI

A thesis submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Facultad de Ciencias Económicas y Empresariales

A mis abuelos, padres, hermana e Irene.

Agradecimientos

Me gustaría expresar mi agradecimiento a mis directores de tesis J. Ignacio Conde Ruiz y Marco Celentani por toda la ayuda prestada. Su apoyo y estímulo intelectual han sido fundamentales para el desarrollo de la tesis doctoral y han tenido un notable impacto en mi formación como economista.

También me gustaría destacar los consejos y la colaboración prestada por Antoine Loeper, Luisa Fuster y Andrés Erosa.

Asimismo, me gustaría dar las gracias a los profesores Juan Rubio Ramírez y a Agustín Casas por su papel de evaluadores externos de la tesis y por sus valiosos comentarios, correcciones y sugerencias de mejora.

Agradezco a Emory University y especialmente a su Departamento de Economía la cálida acogida que me ofrecieron durante la estancia de investigación que realicé en otoño de 2016, así como a los estudiantes del programa de doctorado de dicha universidad por las productivas conversaciones, debates y sugerencias.

También quiero destacar la ayuda recibida por parte de la Comisión Académica del doctorado, con especial agradecimiento a sus directores: Emilio Cerdá Tena y José Carlos Fariñas. Durante todos estos años, su colaboración ha sido fundamental para simplificar los trámites administrativos y en su papel de soporte a la investigación.

A todos mis compañeros de la Fundación de Estudios de Economía Aplicada (FEDEA) por los momentos compartidos durante todo el proceso de elaboración de la tesis, en especial a Carmen, Lucía, Brindusa, Alfonso y Analía.

Finalmente, me gustaría expresar mi agradecimiento al Ministerio de Economía y Competitividad, por haber financiado la tesis doctoral a través de su Programa de Formación de Personal Investigador (FPI) y la estancia de investigación en el extranjero mediante su Programa de Ayudas a la Movilidad Predoctoral.

RESUMEN EJECUTIVO

La tesis doctoral se titula “Ensayos sobre información y predicción” y se compone de un total de cuatro capítulos. El tema central es el estudio del papel que pueden desempeñar ciertos mecanismos de agregación de información, como pueden ser los mercados predictivos o las elecciones, a la hora de mejorar la precisión de los algoritmos predictivos.

El primer capítulo se titula: “Determinantes Económicos de la Alternancia Política: Un estudio con datos de panel para los países de la OCDE” y estudia la relación entre la economía y la alternancia política. La hipótesis inicial es que el proceso estocástico que determina la alternancia política no es independiente de la economía, de manera que la evolución reciente de las variables macroeconómicas sería muy relevante a la hora de explicar los resultados electorales. Detrás de esta situación subyace la “hipótesis de la responsabilidad”, mediante la cual los votantes están pendientes de la información económica ya que consideran que el gobierno es responsable, mediante sus acciones, de la buena o mala situación económica del país. De esta forma, las variables económicas pueden predecir en parte la probabilidad de alternancia política. A lo largo del capítulo se realiza una revisión crítica de la literatura , con especial atención al artículo Brender & Drazen (2008), y se propone la estimación de un modelo de probabilidad de reelección utilizando para ello indicadores macroeconómicos.

Las aportaciones de este capítulo a la literatura son diversas. Por un lado, los resultados obtenidos contradicen a los hallados en la literatura de “voto económico”. Por otro lado, también hay aportaciones metodológicas: la utilización de una tasa de descuento para permitir que los votantes den mayor peso a los datos más recientes, una nueva forma de medición de la alternancia política y la utilización de datos estructurales de déficit que solucionen (o al menos mitiguen) el problema de la multicolinealidad entre las variables independientes.

En contraste con lo observado en otros estudios, el crecimiento del PIB per cápita es la variable más relevante a la hora de explicar la alternancia económica. Concretamente, un incremento de un 1% en el crecimiento ponderado del PIB per cápita a lo largo de la legislatura incrementa la probabilidad de reelección aproximadamente un 8%, ceteris paribus. Al mismo tiempo, el incremento descontrolado del déficit público a lo largo de la legislatura está asociado a una menor probabilidad de reelección. Sin embargo, no existen evidencias de que la variación de la política fiscal en el año previo a la convocatoria electoral afecte a las probabilidades de reelección. Finalmente, varios ciclos políticos son identificados en las variables macroeconómicas, tanto en la evolución del PIB per cápita como en las variables fiscales.

El segundo capítulo se titula: “Determinantes Económicos de la Alternancia Política y Crecimiento Económico en el Largo Plazo”. Este artículo estudia la relación entre el nivel (intensidad) del voto económico y el crecimiento económico en el largo plazo. Se propone la estimación de la sensibilidad de la alternancia política frente al crecimiento económico para cada uno de los países de la OCDE.

La principal aportación metodológica de este capítulo es el empleo de métodos bayesianos (algoritmo de Monte Carlo) por sus mejores propiedades en muestras pequeñas..

Se observa una notable heterogeneidad en las sensibilidades de la alternancia política frente al crecimiento económico. Para ciertos países, el incremento del crecimiento económico a lo largo de la legislatura (percentil 80 en vez de percentil 20) incrementa la probabilidad de reelección un 50\%, mientras que para otros países el incremento es prácticamente nulo. En segundo lugar, se observa que la correlación entre estas sensibilidades y el crecimiento económico en el largo plazo es negativa y estadísticamente significativa. Por último, se repite el análisis pero estimando esta vez la sensibilidad de la alternancia política frente a la política fiscal. En este caso, también existe una correlación negativa entre dicha sensibilidad y el crecimiento económico en el largo plazo. Específicamente, aquellos países cuyos votantes valoran en mayor medida un ciclo presupuestario cóncavo presentan un menor crecimiento en el largo plazo.

El tercer capítulo se titula: "Un algoritmo basado en los mercados predictivos para modelar resultados de fútbol". A lo largo de este trabajo se propone un algoritmo que utiliza datos procedentes de los mercados predictivos (donde los agentes apuestan sobre determinados resultados) para simular una liga, en vez de resultados históricos, como es habitual en la literatura. En este trabajo, se propone un modelo de Poisson con intensidad constante en el tiempo y se utilizan varios métodos de estimación: MCO, MCO ponderado y modelos bayesianos jerarquizados.

Los resultados indican que el algoritmo de estimación propuesto tiene muy buenas capacidades predictivas, siendo capaz de vencer a las predicciones del mercado, salvo en las últimas jornadas del campeonato. Al mismo tiempo, presenta mejores resultados que los modelos tradicionales en la literatura, los cuales se basan en resultados históricos. Por lo tanto, hay evidencias de que el uso de información de los mercados predictivos permite conocer con mayor exactitud las distribuciones de probabilidad de los eventos futuros.

El cuarto capítulo se titula: "Sobre la agregación de información en los mercados predictivos centralizados". A lo largo de este trabajo se pretende estudiar el proceso de agregación de información en los mercados predictivos. En primer lugar, se analiza la eficiencia y calibración de los precios competitivos generados por los mercados predictivos. En segundo lugar, se aplica el modelo propuesto en el tercer capítulo, tras añadirle una serie de refinamientos, con el objetivo de: (1) modelar "in-play prices", (2) evaluar la capacidad de los mercados para agregar información y actualizar los precios competitivos y (3) examinar la existencia de oportunidades de inversión.

Los mercados predictivos en ocasiones tienen dificultades para agregar información, en especial cuando gran cantidad de información es revelada en poco tiempo. El modelo planteado para modelar "in-play prices" presenta un mejor ajuste, medido por la distancia De Finetti, que los precios competitivos, a pesar de disponer de una menor cantidad de información. Por último, el algoritmo de inversión identifica estrategias rentables bajo ciertas circunstancias.

EXECUTIVE SUMMARY

The dissertation is entitled: "Essays on information and prediction" and consists of a total of four chapters. The central issue is the study of the role that certain mechanisms of aggregation of information, such as predictive markets or elections, can play in improving the accuracy of predictive algorithms.

The first chapter of the thesis is entitled: "Economic Determinants of Political Alternation: A Panel Data Analysis of OECD Countries" and studies the relationship between economic performance and political alternation. The initial hypothesis is that the stochastic process that determines the political alternation is not independent of the economy, so that the recent evolution of the macroeconomic variables would be very relevant when explaining the electoral results. We are implicitly considering the "responsibility hypothesis", by which voters are aware of the economic information since they consider that the government is responsible, through its actions, for the good or bad economic situation of the country. In this way, economic variables can predict the probability of political alternation. Throughout the chapter, a critical review of the literature is presented, with special attention to the article Brender & Drazen (2008) and later the estimation of a re-election probability model is proposed, using macroeconomic indicators.

The contributions of this chapter to literature are diverse. On the one hand, the results obtained contradict those found in the "economic vote" literature. On the other hand, there are also methodological contributions: the use of a discount rate to allow voters to give more weight to the most recent data, propose an alternative way of measuring political alternation and the use of structural deficit data to solve the problem of multicollinearity between the independent variables.

In contrast to previous studies, this paper finds that higher growth rates of GDP per capita increase the probability of re-election in OECD countries. In particular, a ceteris paribus increase of 1 percentage point in the weighted average growth rate during the term in office increases the probability of re-election by 8%. At the same time, increases in government deficit over the term in office (excluding last year) decrease the probability of re-election. However, there is no evidence that fiscal policy changes at the end of the legislature affect the re- election chances of the incumbent parties. Finally, several Political Business Cycles and Political Budget Cycles are found. Re-elected governments present, in general terms, higher economic growth rates, a more balanced budget policy and the path of the surplus/deficit over the legislature is more sensible.

The second chapter is entitled: "Economic Determinants of Political Alternation and Long-term Economic Growth". This work studies the relationship between the level (intensity) of the “economic voting” and long-term economic growth. The main objectives of this chapter are: (1) study the relationship between ’economic voting’ and ’long term economic growth’ (2) analyse the long-term welfare implications of politicians’ career concerns and (3) check the presence of country-level heterogeneity in economic voting.

The contributions of this chapter to literature are varied. First, Bayesian Methods (Monte Carlo algorithm) are used because of their better behaviour in small samples. Second, the results found give empirical support to the theoretical models that maintain that the incentive structure of policymakers affects the functioning of institutions, and thus long-term welfare.

There is a substantial heterogeneity in the sensitivities of political alternation to economic growth. It is observed that for some countries an increase in the GDP per capita growth over the legislature (80th percentile instead of the 20th percentile) increases the probability of re-election by 50%, while for other countries the effect is almost zero. Secondly, a statistically significant negative correlation between these sensitivities and long run economic growth is observed. These empirical results represent support for those theoretical models that maintain that the politicians' career concerns (incentive structure) have an impact on the functioning of the institutions, affecting the long-term welfare. Under the presence of strong economic voting, politicians are only concerned with the evolution of the economy in the short term, so they try to manipulate macroeconomic variables generating distortions and neglecting long-term policies. Finally, the analysis is repeated estimating this time the sensitivity of political alternation to fiscal policy for each of the countries. In this case, there is also a negative correlation between the degree of economic voting and long-term economic growth. Specifically, those countries where voters further reward a concave political budget cycle face lower long-run economic growth.

The third chapter is entitled: "A Market-based Algorithm for Predicting Soccer Outcomes". The main objective of this article is to propose a market-based model (Poisson) for predicting soccer outcomes, using information from Betfair Exchange Market. OLS techniques and Bayesian methods are used for the estimation of the model (defensive and offensive parameters).

The results indicate that the proposed estimation algorithm has very good predictive capabilities, being able to beat market predictions, except in the last weeks of the league. At the same time, it presents better results than traditional models in the literature, which are based on past performance. Therefore, there is evidence that the use of information from predictive markets allows us to know more accurately the probability distributions of future events.

The fourth chapter is entitled: "On the aggregation of Information in Centralized Prediction Markets". The objective of this work is to apply the model developed in the previous chapter in order to: (1) check the calibration of competitive prices, (2) evaluate the efficiency of prediction markets, (3) evaluate its ability to add new information and to update the competitive prices and (4) examine the existence of investment opportunities.

Predictive markets sometimes have difficulty adding information, especially when a large amount of information is revealed in a short period of time. The model proposed to model "in-play prices" presents a better adjustment, measured by the "De Finetti" distance, than competitive prices, despite having a smaller amount of information. Finally, the investment algorithm identifies profitable strategies under certain circumstances.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Chapter 2

LIST OF FIGURES

Chapter 4

List of abbreviations

API Application Programming Interface

CA Cyclically Adjusted

CDF Cumulative Distribution Function

CPDS Comparative Political Data Set

CSV Christian Social People`s Party

CV Coefficient of Variation

DIC Deviance Information criterion

DPI Database of Political Institutions

EDF Estimated Distribution Function

EP Economic Performance

FIFA Federation Internationale de Football Association

JAGS Just Another Gibb Sampler

GDP Gross Domestic Product

GFS Government Finance Statistics

HBM Hierarchical Bayesian Model

IDEA Institute for Democracy and Electoral Assistance

IFS International Financial Statistics

IMF International Monetary Fund

JEL Journal of Economic Literature

LPM Linear Probability Model

LR Likelihood Ratio

MC Monte Carlo

MCMC Monte Carlo Markov Chain

MFX Marginal Effect at Mean

ML Maximum Likelihood

OECD Organization of Economic Co-operation and Development

OLS Ordinary Least Squares

PA Political Alternation

PBC Political Budget Cycle

PBS Primary Budget Surplus

PD Posterior Distribution

PDI Polity Democracy Index

PDF Probability Distribution Function

PM Prediction Markets

PMF Probability Mass Function

R & D Research and Development

REML Restricted Maximum Likelihood

RP Relative Measure of Performance

RW Relative Weight

SSE Squared errors of Prediction

UCL UEFA Champions League

UEFA Union des Associations Européennes de Football

UEL UEFA European League

UK United Kingdom

U.S. United States

USA United States of America

WB World Bank

WDI World Development Indicators

WEO World Economic Outlook

ZPC Zárate`s Political Collections

Economic Determinants of Political Alternation: A Panel Data Analysis of OECD Countries

Víctor Hernández García

November 2017

Resumen en Castellano

Este artículo estudia la relación entre la evolución de la economía y la alternancia política en países de la OECD (1975-2014). La revisión del artículo Brender & Drazen (2008) concluye con la identificación de varias limitaciones metodológicas y al mismo tiempo se detecta falta de robustez en algunos de sus resultados. A continuación se proponen varias modificaciones en el modelo con el fin de obtener estimaciones más precisas. En primer lugar, la alternancia política se mide a nivel de partido político en vez de a nivel de líder político. En segundo lugar, con el objetivo de evitar problemas de multicolinealidad entre las variables fiscales y el crecimiento económico, se opta por utilizar datos de déficit público estructural. Por último, se introduce un factor de descuento que permite construir una media ponderada del crecimiento económico a lo largo de la legislatura que otorgue un mayor peso a los datos más recientes. Tras la incorporación de estas novedades metodológicas se estima un modelo de probabilidad de reelección para identificar los determinantes económicos de la alternancia política. En contraste con lo observado en otros estudios (Brender & Drazen, 2008), el crecimiento del PIB per cápita es la variable más relevante a la hora de explicar la alternancia económica. Concretamente, un incremento de un 1 % en el crecimiento ponderado del PIB per cápita a lo largo de la legislatura incrementa la probabilidad de reelección aproximadamente un 8 %, ceteris paribus. Al mismo tiempo, el incremento descontrolado del déficit público a lo largo de la legislatura esta asociado a una menor probabilidad de reelección. Sin embargo, no existen evidencias de que la variación de la política fiscal en el año previo a la convocatoria electoral afecte a las probabilidades de reelección. Finalmente, varios ciclos políticos son identificados en las variables macroeconómicas.

Palabras clave: Voto económico, alternancia política, política fiscal
JEL classification: D72 E32 H61 H62

Economic Determinants of Political Alternation: A Panel Data Analysis of OECD Countries *

Víctor Hernández García†

November 2017

Abstract

This paper studies the relationship between economic performance and political alternation in OECD countries (1975-2014). Firstly, a critical review of Brender & Drazen (2008) is presented, identifying some methodological shortcomings and analysing the (lack of) robustness of some of their results. The correlations identified in Brender & Drazen (2008) are not robust against changes in the sample period or the model specification. Subsequently, several methodological changes are proposed in order to obtain better estimates, such as measuring political alternation at a political party level instead of at a political leader level, using cyclically adjusted primary balances to avoid multicollinearity problems between fiscal variables and GDP and computing a weighted-average of GDP per capita growth. After that, an alternative specification is used to estimate the economic determinants of political alternation. In contrast to previous studies, this paper finds that higher growth rates of GDP per capita increase the probability of reelection in OECD countries. In particular, a ceteris paribus increase of 1 percentage point in the weighted average growth rate during the term in office increases the probability of reelection by 8%. However, there is no evidence that fiscal policy changes at the end of the legislature affect the reelection chances of the incumbent parties. Finally, several Political Business Cycles and Political Budget Cycles are found. Re-elected governments present, in general terms, higher economic growth rates, a more balanced budget policy and the path of the surplus/deficit over the legislature is more sensible.

Keywords: Economic Voting, Political Alternation, Fiscal Policy
JEL classification: D72 E32 H61 H62

*I thank José Ignacio Conde Ruiz and Marco Celentani for their help and support. I gratefully acknowledge the support from the Spanish Ministry of Science and Technology from grants ECO2011-30323-C03-01 & BES-2012-061597.

†E-mail: victorhzgarcia@gmail.com

1.1 Introduction

Modern democracies rely on elections as a device to aggregate individual preferences. In this context, information plays a critical role in determining voting behaviour and election outcomes. The search of the main determinants of election outcomes (and political alternation) is a key topic in Political Economy literature. It is widely documented that the behaviour of politicians is largely affected by the incentive structure within which policy-makers operate. Then identifying the determinants of election results allows us to better understand the political decision-making process and the functioning of political institutions. For instance, if voters take recent evolution of the GDP growth into account when voting, then the existence of political business cycles can be expected in some situations, although this policy is dynamically inefficient.

Although there are many factors that can influence election outcomes (quality of politicians, ideology, foreign policy, scandals, wars, campaign spending...) it is widely documented that economic information plays an important role. A branch of Political Economy, often referred to as "Economic Voting", analyses the relationship between economic performance and election outcomes (See Lewis-Beck & Stegmaier, 2000 and Hibbs, 2005 for a comprehensive review of the literature).

Many empirical studies have found a high correlation between economic performance and election outcomes, in other words, the recent evolution of the economy largely explains the reelection prospects (economic voting hypothesis). However, in the literature, there is no consensus on which are the most relevant economic variables. Different variables have been proposed: real disposable personal income, real GDP per capita growth, unemployment rate, budgetary policy (surplus/deficit), monetary policy (inflation), public debt, taxes, transfers, level of inequality (Gini index), etc.

In this paper, we study the extent to which economic information influences the reelection prospects. The main objective is to re-examine the economic determinants of political alternation, checking whether the correlations identified in the literature between economic variables and political alternation are robust or unstable. To do that we will use data from OECD countries for the period 1960-2014.

The rest of the paper is organised as follows. In Section 1.2, a brief summary of the literature on economic voting is presented. In Section 1.3, a critical review of Brender & Drazen (2008) is carried out to find some methodological shortcomings, such as multicollinearity problems and poorly defined variables. In this part, some regressions are performed in order to reveal the lack of robustness of some of their results. In Section 1.4, an alternative specification is used for predicting political alternation. In Section 1.5, we study the existence of Political Business Cycles and Political Budget Cycles. Finally, conclusions are presented.

1.2 Literature review

In order to summarise the state of the art on ”economic voting”, a selection of some of the most important articles in this field is presented:

Most studies on ”economic voting” are country-level studies. Within this group, some articles focus on the role of taxes as an explanatory factor of electoral results. Besley & Case (1995) focus on USA gubernatorial elections and find that the probability of incumbent defeat is increased by an increase in state taxes. However, this effect is offset (at least in part) if neighbours increase their taxes simultaneously. Vermeir & Heyndels (2006) study municipal elections in Flanders and find incumbents are punished for higher tax rates. The electoral punishment also depends on tax rates in neighbouring municipalities.

Alternatively, several country-level studies focus on the impact of economic growth on election results. Hibbs (2000) proposes the ”Bread and Peace Model”. The per capita disposable income growth and the cumulative numbers of American military personnel killed in action largely explain the results (% vote) for the U.S. presidential elections.

However, few studies address the existence of economic voting in a cross-section of countries. Moreover, the findings obtained when analysing a panel of countries are usually quite unstable. Results sometimes differ from country-to-country or even time-to-time (Lewis-Beck & Stegmaier, 2000).

Powell and Whitten (1993) propose a multivariate model to study the electoral impact in industrialised democracies of several economic and political factors, obtaining that support for right-wing governments is increased by lower inflation and hurt by higher inflation, while inflation is not relevant in the case of left and center governments. Exactly the opposite occurs when analysing the impact of unemployment rate.

Wilkin, Haller & Norpoth (1997) found that ”election-year economic growth influences the vote of the major party in office” in a cross-national research.

Anderson (2000) studies 13 European democracies and finds that economic effects are stronger when the institutional context clarifies who is in charge of policymaking, when the target of credit and blame is large and when citizens have fewer viable alternative choices.

Brender & Drazen (2008) test whether good economic conditions and expansionary fiscal policy help incumbents get reelected in a large panel of democracies. They find that GDP per capita is not a significant variable. At the same time, voters reward budgetary discipline and penalise rising deficits. This article will be extensively analysed and reviewed in the next section.

Jones et al. (2012) provide evidence and microfoundations for the following argument ”voters reward public spending when they can pass the cost on to someone else (e.g., as in Argentina), and punish it otherwise (e.g., as in the United States)”.

Finally, Alesina, Carloni & Lecce (2012) find ”no evidence that governments which quickly reduce budget deficits are systematically voted out of office” in a sample of 19 OECD countries from 1975 to 2008.

1.3 Brender & Drazen (2008). A critical review.

1.3.1 Objectives and model specification

The aim of Brender & Drazen's work is to study the influence of economic growth and fiscal policy on the probability of re-election in a large sample of countries over the period 1960-2001[1]. The authors suggest as a starting point the responsibility hypothesis: "Voters believe that the government is responsible for the evolution of the economy" and also the retrospective voting hypothesis: "Voters reward (or punish) politicians as a function of the good (or bad) evolution of the economic situation over their term in office."

To test these hypotheses, the authors propose the estimation of a reelection probability model. As a general rule[2], the dependent variable (ReelectLeader) receives the value 1 if the incumbent leader [3] is reelected in the elections and a value of 0 if the incumbent leader is defeated[4].

The explanatory variables included in the regression model can be classified into two groups (economic and control variables):

i) **Economic variables**. The model specification includes the following economic variables: real GDP per capita growth and two fiscal variables:

-<u>GDPpcGrowth</u>: Average real GDP per capita growth over the term of office.

$$GDPpcGrowth = 100 * \sqrt[x]{\frac{GDP_0}{GDP_{-x}}} - 1 \tag{1.1}$$

where, GDP_0 is the value of real GDP per capita in the election year; GDP_{-x} is the value of real GDP per capita in the first year of the legislature; x is the number of years in office. Source: "World Development Indicators" (WB).

-<u>SurplusTerm</u>: The change in the average surplus-to-GDP ratio in the two years preceding the elections (not including the election year) compared to the previous two years.

$$SurplusTerm = 1/2 * (B_{-1} + B_{-2}) - 1/2 * (B_{-3} + B_{-4}) \tag{1.2}$$

where, B_{-i} is the surplus as a percentage of GDP i years before the elections. Sources: International Financial Statistics (IMF) and Government Finance Statistics (IMF).

-<u>SurplusLastYear</u>: The change in the surplus-to-GDP ratio in the election year relative to the previous year.

$$SurplusLastYear = (B_0 - B_{-1}) \tag{1.3}$$

where, B_{-i} is the surplus as a percentage of GDP i years before the elections. Sources: International Financial Statistics (IMF) and Government Finance Statistics (IMF).

[1] The authors perform their analysis using both an expanded and a narrow sample. Throughout this article we always focus on the "expanded" sample.

[2] There are some specific cases that are detailed in the Appendix (as the rule chosen when the incumbent leader is not eligible for re-election due to the existence of term limits).

[3] Prime minister in parliamentary systems and president in presidential systems.

[4] Sources: "Database of Political Institutions (DPI)" and Zárate's Political Collections (ZPC).

ii) **Control variables**. The authors use political variables to control for different characteristics of the countries, such as:

-NewDemocracy: A binary variable, for each country in each election year, receiving the value 1 if country is defined as a New Democracy. Otherwise, the country is defined as an Old Democracy and the variable receives a value of 0.

-MajoritarianSystem: A binary variable, for each country in each election year, receiving the value 1 in a country with a majoritarian electoral system, and 0 otherwise.

1.3.2 Main results

Table 1.1 presents the logistic regression output by Brender & Drazen (2008) for developed countries (OECD) over the period 1960-2001. The first column shows the estimated coefficients and p-values in parentheses, while the second column presents the marginal effects computed at the sample means of the data. Greater fiscal discipline (increase in the primary surplus or decrease in the primary deficit) in the election year increases the likelihood of re-election of the ruling leader. At the same time, greater fiscal discipline throughout the rest of the term in office also increases the probability of re-election of the incumbent leader. In other words, voters value budgetary discipline and penalise governments that increase the public deficit. Finally, there is no evidence that "average economic growth over the term in office" is a significant variable in developed countries (OECD), other things equal.

Other things equal, the probability of reelection is higher in New Democracies, while the binary variable that differentiates between majoritarian and proportional systems is not statistically significant.

At the end of the article, the authors make a recommendation to politicians, "Running deficits in an election year is not an effective tool to help reelection and in fact is punished at the polls in developed countries. Politicians, take note!". At the same time, according to authors, there is evidence that fiscal adjustments increase the probability of reelection.

At this point, we may ask ourselves why it is so important to think about Brender & Drazen's results. First of all, their work was published in American Economic Review, has been cited in many articles (358 citations to date according to Google Scholar) and was also cited in various publications of the OECD. In addition, their findings are quite striking, because unlike other studies the authors claim that per capita GDP growth does not influence the probability of re-election, once taken into account the effect of the fiscal and control variables. Moreover, the effect of fiscal discipline during the election year on the probability of re-election is also in some sense counterintuitive.

Once observed these results, some interesting questions arise: Are these correlations robust or unstable? Are they generated by a causal relationship? Are there other correlations between economic performance and political alternation? All these topics will be addressed in the rest of the paper.

Table 1.1: Reelection Probability Model. Brender & Drazen (2008)
OECD (1960-2001)

Dependent Variable:	ReelectLeader	
	β / p-value	Mfx
GDPpcGrowth[a]	-0.008	-0.002
	(0.937)	
SurplusTerm[b]	0.132*	0.033
	(0.096)	
SurplusLastYear[c]	0.352***	0.088
	(0.001)	
NewDemocracy	1.266**	0.316
	(0.033)	
MajoritarianSystem	0.586	0.146
	(0.142)	
Constant	-0.182	
	(0.555)	
Observations	180	
Pseudo R^2	0.071	
LR chi2	15.348	
Prob > chi2	0.009	
Baseline predicted probability	0.489	

[a]GDPpcGrowth: the average growth rate of real per capita GDP during the term.

[b]SurplusTerm: the change in the surplus-to-GDP ratio in the two years preceding the election year, relative to the two previous years

[c]SurplusLastYear: the change in the government surplus ratio to GDP in the election year, compared to the previous year.

***Significant at 1%, **Significant at 5%, *Significant at 10%

1.3.3 Criticism

a) Multicollinearity problems

The "Model of Probability of Reelection" proposed by Brender & Drazen (2008) includes several explanatory variables that have a high correlation between them. In particular, there is a high correlation between fiscal variables (SurplusTerm and SurplusLastYear) and GDPpcGrowth[5].

The existence of approximate multicollinearity can be a problem when estimating and correctly interpreting the model parameters, as it makes it difficult to estimate the individual effects of each of the variables. Given that the value of statistics for contrasts of individual significance is usually small, the probability of not rejecting the null hypothesis is increased and it is more difficult to find statistically significant variables.

As in Alesina et al. (2012), a possible solution to this problem is to work with "cyclically adjusted balances"[6]. The original series of government deficit can be divided into two components: (1) structural deficit: defined as a projection of the deficit assuming that the economy is at its normal level of activity, and (2) cyclical deficit: defined as the part of the deficit related to the economic cycle. The structural deficit data allow the correlation between fiscal variables and GDP growth to be reduced, thereby avoiding (or at least reducing) multicollinearity problems.

b) Fiscal variables are poorly defined

In Brender & Drazen's work, the fiscal variables (SurplusTerm and SurplusLastYear) are not accurately defined because it does not take into account the length of the legislature[7] nor the month in which the elections are held.

The variable SurplusLastYear has the same value if elections are held in January or in December. When elections are held at the end of the year (November, December...), the variable does reflect the true purpose for which it was designed, that is simply to compute the change in the public surplus in the election year with respect to the previous year. However, if elections take place at the beginning of the year, the definition does not fit the time frame required, because it uses surplus/deficit data after the election day (up to 11 months after elections held in January). Hence in reality, it does not measure changes in fiscal policy in the last year of the term in office, but rather in the first year of the next term. This lack of precision undermine the purpose for which the variable was included in the specification and calls into question the model estimations.

After observing the inaccuracy of the variable SurplusLastYear, an alternative definition is proposed. We faced difficulties in dealing with the subject because only annual data is available, as quarterly data of structural deficits are only available for recent years. The following assumption is proposed: "the increase or reduction in the surplus occurs

[5]Note that, in fact, fiscal variables are usually expressed as a percentage of GDP, so variations in GDP growth directly affect fiscal variables.

[6]Bornhorst et al. (2011) describe the methodology to construct "Cyclically Adjusted Balances".

[7]The definition does not take into account the actual length of each legislature.

uniformly throughout the year." It is a fairly strong assumption but it still allows for a more accurate definition to be constructed. The new definition is the following:

$$NewSurplusLastYear = \frac{(m) * (B_0 - B_{-1}) + (12 - m) * (B_{-1} - B_{-2})}{12} \tag{1.4}$$

where, B_{-i} is the government surplus as a percentage of GDP i years before the elections and m is the month in which the elections are held (for example, February=2 and April=4).

An example is presented next in order to illustrate the differences between the two definitions. We assume that the elections take place in the first part of the year to sharpen the differences between the two definitions, for example, in February 2012. The first row of Table 1.2 presents the surplus as a percentage of GDP. The second row shows the variation in the surplus-to-GDP ratio with respect to the previous year. It can be seen that the government in this case implemented an expansive fiscal policy in the months before the elections, while performing a smooth adjustment after the elections.

Table 1.2: Fiscal variables comparison.

	2009	**2010**	**2011**	**2012**
Surplus-to-GDP	-2%	-3%	-4%	-3%
Change in Surplus-to-GDP		-1%	-1%	1%
Last Year Surplus (B & D)				1%
New Last Year Surplus				-0.83%

In our example, according to the definition of Brender & Drazen (2008), the variable SurplusLastYear takes the value 1%, resulting from subtracting the value of the surplus-to-GDP ratio in 2012 (-3%) minus the value of the ratio in 2011 (-4%). According to this definition, the government is improving the situation of public finances (by reducing the deficit by one percentage point in the pre-election year). This definition does not adjust to reality, in fact, the government has increased the budget deficit in 2011 just before the election, while after the election (February 2012) the government has chosen a restrictive budgetary policy and has reduced the deficit by one percentage point.

According to the new definition proposed in this paper, the variable is calculated as follows:

$$NewSurplusLastYear = \frac{(m) * (B_0 - B_{-1}) + (12 - m) * (B_{-1} - B_{-2})}{12} = -0.83\% \tag{1.5}$$

Values: m=2, $B_0 = 1\%, B_1 = -1\%, B_2 = -1\%$

where, B_{-i} is the government surplus as a percentage of GDP i years before the elections and m is the month in which the elections are held.

The new definition is able to capture more accurately the change in fiscal policy in the last twelve months of the legislature. The differences between the two definitions are greater in elections that are held in the first months of the year and are very small when they are held at the end of the year. The new definition allows to better measure the change in public surplus at the end of the term, although its value is approximate due to the unavailability of quarterly or monthly data for the whole sample and therefore the results should be considered with caution.

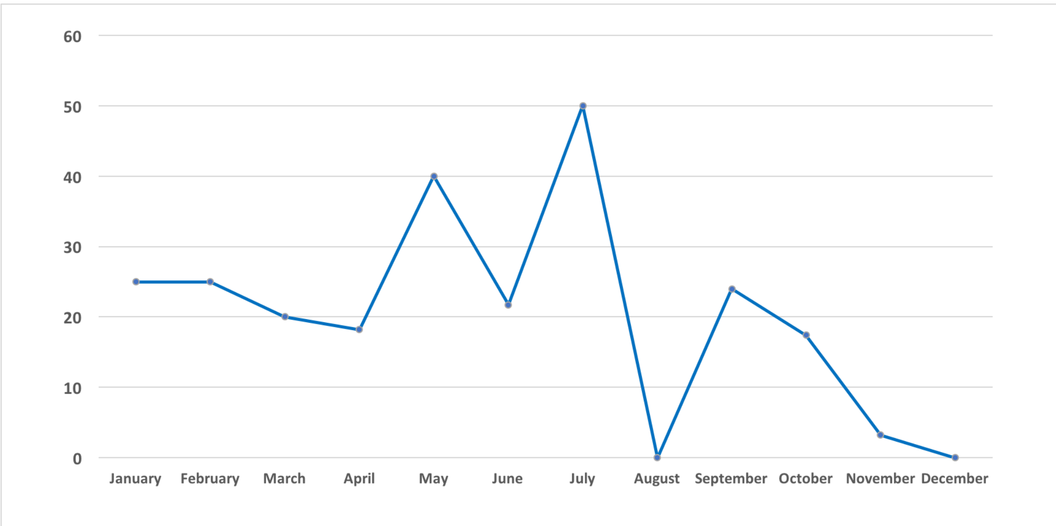
Table 1.3 shows the correlation between both variables (SurplusLastYear and New-SurplusLastYear) depending on the quarter in which the elections are held. As it can be seen, the correlation is really high when elections are held in the last quarter of the year (correlation=0.97), i.e., both definition are almost the same in this period. However, the correlation decreases when the elections are held at the beginning of the year, obtaining a correlation equal to 0.67 and 0.60 for the first and the second quarter, respectively.

Table 1.3: Correlation between SurplusLastYear and NewSurplusLastYear

First Quarter	**Second Quarter**	**Third Quarter**	**Fourth Quarter**
0.67	**0.60**	**0.86**	**0.97**

At the same time, Figure 1.1 shows the percentage of the elections where both definitions have opposite signs. It can be seen that this situation is more common when elections are held at the beginning of the year. Note that in July and August few elections are held, so there is a lot of noise.

Figure 1.1: Percentage of elections in which both definitions have opposite signs



At the same time, the variable "SurplusTerm" does not take into account the actual length of the legislature. The definition is exactly the same for every country and for every election. It does not take into account that terms of office can be different from one country to another nor when elections are called in advance. For instance, Australia has a three-year term, USA, Spain, Denmark... have a four-year term and Mexico, the UK... have a five-year term. The new definition is the following:

$$NewSurplusTerm = \sum_{i=0}^{T-12} \frac{B_i}{(T-12)} \tag{1.6}$$

where, B_i is the primary government budget balance (as a percentage of GDP) during the i month of the legislature and T is the total number of months of the legislature.

c) Definition of "reelection"

Brender & Drazen (2008) analyse political change focusing on re-election of the head of government. In presidential systems, this position is held by the president; while in parliamentary systems, the head of government is the prime minister.

However, there is an alternative approach that measures political change at a political party level. Voters tend to think that the ruling party (or ruling coalition) is responsible for the good (or bad) economic situation. Thus, the alternative variable (ReelectParty[8]) is proposed as a new dependent variable.

d) No discount rate for GDP growth.

Brender & Drazen (2008) propose that the average economic growth rate over the term in office be included as an explanatory variable. There is evidence that voters tend to give more importance to recent events (Paldam & Nannestad, 2000 and Healy & Lenz, 2014). That is, voters (or at least most of them) would have a short time horizon, i.e. they are "myopic". After performing a series of surveys and experiments, Healy & Lenz (2014) claim that voters substitute cumulative growth throughout the term for election-year performance (the end for the whole) mainly because the latter is more easily accessible.

In this context, it appears advisable to replace the arithmetic mean by the weighted mean as the latter gives more weight to data close to the elections, in the same way as Hibbs (2000). This modification allows the impact of economic growth on the probability of re-election to be measured more accurately. The formula used to calculate the "weighted economic growth" is:

$$GDPpcWeighted = \sum_{i=0}^{n} \lambda^i * \Delta GDP_{-i} * (1/\sum_{i=1}^{n} \lambda^i) \tag{1.7}$$

where, ΔGDP_0 is the real GDP per capita growth in the election month; n is the number of months of the legislature; GDP_{-n} is the real GDP per capita growth in the first month of the legislature; λ is the discount rate; $\lambda \in (0, 1)$

[8] The comprehensive definition is described in the Appendix.

1.3.4 Robustness analysis

This section is intended to check whether the results found in Brender & Drazen (2008) are robust. Their reelection probability model will be subjected to several robustness tests.

1.3.4.1 Robustness test against changes in the sample period.

Firstly, we check whether the model is robust against changes in the sample period. Brender & Drazen use a panel containing democratic elections in OECD countries from 1960 to 2001 (180 observations). Keeping constant Brender & Drazen's specification, we re-estimate the regression but extending the sample period until 2014 (75 additional elections). Table 1.4 shows the new regression output:

Table 1.4: Robustness analysis against changes in the sample period

	Original Sample (1960-2001)		Extended Sample (1960-2014)	
Dependent variable:	ReelectLeader β / p-value	Mfx	ReelectLeader β / p-value	Mfx
GDPpcGrowth[a]	-0.008	-0.002	0.097	0.024
	(0.937)		(0.233)	
SurplusTerm[b]	0.132*	0.033	0.174***	0.044
	(0.096)		(0.005)	
SurplusLastYear [c]	0.352***	0.088	0.108	0.027
	(0.001)		(0.112)	
NewDemocracy	1.266**	0.316	1.037*	0.259
	(0.033)		(0.083)	
MajoritarianSystem	0.586	0.146	0.273	0.068
	(0.142)		(0.393)	
Constant	-0.182		-0.259	
	(0.555)		(0.275)	
Observations	180		255	
Pseudo R^2	0.071		0.051	
LR chi2	15.348		16.090	
Prob > chi2	0.009		0.007	
Baseline predicted probability	0.489		0.499	

[a]GDPpcGrowth: the average growth rate of real per capita GDP during the term in office.
[b]SurplusTerm: the change in the surplus-to-GDP ratio in the two years preceding the election year, relative to the two previous years
[c]SurplusLastYear: the change in the government surplus ratio to GDP in the election year, compared to the previous year.

***Significant at 1%, **Significant at 5%, *Significant at 10%,

By expanding the sample period, the variable SurplusLastYear stops being significant. Unlike the original regression, there is no evidence that an improvement in the surplus in the last year of the term in office increases the probability of re-election. It can be concluded that Brender & Drazen's results are not robust against changes in the sample period. The relationship between the change in the public surplus in the election year and the probability of reelection is quite unstable.

1.3.4.2 Robustness test against changes in the model specification.

In this section, the robustness of the Brender & Drazen's results against changes in the model specification is analysed. Three changes are proposed: (1) new definition of political alternation, (2) new definitions of the fiscal variables (SurplusLastYear, SurplusTerm) and (3) using "cyclically adjusted fiscal variables" to avoid multicollinearity problems.

Some comments are required regarding data availability. On the one hand, the primary data (primary budget surplus) used in Brender & Drazen (2008) is not fully available to us. Specifically, we have checked the data sources [9] provided in the article but we have only found primary budget surplus data for 170 of the 180 elections analysed. Primary budget surplus data are not available for several countries in 1960's, although it was available when Brender and Drazen downloaded the data. Moreover, it is not possible to recover the data from the additional materials[10] provided by the authors because only regression variables are available, not the primary data used to compute them. On the other hand, cyclically adjusted balance data are only available since 1970, so it is only possible to include elections from 1975 onwards. The unavailability of data for structural deficit in the 1960s means that this robustness analysis can only be performed on a reduced sample.

Table 1.5 describes the three regressions we will use as a reference for the robustness analysis. Firstly, we have the Original Regression presented in Brender & Drazen (2008). Secondly, we have a reduced version that excludes ten elections from the original sample due to the unavailability of budget surplus data. From now on, we call it "Reduced Sample 1". Finally, we have another reduced version of the original sample, where we exclude elections prior to 1975, so in this case we only compute the regression using 112 elections. From now on, we call it "Reduced Sample 2".

Table 1.5: Original and reduced samples

Name	Period	Obs.
Original Regression in B & D	1960-2001	180
Reduced Sample 1	1960-2001[a]	170
Reduced Sample 2	1975-2001[b]	112

[a]Ten elections corresponding to 1960's decade are excluded.
[b]Structural fiscal data is only available from 1975 onwards.

[9]International Financial Statistics (IFS) & Government Finance Statistics (GFS)
[10]https://www.aeaweb.org/articles?id=10.1257/aer.98.5.2203

As a first step, the original Brender & Drazen model (1960-2001) is re-estimated using the information provided by the authors on their website (database and do file). Additionally, the reelection probability model is also estimated using Reduced Sample 1 and Reduced Sample 2. The regression outputs are presented in Table 1.6. It can be observed that results are almost the same when we use a reduced sample. All the equations identify exactly the same correlations between Political Alternation and Economic Performance. We only observe that the level of significance is lower in the second and third column probably due to the small number of observations. After this step, the robustness tests can be carried out:

Table 1.6: Reelection Probability Model (Brender & Drazen)

	Original Sample		Reduced Sample 1		Reduced Sample 2	
Dependent variable:	ReelectLeader		ReelectLeader		ReelectLeader	
	β / p-value	Mfx	β / p-value	Mfx	β / p-value	Mfx
GDPpcGrowth[a]	-0.008	-0.002	0.050	0.012	0.146	0.0364
	(0.937)		(0.599)		(0.297)	
SurplusTerm[b]	0.132*	0.033	0.139*	0.034	0.142*	0.035
	(0.096)		(0.056)		(0.099)	
SurplusLastYear[c]	0.351***	0.088	0.183**	0.045	0.240*	0.059
	(0.001)		(0.047)		(0.074)	
NewDemocracy	1.266**	0.289	1.037*	0.193	1.176	0.260
	(0.033)		(0.083)		(0.239)	
MajoritarianSystem	0.586	0.145	0.273	0.101	0.045	0.011
	(0.142)		(0.393)		(0.932)	
Constant	-0.182		-0.332		-0.279	
	(0.555)		(0.278)		(0.464)	
Observations	180		170		112	
Pseudo R^2	0.064		0.045		0.073	
LR chi2	13.395		10.79		11.31	
Prob > chi2	0.020		0.055		0.045	

[a]GDPpcGrowth: the average growth rate of real per capita GDP during the term in office.

[b]SurplusTerm: the change in the surplus-to-GDP ratio in the two years preceding the election year, relative to the two previous years

[c]SurplusLastYear: the change in the government surplus ratio to GDP in the election year, compared to the previous year.

***Significant at 1%, **Significant at 5%, *Significant at 10%,

Table 1.7 describes briefly the regressions proposed for the robustness analysis against changes in the model specification.

Table 1.7: Robustness Analysis (Battery of regressions)

	Reelection	Fiscal variables	New Def.[a]	λ^b	Obs.
Rob. Analysis (PART I[c])					
Regression 1	Party	Non-Adjusted	(No, No)	No	180
Rob. Analysis (PART II[d])					
Regression 2	Leader	Non-Adjusted	(No, Yes)	No	170
Regression 3	Leader	Non-Adjusted	(Yes, Yes)	No	170
Rob. Analysis (PART III[e])					
Regression 4	Leader	Cycl. Adjusted	(No, No)	No	112
Regression 5	Leader	Cycl. Adjusted	(Yes, Yes)	No	112

[a] New definition of the fiscal variables (SurplusTerm, SurplusLastYear).
[b] Discount rate for the per capita GDP growth.
[c] New Definition of Political Alternation.
[d] New Definitions of the Fiscal Variables.
[e] Using Cyclically Adjusted Fiscal Balances.

1.3.4.2.1 Part I: Changing the definition of political alternation.

Firstly, Brender & Drazen's regression is replicated but changing the definition of political alternation. Now we use the following rule "there is reelection when the incumbent party (or incumbent coalition) has won again in the elections", instead of focusing on the incumbent leader (prime minister or president). The justification for this modification is to check whether the correlations are maintained when we asume that voters assign responsibility to political parties instead of declaring responsible to the incumbent leader. The new dependent variable is called: ReelectParty[11].

According to ReelectLeader, there were 88 reelections in the Brender & Drazen's sample (48% of the elections). Using ReelectParty there were 107 reelections (59%).

Table 1.8 shows those elections in which both definitions of political alternation differ (a total of 25 elections). If we analyse in detail these elections, it can be seen that in most of the cases (18) occur the following. The prime minister[12] decides not to run again, is replaced by another member of his party and the substitute wins the elections. In this situation, the definition of political alternation proposed by Brender & Drazen (ReelectLeader) considers that there is no reelection. However, according to our definition (ReelectParty) there was reelection because we are focusing on the political party.

For instance, if we look at the election held in Luxembourg in 1984, the leader of the Christian Social People's Party (CSV) Pierre Werner decided not to run again after

[11] The complete definition of the variable is described in the Appendix.
[12] Not subject to term limits.

almost 20 years being prime minister because of his age. He was replaced as a candidate of the CSV by Jacques Santer and the CSV won the election again bordering the absolute majority. In this case, ReelectLeader takes the value 0 (the incumbent retired from the race) and ReelectParty takes the value 1.

Additionally, the other seven times in which there are discrepancies between both definitions relate to complex governments composed of a coalition of parties.

With this alternative specification (Table 1.9), the variable GDPpcGrowth is significant at 5%. At the same time, the variables SurplusTerm and NewDemocracy stop being significant. In addition, the pseudo R^2 increases from 0.064 to 0.08. We can conclude that Brender & Drazen's specification is not robust against changing the definition of the dependent variable.

Table 1.8: List of elections in which both definitions differ

Country	**Elections**	**ReelectLeader**	**ReelectParty**
Australia	1967	0	1
Austria	1983	0	1
Belgium	1968, 1977, 1987	0, 1, 1	1, 0, 0
Finland	1970, 1979	0, 0	1, 1
Germany	1976	0	1
Greece	1996	0	1
Ireland	1992	0	1
Iceland	1983, 1987	0, 0	1, 1
Italy	1972, 1979, 1987, 1992	0, 0, 0, 0	1, 1, 1, 1
Japan	1972, 1976, 1979, 1989	0, 0, 0, 0	1, 1, 1, 1
Luxembourg	1984, 1994	0, 0	1, 1
Netherland	1971	0	1
Turkey	1995	1	0

Table 1.9: Robustness analysis against changes in the definition of political alternation

	Original Sample (1960-2001)		Regression 1	
Dependent variable:	ReelectLeader β / p-value	Mfx	ReelectParty β / p-value	Mfx
GDPpcGrowth[a]	-0.008	-0.002	0.027**	0.651
	(0.937)		(0.018)	
SurplusTerm[b]	0.132*	0.033	0.059	0.0142
	(0.096)		(0.448)	
SurplusLastYear [c]	0.351***	0.088	0.228**	0.065
	(0.001)		(0.041)	
NewDemocracy	1.266**	0.289	0.027	0.006
	(0.033)		(0.965)	
MajoritarianSystem	0.586	0.145	0.029	0.007
	(0.142)		(0.941)	
Constant	-0.182		-0.224	
	(0.555)		(0.503)	
Observations	180		180	
Pseudo R^2	0.064		0.080	
LR chi2	13.395		13.48	
Prob > chi2	0.020		0.0193	

[a]GDPpcGrowth: the average growth rate of real per capita GDP during the term in office.

[b]SurplusTerm: the change in the surplus-to-GDP ratio in the two years preceding the election year, relative to the two previous years

[c]SurplusLastYear: the change in the government surplus ratio to GDP in the election year, compared to the previous year.

***Significant at 1%, **Significant at 5%, *Significant at 10%,

1.3.4.2.2 Part II: Changing the definition of the fiscal variables.

In this part we check the robustness of the Brender and Drazen's model against changes in the definition of the fiscal variables (Regression 2 and 3). Results are shown in Table 1.10.

In Regression 2, we re-estimate the model but modifying the definition of the variable that measure the change in the surplus-to-GDP ratio in the last year, to take into account the month in which the elections are held. The alternative variable is called NewSurplusLastYear[13] and tries to be more precise than the definition used in Brender & Drazen (2008). Applying this change we obtain that the variable NewSurplusLastYear is not significant now. As in the original regression, voters reward governments that reduce the deficit at the beginning and in the middle of the legislature.

In Regression 3, we also modify the definition of the variable that measure the change in the surplus-to-GDP ratio in the term in office (excluding last 12 months). The new variable is called NewSurplusTerm[14]. As in the previous case, the variable NewSurplusLastYear is not significant. In this case the sign of the coefficient is even negative. Moreover, after the change in the definition, the variable NewSurplusTerm is now significant at 1% (instead of 10%). Finally, after the modifications, the pseudo R^2 is greater (0.075 instead of 0.045).

Both regressions show that Brender & Drazen's findings are not robust against changes in the definition of the fiscal variables. Once we apply more precise definitions for the fiscal variables it is obtained that there is no evidence that increasing public spending (or lower taxes) in the last year of the term in office helps or harms the chances of re-election. At the same time, in both regressions, GDPpcGrowth and NewDemocracy are not statistically significant.

[13] Exact definition is given in the Appendix.
[14] The formal definition is given in the Appendix.

Table 1.10: Robustness Analysis. Different specifications I (LOGIT).

	B & D (Red. Sample 1)		Regression 2		Regression 3	
Dependent variable:	ReelectLeader β / p-value	Mfx	ReelectLeader β / p-value	Mfx	ReelectLeader β / p-value	Mfx
GDPpcGrowth[a]	0.050	0.012	0.076	0.019	0.030	0.761
	(0.599)		(0.436)		(0.755)	
SurplusTerm[b]	0.139*	0.034	0.121*	0.302		
	(0.056)		(0.099)			
NewSurplusTerm[c]					0.80	0.020
					(0.004)***	
SurplusLastYear[d]	0.183**	0.088				
	(0.047)					
NewSurplusLastYear[e]			0.177	0.044	-0.058	-0.014
			(0.168)		(0.716)	
NewDemocracy	1.037*	0.193	0.735	0.178	0.561	0.138
	(0.083)		(0.234)		(0.367)	
MajoritarianSystem	0.273	0.101	0.464	0.115	0.445	0.138
	(0.393)		(0.244)		(0.260)	
Constant	-0.332		-0.426		-0.317	
	(0.278)		(0.169)		(0.325)	
Observations	170		170		170	
Pseudo R^2	0.045		0.039		0.075	
LR chi2	10.79		8.96		14.12	
Prob > chi2	0.055		0.111		0.015	

[a] GDPpcGrowth: the average growth rate of real per capita GDP during the term in office.

[b] SurplusTerm: the change in the surplus-to-GDP ratio in the two years preceding the election year, relative to the two previous years

[c] NewSurplusTerm: the change in the surplus-to-GDP over the term in office (excluding the last twelve months). This definition takes into account the real length of the legislature.

[d] SurplusLastYear: the change in the government surplus ratio to GDP in the election year, compared to the previous year.

[e] NewSurplusLastYear: the change in the government surplus ratio to GDP in in the last twelve months of the term in office, compared to the previous twelve months.

***Significant at 1%, **Significant at 5%, *Significant at 10%,

Now, we address the same question but using a different approach. Graphing Political Business Cycles we will show that the correlation between SurplusLastYear and Political Alternation is quite sensitive to the definition used.

Figure 1.2 shows the evolution of the primary budget surplus throughout the legislature and has been computed using the same elections as in Brender and Drazen (2008). We are applying Brender and Drazen's definitions in order to determine when a legislature starts and ends[15]. In this context, normalisation is required since there is great heterogeneity in the duration of the legislatures among countries. To do that, we divide each legislature in ten periods and later we compute the mean value of the variable "primary budget surplus" for each period of each election. Now we can compute the cycle for the variable "primary budget surplus". As we can see, there is a clearly differentiated behaviour according to whether or not there is alternation once the legislature ends. The primary budget surplus is monotonously increasing when the incumbent party (or coalition) is reelected (red line). Note that the trend line (black line) is almost a straight line. At the same time, the primary budget surplus always decreases when the incumbent is not reelected (green line). Thus, looking at Figure 1.2 we obtain the same conclusion as with Brender and Drazen's model: "Rising deficits during the incumbent's term are associated with a lesser probability of reelection".

Figure 1.2: Primary Budget Surplus (PBS). Evolution throughout the legislature using Brender & Drazen's definitions.

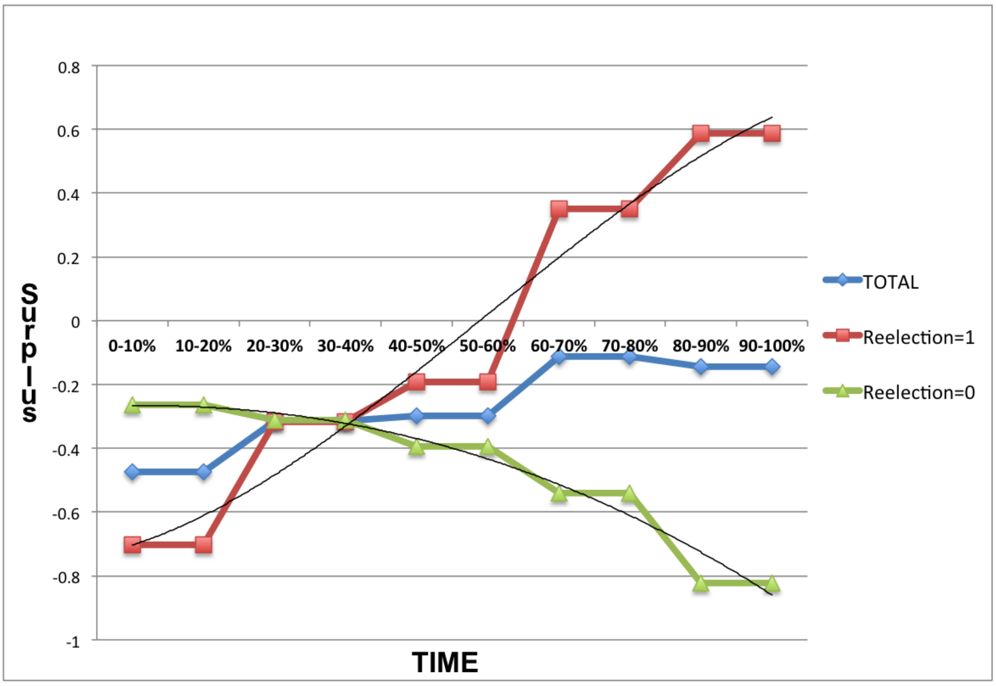
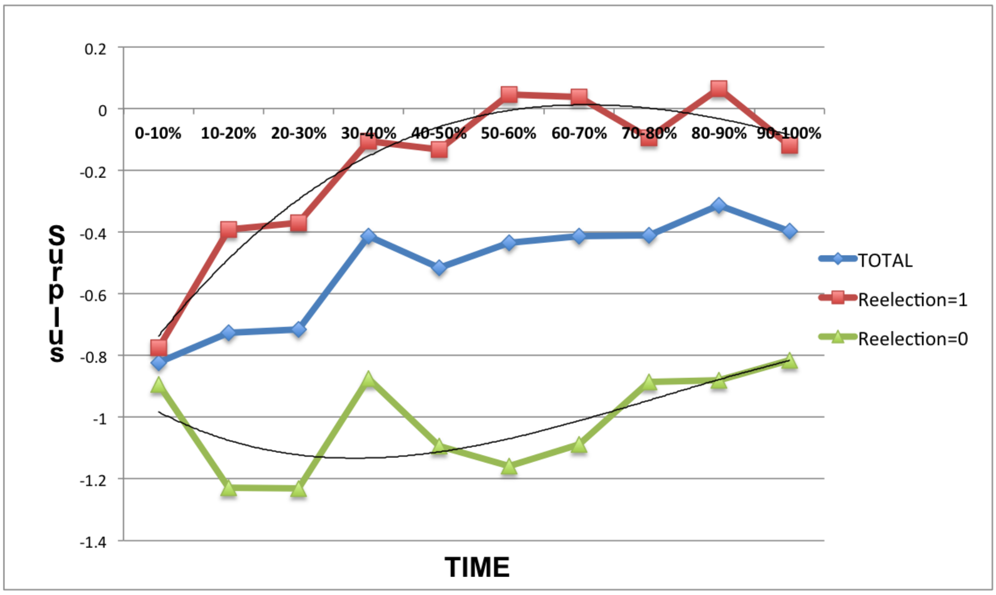


Figure 1.3 presents the same graph as before but now the timing is slightly different because we are taking into account the month when elections are held. In the first half of the legislature, "good governments" improve the primary budget surplus while "bad

[15] In other words, we are considering that every legislature finishes in December.

governments" increase the deficit (or reduce the surplus). However, in contrast with the former case, in the second half of the legislature both curves are not so different. These results are consistent with the robustness analysis performed previously. When we measure precisely the end of the legislatures, it is not a clearly differentiated path according to whether or not there is alternation once the legislature ends. Note that the trend line (black line) is clearly concave for "good governments" and convex for "bad governments".

Figure 1.3: Primary Budget Surplus (PBS). Evolution throughout the legislature considering the month when elections are held.



1.3.4.2.3 Part III: Using cyclically adjusted fiscal data.

In Regression 4 we use cyclically adjusted fiscal variables (instead of non-cyclically adjusted). The new variables are called: SurplusLastYearCA and SurplusTermCA. Everything else remains identical to the original regression. The results change significantly (Table 1.11). Economic growth becomes a significant variable (10%) and the sign of the coefficient is positive, implying that higher economic growth over the term in office increases the probability of reelection, other things equal. Furthermore, the fiscal variables are not significant at usual levels. It can be concluded that eliminating multicollinearity problems (using cyclically adjusted fiscal variables) there is empirical evidence that economic growth is a significant variable.

In Regression 5 we continue using cyclically adjusted fiscal variables and we also apply the new definitions for the fiscal variables. With this new specification GDPpcGrowth keeps being significant at 10%. The same occurs with the variable NewSurplusTermCA, which measure the change in the cyclically adjusted surplus over the term in office (excluding last 12 months). There is no evidence that fiscal discipline during last year of the legislature helps incumbent being reelect. In fact, the coefficient of the variable NewSurplusLastyearCA is negative. In addition, the pseudo $R2$ increases from 0.049 to 0.063.

Table 1.11: Robustness Analysis. Different specifications II (LOGIT).

	B & D (Red. Sample 2)		Regression 4		Regression 5	
Dependent variable:	ReelectLeader		ReelectLeader		ReelectLeader	
	β / p-value	Mfx	β / p-value	Mfx	β / p-value	Mfx
GDPpcGrowth[a]	0.146	0.0364	0.279*	0.069	0.276 *	0.069
	(0.297)		(0.061)		(0.075)	
SurplusTerm[b]	0.142*	0.035				
	(0.099)					
SurplusTermCA[c]			0.101*	0.025		
			(0.055)			
NewSurplusTermCA[d]					0.519*	0.129
					(0.089)	
SurplusLastYear[e]	0.240*	0.059				
	(0.074)					
SurplusLastYearCA[f]			0.080	0.020		
			(0.284)			
NewSurplusLastYearCA[g]					-0.071	-0.017
					(0.743)	
NewDemocracy	1.176*	0.260	0.584	0.140	0.539	0.130
	(0.239)		(0.503)		(0.540)	
MajoritarianSystem	0.045	0.011	0.073	0.018	0.100	0.025
	(0.932)		(0.890)		(0.844)	
Constant	-0.279		-0.581		-0.608	
	(0.464)		(0.116)		(0.108)	
Observations	112		112		112	
Pseudo R^2	0.073		0.049		0.063	
LR chi2	11.31		6.58		8.69	
Prob > chi2	0.045		0.254		0.122	

[a]GDPpcGrowth: the average growth rate of real per capita GDP during the term in office.

[b]SurplusTerm: the change in the surplus-to-GDP ratio in the two years preceding the election year, relative to the two previous years

[c]SurplusTermCA: the change in the surplus-to-GDP over the term in office (excluding the last twelve months). This definition takes into account the real length of the legislature.

[d]NewSurplusTermCA: the change in the surplus-to-GDP over the term in office (excluding the last twelve months). This definition takes into account the real length of the legislature.

[e]SurplusLastYear: the change in the government surplus ratio to GDP in the election year, compared to the previous year.

[f]SurplusLastYearCA: the change in the government surplus ratio to GDP in in the last twelve months of the term in office, compared to the previous twelve months.

[g]NewSurplusLastYearCA: the change in the government surplus ratio to GDP in in the last twelve months of the term in office, compared to the previous twelve months.

***Significant at 1%, **Significant at 5%, *Significant at 10%,

1.3.4.2.4 Summary: Robustness analysis against changes in the specification.

Table 1.12 summarises the results obtained from the robustness analysis, indicating which variables are significant in each regression. Correlations found in Brender & Drazen's work (2008) do not seem very robust. Recall that the authors claimed to have found evidence that a higher surplus (or lower deficit) in the election year increased the chances of re-election in the OECD countries. By contrast, economic growth was not a statistically significant variable in their model of re-election probability.

However, when introducing some reasonable changes in the model's specification, very different results are obtained and new correlations are uncovered. Real GDP per capita growth is statistically significant in most of the regressions and positively affect the chances of re-election. Maintaining stable government finances over the term of office also increases the chances of re-election as in Brender & Drazen (2008). At the same time, there is no evidence that increasing the surplus (or reducing deficit) during the election year is positively valued by voters. In summary, voters reward governments that achieve high levels of economic growth and reduce the structural deficit at the beginning and in the middle of the term of office, but not at the end.

Table 1.12: Summary of the robustness analysis: Which variables are significant?

	GDPpcGrowth	**SurplusTerm**	**SurplusLastYear**	**Obs.**
R. Analysis (PART I[a])				
B & D (Original Model)	No	Yes*	Yes***	180
Regression 1	Yes**	No	Yes**	180
R. Analysis (PART II[b])				
B & D (Reduced Sample 1)	No	Yes*	Yes**	170
Regression 2	No	Yes*	No	170
Regression 3	No	Yes***	No	170
R. Analysis (PART III[c])				
B & D (Reduced Sample 2)	No	Yes*	Yes*	112
Regression 4	Yes*	Yes*	No	112
Regression 5	Yes*	Yes*	No	112

[a] New Definition of Political Alternation
[b] New Definitions of the Fiscal Variables
[c] Using Cyclically Adjusted Fiscal Balances

It is possible that the lack of significance of the economic growth variable in Brender & Drazen's regression was caused by the inclusion of non-cyclically adjusted fiscal variables, highly correlated with GDP. This way, the individual effect of economic growth on the likelihood of re-election is not properly estimated in Brender & Drazen (2008). At the same time, there is evidence that the correlation between SurplusLastYear and Political Alternation is very sensitive to the definition of SurplusLastYear. It is necessary to take into account the month when elections are held in order to measure accurately the correlation.

1.4 Alternative specification to identify the economic determinants of political alternation.

Once the lack of robustness of the results found by Brender & Drazen is shown, the purpose of this section is to propose an alternative specification in order to identify new correlations between democracy (election results) and economic performance. Another objective is to test the hypothesis that the growth of real GDP per capita is a statistically significant variable when estimating the probability of re-election.

Next, a brief overview of variables and data sources used in the new specification: [16]

- **ReelectParty:** Dependent Variable. It was constructed from: "Database of Political Institutions", (World Bank).
- **GDPpcWeighted:** Weighted economic growth over the term in office. It is used the same formula as in Hibbs (2000). Discount rate: $\lambda = 0.95$. Real GDP per capita data are from the "World Developments Indicators" (World Bank).
- **SurplusTermCA & NewSuplusLastYearCA:** Cyclically Adjusted Fiscal Variables. They was collected from "Economic Outlook Database" (OCDE).
- **Duration:** Variable that measures the number of years that the incumbent party has been governing the country. Alesina, Carloni & Lecce (2012) suggest that the longer the government has been in office, the higher its probability of defeat. Source: "Database of Political Institutions" (World Bank).
- **Control variables.** The same binary control variables as in Brender & Drazen (2008): NewDemocracy and MajoritarianSystem.

The sample period is from 1975 to 2014 and only democratic elections are included. It has not been possible to include before 1975 elections because structural deficit data are only available since 1970, and in order to include an election it is necessary to have data from the beginning of the legislature.

Table 1.13 shows the regression output, using the new specification. Unlike Brender & Drazen (2008), an increase in the GDP per capita growth rate increases the probability of reelection in OECD countries. A ceteris paribus increase of 1 percentage point in the average growth rate during the term increases the probability of reelection by 8%. At the same time, increases in public deficit over the term in office decrease the probability of re-election. Voters penalise persistent budgetary imbalances. However, there is no evidence that fiscal policy changes (increase or decrease in the government surplus) in the last year of the term in office affect the chances of re-election of incumbent parties.

[16] A comprehensive description of the variables and data sources is presented in the Appendix.

Table 1.13: Alternative specification (LOGIT)
OECD (1970-2014)

	1970-2007		1970-2014	
Dependent variable:	ReelectParty		ReelectParty	
	β / p-value	Mfx	β / p-value	Mfx
GDPpcWeighted[a]	0.344**	0.076	0.350***	0.079
	(0.014)		(0.001)	
NewSurplusLastYearCA[b]	-0.299	-0.066	-0.210	-0.048
	(0.142)		(0.173)	
SurplusTermCA[c]	0.504*	0.111	0.350*	0.079
	(0.077)		(0.087)	
NewDemocracy	-0.305	-0.067	-0.304	-0.069
	(0.743)		(0.740)	
MajoritarianSystem	0.841*	0.185	0.803**	0.183
	(0.056)		(0.047)	
Duration [d]	-0.320*	-0.070	-0.423**	-0.096
	(0.079)		(0.015)	
Constant	0.880		1.209**	
	(0.128)		(0.016)	
Observations	165		194	
Pseudo R^2	0.090		0.108	
LR chi2	19.104		27.454	
Prob > chi2	0.004		0.000	
Baseline predicted probability	0.655		0.634	

[a]GDPpcWeighted: Weighted economic growth over the term in office. Discount rate ($\lambda = 0.95$).

[b]NewSurplusLastYearCA: the change in the surplus-to-GDP ratio in the last twelve months of the term in office (cyclically adjusted data.)

[c]SurplusTermCA: the change in the surplus-to-GDP ratio in the two years preceding the election year, relative to the two previous years (cyclically adjusted data)

[d]Duration: Variable that measures the number of years that the incumbent party has been governing the country.

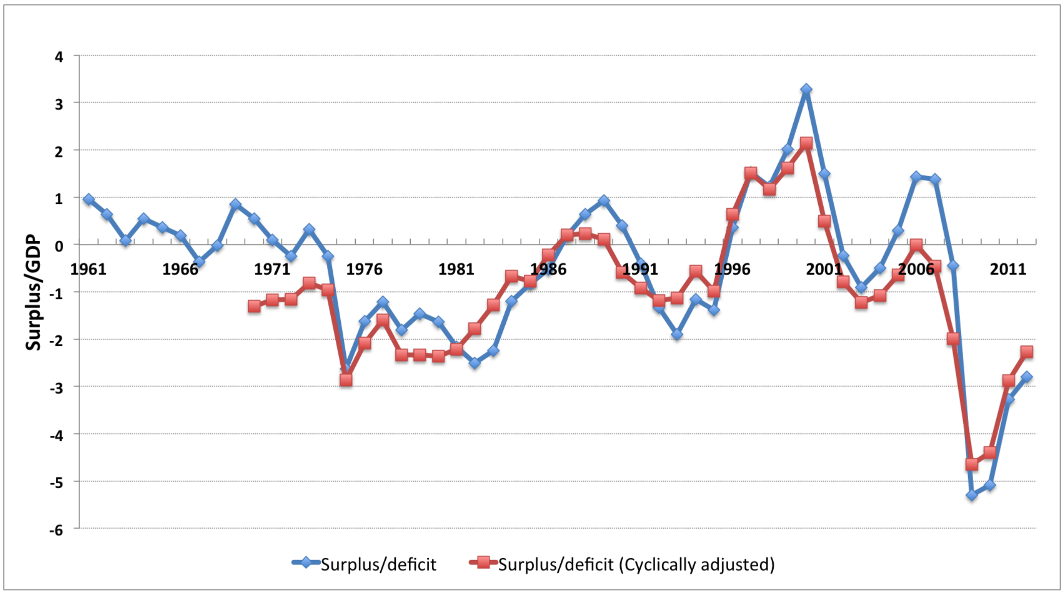
***Significant at 1%, **Significant at 5%, *Significant at 10%

The probability of reelection is higher in Proportional Systems, while the binary variable that differentiates between new and old democracies is not statistically significant. Finally, the more years the incumbent party has been in office (variable "Duration"), the higher the probability of its defeat, other things equal.

Now we check the robustness of the new specification against extreme events. The global economic crisis that began in 2008 caused significant declines in GDP and strong increases in government deficits in virtually all OECD countries.

Figure 1.4 shows the evolution of government surplus (blue: non-cyclically adjusted, red: cyclically adjusted) for the OECD weighted average[17]. If we identify the five worst deficit data in the time series (1975, 2009, 2010, 2011 and 2012), four of them took place in the period (2008-2014), which highlights the magnitude of the economic crisis.

Figure 1.4: Sovereign Debt Crisis in OECD (2008-2012)



Source: "Economic Outlook Database" (OECD)

We re-estimate the model only for the period 1975-2007 (excluding the 29 elections held during the economic crisis). Results are shown in the first column of the Table 1.13. The independent variables had extreme values in that period, so it is very interesting to test the robustness of the model to extreme events.

We can conclude that excluding the crisis period the estimation results do not change. Only a small decrease in the level of significance of estimated coefficients is observed, which may be explained by the reduction in the number of elections (N=165).

[17]Weights are calculated using GDP.

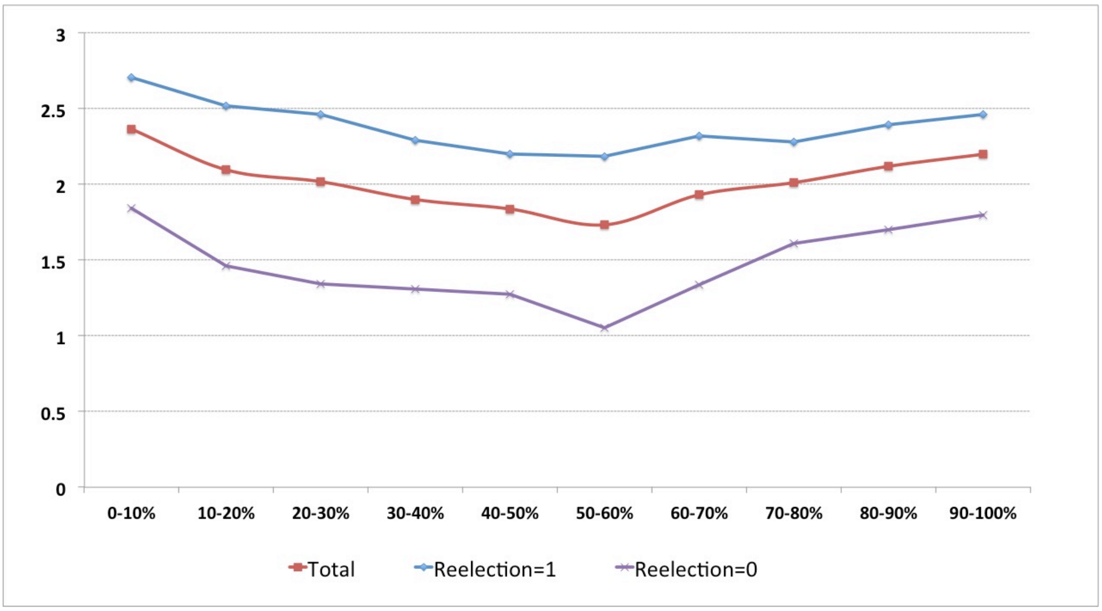
1.5 Identifying Political Cycles

In this section, we can see the evolution of different macroeconomic indicators throughout the legislature for the different countries of the OECD. The aim of this section is to identify the existence of Political Business Cycles and Political Budget Cycles. These graphs, which collect observations (elections) from the period 1975-2014, are divided into 10 intervals - so if a legislature lasted 40 months the first interval would be constructed from the per capita GDP growth rate for the first four months, the second from months 5 to 8, etc ...

Likewise, the graphs show three series: "Total", which collects all the observations of the sample; "Reelection = 1", which includes only those in which at the end of the legislature there is no change of government; and "Reelection = 0", which includes those in which there is a change of government at the end of the legislature. In this section, we measure political alternation focusing on the party.

Figure 1.5 shows the evolution of per capita GDP growth. From these series we can derive several observations. First, economic growth in all scenarios is reduced at the beginning of the legislature and begins to rise in the 60-70% interval. Second, it can be observed that the growth rate is higher for those cases in which there is no change of government at the end of the legislature, whereas in cases where voters opt for another party at the end of the legislature, growth rates are considerably more discrete - the difference between the two scenarios moves at around 1% growth. Finally, we can see that the volatility in the series is somewhat greater for those cases in which there is change of government at the end of the legislature. These results are consistent with the existence of economic voting.

Figure 1.5: Political Business Cycle (GDP Growth). OECD (1975-2014)



This seems to point to several hypotheses. Firstly, voters clearly seem to reward good economic performance of governments, so that governments which have been in power over times of greater expansion are re-elected in greater proportion than those who have governed during weak growth stages. Secondly, in all scenarios we can observe how there is an acceleration in growth rates at the end of the legislature. This may be because governments carry out public expenditure growth or tax reduction to please voters as the end of legislature approaches and thereby contribute to their re-election. The fiscal stimulus would lead to higher growth rates at the end of the term of office. Another hypothesis is that certain governments behave strategically and call elections when growth rates are particularly robust. This would explain the greater growth present in all scenarios.

Figure 1.6 we show the evolution of non cyclically adjusted public surplus. As we can see, there is a clearly differentiated behaviour according to whether or not there is alternation in power once the legislature ends.

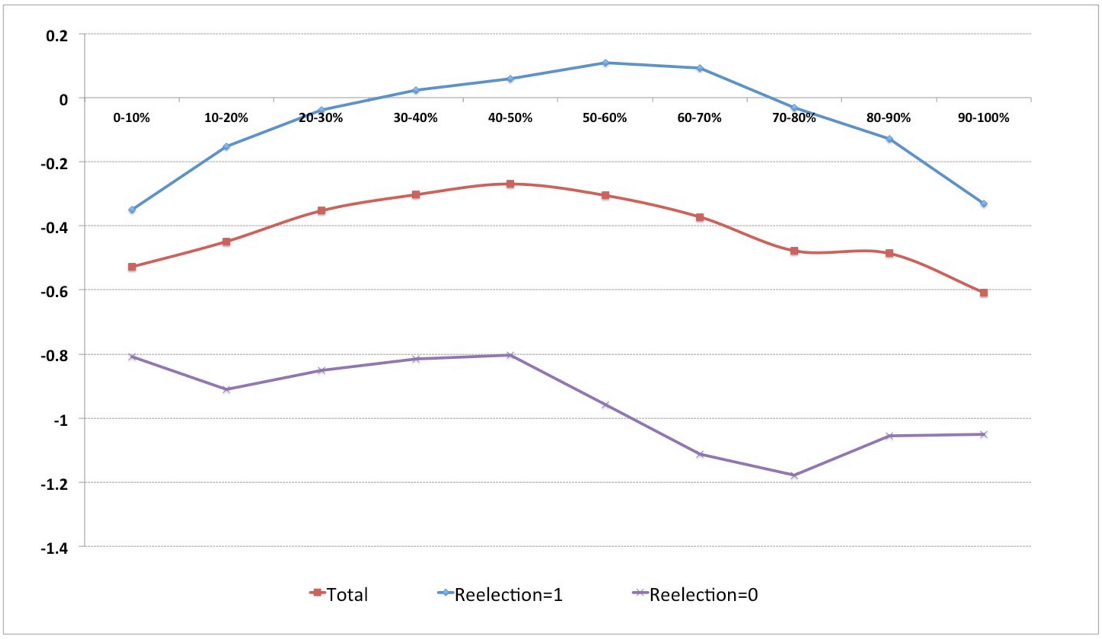
The main observations derived from the following graph could be the following. Firstly, from the basis of all observations, at the beginning of the legislature a reduction in deficit takes place. However, from the 30-40% interval the deficit begins to expand and, except for a very slight reduction in the period 80-90%, progresses throughout the legislature, especially in the last interval. Second, those countries in which a change of government takes place at the end of the legislature present a greater deficit. This gap widens throughout the legislature and only slightly shortens in the last two intervals.

Figure 1.6: Political Budget Cycle (non-cyclically adjusted surplus). OECD (1975-2014)

The graph analysed reinforces some of the hypotheses that we have indicated previously. In the first place, it seems that voters reward the good budgetary behaviour of governments, so that those who have maintained a surplus situation are more likely to be re-elected. Secondly, except for those cases in which governments are not re-elected, in which case the deficit is expanded during most of the legislature, we can observe how governments reduce the deficit at the beginning of the legislature and then increase it at the end of it. This seems to connect with the previous hypothesis that governments carry out fiscal expansions at the end of the legislature to stimulate economic growth and thereby contribute to their re-election. This expansion of the end of the legislature is especially intense in the last interval.

In Figure 1.7 we show the evolution of the cyclically adjusted (structural) surplus. The evolution of this indicator is not very different from the previous case, except for those governments that are voted out. Again, the governments that are re-elected show a more balanced budget policy and again for cases where there is a reelection and for the aggregate of observations we can see how the reduction of the structural deficit in the first measures of the legislature derives in a fiscal expansion as its conclusion approaches. However, the evolution of this indicator varies somewhat with respect to those governments that are not re-elected. As we can see, the structural deficit remains stable until the middle of the legislature, when there is an expansion followed by a setback in the last two intervals. Another interesting observation is that governments that are re-elected and the aggregate of governments leave the structural deficit more or less the same as at the beginning of the legislature, whereas in cases where a change of government takes place the deficit increases remarkably.

Figure 1.7: Political Budget Cycle (cyclically adjusted surplus). OECD (1975-2014)



1.6 Conclusions

Throughout this article it has been demonstrated that the correlations found in Brender & Drazen (2008) were little robust. When some reasonable changes are introduced in the model specification, in the definition of variables and in the sample period, the results obtained have varied and some of the correlations have disappeared.

Firstly, the model's robustness is tested against changes in the sample period. The sample period is extended from 2001 to 2014, keeping the rest of the specification exactly as in the original work. On performing this simple robustness exercise, one of the two correlations disappear. In particular, deficits in the last year of the term become not significant; there is no evidence that increasing the deficit in the last year in office reduces the likelihood of re-election.

Secondly, the robustness of the model is tested against changes in the model specification and against changes in the definition of the variables. Three changes are proposed (1) new definition of political alternation, (2) new definitions of the fiscal variables and (3) substituting non-cyclically adjusted fiscal data by cyclically adjusted data in order to avoid multicollinearity problems. Considering again the model with these changes it is shown that the deficit in the last year of the term in office is not significant and that economic growth positively affects the probability of re-election in most of the regressions.

Subsequently, an alternative specification for the reelection probability model is proposed, including the changes suggested in the previous section and also adding a variable that measures the years that the ruling party has been in office (Duration). With this alternative specification the model is estimated for the period 1975-2014. The results are as follows:

- Unlike Brender & Drazen (2008), an increase in the GDP per capita growth rate increases the probability of reelection in OECD countries. In particular, a ceteris paribus increase of 1 percentage point in the average growth rate during the term increases the probability of reelection by 8% [18].

- At the same time, increases in government deficit over the term in office (excluding last year) decrease the probability of re-election, but increases in the last year have no impact.

- The more years the incumbent party has been in office (variable "Duration"), the higher the probability of its defeat, other things equal.

It is important to note that the new correlations identified in this article should not be considered as conclusive evidence. Unstable results are common in cross-national studies about "economic voting" and it is worth further investigation about this field of research.

In the last part of the paper, several Political Business Cycles are characterised. On the one hand, the variable GDP growth decrease at the beginning of the legislature and

[18] Marginal effect at mean (MEM).

begins to rise in the 60-70% interval regardless of the presence or absence of alternation. The path (cycle) is the same for "good" and "bad" governments. However, if we look at levels, it can be seen that the growth rate is higher when the government is reelected.

Regarding budget surplus, a different path is observed for each kind of governments. The "good governments" increase the surplus at the beginning and in the middle of the legislature and reduce it before elections. On the contrary, "bad" governments increase the deficit from the beginning of the legislature and later the path is less clear. Finally, differences in levels are observed, the governments that are re-elected show in general a more balanced budget policy.

1.7 References

Anderson, C. J. (2000). Economic voting and political context: a comparative perspective. Electoral Studies, 19(2), 151-170.

Alesina, A., Ozler, S., Roubini, N., & Swagel, P. (1996). Political instability and economic growth. Journal of Economic growth, 1(2), 189-211.

Alesina, A., Carloni, D., & Lecce, G. (2012). The electoral consequences of large fiscal adjustments. In Fiscal Policy after the Financial Crisis (pp. 531-570). University of Chicago Press.

Beck, T., Clarke, G., Groff, A., Keefer, P., & Walsh, P. (2001). New tools in comparative political economy: The Database of Political Institutions. the world bank economic review, 15(1), 165-176.

Besley, Timothy & Case, Anne, (1995). "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition," American Economic Review, American Economic Association, American Economic Association, vol. 85(1), pages 25-45, March.

Bornhorst, F., Dobrescu, G., Fedelino, A., Gottschalk, J., & Nakata, T. (2011). When and How to Adjust Beyond the Business Cycle?: A Guide to Structural Fiscal Balances. International Monetary Fund, Fiscal Affairs Department.

Brender, A., & Drazen, A. (2005). Political budget cycles in new versus established democracies. Journal of monetary Economics, 52(7), 1271-1295.

Brender, Adi, & Allan Drazen. 2008. "How Do Budget Deficits and Economic Growth Affect Reelection Prospects? Evidence from a Large Panel of Countries." American Economic Review, 98(5): 2203-20.

Canes-Wrone, & Jee-Kwang Park. (2012). "Electoral Business Cycles in OECD Countries". American Political Science Review 106(1): 103-22.

Fair, R. C. (1978). The effect of economic events on votes for president. The Review of Economics and Statistics, 159-173.

Grilli, V., Masciandaro, D., Tabellini, G., Malinvaud, E., & Pagano, M. (1991). Political and monetary institutions and public financial policies in the industrial countries. Economic policy, 342-392.

Healy, A., & Lenz, G. S. (2014). Substituting the End for the Whole: Why Voters Respond Primarily to the Election Year Economy. American Journal of Political Science, 58(1), 31-47.

Hibbs, Douglas A, Jr, 2000. " Bread and Peace Voting in U.S. Presidential Elections," Public Choice, Springer, vol. 104(1-2), pages 149-80, July.

Hibbs, D. A. (2005). Voting and the Macroeconomy. Centrum fr forskning om offentlig sektor (CEFOS).

Jones M. P., Meloni, O., & Tommasi, M. (2012). Voters as fiscal liberals: Incentives and accountability in federal systems. Economics Politics, 24(2), 135-156.

Lewis-Beck, M. S. & Paldam, M. 2000. Economic voting: an introduction. Electoral Studies, 19: 113-21.

Lewis-Beck, M. S. & Stegmaier, M. 2000. Economic determinants of electoral outcomes. Annual Review of Political Science, 3: 183-219.

Lewis-Beck, M. S. & Fraile-Maldonado, M. 2010. Economic Voting in Spain: A 2000 Panel Test. Electoral Studies: An International Journal, ISSN 0261-3794, Vol. 29, N. 2, pags. 210-220.

Marshall, M. G., Jaggers, K., & Gurr, T. R. (2010). Polity IV Data Series Version 2010. College Park, MD: University of Maryland. Retrieved from http://www. systemicpeace. org/polity/polity4. htm.

Paldam, M., & Nannestad, P. (2000). What do voters know about the economy?: A study of Danish data, 1990-1993. Electoral Studies, 19(2), 363-391.

Powell Jr, G. B., & Whitten, G. D. (1993). A cross-national analysis of economic voting: taking account of the political context. American Journal of Political Science, 391-414.

Romer, T. (2012). Comments on Alesina, Carloni, and Lecce,"The Electoral Consequences of Large Fiscal Adjustments".

Vermeir, J., & Heyndels, B. (2006). Tax policy and yardstick voting in Flemish municipal elections. Applied Economics, 38(19), 2285-2298.

Wilkin, S., Haller, B., & Norpoth, H. (1997). From Argentina to Zambia: a worldwide test of economic voting. Electoral Studies, 16(3), 301-316.

1.8 Statistical Appendix

A. Databases

- Database of Political Institutions (DPI), World Bank.
- International Financial Statistics (IFS), International Monetary Fund.
- Government Finance Statistics (GFS), International Monetary Fund.
- World Development Indicators (WDI), World Bank.
- Political Finance Database, Institute for Democracy and Electoral Assistance (IDEA).
- Polity IV, University of Maryland, Center for International Development and Conflict Management.
- Economic Outlook Database (OCDE).
- World Economic Outlook (WEO), International Monetary Fund.
- A Historical Public Debt Database, International Monetary Fund.
- World Political Leaders, Zárate's Political Collections
- CPDS I, University of Bern.

B. Variable Definitions

i) Brender & Drazen's specification (Tables 1.1, 1.4, 1.5 & 1.6)

DEPENDENT VARIABLE

-ReelectLeader: See comprehensive definition in Brender & Drazen (2008), pag. 2205-2206 (Variable "Reelect").

ECONOMIC VARIABLES

-GDPpcGrowth: Average real GDP per capita growth over the term in office.

$$GDPpcGrowth = 100 * \sqrt[x]{\frac{GDP_0}{GDP_{-x}} - 1} \tag{1.8}$$

where, GDP_0 is the value of Real GDP per capita in the election year; GDP_{-x} is the value of Real GDP per capita in the first year of the legislature; x is the number of years in office. Source: "World Development Indicators" (WB).

FISCAL VARIABLES

-SurplusTerm: The change in the average surplus-to-GDP ratio in the two years preceding the elections (not including the election year) compared to the previous two years.

$$SurplusTerm = 1/2 * (B_{-1} + B_{-2}) - 1/2 * (B_{-3} + B_{-4}) \tag{1.9}$$

where, B_{-i} is the surplus as a percentage of GDP i years before the elections. Source: International Financial Statistics (IMF) and Government Finance Statistics (IMF).

-SurplusLastYear: The change in the surplus-to-GDP ratio in the election year relative to the previous year.

$$SurplusLastYear = (B_0 - B_{-1}) \tag{1.10}$$

where, B_{-i} is the surplus as a percentage of GDP i years before the elections. Source: International Financial Statistics (IMF) and Government Finance Statistics (IMF).

CONTROL VARIABLES

-NewDemocracy: A binary variable, for each country in each election year, receiving the value 1 if country is defined as a New Democracy. Otherwise, the country is defined as an Old Democracy and the variable receives a value of 0. Source: Database of Political Institutions (DPI), World Bank.

-MajoritarianSystem: A binary variable, for each country in each election year, receiving the value 1 in a country with a majoritarian electoral system, and 0 otherwise. Source: Database of Political Institutions (DPI), World Bank.

ii) Alternative specification (Tables 1.9, 1.10 & 1.11)

DEPENDENT VARIABLE

-ReelectParty: The definition is as follow:

a) Case 1: The old and the new government are formed by a single party.

-ReelectParty=1 if the incumbent party wins the elections

-ReelectParty=0 if the incumbent party is defeated

b) Case 2: The old government was formed by a single party and the new government is formed by a coalition of parties.

-ReelectParty=1 if the incumbent party is member of the coalition formed after the elections and has more than 60% of the seats in the new coalition.

-ReelectParty=0 otherwise

c) Case 3: The old government was formed by a coalition of parties and the new government is formed by a single party.

-ReelectParty=1 if the winning party was part of the ruling coalition before the election and had more than 60% of the seats.

-ReelectParty=0 otherwise

d) **Case 4**: The old and new government are formed by a coalition of parties.

-ReelectParty=1 if the parties that were members of the previous coalition have more than 60% of the seats in the new coalition and the parties who are members of the new coalition had more than 60% of the seats in the previous coalition

-ReelectParty=0 otherwise

ECONOMIC VARIABLES

-GDPpcWeighted: Weighted economic growth rate over the term in office.

$$GDPpcWeighted = \sum_{i=0}^{n} \lambda^i * \Delta GDP_{-i} * (1/\sum_{i=1}^{n} \lambda^i) \tag{1.11}$$

where, ΔGDP_0 is the real GDP per capita growth in the month in which the election was held; n is the number of months of the legislature; GDP_{-n} is the real GDP per capita growth in the first month of the legislature; λ is the discount rate; $\lambda = 0.95$

FISCAL VARIABLES

-SurplusTermCA: The change in the average surplus-to-GDP ratio in the two years preceding the elections (not including the election year) compared to the previous two years (cyclically adjusted data).

$$SurplusTerm = 1/2 * (B_{-1} + B_{-2}) - 1/2 * (B_{-3} + B_{-4}) \tag{1.12}$$

where, B_{-i} is the surplus as a percentage of GDP i years before the elections. Source: International Financial Statistics (IMF) and Government Finance Statistics (IMF).

-SurplusLastYearCA: The change in the surplus-to-GDP ratio in the election year relative to the previous year (cyclically adjusted data).

$$SurplusLastYear = (B_0 - B_{-1}) \tag{1.13}$$

where, B_{-i} is the surplus as a percentage of GDP i years before the elections. Source: International Financial Statistics (IMF) and Government Finance Statistics (IMF).

-NewSurplusTermCA: The change in the surplus-to-GDP ratio throughout the term in office, excluding the last twelve months (cyclically adjusted data).

$$NewSurplusTerm = \sum_{i=0}^{T-12} \frac{B_i}{(T-12)} \tag{1.14}$$

where, B_i is the primary government budget balance (as a percentage of GDP) during the i month of the legislature and T is the total number of month of the legislature.

-NewSurplusLastYearCA: The change in the surplus-to-GDP ratio in the last twelve months of the term in office (cyclically adjusted data).

$$NewSurplusLastYear = \frac{(m) * (B_0 - B_{-1}) + (12 - m) * (B_{-1} - B_{-2})}{12} \tag{1.15}$$

where, B_{-i} is the government surplus as a percentage of GDP i years before the elections and m is the month in which the elections are held (for example, February=2 and April=4).

OTHER VARIABLES

-Duration: number of years that the incumbent party has been governing the country.

Economic Determinants of Political Alternation and Long-term Economic Growth

Víctor Hernández García

November 2017

Resumen en Castellano

Este artículo estudia la relación entre el nivel (intensidad) de voto económico y el crecimiento económico en el largo plazo. En primer lugar, la sensibilidad de la alternancia política frente al crecimiento económico es estimada para cada uno de los países de la OCDE utilizando para ello métodos bayesianos (algoritmo de Monte Carlo). El análisis de las distribuciones posteriores de probabilidad muestra la existencia de una notable heterogeneidad entre los países analizados. Para ciertos países, el incremento del crecimiento económico a lo largo de la legislatura (percentil 80 en vez de percentil 20) incrementa la probabilidad de reelección un 50 %, mientras que para otros países el incremento es prácticamente igual a cero. En segundo lugar, se observa que la correlación entre estas sensibilidades y el crecimiento económico en el largo plazo es negativa y estadísticamente significativa. Estos resultados empíricos son consistentes con los modelos teóricos que mantienen que la estructura de incentivos de los políticos afecta al funcionamiento de las instituciones, y por lo tanto al bienestar en el largo plazo. Bajo la presencia de fuerte voto económico, los polcos se preocupan especialmente de la evolución de la economía en el corto plazo, por lo que intentan manipular las variables macroeconómicas generando distorsiones e ignorando las políticas de largo plazo. Por último, se repite el análisis anterior pero estimando esta vez la sensibilidad de la alternancia polca frente a la política fiscal. En este caso, también existe una correlación negativa entre dicha sensibilidad y el crecimiento económico en el largo plazo. Específicamente, aquellos países cuyos votantes valoran en mayor medida un ciclo presupuestario cóncavo presentan un menor crecimiento en el largo plazo.

Palabras clave: Voto económico, alternancia política, política fiscal
JEL classification: D72 E32 H61 H62

Economic Determinants of Political Alternation and Long-term Economic Growth*

Víctor Hernández García†

November 2017

Abstract

This paper studies the relationship between the economic determinants of political alternation and long-term economic growth. Firstly, the sensitivity of political alternation to economic growth is estimated for each of the OECD countries using bayesian methods. The analysis of the approximate posterior distributions show a large heterogeneity between countries. It is observed that for some countries an increase in the GDP per capita growth over the legislature (80th percentile instead of the 20th percentile) increases the probability of reelection by 50%, while for other countries the effect is almost zero. Secondly, a statistically significant negative correlation between these sensitivities and long run economic growth is observed. These empirical results represent support for those theoretical models that maintain that the politicians' career concerns (incentive structure) have an impact on the functioning of the institutions, affecting the long-term welfare. Under the presence of strong economic voting, politicians are only concerned with the evolution of the economy in the short term, so they try to manipulate macroeconomic variables generating distortions and neglecting long-term policies. Finally, the analysis is repeated estimating this time the sensitivity of political alternation to fiscal policy for each of the countries. In this case, there is also a negative correlation between the degree of economic voting and long-term economic growth. Specifically, those countries where voters further reward a concave political budget cycle face lower long-run economic growth.

Keywords: Politician Incentives, Political Business Cycle, Political Budget Cycle, Bayesian Inference

JEL classification: D72 E32 H3 H6

---

*I thank José Ignacio Conde Ruiz and Marco Celentani for their help and support. I gratefully acknowledge the support from the Spanish Ministry of Science and Technology from grants ECO2011-30323-C03-01 & BES-2012-061597.

†E-mail: victorhzgarcia@gmail.com

2.1 Introduction

Economic voting theory claims that the stochastic process that determines election outcomes is not independent of the economic performance. During the last years, economic voting theory has received considerable empirical support. There is strong evidence that the economy plays a great role in political elections (Lewis-Beck and Stegmaier, 2000).

In Hernández (2017c) the main objective is to identify the economic determinants of political alternation in developed countries. A reelection probability model is estimated in order to find the main economic determinants, i.e., the most relevant macroeconomic variables. To do that democratic elections from 25 OECD countries are included in the pooled regression.

In the present paper the approach and the objectives are slightly different. In this case the main objective is to analyse the relationship between the degree of economic voting and long-term economic growth. So in this case we want to study the presence of economic voting at a country level (one regression per country) instead of estimating a regression with data from all the countries.

The starting hypothesis is the presence of strong "economic voting heterogeneity" between countries. While in some countries voters are very sensitive to economic performance and incumbent administrative competencies, in other countries, the short-term evolution of the economy explains only a small part of the election outcomes. Our final conjecture is that these differences can be related with the long-term economic growth, so it can be expected a negative correlation between sensitivity of political alternation to economic performance and long run economic growth.

It is widely documented that voter behaviour largely influences the actions, policies and incentives of the politicians (Persson and Tabellini (2000)). If the incumbent believes that voters give a disproportionate weight to short-term economic performance, there is an incentive to neglect long-term policies (education, infrastructures, R & D, long-term welfare...) and focus on improving the performance of the economy in the short term[1]. This behaviour is known in the literature as short-term bias (Bonfiglioli and Gancia (2013)) and can be dynamically inefficient due to distortions. However, if voters ignore macroeconomic variables when voting and only react to other political issues (honesty, ideology, rights and freedoms, long-term policies...), the incentive structure is completely different and the short-term bias is likely to disappear. In this case, it would not make sense to focus on improving economic growth in the short term nor modifying the fiscal policy because the manipulations may generate distortions in the economy and difficult re-election prospects in future elections. So this is why we expect a negative correlation between sensitivity of political alternation to economic performance and long-term economic growth.

[1] Unless the polls forecast a large gap between the incumbent and the opposition.

The rest of the paper is organised as follows. Section 2.2 summarises the literature on politicians' career concerns and economic voting. Section 2.3 introduces the database. Section 2.4 introduces the methodology to estimate the correlation between sensitivity of political alternation to economic performance and long run economic performance. In Section 2.5 the bayesian inference is presented. In Section 2.6 the robustness of the proposed models is tested. Finally, conclusions are drawn.

2.2 Literature review

The present study is related to the literature on politicians' career concerns (for a review see Persson and Tabellini (2000)). Celentani and Conde-Ruiz (2008) propose a dynamic model in which politicians' career concerns have countervailing consequences on welfare. On the one hand, they are constructive because they increase the probability of reelection of competent incumbents. On the other hand, politicians' career concerns may promote that incumbent distort policies in order to increase reelection prospects. Bonfiglioli & Gancia (2013) propose a model of electoral accountability and obtain that when politicians' ability is ex ante unknown, elections improve political accountability and selection. However, incumbents underinvest in costly policies (with future returns) because they prefer to signal high ability in order to increase reelection chances.

However, few studies address this topic from a empirical point of view. In fact, to the best of our knowledge, no empirical work has studied the relationship between politicians' career concerns and long-term economic growth.

Our work is also related to economic voting literature (for a comprehensive review of the literature see Lewis-Beck and Stegmaier (2000) and Hibbs (2005)). It is worth mentioning the following references:

Grilli et al. (1991) claim that the stochastic process that determines the political alternation is independent of the economy. At the same time, political alternation affects economic performance. Using simple procedures, they found a negative correlation between average government durability and debt accumulation in OECD countries. In the same vein, Alesina et al. (1996) find that in countries and time periods with a high propensity of government collapse, growth is significantly lower than otherwise.

Alesina, Carloni & Lecce (2012) find "no evidence that governments which quickly reduce budget deficits are systematically voted out of office" in a sample of 19 OECD countries from 1975 to 2008. Aisen and Veiga (2013) find that political instability adversely affects growth by lowering the rates of productivity growth and physical and human capital accumulation. Hernández (2017c) finds that higher growth rates of GDP per capita increase the probability of reelection in OECD countries. However, there is no evidence that fiscal policy changes at the end of the legislature affect the re-election chances of the incumbent parties.

Finally, classical studies on the determinants of long-term growth do not include these kind of factors (economic voting or politicians' career concerns) as explanatory variables (Barro (1996); Doppelhofer et al. (2000)).

2.3 Data sources

We are interested in investigating the relationship between economic voting and long-term economic performance in developed countries, so in this study we focus only on OECD countries.

The database used in this article is based on information from various sources[2]. GDP growth rates and population data is collected from "World Development Indicators" (WDI)[3] and OECD Statistics. Fiscal data are taken from Economic Outlook Database (OECD). Political variables are taken from "Database of Political Institutions" (DPI). Finally, information on the level of democracy is taken from "Polity Democracy Index" (Marshall et al., 2010). The combination of sources allows us to include elections from 1975 to 2015[4].

We want to restrict our sample only to democratic elections, so we include only the elections in which the country has a positive score in the Polity Democracy Index (Marshall et al., 2010)[5]. Moreover, the first democratic election is excluded because it is imposible to determine if there has been political alternation due to the lack of an incumbent party.

Additionally, we only include countries with at least 6 democratic elections in the sample period (1975-2015) in order to guarantee a minimum of accuracy in the estimation of the sensitivity of the political alternation to economic performance. Estimating a regression (two coefficients) with five or less elections is not recommended. The following new democracies of Eastern Europe are excluded because they do not meet this requirement: Czech Republic, Estonia, Letonia, Eslovenia, Hungary, Poland and Slovak Republic.

At the same time, Mexico[6] and Turkey are excluded because of their low income. In spite of being members of OECD, their income is more than three times lower than the average OECD income.

Finally, Switzerland is excluded due to its complex executive power, the Swiss Federal Council, which is formed by seven member belonging to different parties. Traditionally the presidency rotates among members in mandates of one year and under the approval of the Federal Assembly. In this context, it is not easy to identify an incumbent party in order to assign responsibility for the progress of the national economy.

After applying these criteria, it is possible to include a total of 283[7] elections corresponding to 25 OECD countries. The number of observations (elections) range from 6 (Czech Republic and Chile) to 16 (Australia). The list of countries and elections included in the analysis is detailed in the Appendix (Table 2.A11). The list of excluded countries is also shown (Table 2.A12).

[2] A comprehensive description of the data sources is presented in the Appendix (Section B).

[3] WDI is the primary World Bank collection of development indicators.

[4] DPI data begin in 1975.

[5] We require that the country has a positive score in every year of the legislature, not only in the election year.

[6] Additionally, Mexico does not fullfil the requirement of having at least six democratic elections according to the Polity Democracy Index (PDI).

[7] The sample size decreases to n=200 when fiscal variables are included as explanatory variables.

2.4 Methodology: Bayesian Econometric Methods

The main objective of this paper is to test the following conjecture: ”it is expected a negative correlation between the sensitivity of political alternation to economic performance and long run economic growth in OECD countries”. To do this, we propose to apply Bayesian methods due to its better properties in small samples instead of using the standard maximum likelihood estimator.

Several recent studies have noted that Bayesian Methods (Monte Carlo simulation) are better tan frequentist maximum likelihood (ML) methods when dealing with small samples (McNeish, 2016). Baldwin & Fellingham (2013) compare the performance of an adjusted restricted maximum likelihood (REML) analysis to a Bayesian analysis in the context of multilevel models. They find that Bayesian Methods can produce more accurate inferences about the parameters. Doron & Gaudreau (2014) claim that several reasons favoured the use of the Bayesian estimator rather tan the more traditional maximum likelihood estimator, among which stands out its best performance in small samples.

Moreover, using bayesian methods offers extra advantages. For instance, after running regressions it is possible to calculate the probability of a parameter being positive or negative using the posterior probability distribution. This feature greatly simplifies testing conjectures.

In the first phase of the research, the objective is to propose a reelection probability model in order to estimate the ”sensibility of political alternation to economic performance” for each of the OECD countries. A total of 25 probit[8] regressions (one for each country) will be estimated using bayesian methods (MC algorithm).

The dependent variable ”ReelectParty” measures political alternation at a political party level[9]. The variable is equal to 1 if the ruling party (or ruling coalition) is re-elected and equal to 0 otherwise[10]. It is computed from: ”Database of Political Institutions” (DPI).

At the same time, we can not include more than one explanatory variable in the model at the same time due to small number of observations[11]. Throughout the article several variables will be proposed as explanatory variables. All these variables are intended to be a measure of economic performance, summarising the good or bad economic management. We believe that economic growth is the variable that best captures the evolution of the economic situation of a country[12]. Additionally, other explanatory variables will be included in the specification, such as fiscal variables: cyclically adjusted budget surplus and non-cyclically adjusted budget deficit.

[8] We propose a Probit model because the dependent variable (Political Alternation) is a binary variable. We also estimate a Logit and a Linear Probability Model (LPM) to check if the results are robust.

[9] There is an alternative approach that measures political alternation at a leader level, as in Brender & Drazen (2008).

[10] The complete definition of the variable is described in the Appendix (II. Variable Definitions).

[11] The mean number of elections per country is only 11.

[12] We have not used ”Disposable Personal Income” due to data unavailability for the sample period.

The econometric model is as follows:

$$ReelectParty = F(\beta_0 + \beta_1 * EconomicPerformance + u) \tag{2.1}$$

where, **F** is the Cumulative Distribution Function (CDF) of the Standard Normal Distribution and **u** is the standard error term.

Bayesian inference utilises prior probabilities and data to estimate the posterior probability of the model parameters (instead of a point estimate). In Bayesian econometrics, the parameters are random variables whose distribution is unknown. The researcher uses data and a subjective distribution of the parameters (prior distribution) and after the estimation an estimated distribution (posterior distribution) is obtained.

The algorithm proposed for the estimation is a variant of the Monte Carlo algorithm. Monte Carlo simulation is repeated 200,000 times (trials) to trace out the distribution. After the estimation, we will have obtained 200.000 estimated values for β_0 y β_1, which allows us to calculate the estimated distributions (posterior distribution) of β_0 y β_1.

Note that in a Probit model the parameters can not be directly interpreted as marginal effects. Then now we need to propose a methodology to compute the marginal effect of the explanatory variable, which we call "sensitivity of political alternation to economic performance". At this point, some possibilities arise. An easy way is computing the marginal effect at the mean of the independent variable. The disadvantage of this option is that it only takes into account a part of the distribution. We consider that is better to measure the change in reelection probability when we increase the independent variable from the 20th percentile to the 80th percentile[13]. This way we are considering a large part of the distribution (excluding both extremes).

Once the posterior distributions of the sensitivities are obtained, we can calculate the following variables for each of the OECD countries: (1) the mean and median of the distribution, (2) the probability of the sensitivity being positive or negative and (3) the probability that the sensitivity is between two values.

In the second phase of the study, we want to study the correlation between the "sensitivities of political alternation to economic performance" previously estimated and the "long run economic growth". The correlation will be computed doing a simulation and generating random draws from the posterior distributions of the sensitivities. After that it will be easy to test the initial conjecture computing the probability of the correlation being negative.

[13]The percentiles are computed using GDP growth data from each country.

2.5 Results

As we have argued previously, due to the small number of observations it is not advisable to use more than one explanatory variable at the same time. In this section several macroeconomic variables are proposed as independent variables: GDP growth (Section 2.5.1) and fiscal variables (Section 2.5.2).

2.5.1 Using GDP growth as a measure of economic performance.

The evolution of the real per capita GDP growth may be a good indicator of the economic situation of a country. The aim is to compute the sensitivity of political alternation to economic growth for each country. To do that, the following econometric model is proposed:

$$ReelectParty = F(\beta_0 + \beta_1 * WGDPpcGrowth + u) \tag{2.2}$$

where, **F** is the Cumulative Distribution Function (CDF) of the Standard Normal Distribution and the variable **u** is the standard error term.

Next, a brief overview of the variables is presented:

ReelectParty: Dummy variable that measures political alternation[14].

WGDPpcGrowth: The weighted average of real per capita GDP growth over the term in office. The weighting factor is a discount rate that gives more weight to the growth data closer to the election. The introduction of this discount rate would be justified by the myopia of voters. There is evidence that voters (or at least most of them) would have a short time horizon (see Lewis-Beck and Paldam, 2000, Paldam and Nannestad, 2000 & Healy & Lenz, 2014). The formula is as follows (Hibbs, 2000):

$$GDPpcWeighted = \sum_{i=1}^{n} \lambda^i * \Delta GDP_{-i} * (1/\sum_{i=1}^{n} \lambda^i) \tag{2.3}$$

where, ΔGDP_0 is the real GDP per capita growth in the month in which the election was held; n is the number of months of the term in office; GDP_{-n} is the real GDP per capita growth in the first month of the term in office; λ is the discount rate ($\lambda = 0.95$[15]).

In order to estimate the econometric model, two alternatives are proposed (Model 1 and Model 2). Both are Probit models, include a discount factor for the explanatory variable and they are estimated using MC algorithm. The only different is the window chosen for the independent variable. In Model 1, the variable WGDPpcGrowth is computed using the last 36 months of the legislature, while in Model 2 only the last 24 months are taken into account.

[14] Detailed definition is shown in the Appendix.

[15] Exogenous value taken from Hibbs (2000) and Hernández (2017c).

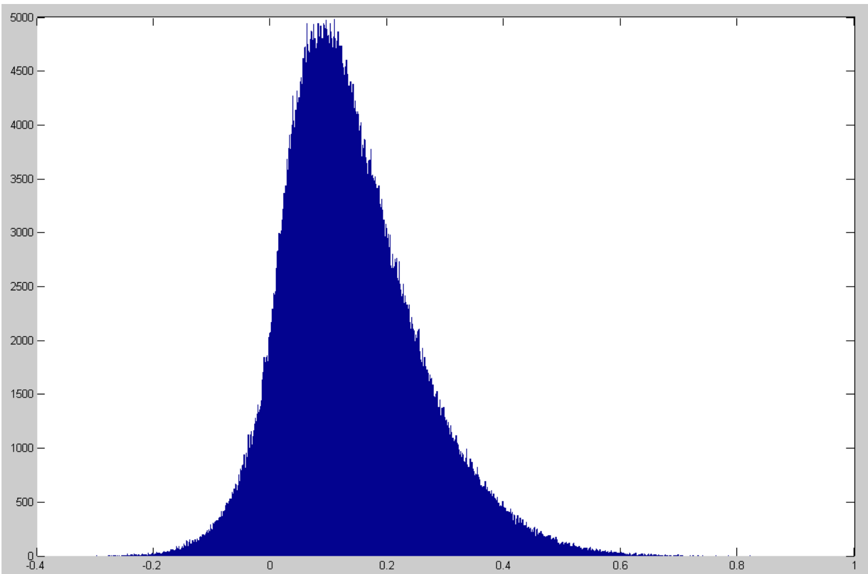
Table 2.1: Estimating the sensitivity of Political Alternation to Economic Growth

	Model 1	**Model 2**
Functional Form	Probit	Probit
Dependent Var.	ReelectParty	ReelectParty
Independent Var.	WGDPpcGrowth	WGDPpcGrowth
Window[a]	Last 36 months	Last 24 months
Discount Rate	$\lambda = 0.95$	$\lambda = 0.95$
Technique	Bayesian Methods (MC algorithm)	Bayesian Methods (MC algorithm)

[a]Months used to compute WGDPpcGrowth.

As an example, Figure 2.1 shows the posterior probability distribution of the sensitivity of the political alternation to economic growth for the United Kingdom (Model 1). Note that Bayesian provides posterior probability distributions for the parameters instead of point estimates, as in classical frequentist models. In the case of the United Kingdom the median of the posterior distribution is equal to 0.105. This sensitivity can be interpreted as: "if we increase the value of the variable WGDPpcGrowth from the 20th percentile to the 80th percentile, the probability of reelection increases by 10.5%". At the same time, the probability of the marginal effect of WGDPpcGrowth being positive is equal to 83%.

Figure 2.1: Estimated Distribution Function of the Marginal Effect of WGDPpcGrowth (UK, Model 1)



Median of the EDF: 0.105, Mean of the EDF: 0.1167
Prob(Marg. Effect of WGDPpcGrowth$>$ 0) = 83%

Now we are able to estimate the "sensitivity of political alternation to economic growth" for every OECD country. Table 2.2 shows the median value of the posterior distributions. There exist a great variation in the degree of economic voting across countries, as in Paldam (1991). Due to the small size of the sample, the sensitivity is only significant for 6 countries.

Table 2.2: Bayesian inference of the sensitivities

Country	Model 1		Model 2	
	Sensitivity[a]		Sensitivity	
	Median of the PD[b]	P-value[c]	Median of the PD	P-value
Australia	0.059	0.61	0.033	0.56
Austria	-0.038	0.57	-0.124	0.68
Belgium	0.004	0.51	-0.004	0.51
Canada	0.216	0.77	0.182	0.71
Chile	-0.269	0.62	-0.263	0.63
Czech Republic	0.481	0.77	0.512	0.77
Germany	0.114	0.68	0.026	0.54
Denmark	0.378*	0.92	0.358*	0.93
Spain	0.364	0.88	0.370	0.85
Finland	0.094	0.62	0.117	0.64
France	0.462**	0.95	0.566**	0.97
UK	0.105	0.83	0.156	0.69
Greece	0.384**	0.95	0.412**	0.96
Ireland	0.155	0.67	0.105	0.62
Iceland	0.009	0.51	-0.025	0.53
Israel	0.032	0.54	0.076	0.59
Italy	0.419	0.93	0.295	0.87
Japan	0.383**	0.95	0.391**	0.97
Luxembourg	0.080	0.57	0.029	0.53
Netherlands	0.305	0.86	0.319	0.86
Norway	0.022	0.52	0.013	0.51
New Zealand	0.385	0.89	0.312*	0.90
Portugal	0.448*	0.93	0.339	0.86
Sweden	0.422	0.89	0.369	0.88
USA	0.539*	0.94	0.513**	0.95

***Significant at 1%, ** Significant at 5% and * Significant at 1%.

[a]Sensitivity of Political Alternation to Economic Growth, defined as the change in reelection probability when we increase the independent variable from the 20th percentile to the 80th percentile.

[b]Posterior Distribution.

[c]Defined as the probability of being positive (negative) if the median value is positive (negative).

Table 2.3 presents the average real per capita GDP growth for each OECD country during the period 1970-2015. The countries with higher growth rates are Ireland, Chile and Luxembourg. At the same time, the growth rates are specially volatile in Czech Republic, Greece and Chile. Additionally, the correlation between the long-term economic growth and the GDP level in 1970[16] is equal to -0.12, indicating that the process of convergence is not very intense during this period.

Table 2.3: Real GDP per capita Growth (1970-2015)

Country	**Real GDP per capita Growth**[a]					
	Average	**Rank.**	**Standard Deviations**[b]	**Rank.**	**CV**[c]	**Rank.**
Australia	1.67	16	1.73	25	0.90	22
Austria	2.00	10	1.98	22	0.82	24
Belgium	1.80	14	2.07	19	0.91	21
Canada	1.59	21	2.09	16	1.08	15
Chile	2.72	2	4.42	1	1.66	3
Czech Republic	1.64	18	3.91	3	2.28	1
Denmark	1.46	22	2.35	13	1.24	10
Finland	2.02	9	3.18	10	1.29	7
France	1.62	20	1.93	24	0.90	23
Germany	1.87	12	2.01	21	1.06	16
Greece	1.17	25	4.41	2	1.91	2
Iceland	2.37	4	3.82	5	1.49	4
Ireland	3.47	1	3.45	8	0.98	20
Israel	2.09	6	3.32	9	1.23	11
Italy	1.45	23	2.68	12	1.27	9
Japan	2.05	8	3.66	6	1.16	12
Luxembourg	2.53	3	3.51	7	1.36	6
Netherlands	1.65	17	2.17	15	1.03	18
New Zealand	1.38	24	2.08	18	1.48	5
Norway	2.30	5	1.95	23	0.78	25
Portugal	2.06	7	3.83	4	1.28	8
Spain	1.83	13	2.96	11	1.11	13
Sweden	1.64	19	2.25	14	1.09	14
UK	1.89	11	2.08	17	1.04	17
USA	1.78	15	2.02	20	0.98	19

[a]Source: Economic Outlook Database (OECD)
[b]Short-term standard deviation
[c]CV: Coefficient of Variation

[16]Constant 2010 US.

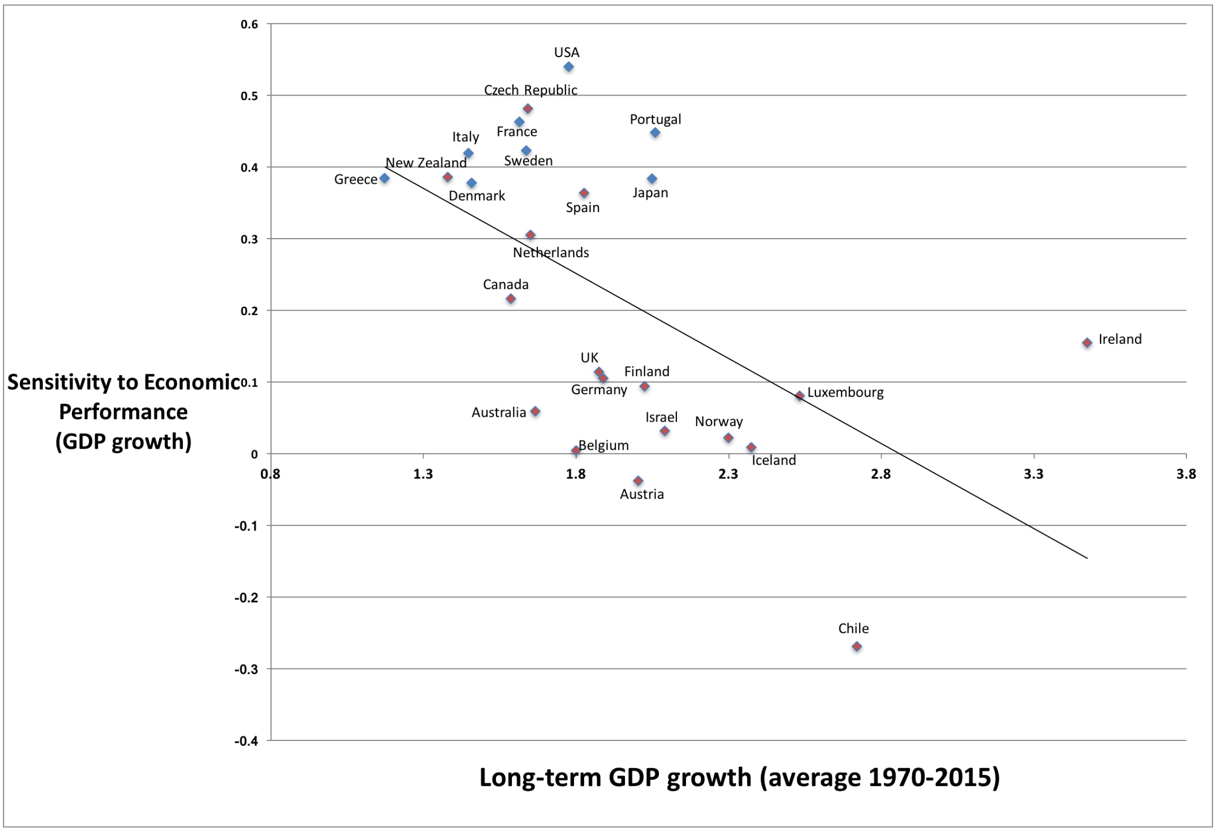
Now, we can study the correlation between the sensitivity of PA to EP and long run economic growth.

Figure 2.2 shows the relationship between long-term GDP growth (x-axis) and sensitivity of Political Alternation to Economic Growth (y-axis). The values of the sensitivities correspond to the median values of the posterior distribution estimated from model 1[17]. The values of the long-term GDP growth are taken from Table 2.3. The correlation is directly computed and is equal to -0.51, so is clearly negative. This procedure do not use all the posterior distributions, but only the median value.

An alternative method to compute the correlation is to generate random draws from the posterior distributions of the sensitivities. The procedure is as follows. For each country, a random draw is generated from the posterior distribution of the sensitivities. Later, we compute the correlation between the sensitivities and long-term economic growth using values from Table 2.3. We repeat this process 200.000 times. Finally, we have calculated 200.000 values for the correlation (posterior distribution).

Figure 2.3 shows the posterior distribution for the correlation. The probability of the correlation being negative is 94%[18], so it is statistically significant at 10%.

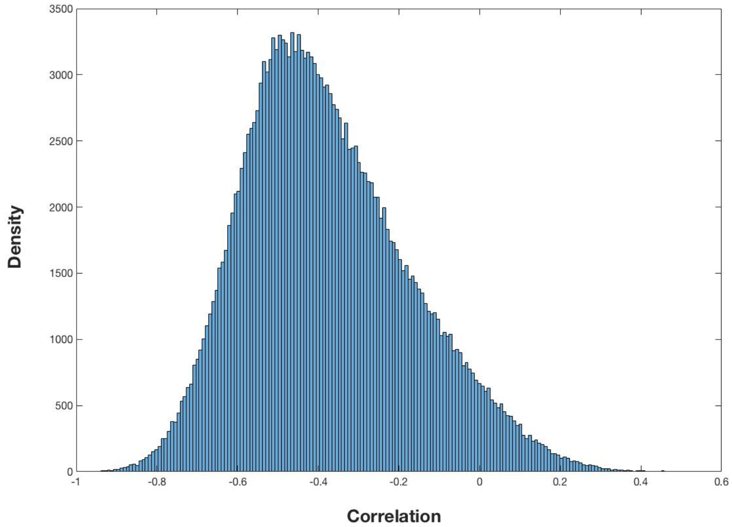
Figure 2.2: Correlation between "Sensitivity" and "long-run GDP Growth" (Model 1)



[17]Results for Model 2 are presented in the Appendix (Figure 2.A6).

[18]The probability is equal to 91% when model 2 is used. Result are shown in the Appendix (Figure 2.A7).

Figure 2.3: Histogram of the correlation(Model 1)



Median of the EDF: -0.2512, Mean of the EDF: -0.2482
Prob(corr<0)=94%

In Hernández (2017c) several opportunistic political business cycles are identified. The economic growth decreases in the first half of the legislature and begins to rise when 60% of the legislature has passed. The same shape (u-shaped) is observed both when the government is re-elected and when it is defeated. However, if we look at levels, it can be seen that the growth rate is higher when the government is reelected.

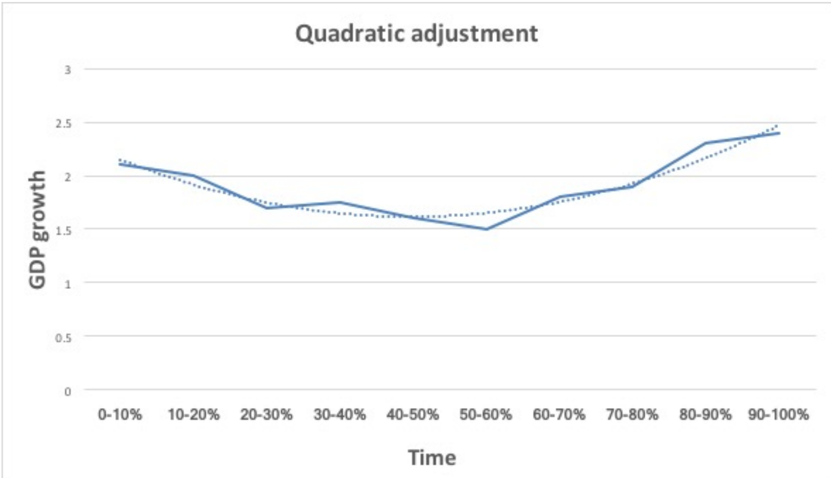
In this part, we study the sensitivity of political alternation to the degree of convexity/concavity of the political business cycle. Firstly, we use a quadratic function to adjust the political business cycle for each election of each country. Doing this exercise we obtain three coefficients for each election: (1) a constant, (2) a linear parameter and (3) a quadratic term. Note that if the quadratic parameter is positive (negative), the adjusted curve is convex (concave). Figure 2.4 shows the quadratic adjustment for a given election (convex path). Secondly, we used the 283 estimated quadratic terms (one for each election) to estimate the following model (Model 3) for each country:

$$ReelectParty = F(\beta_0 + \beta_1 * QuadraticGDP + u) \quad (2.4)$$

where QuadraticGDP is simply the quadratic term obtained when we use a quadratic function to adjust the political business cycle.

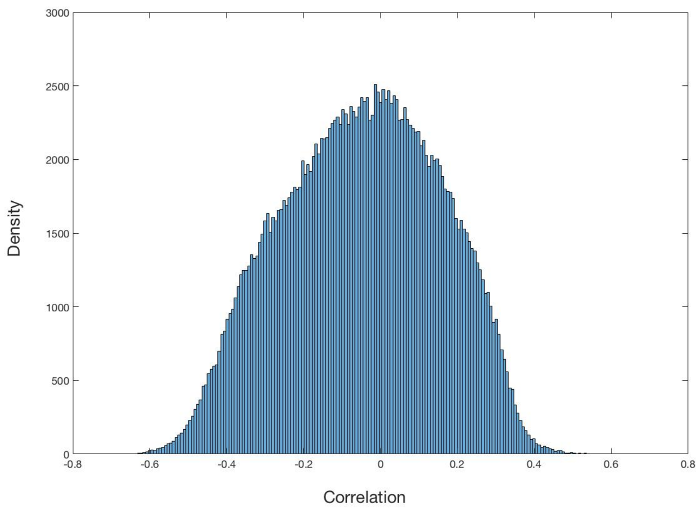
At this point we have estimated the sensitivity of political alternation to the degree of convexity/concavity of the political business cycle for each country, so we can compute the correlation between this sensitivity and long-term economic growth. Results are shown in Figure 2.5. It can be seen that the correlation is not very strong in this case. The probability of being negative is only 60%, what means that it is not statistically significant. There is no evidence that the sensitivity of political alternation to the degree of convexity/concavity of the political business cycle is correlated with long-term economic performance. This result are consistent with (Hernández (2017c)) given that no differences in the shape of the political business cycle were found between reelected and defeated incumbents.

Figure 2.4: Quadratic adjustment of the GDP growth curve.



Constant=2.45, linear=-0.33 , quadratic=0.04

Figure 2.5: Histogram of the correlation (Model 3)



Median of the EDF: -0.08, Mean of the EDF: -0.07
Prob(corr<0)=60%

2.5.2 Using budget surplus as a measure of economic performance.

Variations in public expenditure and taxation have a direct and immediate effect on voters' welfare. A few studies have identified political budget cycles, which can be an evidence of political manipulation.

Alesina et. al (1992) observe evidence of "political budget cycles" and "political monetary cycles" but they find little evidence of pre-electoral effects of economic outcomes, in particular on GDP growth and unemployment. This result is consistent with the notion that it is easy to manipulate policy instruments, while it is more difficult to control policy outcomes.

Hernández (2017c) studies OECD countries and identifies several political budget cycles. Moreover, a different path is observed depending on whether or not there is political alternation. When the government is re-elected the budget surplus increases at the beginning and in the middle of the legislature and slightly decreases just before elections. In this case the path is concave (inverted-u shaped). On the contrary, for those elections in which political alternation takes place the budget surplus decreases during the first half of the legislature and later the path is less clear (convex curve). Finally, differences in levels are observed, the governments that are re-elected show in general a more balanced budget policy.

The evolution of the fiscal variables can be considered as a measure of economic performance, then voters can use this information when deciding their vote. In this part, we repeat the analysis performed in Section 2.5.1 but now using fiscal variables as a measure of economic performance.

Firstly, we compute the sensitivity of political alternation to the mean level of budget surplus over the legislature using the following model[19].

$$ReelectParty = F(\beta_0 + \beta_1 * SurplusLevel) + u) \tag{2.5}$$

where SurplusLevel is the mean value of the budget surplus throughout the legislature. Model 4 uses non-cyclically adjusted budget surplus and Model 5 uses cyclically adjusted budget surplus[20].

Now we can compute the correlation between the estimated sensitivities and long-term economic growth. Table 2.6 shows the results. In both models, it is obtained a statistically significant negative correlation between long-term economic growth and the sensitivity of political alternation to the mean level of budget surplus over the legislature. There is evidence that in those countries where voters give more weight to fiscal policy, long-term economic performance is worse. A possible explanation is related to politicians' career concerns literature. When voters value a good fiscal performance, politicians have incentive to manipulate fiscal variables in order to increase reelection prospects.

[19]The number of elections included in the sample decrease until 201 due to fiscal data availability.

[20]Data is taken from Economic Outlook Database (OECD).

Table 2.4: Estimating the sensitivity of Political Alternation to Fiscal Policy

	Model 4	**Model 5**
Functional Form	Probit	Probit
Dependent Var.	ReelectParty	ReelectParty
Independent Var.	Budget Surplus (level)	CA Budget Surplus (level)
Technique	Bayesian Methods (MC algorithm)	Bayesian Methods (MC algorithm)

Secondly, we compute the sensitivity of political alternation to the degree of convexity/concavity of the political budget cycle, following the same methodology as in Section 2.5.1. Firstly, we use a quadratic function to adjust the political budget cycle for each election of each country, obtaining a constant, a linear term and a quadratic term. Secondly, we used the estimated quadratic terms (one for each election) to estimate the following reelection probability model:

$$ReelectParty = F(\beta_0 + \beta_1 * QuadraticSurplus + u) \tag{2.6}$$

where QuadraticGDP is simply the quadratic parameter obtained when we use a quadratic function to adjust the political budget cycle.

As in the previous case, two models are proposed, whose characteristics are detailed in the Table 2.5. Model 6 uses non-cyclically adjusted budget surplus and Model 7 uses cyclically adjusted budget surplus. In both models, it is obtained a statistically significant negative correlation between long-term economic growth and the sensitivity of political alternation to the degree of convexity/concavity of the budget surplus curve (Table 2.6). There is evidence that in those countries where voters reward concavity, long-term economic performance is worse. The possible explanation is similar to the previous case. When a more concave political budget cycle increases the probability of reelection, politicians have incentive to manipulate fiscal variables in order to increase reelection prospects.

Table 2.5: Estimating the sensitivity of Political Alternation to the degree of convexity/concavity of the political budget cycle

	Model 6	**Model 7**
Functional Form	Probit	Probit
Dependent Var.	ReelectParty	ReelectParty
Independent Var.	Budget Surplus (quadratic)	CA Budget Surplus (quadratic)
Technique	Bayesian Methods (MC algorithm)	Bayesian Methods (MC algorithm)

Table 2.6: Correlation between economic voting and long-term economic growth

	Model 4	**Model 5**	**Model 6**	**Model 7**
Probability of the correlation being negative	0.88	0.90	0.985	0.91

2.6 Robustness Analysis

In this section we examine the robustness of these findings to various different specifications. For simplicity, we only focus on Model 1.

Firstly, we test the robustness of the model against changes in the functional form. Model 8 proposes the estimation of a Logit model instead of a Probit model. Everything else remain exactly the same as in Model 1. At the same time, Model 9 proposes the Linear Probability Model (LPM). The probability of the correlation between sensitivity of political alternation to GDP growth and long-term GDP growth being negative is presented in Table 2.7 (last row). It can be observed that results are not altered when we modify the functional form of the model. The correlation continues being negative and statistically significant.

Table 2.7: Part I: Changing the functional form

	Model 8	**Model 9**
Functional Form	Logit	LPM
Dependent Var.	ReelectParty	ReelectParty
Independent Var.	WGDPpcGrowth	WGDPpcGrowth
Window	Last 36 months	Last 36 months
Discount Rate	$\lambda = 0.95$	$\lambda = 0.95$
Estimation Method	Bayesian Methods (MC algorithm)	Bayesian Methods (MC algorithm)
Prob. of Corr<0[a]	0.96	0.93

[a]Probability of the correlation between sensitivity of political alternation to GDP growth and long-term GDP growth being negative.

Secondly, we test the robustness of the model when we modify the window used to compute the weighted average of the GDP growth. In Model 10, we compute the weighted average of GDP growth considering only the last 12 months. In Model 11 the entire legislature is used. The correlation continues being negative in both models, but in Model 11 the correlation is not statistically significant at 10%, although it is very close.

Table 2.8: Part II: Modifying the window for WGDPpcGrowth.

	Model 10	**Model 11**
Functional Form	Probit	Probit
Dependent Var.	ReelectParty	ReelectParty
Independent Var.	WGDPpcGrowth	WGDPpcGrowth
Window	Last 12months	The entire legislature
Discount Rate	$\lambda = 0.95$	$\lambda = 0.95$
Estimation Method	Bayesian Methods (MC algorithm)	Bayesian Methods (MC algorithm)
Prob. of Corr<0[a]	0.92	0.89

[a]Probability of the correlation between sensitivity of political alternation to GDP growth and long-term GDP growth being negative.

Thirdly, we test the robustness of the model against changes in the estimation methods. We estimate the model using maximum likelihood (ML) method instead of bayesian methods (Model 12). In this case, we obtain point estimates for the sensitivities. The correlation is equal to -0.42. There is evidence that the correlation is negative (the confident interval do not include the zero and the p-value is equal to 0.08).

Table 2.9: Part III: Changing the estimation method

	Model 12
Functional Form	Probit
Dependent Var.	ReelectParty
Independent Var.	WGDPpcGrowth
Window	Last 36 months
Discount Rate	$\lambda = 0.95$
Estimation Method	Maximum Likelihood
Correlation[a]	-0.42*
p-value[b]	0.08

[a]Correlation between sensitivity of political alternation to GDP growth and long-term GDP growth
[b]In this case, p-value is reported.

Finally, we test the robustness of the model against changes in the definition of political alternation (Model 13). As an alternative definition, we use the variable proposed in Brender & Drazen (2008), that is ReelectLeader. Both definitions are different about 15% of the elections. It can be observed that the probability of the correlation being negative is almost the same (0.92 instead of 0.94).

Table 2.10: Part IV: Changing the definition of political alternation

	Model 13
Functional Form	Logit
Dependent Var.	ReelectLeader
Independent Var.	WGDPpcGrowth
Window	Last 36 months
Discount Rate	$\lambda = 0.95$
Estimation Method	Bayesian Methods (MC algorithm)
Prob. of Corr<0[a]	0.92

[a]Probability of the correlation between sensitivity of political alternation to GDP growth and long-term GDP growth being negative.

After performing the robustness analysis we can conclude that findings are not sensitive against changes in the functional form, in the estimation method, in the months used to compute WGDPpcGrowth and in the definition of political alternation. The probability of the correlation being negative ranges from 89% to 96%, then it is quite stable.

2.7 Conclusions

This paper studies the relationship between the economic determinants of political alternation and long-term economic growth in OECD countries (1975-2015). We are interested in whether politicians' career concerns can influence long-term welfare.

Firstly, the sensitivity of political alternation to economic growth is estimated for each of the OECD countries using bayesian methods (MC algorithm). Regression outputs show a large heterogeneity between countries. It is observed that for some countries an increase in the GDP per capita growth over the legislature (80th percentile instead of the 20th percentile) increases the probability of reelection by 50%, while for other countries the effect is almost zero.

Secondly, a statistically significant negative correlation between this sensitivity and the long-run economic growth is observed. These empirical results represent support for theoretical models that maintain that the politicians' career concerns (incentive structure) have an impact on the functioning of the institutions, affecting the long-term welfare. Under the presence of strong economic voting, politicians are only concerned with the evolution of the economy in the short term (long-term bias), so they try to manipulate macroeconomic variables generating distortions and neglecting long-term policies.

Thirdly, the previous analysis is repeated focusing this time on fiscal variables. In this case, there is also a negative correlation between the sensitivity of political alternation to fiscal policy and long-term growth. Specifically, those countries where voters reward a more balanced fiscal policy face lower long-run economic growth in average. Additionally, it is also observed that those countries where voters reward a more concave political budget cycle face lower long-run economic growth. Again politician incentives affect long-term welfare.

Finally, robustness tests indicate that findings are robust against changes in the functional form, in the estimation method, in the months used to compute the variable "WGDPpcGrowth" and in the definition of political alternation.

2.8 References

Aisen, A., & Veiga, F. J. (2013). How does political instability affect economic growth?. European Journal of Political Economy, 29, 151-167.

Alesina, A., Cohen, G. D., & Roubini, N. (1992). Macroeconomic policy and elections in OECD democracies. Economics Politics, 4(1), 1-30.

Alesina, A., Ozler, S., Roubini, N., & Swagel, P. (1996). Political instability and economic growth. Journal of Economic growth, 1(2), 189-211.

Alesina, A., Carloni, D., & Lecce, G. (2012). The electoral consequences of large fiscal adjustments. In Fiscal Policy after the Financial Crisis (pp. 531-570). University of Chicago Press.

Baldwin, S. A., & Fellingham, G. W. (2013). Bayesian methods for the analysis of small sample multilevel data with a complex variance structure.Psychological methods,18(2), 151.

Barro, R. J. (1996). Determinants of economic growth: a cross-country empirical study (No. w5698). National Bureau of Economic Research.

Beck, T., Clarke, G., Groff, A., Keefer, P., & Walsh, P. (2001). New tools in comparative political economy: The Database of Political Institutions. the world bank economic review, 15(1), 165-176.

Besley, Timothy & Case, Anne, (1995). "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition," American Economic Review, American Economic Association, American Economic Association, vol. 85(1), pages 25-45, March.

Bonfiglioli, A., & Gancia, G. (2013). Uncertainty, electoral incentives and political myopia. The Economic Journal, 123(568), 373-400.

Brender, Adi, & Allan Drazen. (2008). "How Do Budget Deficits and Economic Growth Affect Reelection Prospects? Evidence from a Large Panel of Countries." American Economic Review, 98(5): 2203-20.

Celentani, M, & Conde-Ruiz J. I. (2008). "Politicians career concerns with political uncertainty". Working paper.

Doppelhofer, G., Miller, R. I., & Sala-i-Martin, X. (2000). Determinants of long-term growth: A Bayesian averaging of classical estimates (BACE) approach. (No. w7750). National bureau of economic research.

Doron, J., & Gaudreau, P. 2014. A point-by-point analysis of performance in a fencing match: Psychological processes associated with winning and losing streaks.Journal of Sport and Exercise Psychology,36(1), 3-13.

Fair, R. C. (1978). The effect of economic events on votes for president. The Review of Economics and Statistics, 159-173.

Grilli, V., Masciandaro, D., Tabellini, G., Malinvaud, E., & Pagano, M. (1991). Political and monetary institutions and public financial policies in the industrial countries. Economic policy, 342-392.

Healy, A., & Lenz, G. S. (2014). Substituting the End for the Whole: Why Voters Respond Primarily to the ElectionYear Economy. American Journal of Political Science, 58(1), 31-47.

Hernández, V. (2017c). Economic Determinants of Political Alternation: A Panel Data Analysis of OECD Countries. Working Paper.

Hibbs, Douglas A, Jr, 2000. " Bread and Peace Voting in U.S. Presidential Elections," Public Choice, Springer, vol. 104(1-2), pages 149-80, July.

Hibbs, D. A. (2005). Voting and the Macroeconomy. Centrum fr forskning om offentlig sektor (CEFOS).

Lewis-Beck, M. S. & Paldam, M. 2000. Economic voting: an introduction. Electoral Studies, 19: 113-21.

Lewis-Beck, M. S. & Stegmaier, M. 2000. Economic determinants of electoral outcomes. Annual Review of Political Science, 3: 183-219.

Marshall, M. G., Jaggers, K., & Gurr, T. R. (2010). Polity IV Data Series Version 2010. College Park, MD: University of Maryland. Retrieved from http://www. systemic-peace. org/polity/polity4. htm.

McNeish, D. (2010). On using Bayesian methods to address small sample problems.Structural Equation Modeling: A Multidisciplinary Journal,23(5), 750-773.

Paldam, M. (1991). How robust is the vote function? A study of seventeen nations over four decades. Economics and politics: the calculus of support, 9-31.

Paldam, M., & Nannestad, P. (2000). What do voters know about the economy?: A study of Danish data, 19901993. Electoral Studies, 19(2), 363-391.

Persson, T., & Tabellini, G. (2000). Political economics: explaining public policy.

2.9 Statistical Appendix

A. Additional Tables

Table 2.A11: Detailed sample

Country	Elections in the Sample	Re-election / Total
Australia	1972(0), 1974(1), 1975(0), 1977(1), 1980(1), 1983(0), 1984(1), 1987(1), 1990(1), 1993(1), 1996(0) 1998(1), 2001(1), 2004(1), 2007(0), 2010(1), 2013(0)	11/17
Austria	1971(1), 1975(1), 1979(1), 1983(1), 1986(0), 1990(1), 1994(1), 1995(1), 1999(0), 2002(1), 2006(0) 2008(1), 2013(1)	10/13
Belgium	1971(1), 1977(0), 1978(1), 1981(0), 1985(1), 1987(0), 1991(1), 1995(1), 1999(0), 2003(1), 2007(0) 2010(1), 2014(0)	7/13
Canada	1972(1), 1974(1), 1979(0), 1980(0), 1984(0), 1988(1), 1993(0), 1997(1), 2000(1), 2004(1), 2006(0) 2008(1), 2011(1), 2015(0)	8/14
Chile	1993(0), 2000(1), 2005(1), 2009(0), 2013(0)	2/5
Czech Rep.	1996(1), 1998(0), 2002(1), 2006(0), 2010(0), 2013(0)	2/6
Germany	1976(1), 1980(1), 1987(1), 1990(1), 1994(1), 1998(0), 2002(1), 2005(0), 2009(1), 2013(1)	8/10
Denmark	1971(0), 1975(0), 1977(1), 1979(1), 1981(1), 1984(1), 1987(1), 1988(1), 1990(1), 1994(1), 1998(1) 2001(0), 2005(1), 2007(1), 2011(0), 2015(0)	11/15
Spain	1979(19, 1982(0), 1986(1), 1989(1), 1993(1), 1996(0), 2000(1), 2004(0), 2008(1), 2011(0)	6/10
Finland	1975(0), 1979(1), 1983(1), 1987(0), 1991(0), 1995(0), 1999(1), 2003(1), 2007(0), 2011(1), 2015(0)	5/11
France	1973(1), 1978(1), 1981(0), 1986(0), 1988(0), 1993(0), 1997(0), 2002(0), 2007(1), 2012(0)	3/10
UK	1974(0), 1979(0), 1983(1), 1987(1), 1992(1), 1997(0), 2001(1), 2005(1), 2010(0), 2015(1)	6/10
Greece	1977(1), 1981(0), 1985(19, 1993(0), 1996(19, 2000(1), 2004(0), 2007(1), 2009(0), 2012(0), 2015(0)	5/11
Ireland	1977(0), 1981(0), 1982(1), 1987(0), 1989(1), 1992(1), 1997(0), 2002(1), 2007(1), 2011(0)	5/10
Iceland	1974(0), 1978(0), 1979(0), 1983(1), 1987(1), 1991(0), 1995(1), 1999(1), 2003(1), 2007(1), 2009(0) 2013(0)	6/12
Israel	1977(0), 1981(1), 1984(0), 1988(1), 1992(0), 1996(0), 1999(0), 2001(1), 2003(1), 2009(0), 2013(0) 2015(1)	5/12
Italy	1972(1), 1976(1), 1979(1), 1983(1), 1987(1), 1992(1), 1996(0), 2001(0), 2006(0), 2008(0), 2013(0)	6/11
Japan	1972(1), 1976(1), 1979(1), 1983(1), 1986(1), 1990(1), 1993(0), 2000(1), 2003(1), 2005(1), 2009(0) 2012(0), 2014(1)	11/14
Luxembourg	1974(0), 1979(0), 1984(1), 1989(1), 1994(1), 1999(1), 2004(1), 2009(1), 2013(0)	6/9
Netherlands	1971(1), 1977(0), 1981(1), 1982(1), 1986(1), 1989(1), 1994(0), 1998(1), 2002(1), 2003(0), 2006(1) 2010(0), 2012(1)	9/13
Norway	1977(1), 1981(0), 1985(1), 1989(0), 1993(1), 1997(0), 2001(1), 2005(0), 2009(1)	5/9
New Zealand	1981(1), 1984(0), 1987(1), 1990(0), 1993(1), 1996(1), 1999(0), 2002(1), 2005(1), 2008(0), 2011(1) 2014(1)	8/12
Portugal	1980(0), 1983(0), 1985(0), 1987(1), 1991(1), 1995(0), 1999(1), 2002(0), 2005(0), 2009(1), 2011(0) 2015(0)	4/12
Sweden	1973(1), 1976(0), 1979(0), 1982(0), 1985(1), 1988(1), 1991(0), 1994(0), 1998(1), 2002(1), 2006(0) 2010(1), 2014(0)	6/13
USA	1972(1), 1976(0), 1980(0), 1984(1), 1988(1), 1992(0), 1996(1), 2000(0), 2004(1), 2008(0), 2012(1)	
TOTAL		158/284

Table 2.A12: List of excluded countries

Country	Excluded elections[a]	Available elections
Estonia	1990 (first), 1992 (data), 1995 (data)	1999, 2003, 2007, 2011, 2015
Hungary	1990 (first), 1994 (data)	1998, 2002, 2006, 2010, 2014
Latvia	1990 (first), 1993 (data), 1995 (data)	1998, 2002, 2006, 2010, 2011
Mexico	1976 (demo), 1982 (demo), 1988 (demo), 1994 (demo)	2000, 2006, 2012
Poland	1990 (first)	1995, 2000, 2005, 2010, 2015
Slovakia	1990 (first), 1992 (data), 1994 (data)	1998, 2002, 2006, 2010, 2012
Slovenia	1990 (first), 1992 (data), 1996 (data)	2000, 2004, 2008, 2011, 2014
South Korea	1978 (demo), 1979 (demo), 1980 (demo), 1981 (demo), 1987 (first)	1992, 1997, 2002, 2007, 2012
Switzerland[b]	-	-
Turkey[c]	-	-

[a] Exclusion reasons: (1) first democratic election (first), (2) the "Polity Democracy Index" indicates lack of democracy (demo) and (3) data inavailability (data).

[b] Excluded due to its complex executive power.

[c] Excluded because of its low per capita income.

Figure 2.A6: Correlation between "Sensitivity" and "long-run GDP Growth" (Model 2)

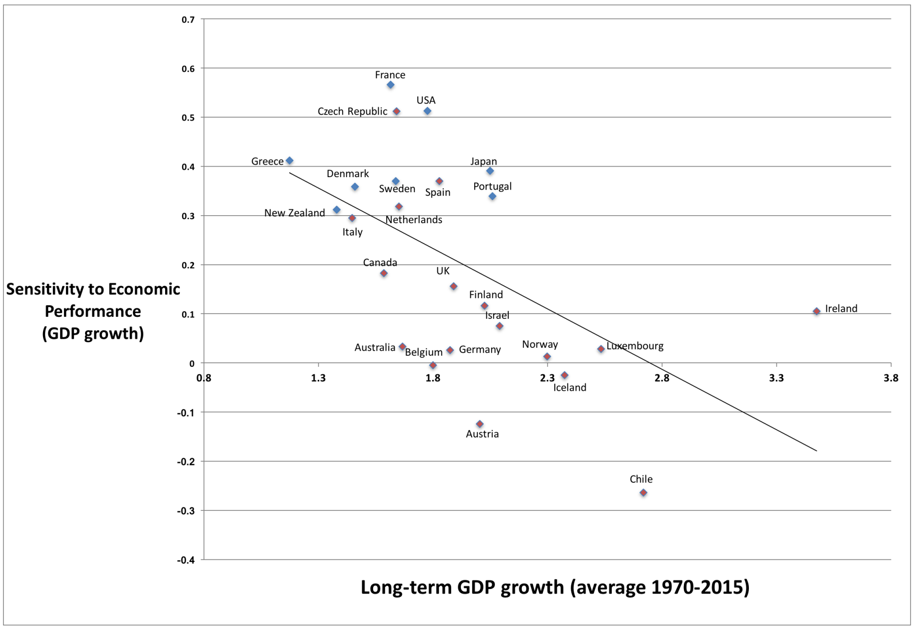
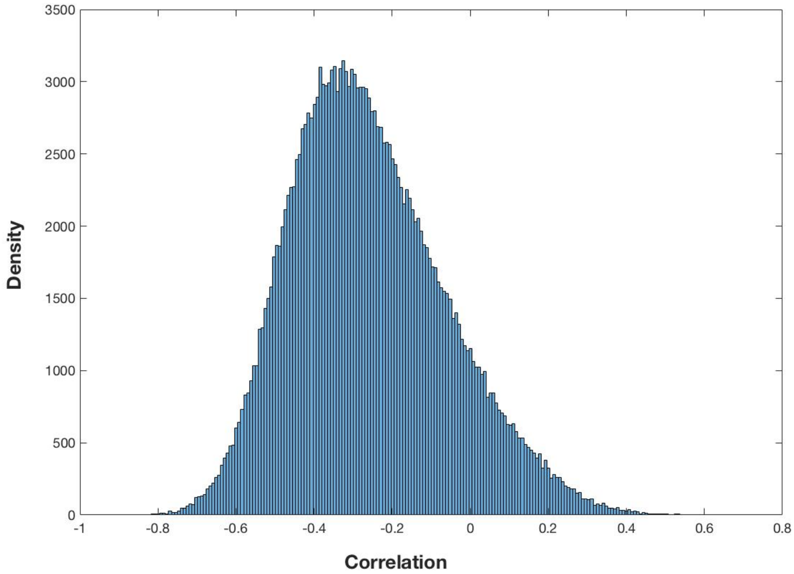


Figure 2.A7: Histogram of the correlation (Model 2)



Median of the EDF: -0.2512, Mean of the EDF: -0.2482
Prob(corr<0)=90%

B. Databases

- Database of Political Institutions (DPI), World Bank.
- International Financial Statistics (IFS9, International Monetary Fund.
- World Development Indicators (WDI), World Bank.
- Political Finance Database, Institute for Democracy and Electoral Assistance (IDEA).
- Polity IV, University of Maryland, Center for International Development and Conflict Management.
- Economic Outlook Database (OCDE).
- World Economic Outlook (WEO), International Monetary Fund.
- A Historical Public Debt Database, International Monetary Fund.
- World Political Leaders, Zárate's Political Collections

C. Variable Definitions

DEPENDENT VARIABLE: ReelectParty.

The definition is as follows:

a) Case 1: The old and the new government are formed by a single party.

-ReelectParty=1 if the incumbent party wins the elections

-ReelectParty=0 if the incumbent party is defeated

b) Case 2: The old government was formed by a single party and the new government is formed by a coalition of parties.

-ReelectParty=1 if the incumbent party is member of the coalition formed after the elections and has more than 60% of the seats in the new coalition.

-ReelectParty=0 otherwise

c) Case 3: The old government was formed by a coalition of parties and the new government is formed by a single party.

-ReelectParty=1 if the winning party was part of the ruling coalition before the election and had more than 60% of the seats.

-ReelectParty=0 otherwise

d) Case 4: The old and new government are formed by a coalition of parties.

-ReelectParty=1 if the parties that were members of the previous coalition have more than 60% of the seats in the new coalition and the parties who are members of the new coalition had more than 60% of the seats in the previous coalition

-ReelectParty=0 otherwise

SOURCES: The variable "ReelectParty" is constructed from "Database of Political Institutions" (DPI). This database provides annual data for the period 1975-2015 (updated January 2017). We used the following DPI variables:

- SYSTEM: Parliamentary (2), Assembly-elected President (1), Presidential (0) [21]
- TOTALSEATS: total seats of the parliament.
- To identify when presidential/parliamentary elections were held: LEGELEC: 1 if there was a parliamentary election in this year (for parliamentary countries), DATELEG: month when parliamentary elections were held, EXELEC: 1 if there was a presidential election in this year (for presidential countries), DATEEXEC: month when presidential elections were held.
- To identify the largest government party each year: GOV1ME (name of the party), GOV1SEAT (seats)
- To identify the 2nd government party: GOV2ME (name of the party), GOV2SEAT (seats)

[21] There are only three presidential countries in the sample (USA, France & Mexico)

- To identify the 3rd government party: GOV3ME (name of the party), GOV3SEAT (seats)
- To identify other government parties: GOVOTH (number of parties), GOVOTHST (total seats)
- To identify the largest opposition party each year: OPP1ME (name of the party), OPP1SEAT (seats)
- To identify the 2nd opposition party: OPP2ME (name of the party), OPP2SEAT (seats)
- To identify the 3rd opposition party: OPP3ME (name of the party), OPP3SEAT (seats)
- To identify other opposition parties: OPPOTH (name of the party), OPPOTHST (seats)

WGDPpcGrowth

The definition is as follows:

-WGDPpcGrowth: Weighted real per capita GDP growth rate over the legislature.

$$WGDPpcGrowth = \sum_{i=1}^{n} \lambda^i * \Delta GDP_{-i} * (1/\sum_{i=1}^{n} \lambda^i) \tag{2.7}$$

where, ΔGDP_0 is the real GDP per capita growth in the month in which the election was held; n is the number of months of the term in office; GDP_{-n} is the real GDP per capita growth in the first month of the term in office; λ is the discount rate (monthly)

EXAMPLE: USA (2012 presidential election)

-Date (2008 election): November 4, 2008
-Date (2012 election): November 6, 2012
-Duration of the term in office: 47 months (excluding November 2008 & November 2012)

Table 2.A13: Example: USA Presidential Election (2012)

	2008	2009				2010				2011				2012				
	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	
GDPpcGrowth	-9.7%	-6.1%	-1.2%	0.6%	3.2%	1.5%	1.4%	1.9%	1.6%	-0.6%	1.7%	0.4%	3.2%	1.1%	0.4%	2.2%	-0.5%	
Number of months	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	1	**47**

Table 2.A14: Values of ΔGDP_{-i}

	ΔGDP_{-47}	ΔGDP_{-46}	ΔGDP_{-45}	ΔGDP_{-44}	ΔGDP_{-43}	...	ΔGDP_{-5}	ΔGDP_{-4}	ΔGDP_{-3}	ΔGDP_{-2}	ΔGDP_{-1}
Value	-9.7%	-6.1%	-6.1%	-6.1%	-1.2%	-	0.4%	2.2%	2.2%	2.2%	-0.5%
Month	Dec-2008	Jan-2009	Feb-2009	Mar-2009	Apr-2009		Jun-2012	Jul-2012	Aug-2012	Sep-2012	Oct-2012

OTHER VARIABLES: Democracy

-Democracy: Dummy variable that measures the existence of democracy. This variable is equal to 1 if the polity2 variable (Polity IV database) is positive and is equal to 0 otherwise.

SOURCES: The variable "Democracy" is constructed from "Polity IV Database" (University of Maryland).

A Market-based Algorithm for Predicting Soccer Outcomes

Víctor Hernández García

November 2017

Resumen en Castellano

El principal objetivo de este estudio es proponer un algoritmo que utiliza datos procedentes de los mercados predictivos para predecir los resultados de los partidos de fútbol. Los mercados predictivos son un campo de pruebas adecuado para testar la validez de nuevas técnicas estadísticas, ya que todos los participantes observan la información en tiempo real y con pocas asimetrías, lo eventos tienen una fecha clara de terminación y el resultado no esta influido por la propia actividad de los mercados, como si sucede en el mundo de las finanzas. En este trabajo, se propone un modelo de Poisson con intensidad constante en el tiempo y se utilizan datos de Betfair Exchange. Al mismo tiempo, varios métodos de estimación son propuestos: OLS, OLS ponderado y modelos bayesianos jerarquizados. La capacidad de predicción de los modelos es evaluada utilizando la medida De Finetti, que es definida por la distancia euclídea entre la predicción y la realización. Los resultados indican que el modelo propuesto tiene muy buenas capacidades predictivas, siendo capaz de vencer a las predicciones de los mercados, salvo en las últimas jornadas del campeonato. Al mismo tiempo, presenta mejores resultados que los modelos tradicionales en la literatura, los cúales se basan en resultados pasados. Por lo tanto, es posible concluir que el uso de información de los mercados predictivos permite obtener estimaciones más precisas de las distribuciones de probabilidad de los eventos futuros.

Palabras clave: Predicción, Mercados Predictivos, Métodos Bayesianos, MCMC
JEL classification: C5 C6 D4 D8

A Market-based Algorithm for Predicting Soccer Outcomes *

Víctor Hernández García†

November 2017

Abstract

The main objective of this study is to develop a market-based forecasting algorithm for predicting soccer outcomes. A football model with defensive and offensive parameters is estimated using information from prediction markets (Betfair Exchange Market). Prediction markets are an interesting testing ground to test the validity of new statistical techniques, since all participating agents observe the relevant information without asymmetries, the events have a clear completion date and the final outcome is not influenced by market activity. A time-homogeneous independent Poisson model is estimated applying the following techniques: OLS, Weighted OLS and hierarchical bayesian methods. The prediction capability of the proposed methodology is tested using "De Finetti" measure, which is defined as the Euclidean distance between the prediction and the realisation. The results obtained indicate that the models have good predictive properties, being able to improve market predictions, except in the last weeks of the season. Moreover, they also generate better predictions than models based on past performance. So, it can be concluded that the use of information from prediction markets can lead to more accurate estimates of the probability distribution of future events.

Keywords: Forecasting, Prediction Markets, Bayesian Methods, MCMC
JEL classification: C5 C6 D4 D8

*I thank José Ignacio Conde Ruiz and Marco Celentani for their help and support. I gratefully acknowledge the support from the Spanish Ministry of Science and Technology from grants ECO2011-30323-C03-01 & BES-2012-061597..

†E-mail: victorhzgarcia@gmail.com

3.1 Introduction

The main objective of this study is to propose a market-based methodology for predicting the outcome of future events. High-frequency data from prediction markets are used as primary information. A prediction market is created for the purpose of trading the outcome of events, so equilibrium prices can be interpreted as the market belief about the probability of the different events. In these markets, the payoff is a binary option, in other words, agents get a fixed amount of money or nothing at all. Some examples of prediction markets are: futures markets, stock markets, political markets and (sport) betting markets.

Our initial hypothesis is that the use of information from prediction markets can improve the prediction capability of the models. Then we will examine whether prediction markets data can provide useful information about the probability distribution of future events.

In this work, we focus on sport betting markets instead of financial markets mainly for the following reasons. Firstly, the events have a clear completion date and the outcome is perfectly observed by all agents. Unlike financial markets, in betting markets the final outcome is not influenced by market activity and all participating agents observe the relevant information without asymmetries. Finally, betting markets are easily accesible and have a large volume of transactions. For the reasons given before, betting markets are an interesting testing ground to test the validity of new models and statistical techniques. To the best of our knowledge, none of the models proposed so far in the sport economics literature have used information from these markets.

While we will use the Spanish Football League (LIGA) data as an illustrative example, the main goal of this paper is to lay out the methodology to systematically carry out inference from predictive markets data. This framework could be applied to a variety of structural models, sports and data sets.

The remainder of this paper is organized as follows. Firstly, in Section 3.2, a review of the sport economics literature is presented. In Section 3.3, we describe our database, which include high-frequency betting data from Betfair Exchange. In section 3.4, we introduce and justify the five assumptions we will use to develop the models. On the one hand, we just incorporate some of the classical assumptions used in the literature, such as time-homogeneous Poisson and independence assumptions. On the other hand, we include some additional refinements, proposing an hypothetical partitioning of goal scoring intensity and allowing for fluctuations in team's performance over the season. Section 3.5 describes the methodology and the models. Several estimation methods are proposed: (1) (weighted) OLS, (2) hierarchical bayesian model with normal random shocks and (3) hierarchical bayesian model with autoregressive shocks. In Section 3.6, these methods are applied to calculate predictions for single matches and to simulate Liga 2013-2014. The predictive power of the models, relative to Market predictions and models based on past performance, is tested using De Finetti measure. Finally, in Section 3.7, conclusions are presented.

3.2 Literature review

Most studies on "modelling soccer data" are based only on "historical outcomes". Maher (1982) uses a poisson model for predicting football scores using past performance. An iterative ML (Newton Raphson) is used to estimate the following parameters: home defense, home attack, away defense and away attack. Dixon & Coles (1997) proposes a poisson dynamic model that accounts for fluctuations in performance of individual teams. The model gives relatively more weight to the most recent results. Karlis & Ntzoufras (2003) relaxes the independent assumption and use a bivariate poisson distribution to predict soccer and water-polo scores using a expectation-maximization algorithm. Brillinger (2008) applies a trinomial model in order to model win, draw and loss probabilities using the outcome results. Everson & Goldsmith-Pinkham (2008) proposes a bayesian hierarchical framework with Poisson additivity and home-field advantages to predict goal scoring. Finally, Karlis & Ntzoufras (2009) has employed the Skellam's distribution to model the goal difference between home and away teams.

Some studies incorporate more information to the models, apart from the scores of matches already played. Dyte & Clark (2000) presents a log-linear Poisson for predicting the distribution of scores in international soccer matches taking the FIFA ratings as covariates. Bueno et al. (2010) proposes a bayesian methodology for predicting soccer outcomes based on experts' opinions and past performance. Langseth (2013) extends the classical soccer models by taking a vast amount of data into account, including fired shots, shots on target, etc. Then a data intensive forecasting model is used in order to beat the bookie.

Several articles refer to betting markets when modelling soccer data. Dixon & Robinson (1998) proposes a two-dimensional birth process[1] to investigate setting prices in the spread betting market. Their model exploits only each teams' past goal scores and the goal times within each match. Kuypers (2010) introduces a model of bookmaker behaviour and look for profitable opportunities in soccer betting markets using an Ordered Probit.

However, to the best of our knowledge, none of the proposed models use predictive markets (betting data) as input to estimate the models. When modelling the stochastic processes that determine the goals scored by the different teams, we think that competitive markets provide a more complete information than the final scores that, after all, are a simple realisation of each stochastic process. The outcome of a particular match may depend greatly on the luck and not really on the quality of both teams. Therefore, the use of high-frequency information from predictive markets is the main contribution of this study to the literature.

Finally, this article is also very related to Prediction Markets literature (See Wolfers & Zitzewitz, 2006 for a review), which has grown significantly over the last years taking advantage of the availability of detailed data sets.

[1] A birth-death process is a Markov process with only two types of state transitions: "births" (state variable increases by one) and "deaths" (state variable decreases by one). See Moller & Sorensen (1994).

3.3 Data.

Betting markets have experienced a remarkably growth during the last decades principally due to deregulation, globalization and the emergence of online gambling. At present, according to the price setting mechanism, two types of betting markets can be distinguished: betting exchanges and fixed-odds markets.

Functioning of betting exchanges is quite similar to financial markets. In these markets, agents exchange odds in real time on sporting events and such transactions generate a competitive price at any given time. It is posible to buy and to sell an outcome, allowing backing and laying strategies. Bookmakers are simply intermediaries who fulfil the mission of bringing the different agents into contact and they generate revenue charging commissions on net winnings. In some cases, bookmakers use betting exchanges to hedge liabilities.

At the same time, in fixed-odds markets, bookmakers set prices to maximize profits and to achieve a balanced book. In this case, bookmakers offer odds and usually generate revenue by setting unfair odds (the sum of the implied probabilities is always greater than 1). In this market, punters have to bet against the bookmaker.

In this work, we only focus on betting Exchange for several reasons. Firstly, we are interested in competitive equilibrium prices because we are trying to take advantage of the market consensus. Secondly, odds are not altered by commissions as in the fixed-odds market case. Thirdly, it is possible to monitor the behaviour of both sides of the market (back and lay). Finally, in betting exchanges there are no maximum bet limits.

Below is a summary of the data (Table 3.1). Our dataset include fully time-stamped historical Betfair Exchange Market price data[2]. The database covers all 380 matches of the Liga 2013-2014 and includes pre-play and in-play data for several markets (Match Odds[3], Correct Score[4], Over-Under betting, Half Time Score, Next Goal...). The Liga 2013-2014 is the latest available, as in November 2014 Betfair updated the application for downloading data[5]. The frequency is one observation per second and the following information is included for each selection:

- Best three back odds and its volume.
- Best three lay odds and its volume.
- Last price matched.
- Total matched (including both back and lay transactions).
- The status of the market (active or suspended).
- Timestamp.

[2]The data was downloaded from www.fracsoft.com

[3]Match odds is a 3-way market where punters can bet either home, away or draw.

[4]Correct Score is a 17-way market where punters can bet on the correct score (0-0,1-0,0-1...)

[5]Betfair API 6.0

Table 3.1: Descriptive Statistics (Betfair Exchange)

	Match Odds			Correct score			Other markets [a]		
	Total	**Mean**	**Median**	**Total**	**Mean**	**Median**	**Total**	**Mean**	**Median**
Observations (Total) [b]	22.8	0.6	0.4	136.8	0.36	0.16	50	0.13	0.08
Observations (Pre-play)	15.5	0.4	0.31	93	0.24	0.18	34	0.09	0.07
Observations (In-play)	7.3	0.2	0.15	43.8	0.12	0.10	16	0.04	0.03
Observations (Active)	20.1	0.53	0.52	126	0.33	0.33	45	0.11	0.12
Total Matched [c]	802	2.1	0.5	106	0.28	0.19	76	0.2	0.2
No. of matches	380			380			380		

[a]They include: half time correct score, over/under markets, correct score away, correct score home, next goal and both teams to score.

[b]The data are expressed in millions of observations.

[c]The data are expressed in millions of euros.

- Pre-play or in-play

It can be observed that match odds markets are the most liquid ones, matching 2.1 millions of euros per match on overage. At the same time, correct score markets rank in the second place, according to the number of bets placed, with 0.28 millions of euros per match on average. Finally, the rest of the markets are much less liquid, so in these markets competitive prices do not reveal much information.

3.4 Model Assumptions.

The objective of this section is to introduce the five assumptions we will used to develop the forecasting models. Before that, some notation is required. In a match between teams indexed i (home team) and j (away team), we define the following variables:

- Let $X^H_{i,j}$ represents the goals scored by team i against team j, or equivalently, the goals conceded by team j against team i.
- Let $X^A_{i,j}$ represents the goals conceded by team i against team j, or equivalently, the goals scored by team j against team i.

3.4.1 The Poisson Assumption:

Assuming Poisson distributions for home and away goals is a classical assumption in the literature for predicting football scores (Everson & Goldsmith-Pinkham, 2008; Bueno et al., 2010). The Poisson distribution is a member of the exponential family and is actually a limiting case of a Binomial distribution when the number of trials grows and the success probability tends to zero. In a Poisson distribution, interarrival times follow a exponential distribution. Poisson distribution has a number of advantages which explain its wide use in empirical works, among them its simplicity (it is specified only for one parameter) and analytical tractability.

Before assuming the Poisson Assumption, some tests are carried out to check whether our data is well-approximated by a simple Poisson distribution.

Testing the Poisson Assumption:

i) **Variance/mean ratio test**. One of the most important characteristic of the Poisson distribution is that the variance equals the mean. We can use this property to check the validity of the Poisson assumption. According to (Simonoff, 2013), a comparison of the sample mean and variance would provide a valid way to check the validity of the Poisson assumption, given the inflation of variance of the negative binomial and other mixed-Poisson distributions. Following the methodology described in Karlis & Ntzoufras (2000), we calculate the variance to mean ratio[6] for all 20 Liga teams (Figure 3.1). Note that 90-degrees line correspond to Poisson distribution, while points above the line are indications of overdispersion relative to the Poisson distribution. It can be observed that 12 of the 20 points (60%) are above the line, as their variance to mean ratio is greater than one[7]. Graphically, it is not possible to detect any pattern of deviation. For instance, under a negative binomial distribution, the variance is a linear function of the mean. In the case of the Poisson Inverse Gaussian, the variance is a quadratic function of mean.

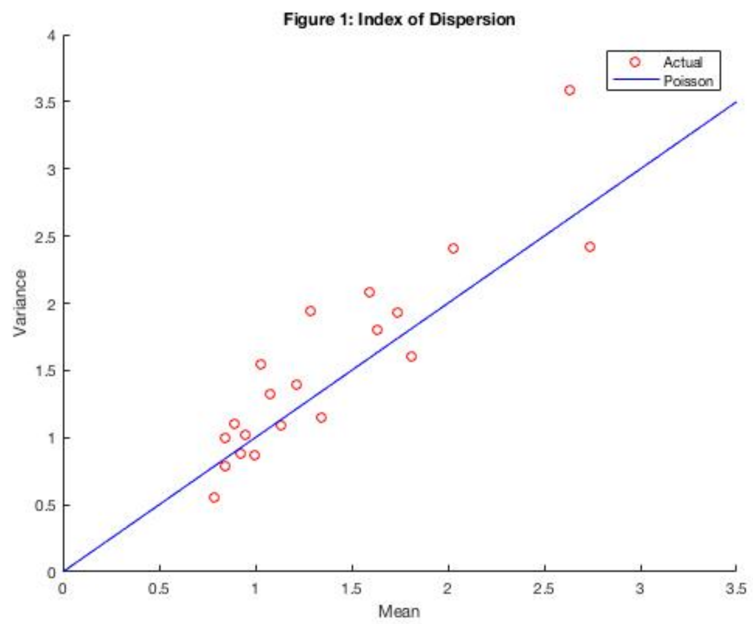


Figure 3.1: Index of Dispersion (goal variance / goal mean)

Now, using the estimations of the variance to mean ratios, we perform a test to check if there are evidence against the Poisson distribution. Under the null hypothesis (i.e. the data follow a Poisson distribution) we expect that half of the teams present underdispersion and the other half present overdispersion. Then we can compute a classical Proportion Test[8] using this information. The confident interval includes 0.5 and p-value equals 0.37, so we can not reject the Poisson Assumption.

[6] Also known as index of dispersion.

[7] Karlis & Ntzoufras (2000) obtained a similar proportion. In their study the 58.1% of the teams showed overdispersion

[8] In Stata, the command is prtest

ii) **Rootogram**. Everson & Goldsmith-Pinkham (2008) proposes an alternative approach to check the validity of the Poisson Assumption. Following the methodology described in their article, we develop a Rootogram of Liga 2013-2014 goal scoring. Firstly, we compute the square roots of the counts of individual goals for the 380 matches played during Liga 2013-2014 (760 observations). Secondly, we compute the square roots of the expected Poisson counts for 760 draws from a theoretical Poisson distribution (with parameter equal to 1.357, i.e., the goal average during the season). Results are displayed using a Rootogram (Figure 3.2)[9]. In general, we can conclude that our soccer data are consistent with a Poisson distribution, although there are slight deviations between both rootograms. On the one hand, there are slight evidence that the true goal distribution has more density at zero. Several studies (See Karlis & Ntzoufras (2003)) obtained the same conclusion, proposing the use of a zero-inflated poisson distributions instead of a Poisson distribution. At the same time, the actual count for seven goals is greater than the theoretical expected count. In both cases, deviations do not seem significant.

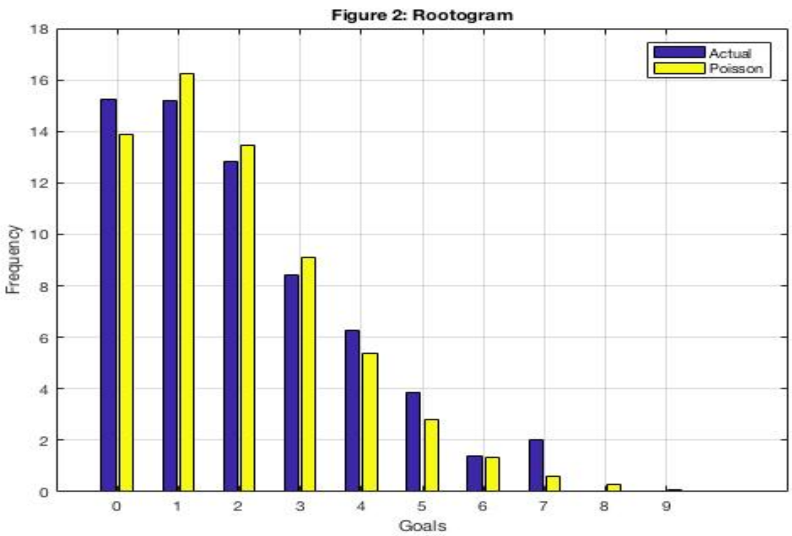


Figure 3.2: Rootogram

After ruling several tests, it seems that the Poisson distribution is appropiate and fits the data well. So, henceforth we assume that in each match home-team and away-team goals follow a Poisson distribution with parameters $\lambda^H_{i,j}$ and $\lambda^A_{i,j}$, respectively.

$$\text{Assumption 1:} \quad X^H_{i,j} \sim Poisson(\lambda^H_{i,j} * t) \quad ; \quad X^A_{i,j} \sim Poisson(\lambda^A_{i,j} * t) \tag{3.1}$$

In the above expression t stands for "match time". This variable measure the percentage of the time[10] elapsed from the beginning of the match and ranges from 0 (the start of the match) to 1 (the end of te match). Then in the middle of a match, t equals 0.5.

[9] In the Appendix, results are shown using home and way team goal distribution (figures 3.A13 and 3.A14), obtaining similar conclusions.

[10] Given as a fraction of unity.

3.4.2 The Independence Assumption:

The inclusion of this assumption greatly reduces the complexity of the models and reduces the computation times of the inference process. In the literature, the decision of including the assumption of independence when modelling soccer data is widely debated. On the one hand, some authors include this assumption (Keller (1994), Ridder et al. (1994), Everson & Goldsmith-Pinkham (2008), Suzuki et al. (2010)). On the other hand, other authors suggest the use of more flexible models that allow to model possible dependencies. Maher (1982) suggests the Bivariate Poisson distribution. Dixon & Coles (1997) obtained that the assumption of independence between scores is reasonable (except for the scores 0-0, 1-0, 0-1 and 1-1) and proposed a bivariate inflation distribution. Karlis & Ntzoufras (2000) found evidence in favour of the dependence of the two variables, although in magnitude is very small, and proposes a Negative Binomial model.

The "Pearson Chi-square" independence test is then carried out using the final scores of the 380 matches corresponding to the League 2013-2014, in order to identify the existence of possible dependencies and to assess whether the assumption of independence is appropriate. The Pearson Chi-square test is able to check the association between two nominal (categorical) variables. The null hypothesis states that knowing the level of Variable X does not help you predict the level of Variable Y. That is, the variables are independent. At the same time, the alternative hypothesis is that knowing the level of Variable X can help you predict the level of Variable Y. Figure 3.3 presents a two-way contingency table for scores and the "Pearson Chi-square" independence test:

	Away_team						
Home_team	0	1	2	3	4	5	Total
0	32	33	19	7	2	3	96
1	41	31	27	6	3	1	109
2	24	26	22	9	2	2	85
3	18	17	7	1	1	0	44
4	12	11	5	3	0	0	31
5	6	3	0	0	0	0	9
6	1	1	0	0	0	0	2
7	3	0	0	1	0	0	4
Total	137	122	80	27	8	6	380

Pearson chi2(35) = 25.7251 Pr = 0.874

Figure 3.3: Pearson Chi-square test for independence

The test finds no evidence against the assumption of independence between both variables (p-value is much greater than 0.05). In view of the results, from now on we apply the Independence Assumption:

$$\text{Assumption 2:} \qquad X_{i,j}^{H}(t) \quad \perp \quad X_{i,j}^{A}(t) \qquad i \neq j \tag{3.2}$$

3.4.3 Additive Stochastic Process Assumption: Composite Poisson.

Additive Property of Poisson Distributions: Let $N_1(t)$ and $N_2(t)$ be two independent Poisson processes with parameters λ_1 and λ_2, respectively, and let $N(t) = N_1(t) + N_2(t)$. It follows that $N(t)$ is a Poisson process with parameters $\lambda_1 + \lambda_2$.

Using the additive property, as in (Everson & Goldsmith-Pinkham, 2008), we propose a hypothetical partitioning (Assumption 3) to model both team total scores, $X_{i,j}^H(t)$ and $X_{i,j}^A(t)$, as the sum of two independent poisson processes. Then we are assuming the following Composite Poisson model:

Home team (i) goals:

$$X_{i,j}^H(t) = O_i^H(t) + D_j^A(t) \tag{3.3}$$

- Offensive parameter (home team): $O_i^H(t) \sim Poisson(\theta_i^H * t)$
- Defensive parameter (away team): $D_j^A(t) \sim Poisson(\delta_j^A * t)$

Away team (j) goals:

$$X_{j,i}^A(t) = O_j^A(t) + D_i^H(t) \tag{3.4}$$

- Offensive parameter (away team): $O_j^A(t) \sim Poisson(\theta_j^A * t)$
- Defensive parameter (home team): $D_i^H(t) \sim Poisson(\delta_i^H * t)$.

Note that it is required, $\theta_i^H > 0$, $\delta_j^A > 0$, $\theta_j^A > 0$ and $\delta_i^H > 0$.

We also assume:

$$O_i^H(t) \perp D_j^A(t) \tag{3.5}$$

$$O_j^A(t) \perp D_i^H(t) \tag{3.6}$$

Applying Assumption 2, we get:

$$O_i^H(t) \perp D_j^A(t) \perp O_j^A(t) \perp D_i^H(t) \tag{3.7}$$

Applying the additive property of poisson distributions, we get the following equalities, which will greatly facilitate inference:

$$\lambda_{i,j}^H = \theta_i^H + \delta_j^A \tag{3.8}$$

$$\lambda_{i,j}^A = \theta_j^A + \delta_i^H \tag{3.9}$$

3.4.4 Time-homogeneous Poisson Process:

There is strong evidence that the probability of a goal being scored steadily increases over the course of the match, perhaps because of tiredness of players (Dixon & Robinson (1998)). Figure 3.4 shows the sample cumulative distribution of goals (Liga 2013-2014).

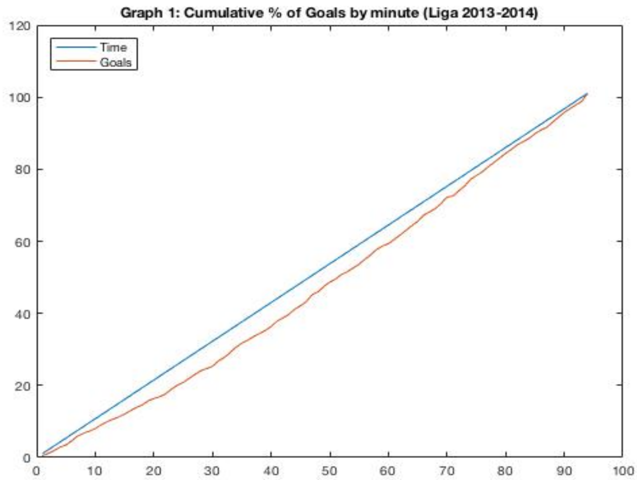


Figure 3.4: Sample cumulative % of goals by minute (LIGA 2013-2014)

Nevertheless, in this study we assume, for simplicity, that the goal score intensity does not depend on the time[11].

$$\text{Assumption 4:} \qquad \lambda_{i,j}^{H}(t) = \lambda_{i,j}^{H} \quad \forall t; \qquad \lambda_{i,j}^{A}(t) = \lambda_{i,j}^{A} \quad \forall t \tag{3.10}$$

3.4.5 Team quality:

In the literature, there is an intense debate about how to model team quality fluctuations. Some authors assume, generally for simplicity, that the team quality is constant over the season. At the same time, other studies allow quality fluctuations. Dixon & Coles (1997) accounting for fluctuations in performance of individual teams. Bueno et al. (2010) also assumes that team quality is continually evolving. In our case, we will use both assumptions depending on the models:

Assumption 5. i: Parameters (offensive and defensive) are constant over the season. This assumption is less realistic but simplifies calculations and computation times. We will use this assumption in the models presented in section 3.5.2.1 (OLS estimation) and 3.5.2.2 (Hierarchical Bayesian Model with random shocks).

Assumption 5. ii: Parameters (offensive and defensive) change throughout the season. This assumption is more flexible and realistic, because it allows fluctuations in "the quality" of individual teams. This assumption will be used in section 3.5.2.3 (Hierarchical Bayesian Model with autoregressive shocks).

[11] A refinement is introduced in Hernandez (2017b) to adjust match time according to historical goal distribution.

3.5 Methodology. Parameter inference.

In this section we define a methodology to estimate the offensive and defensive parameters (quality indicators) of the different teams using information from prediction markets. Remember that for each team we need to estimate a total of four parameters: home attack, home defense, away attack and away defense:

The inference process consists of two phases. In the first phase, we use information from the betting markets to estimate the parameters ($\lambda's$) that characterise the Poisson processes in each of the 380 games of the season. Since there are two stochastic processes in each match, one for each team, we need to obtain a total of 760 values. It should be noted that in the rest of studies, this phase does not call for any calculation, since historical results (or expert opinions) are directly used as primary information. In the second phase, the information collected in the first phase is used to estimate the offensive and defensive capacity of the teams ($\theta's$ and $\delta's$).

3.5.1 Step 1: Obtaining input data ($\hat{\boldsymbol{\lambda}}^H_{i,j}$, $\hat{\boldsymbol{\lambda}}^A_{i,j}$)

First, it is necessary to calculate the inverse of the betting odds in order to obtain a probability measure. It is easy to check that for any match the sum of the odds of the three possible outcomes (home win, draw or away win) is more than 1, indicating that odds are not fair [12]. Next, we proceed to normalise the odds, so that the sum is equal to 1 (100%).

At this point we can already estimate the parameters. In this part, we make use of: Assumption 1 (Poisson distribution), Assumption 2 (independence) and Assumption 3 (Time-homogeneous Poisson distribution). We will also need the Probability Mass Function (PMF)[13] for a Poisson distribution, defined as:

$$\text{P(k events in the interval)} = \frac{e^{-\lambda}\lambda^k}{k!} \tag{3.11}$$

where λ is the average number of events per interval and k takes values 0, 1, 2...

If we apply this formula to the soccer context, at the beginning of a match the probability of observing a particular final score (k,m) is defined by the following joint mass probability function:

$$\text{P(team i scores k goals, team j scores m goals)} = \frac{e^{-(X^H_{i,j})}(X^H_{i,j})^k}{k!} \times \frac{e^{-(X^A_{i,j})}(X^A_{i,j})^m}{m!} \tag{3.12}$$

where $X^H_{i,j}$ and $X^A_{i,j}$ are the goal scoring parameters of team i and j, respectively, when playing at team i home. k takes values 0, 1, 2... and m takes values 0, 1, 2...

[12] In fact, sample mean is about 1.018

[13] Discrete equivalent of a pdf.

Notation: In every moment of a match, let w be the current home team score and q the current away team score. Let t be match time (% of total time elapsed). Let Θ_H^t be the set of possible final home team scores conditioned to the current score (at time t). Analogously, let Θ_A^t be the set of possible final away team scores conditioned to the current score (at time t). For example, if the score at time t is (w=1, q=2), $\Theta_H^t = [1, 2, ..., \infty]$ and $\Theta_A^t = [2, 3, ..., \infty]$. Let $\hat{\boldsymbol{\lambda}}_{i,j,t}^H$ be the estimated goal scoring parameter of team i against team j in time t. Let $\hat{\boldsymbol{\lambda}}_{i,j,t}^A$ be the estimated goal scoring parameter of team j against team i at time t. Let $a_{i,j}^t$ and $b_{i,j}^t$ be the Betfair Exchange implied probability of draw and local victory at time t, respectively.

Using the implicit probabilities of Betfair Exchange and the PMF formula, we can solve the following implicit system of equations[14] for every observation[15] of every match:

Equation 1:

$$0 = \sum_{k \subseteq \Theta_A^t} \sum_{\substack{m=k \\ m \subseteq \Theta_H^t}} \frac{e^{-(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^H}{1-t})}(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^H}{1-t})^{(k-w)}}{(k-w)!} \times \frac{e^{-(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^A}{1-t})}(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^A}{1-t})^{(m-q)}}{(m-q)!} - a_{i,j}^t \qquad \substack{t \subseteq [0,1] \\ i=j=1,...,20} \tag{3.13}$$

Equation 2:

$$0 = \sum_{k \subseteq \Theta_A^t} \sum_{\substack{m<k \\ k \subseteq \Theta_A^t}} \frac{e^{-(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^H}{1-t})}(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^H}{1-t})^{(k-w)}}{(k-w)!} \times \frac{e^{-(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^H}{1-t})}(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^H}{1-t})^{(m-q)}}{(m-q)!} - b_{i,j}^t \qquad \substack{t \subseteq [0,1] \\ i=j=1,...,20} \tag{3.14}$$

The system is perfectly identified, so every pair ($a_{i,j}^t$ and $b_{i,j}^t$) identify one and only one pair ($\hat{\boldsymbol{\lambda}}_{i,j,t}^H$, $\hat{\boldsymbol{\lambda}}_{i,j,t}^A$). Solving all the systems of equations, we find the pair of unknown ($\hat{\boldsymbol{\lambda}}_{i,j,t}^H$, $\hat{\boldsymbol{\lambda}}_{i,j,t}^A$) for every second of every match, so we have a entire distribution for every match and every team (ranging from 5600 to 5800 depending of the match duration).

From $\hat{\lambda}_t$'s to $\hat{\lambda}$'s: Once we have estimated ($\hat{\boldsymbol{\lambda}}_{i,j,t}^H$, $\hat{\boldsymbol{\lambda}}_{i,j,t}^A$), $\forall$ t $\subseteq$ [0,1], we have to find the way to summarize all this information (the entire distribution) into a single value. Three possible methods are considered:

- Value at the beginning of the match. A simple option is to use match odds at the beginning of the match. But, with this method we are losing information, because we are not considering the entire distribution.
- Median value along all the match. A second option is to use the odds of the entire match and calculate the median value of the parameters. Its main advantage is that it collects information from the entire match. In turn, it has the disadvantage of being somewhat conditioned to the circumstances of the match, such as strategic playing, red cards, injuries, refereeing, etc.

[14] We solve the implicit systems of equations using FSOLVE command in Matlab.

[15] The frequency is one observation per second.

- The median value between the beginning of the match and 80th minute (truncated median). This is an intermediate option that excludes the final minutes of the match, in which the strategic factor can be very significant.

In principle, there are reasons to consider that the last option is the most appropriate one, since it presents certain advantages over the others, so it will be the measure used throughout the article. However, all other methods will be used as an indicator of robustness in order to verify that the conclusions are maintained.

Finally, we have the input data ($\hat{\boldsymbol{\lambda}}^H_{i,j}$, $\hat{\boldsymbol{\lambda}}^A_{i,j}$, i=j=1,...,20) that we were looking for. In the next stage, we will use this information to estimate the models instead of past scores, FIFA rankings or experts' opinions, the primary data normally used in the soccer literature.

3.5.2 Step 2: Estimation Methods (Inference)

The goal of this section is to estimate the four quality parameters: home attack, home defense, away attack and away defense. Several estimation methods are introduced. Firstly, we use classical methods: ordinary least squares and weighted ordinary least squares. Secondly, hierarchical bayesian methods are used. Bayesian inference uses prior probability distributions and data to estimate the posterior probability distributions, rather than a point estimate, as in classical models. Below, econometric models are described:

3.5.2.1 (Weighted) Ordinary Least Squares (OLS)

<u>Econometric model</u>:

$$\text{Home team:} \qquad \hat{\boldsymbol{\lambda}}^H_{i,j} = \theta^H_i + \delta^A_j + u^H_{i,j} \qquad i = j = 1, 2, ..., 20; \quad i \neq j \tag{3.15}$$

$$\text{Away team:} \qquad \hat{\boldsymbol{\lambda}}^A_{i,j} = \theta^A_j + \delta^H_i + u^A_{i,j} \qquad i = j = 1, 2, ..., 20; \quad i \neq j \tag{3.16}$$

First of all, we have to normalize two parameters, one from the home team equation and the other one from the away team equation. Otherwise, we will have multiple solutions, as shown below:

Imagine we have estimated the eighty parameters without applying normalizations. Then ($\hat{\boldsymbol{\theta}}^H_i$, $\hat{\boldsymbol{\delta}}^H_i$, $\hat{\boldsymbol{\theta}}^A_i$, $\hat{\boldsymbol{\delta}}^A_i$, i=j=1,...,20) forms a solution to the OLS problem. Let r and s be any real constants. Now we can construct the following parametes:

$$\bar{\boldsymbol{\theta}}^H_i = \hat{\boldsymbol{\theta}}^H_i + r; \qquad \bar{\boldsymbol{\delta}}^A_j = \hat{\boldsymbol{\delta}}^A_j \quad r \tag{3.17}$$

$$\bar{\boldsymbol{\theta}}^A_j = \hat{\boldsymbol{\theta}}^A_j + s; \qquad \bar{\boldsymbol{\delta}}^H_i = \hat{\boldsymbol{\delta}}^H_i \quad s \tag{3.18}$$

Then, for any real values of s and r, ($\hat{\boldsymbol{\theta}}_i^H$, $\hat{\boldsymbol{\delta}}_i^H$, $\hat{\boldsymbol{\theta}}_i^A$, $\hat{\boldsymbol{\delta}}_i^A$, i=j=1,...,20) is also a solution for the problem. So, before estimating the model, we normalize two parameters (one from each equation).

Two conditions are required in order to identify the model:

$$\text{Condition 1:} \qquad \text{(Number of matches} \times 2) > \text{Number of parameters} \tag{3.19}$$

Then, we need at least four complete rounds (weeks). Remember that we have to estimate 78 parameters[16] and a round has ten matches. So, it is not possible to estimate the model before fifth round[17].

$$\text{Condition 2:} \qquad \text{Existence of a link between all teams.} \tag{3.20}$$

Our goal, once the model is estimated, is to have a measure to compare the relative quality of any pair of teams. This relative quality can be revealed directly (from the games between both teams) or indirectly (from the games against common rivals). In any case, there must be a link between the two teams. Formally, let Δ_i be a set including the following elements: the rivals that have faced the team i so far (in the present season), the rivals that have faced team i's rivals so far and so on ad infinitum. In order to guarantee model identification, we require that, $\forall\ i \neq j$, at least one of the following conclusions is true: sets Δ_i and Δ_j have at least one element in common or team i faced team j.

The following estimation methods are proposed:

i) Ordinary Least Square (OLS) regression. In this case, all the matches weigh exactly the same. We are implicitly assuming Assumption 5.i, i.e, parameters are constant over the season.

ii) Weighted Ordinary Least Square regression. We introduce a exponential discount factor (β) in order to give more weight to recent data. This is a simplistic approach to allow fluctuations in performance over the season (Assumption 5.ii). Relative weight (RW) between two different rounds, indexed x and z, is defined as follows:

$$RW_{x,z} = \frac{\beta^z}{\beta^x} \tag{3.21}$$

When choosing the value of β, there are two possibilities. On the one hand, we can take the estimated discount factor from previous studies. Dixon & Coles (1997) proposes an exponential weighting function ϕ in order to give more value to recent information. On the other hand, an alternative approach is to introduced an endogenous discount rate in the model specification. This way, the value of β is estimated simultaneously with the other model coefficients.

In our study, we consider that the second approach is more convenient, because there are very few studies that apply weighted ordinary least squares in the context of modelling soccer data, so we have few references.

[16] We have twenty teams, four parameters per team and we have to exclude the two normalized parameters.

[17] Unless previous season games are used.

3.5.2.2 Hierarchical Poisson-Bayesian Model with Normal random effects priors and Uniform-Gamma hyperpriors (HBM 1).

Below, a two-level hierarchical bayesian model is proposed to model soccer data. From now on, we refer to this model as HBM 1. Subscript i is for home teams (i=1,...,20), subscript j is for away teams (j=1,...,20) and subscript k is for matches (k=1,...,38[18]).

<u>Level 1 (Structural model).</u> It contains the central part of the model.

$$X_{i,j}^H \mid \theta_i^H, \delta_j^A \sim Poisson((\theta_i^H + \delta_j^A) * t) \qquad i = j = 1, 2, ..., 20; \qquad i \neq j \tag{3.22}$$

$$X_{i,j}^A \mid \theta_j^A, \delta_i^H \sim Poisson((\theta_j^A + \delta_I^H) * t) \qquad i = j = 1, 2, ..., 20; \qquad i \neq j \tag{3.23}$$

$$X_{i,j}^H \mid \theta_i^H, \delta_j^A \perp X_{i,j}^A \mid \theta_j^A, \delta_i^H \qquad i = j = 1, 2, ..., 20; \qquad i \neq j \tag{3.24}$$

Note that the estructural model contain the following four assumptions:

- Poisson Assumption (Assumption 1). Given defensive and offensive parameters, goal scoring is modelled using a Poisson distribution.
- Independence Assumption (Assumption 2). Both goal scoring processes are independent.
- Additive Poisson property (Assumption 3). We use this property to decompose the global processes into two parts: ${}^{H}_{i,j} = \theta_i^H + \delta_j^A$ and ${}^{A}_{i,j} = \theta_j^A + \delta_i^H$
- Time homogeneous Poisson process (Assumption 4).

<u>Level 2 (Hierarchical Normal Random Effect priors).</u> In this level, the parameters of the estructural model (θ's and δ's) are modelled. We are assuming Normal Random Effect priors for these parameters. Doing this, we are implicitly using the "assumption Parameters are constant over the season" (Section 4.5.1). Shocks are normally distributed[19] and have zero mean.

$$log(\theta_i^H) = \alpha_i + \epsilon_{i,k} \qquad i = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.25}$$

$$log(\theta_j^A) = \tau_j + \epsilon_{j,k} \qquad j = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.26}$$

$$log(\delta_i^H) = \rho_i + \epsilon_{i,k} \qquad i = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.27}$$

$$log(\delta_i^H) = \omega_j + \epsilon_{j,k} \qquad i = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.28}$$

$$\epsilon_{i,k}, \epsilon_{j,k} \sim Normal(0, \sigma) \quad \forall \quad 1 < i \leq 20; \quad 1 < j \leq 20; \quad 0 < k \leq 20 \tag{3.29}$$

[18] Every team plays 19 matches at home and 19 away.

[19] We have also considered Gamma distributed random shocks, which are mathematically convenient (the gamma distribution is the conjugate prior distribution for poisson), but may be restrictive.

Hyperpriors distributions. Now we are defining the priors (also known as hyperpriors) for the parameters of the level 2. Following the literature (Best et al., 2011), we use non-informative Uniform-Gamma hyperpriors.

$$\alpha_i \sim Uniform(\quad inf, +inf) \qquad i = 1, 2, ..., 20 \tag{3.30}$$

$$\tau_j \sim Uniform(\quad inf, +inf) \qquad j = 1, 2, ..., 20 \tag{3.31}$$

$$\rho_i \sim Uniform(\quad inf, +inf) \qquad i = 1, 2, ..., 20 \tag{3.32}$$

$$\omega_j \sim Uniform(\quad inf, +inf) \qquad j = 1, 2, ..., 20 \tag{3.33}$$

$$\sigma \sim Gamma(0.01, 0.00001) \tag{3.34}$$

Parameter inference: Once the model is described, the objective is to estimate the parameters (θ's and δ's). The bayesian estimation process is summarized below:

1. **Input Data**: We use the $\hat{\ }'s$ we have estimated in Section 5.1 as input data. Then, we can use the following expressions in order to estimate the model:

$$\hat{\boldsymbol{\lambda}}^H_{i,j} = \theta^H_i + \delta^A_j \tag{3.35}$$

$$\hat{\boldsymbol{\lambda}}^A_{i,j} = \theta^A_j + \delta^H_i. \tag{3.36}$$

2. **Normalization**: As in previous section, before estimating the model we normalize two parameters, one from each equation.

3. **Estimation algorithm**: Closed form bayesian estimation is not possible in this case. A Gibbs Sampler algorithm is proposed[20]. This algorithm is a variant of Monte Carlo Markov Chain and is widely used in Applied Bayesian Modelling. It has also been used in the context of modelling soccer data (Everson & Goldsmith-Pinkham (2008)). Gibbs sampler has the advantage that is applicable when the joint distribution is not known explicitly or when is really hard to sample from directly.

4. **Sofware**: JAGS (Just another Gibbs sampler) is a statistical software specifically designed for analysis of Bayesian hierarchical models using Markov Chain Monte Carlo (and its variants). So it is the perfect programming environment to estimate this kind of models. The algorithm was implemented calling JAGS from Matlab using MATJAGS interface.

5. **Results**: Finally, we obtain the posterior probability distribution of the parameters ($\hat{\theta}$'s and $\hat{\delta}$'s). Using this methodology, we are able to estimate the four parameters for each team: home attack, home defense, away attack and away defense.

[20]Comprehensive information about Gibbs Sampler algorithm is provided in: Casella & George (1992) and Gelman et al. (2014).

3.5.2.3 Hierarchical Poisson-Bayesian Model with autoregressive priors and Uniform-Gamma hyperpriors distribution (HBM 2).

Below, another two-level hierarchical bayesian model is proposed to model soccer data. From now on, we refer to this model as HBM 2. Unlike the previous model, in this case it is allowed that the parameters change over the season. Now we focus on level 2 and the hyperpriors, given the structural model (Level 1) is exactly the same as in Section 3.5.2.2:

Level 2 (Autoregressive prior distributions)

In this case, temporal dependence of the data is specifically taken into account. Shocks are modelled as first-order autoregressive processes, following the methodology described in (Best et al., 2011). This scheme allows for fluctuations in the parameters over the season (Assumption 5.i).

$$\theta^H_{i,k} = \alpha_i + \rho * (\theta^H_{i,k} \quad \alpha_{t-1,i}) + \epsilon_{t,i} \qquad i = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.37}$$

$$\theta^A_{j,k} = \nu_j + \rho * (\theta^A_{j,k} \quad \nu_{t-1,j}) + \epsilon_{j,k} \qquad j = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.38}$$

$$\delta^H_{i,k} = \epsilon_i + \rho * (\delta^H_{i,k} \quad \epsilon_{t-1,i}) + \epsilon_{i,k} \qquad i = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.39}$$

$$\delta^A_{j,k} = \pi_j + \rho * (\theta^A_{j,k} \quad \pi_{t-1,j}) + \epsilon_{j,k} \qquad i = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.40}$$

$$\epsilon_{i,k}, \epsilon_{j,k} \sim Normal(0, \sigma) \quad \forall \quad 1 < i \leq 20; \quad 1 < j \leq 20; \quad 0 < k < 20 \tag{3.41}$$

Hyperpriors distributions

Hyperpriors for α, ν, ϵ and π are non-informative. At the same time, we restrict κ values to be between 0 and 1 and we propose a Gamma hyperprior for the parameter σ.

$$\alpha_i \sim Uniform(\quad inf, +inf) \qquad i = 1, 2, ..., 20 \tag{3.42}$$

$$\nu_j \sim Uniform(\quad inf, +inf) \qquad j = 1, 2, ..., 20 \tag{3.43}$$

$$\epsilon_i \sim Uniform(\quad inf, +inf) \qquad i = 1, 2, ..., 20 \tag{3.44}$$

$$\pi_j \sim Uniform(\quad inf, +inf) \qquad j = 1, 2, ..., 20 \tag{3.45}$$

$$\kappa \sim Uniform(0, 1) \tag{3.46}$$

$$\sigma \sim Gamma(0.001, 0.001) \tag{3.47}$$

Parameter inference: The estimation process is analogous to previous section. Nonetheless, now the output is totally different because we obtain a different posterior probability distribution for each week.

3.6 Application: LIGA 2013-2014

Once we have designed the methodology to model soccer data using prediction markets information, we just apply this approach to Spanish Football League data (Liga 2013-2014) in order to check its validity and its prediction properties.

3.6.1 Parameter estimates:

Using Betfair Betting Exchange data from the first "k" weeks of the season, we generated estimates for each team's parameters. We repeat this process for k=5,..., 38. Remember that we need at least 40 matches (4 Weeks) to estimate the models, that is the reason we start at Week 5.

As an illustrative example, Table 3.2 shows OLS estimations using information from the first half of the season (until week 19). Note that OLS method provides point estimates instead of predictive distributions (Bayesian case), computing the "best guess" of an unknown population parameter.

Table 3.2: OLS estimates

	Quality index (week by week)							
	OLS estimation				Weighted OLS			
Team	Off. (Home)	Off. (Away)	Def. (Home)	Def. (Away)	Off. (Home)	Off. (Away)	Def. (Home)	Def. (Away)
Barcelona	2.70	2.30	-0.54	-0.29	1.93	1.82	-1.64	-1.39
Real Madrid	2.63	2.28	-0.41	-0.22	1.94	1.83	-1.46	-1.27
Atlético	1.82	1.58	-0.51	-0.30	1.58	1.47	-1.60	-1.34
Athletic	1.12	1.11	-0.18	0.21	1.18	1.07	-1.14	-0.88
Villarreal	1.24	1.13	-0.29	0.18	1.24	1.13	-1.26	-0.91
Sevilla	1.31	1.17	-0.19	0.21	1.25	1.14	-1.15	-0.88
Real Sociedad	1.10	1.14	-0.15	0.36	1.23	1.12	-1.11	-0.79
Valencia	1.33	1.20	-0.23	0.24	1.28	1.17	-1.18	-0.86
Espanyol	0.98	0.91	-0.10	0.39	0.98	0.87	-1.09	-0.74
Getafe	0.84	0.84	0.09	0.44	0.93	0.82	-0.96	-0.67
Granada	0.96	0.95	0.00	0.41	0.93	0.82	-0.92	-0.70
Málaga	0.79	0.89	0.00	0.52	1.01	0.90	-0.98	-0.72
Levante	0.73	0.81	0.09	0.63	0.85	0.74	-0.89	-0.61
Osasuna	0.75	0.86	0.08	0.54	0.88	0.77	-0.92	-0.65
Celta	0.84	0.95	0.02	0.60	1.00	0.89	-0.96	-0.65
Almería	0.84	0.81	0.18	0.62	0.82	0.71	-0.84	-0.59
Elche	0.77	0.78	0.09	0.49	0.82	0.71	-0.91	-0.69
Rayo Vallecano	0.84	0.94	0.17	0.64	1.01	0.90	-0.86	-0.60
Valladolid	0.76	0.82	0.03	0.44	0.84	0.73	-0.93	-0.68
Betis	0.93	1.03	-0.02	0.42	1.08	0.97	-1.01	-0.75

Figure 3.5 shows Bayesian posterior distributions for some selected teams's attack parameters. (Real Madrid, Barcelona, Sevilla and Atlético). Parameters have been estimated using model BHM 1 and Gibbs sampling (100.000 draws). As in the previous case, algorithms only include information from the first half of the season, so it is updated up to Week 19.

Figure 3.6 shows Bayesian posterior distributions for Real Madrid's attack parameter for the whole season. Parameters have been estimated using model BHM 2 and Gibbs sampling. In this case, temporal dependence of the data is modelled and posterior distributions evolve over the season.

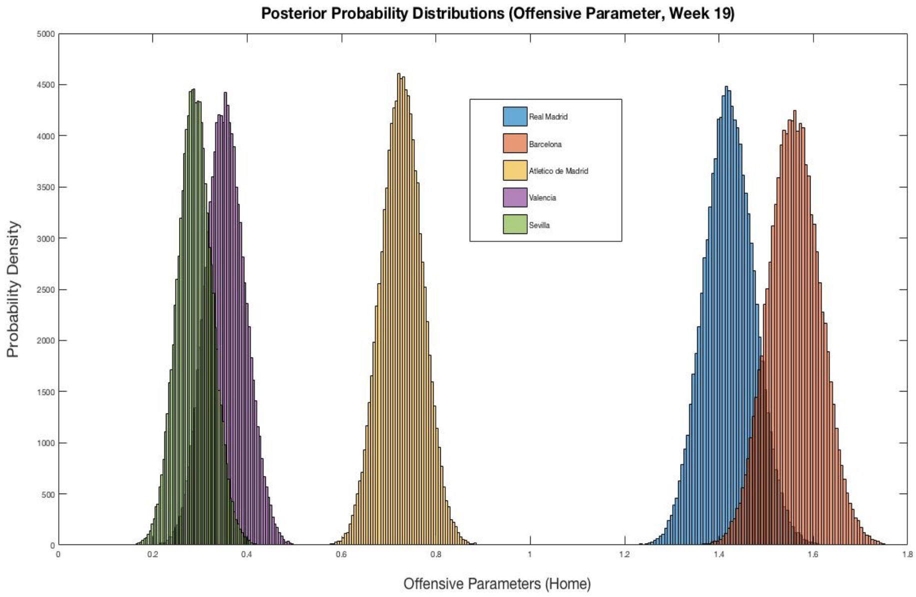


Figure 3.5: BHM 1: Posterior Probability Distributions for the home attack parameters of Real Madrid, Barcelona, Atlético, Sevilla y Valencia (Updated at Week 19).

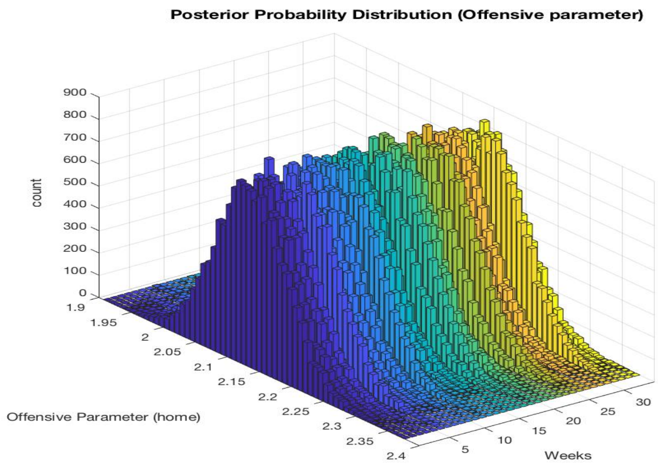


Figure 3.6: BHM 2: Posterior Probability Distributions for the home attack parameter of Real Madrid (from Week 5 to Week 38).

3.6.2 Predictions for single matches (short-term forecast):

Once we have estimated the parameters, we are able to predict any single match. Firstly, let's introduce some notation: let $P^H_{i,j}$ be the probability of local win (team i), let $P^D_{i,j}$ be the probability of draw and let $P^A_{i,j}$ be the probability of away win (team j). Now we can define the set of possible forecasts ($\Gamma_{i,j}$) as:

$$\Gamma_{i,j} = [(P^H_{i,j}, P^D_{i,j}, P^A_{i,j}) \subseteq [0,1]^3 \ : \ (P^H_{i,j} + P^D_{i,j} + P^A_{i,j}) = 1] \tag{3.48}$$

In the OLS framework, we use the point estimates (θ's and δ's) in order to get the predictions. For a given match played by team i and team j, the OLS prediction is computed as follows:

Goal scoring processes:

$$\text{Home Goal Scoring (team i):} \qquad X^H_{i,j} \sim Poisson(\hat{\boldsymbol{\theta}}^H_i + \hat{\boldsymbol{\delta}}^A_j) \tag{3.49}$$

$$\text{Home Goal Scoring (team j):} \qquad X^A_{i,j} \sim Poisson(\hat{\boldsymbol{\theta}}^A_j + \hat{\boldsymbol{\delta}}^H_i) \tag{3.50}$$

Once we know both goal scoring processes, we use the joint probability distribution in order to compute the probability of all possible outcomes for the match. Then we compute match probabilities as follows:

$$P^H_{i,j} = \sum_{x=1}^{\infty}\sum_{y=0}^{x-1} \ P(X^H_{i,j} = x) \times P(X^A_{i,j} = y) \tag{3.51}$$

$$P^D_{i,j} = \sum_{x=0}^{\infty} \ P(X^H_{i,j} = x) \times P(X^A_{i,j} = x) \tag{3.52}$$

$$P^A_{i,j} = \sum_{y=1}^{\infty}\sum_{x=0}^{y-1} \ P(X^H_{i,j} = x) \times P(X^A_{i,j} = y) \tag{3.53}$$

In the Bayesian context, the procedure is a little more complex since we have posterior probability distributions instead of point estimates. Then the posterior sampling procedure is described as follows:

1. We randomly choose one value for each parameter from the posterior probability distributions ($\hat{\boldsymbol{\theta}}^H_{i,h}$, $\hat{\boldsymbol{\delta}}^A_{j,h}$, $\hat{\boldsymbol{\delta}}^A_{j,h}$, $\hat{\boldsymbol{\theta}}^H_{i,h}$).
2. We compute: $\hat{\boldsymbol{\theta}}^H_{i,h} + \hat{\boldsymbol{\delta}}^A_{j,h}$ and $\hat{\boldsymbol{\theta}}^A_{j,h} + \hat{\boldsymbol{\delta}}^H_{i,h}$.
3. Obtain $P^H_{i,j,h}, P^D_{i,j,h}$, and $P^A_{i,j,h}$, as in the OLS case.
4. Repeat steps n times (for h=1,...,n[21]).
5. Compute the median value:

[21] In our study, n=100,000

$$P^H_{i,j} = \frac{\sum_{h=1}^{n}(P^H_{i,j,h})}{n} \tag{3.54}$$

$$P^D_{i,j} = \frac{\sum_{h=1}^{n}(P^D_{i,j,h})}{n} \tag{3.55}$$

$$P^A_{i,j} = \frac{\sum_{h=1}^{n}(P^A_{i,j,h})}{n} \tag{3.56}$$

Once we have developed a systematic method to compute single match probabilities from predictive markets, we check its performance. In the literature (Everson & Goldsmith-Pinkham, 2008), a normal method used to evaluate the goodness of a prediction is to calculate the De Finnetti distance (De Finetti, 1972) which is the Euclidean distance between the point correspondent to the outcome[22] and that one correspondent to the forecast. For instance, consider predicted match probabilities (0.2, 0.3, 0.5). If the actual outcome is (0,1,0), De Finetti measure for that prediction equals 0.78 (0.2^2+0.7^2+0.5^2). Note that if we assume complete uncertainty about the final outcome (1/3, 1/3, 1/3), i.e., equiprobable predictor, De Finetti distance always equals 2/3. The historical-mean predictor[23] is also considered. Now, we can use these values as reference to evaluate our prediction model performance. If our model is not able to win equiprobable and historical-mean predictors, our model is not useful at all.

Below, we carry out an exercise to compare Betfair Exchange predictions and the predictions offered by our models. Firstly, we use our models to compute match probabilities week by week, starting on Week 5 and finishing on Week 38. So, we have to forecast ten matches per week. When calculating match probabilities for some week, we use information from the beginning of the season until the previous week. We do not use information of the current week in order to be on equal terms with markets. Secondly, we compute implied match probabilities from Betfair Exchange[24] for all matches of the season. The data is collected at the beginning of the game. Finally, we compute the sum of all De Finetti distances week by week for both markets prediction and our models' predictions. In order to have references we also include De Finetti's distances for the equiprobable predictor and for the historical-mean predictor.

Results are shown in Table 3.3. If we consider the whole season, market predictions are a bit better than models' predictions according to "De Finetti distance". Remember that a smaller De Finetti distance indicates better fit. At the same time, all the models perform much better than equiprobable and historical-mean predictors.

However, if we evaluate predictions only from Week 5 to Week 32, excluding last 6 Weeks of the season, models' predictions are more accurate than market ones, because all of them present a smaller De Finetti measure in the subtotal (J5-J32). In conclusion, our forecasting models perform really well throughout the season and are able to beat betting markets, except in the last weeks.

[22]There are three possible outcomes: (1,0,0), (0,1,0) and (0,0,1).

[23]This predictor uses the historical means (from 1998 to 2016) for the match probabilities (.48, .25, .27).

[24]We calculate the inverse of the odds and later we normalise values to sum 1.

Table 3.3: A comparison of prediction capabilities (De Finetti Measure).

	De Finetti Distance (week-by-week)						
	OLS	**Weighted OLS**	**BHM 1**	**BHM 2**	**Betfair Exchange**	**Equiprob. Predictor**	**Hist. Predictor**
Week 5	4.94	5.13	5.15	5.47	5.13	6.67	6.73
Week 6	2.99	3.09	3.29	3.06	2.88	6.67	5.38
Week 7	6.62	6.43	6.81	6.62	6.91	6.67	7.06
Week 8	4.07	4.22	4.13	4.16	4.24	6.67	5.38
Week 9	8.57	8.29	8.24	8.34	8.21	6.67	7.06
Week 10	4.78	4.92	5.02	4.87	5.07	6.67	5.75
Week 11	6.67	6.62	6.51	6.66	6.92	6.67	7.55
Week 12	4.49	4.39	4.48	4.49	4.55	6.67	6.16
Week 13	5.94	6.00	5.99	5.97	6.02	6.67	7.10
Week 14	4.19	4.22	4.31	4.20	4.37	6.67	6.16
Week 15	5.57	5.54	5.65	5.58	5.48	6.67	6.24
Week 16	5.54	5.60	5.48	5.48	5.72	6.67	5.87
Week 17	5.09	5.06	5.07	5.17	5.00	6.67	7.02
Week 18	3.33	3.39	3.48	3.44	3.39	6.67	5.38
Week 19	5.92	5.82	5.99	5.97	5.75	6.67	6.24
Week 20	6.12	6.19	6.15	6.23	6.30	6.67	6.77
Week 21	5.05	5.04	5.01	4.94	5.08	6.67	5.79
Week 22	6.83	6.70	6.83	6.76	6.86	6.67	6.28
Week 23	5.71	5.73	5.58	5.64	5.67	6.67	5.87
Week 24	5.79	5.64	5.63	5.66	5.59	6.67	6.65
Week 25	6.98	6.88	6.90	6.92	6.73	6.67	6.69
Week 26	5.74	5.73	5.71	5.81	5.81	6.67	5.91
Week 27	6.40	6.31	6.38	6.36	6.58	6.67	6.20
Week 28	4.33	4.21	4.43	4.39	4.04	6.67	6.28
Week 29	5.78	5.80	5.88	5.84	5.90	6.67	6.65
Week 30	5.34	5.33	5.30	5.37	5.19	6.67	6.24
Week 31	6.49	6.51	6.42	6.47	6.70	6.67	7.06
Week 32	3.87	3.89	3.99	3.63	3.89	6.67	5.79
Subtotal (J5-32)	153.11	152.67	153.82	153.47	153.98	186.67	177.25
Week 33	5.34	5.27	5.35	5.43	5.11	6.67	6.28
Week 34	5.45	5.36	5.34	5.35	5.12	6.67	5.87
Week 35	5.42	5.29	5.45	5.44	4.69	6.67	6.61
Week 36	9.79	9.84	9.67	9.49	9.46	6.67	7.14
Week 37	9.11	9.07	8.77	8.82	8.85	6.67	6.32
Week 38	5.57	5.42	5.38	5.62	4.80	6.67	6.24
Subtotal (J33-J38)	40.67	40.25	39.97	40.14	38.03	40.00	38.46
Total	193.78	192.92	193.79	193.60	192.01	226.67	215.71

Finally, if we observe De Finetti distances in the last part of the season, it can be concluded that forecasting models perform really bad, since they are even beaten by predictors. This results should not be a matter of concern since last weeks are a little bit weird and are really hard to model. Quality is not as relevant as before and other factors become important, such as motivation, incentives, position in the table or even illegal bonuses for victory or defeat. At the end of the season, some teams do not play anything while other teams have a lot at stake.

An example can be used to illustrate this fact. In the last week of the 2013-2014 Liga, Rayo Vallecano faced Getafe. The OLS model predictions for this match are[25]: Rayo Vallecano wins the match (50%), draw probability (35%) and Getafe wins the match (15%). However, the implied probabilities of the Betfair Exchange Market were quite different: Rayo Vallecano wins the match (17%), draw probability (22%) and Getafe wins the match (61%). These market probabilities are not consistent with the relative strength of Rayo Vallecano (position 13 in the table) and Getafe (position 18 in the table). But markets have more information and they took into account the different level of motivation between both teams. Rayo Vallecano was in a safe place and it had no incentives to win and Getafe was fighting to not be relegated. Then markets incorporate all this information and update match probabilities according it.

Table 3.4 shows the mean distance per match. It can be observed that the mean distance deeply increase in the last part of the season, not only for the models considered but also for markets' predictions. There is therefore evidence that the last matches of the championship are more difficult to forecast in general. However, the increase is lower in Betfair predictions, because of the amount of information available to markets.

Table 3.4: De Finetti Measure: Mean per match

	De Finetti Measure (mean per match)						
	OLS	**Weighted OLS**	**BHM 1**	**BHM 2**	**Betfair Exchange**	**Equiprob. Predictor**	**Hist. Predictor**
Subtotal (5-32)	0.547	0.545	0.550	0.548	0.550	0.667	0.633
Subtotal (33-38)	0.678	0.671	0.666	0.669	0.634	0.667	0.641
Total	0.570	0.567	0.570	0.569	0.565	0.667	0.634

Table 3.5 shows the percentage of matches in which the historical-mean predictor loses. These percentages are higher than those obtained in Bueno et al. (2010), i.e., 62.5% of matches, although both results are not directly comparable since their study uses World Cup data instead of Spanish League data.

Table 3.5: Comparison

	% of matches in which the historical-mean predictor loses				
	OLS	**Weighted OLS**	**BHM 1**	**BHM 2**	**Betfair Exchange**
Subtotal (5-32)	66.8%	65.0%	65.0%	65.3%	63.6%
Subtotal (33-38)	56.7%	55.0%	58.3%	55.0%	63.3%
Total	65.0%	63.53%	63.8%	63.5%	63.5%

[25] Weighted OLS and Bayesian predictions are quite similar.

Figure 3.7 shows correlation between Betfair Exchange and OLS predictions week by week. It can be easily observed that the correlations are really high (close to one) all season except last four weeks. From week 34, correlations deeply decrease, evidencing that the last part of the season is anomalous. Figure 3.8 repeats the analysis for the Bayesian predictions, observing exactly the same pattern.

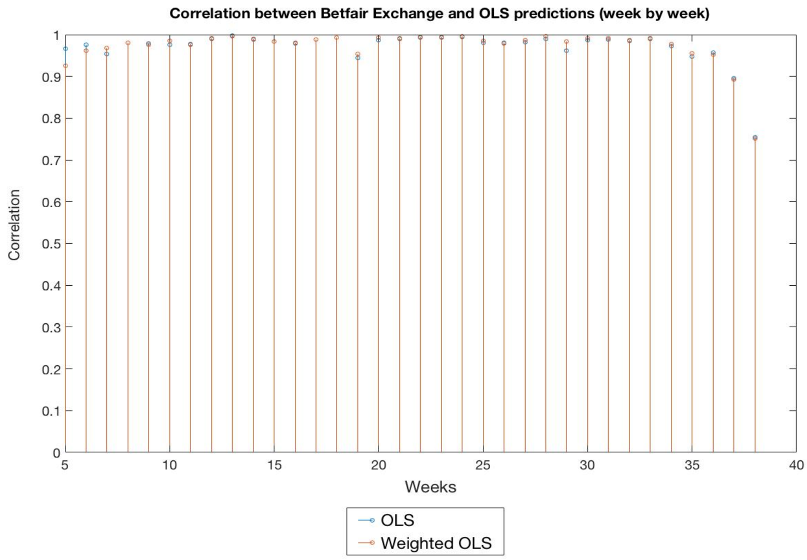


Figure 3.7: Correlation between Betfair Exchange and OLS predictions (week by week).

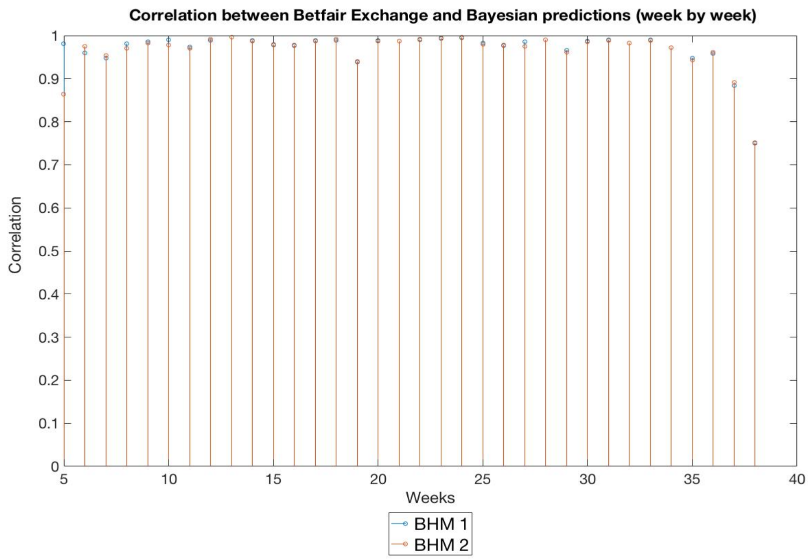


Figure 3.8: Correlation between Betfair Exchange and Bayesian predictions (week by week).

3.6.3 Past Performance vs Prediction Markets data (short term).

At the beginning of the present article we enunciated the following hypothesis: "using prediction markets data instead of past results can improve the predictive power of the soccer models". In this section, we propose an exercise to check whether there is evidence supporting this hypothesis.

The procedure is as follows. As in the literature (Everson & Goldsmith-Pinkham, 2008), we use past results as primary information for estimating the defensive and offensive parameters. Later, we use these estimates to compute match probabilities week by week, starting at Week 20 and finishing at Week 38. We start at Week 20 in order to have enough past results to estimate the models. Finally, we use actual results to compute De Finetti distances, as in Section 3.6.2.

Results are reported in Table 3.6. It can be observed that models based on prediction market data beat models based on past results, according to De Finetti measure. For instance, the sum of the De Finetti distances for the second half of the season is 121.3 when we use past performance, while it is less than 115 when we use prediction markets data.

If we exclude the last weeks of the season and focus only between Week 20 and Week 32, conclusions are maintained. In this case, the model based on past results performs better than predictors but worse than models based on predictive markets.

Table 3.6: Past Performance vs Prediction Markets data

	De Finetti Measure (week-by-week)							
	OLS	**Weighted OLS**	**BHM 1**	**BHM 2**	**Betfair Exchange**	**Past Performance**	**Equiprob. Predictor**	**Hist. Predictor**
Week 20	6.12	6.19	6.15	6.23	6.30	6.55	6.67	6.77
Week 21	5.05	5.04	5.01	4.94	5.08	4.92	6.67	5.79
Week 22	6.83	6.70	6.83	6.76	6.86	8.03	6.67	6.28
Week 23	5.71	5.73	5.58	5.64	5.67	8.35	6.67	5.87
Week 24	5.79	5.64	5.63	5.66	5.59	5.10	6.67	6.65
Week 25	6.98	6.88	6.90	6.92	6.73	5.35	6.67	6.69
Week 26	5.74	5.73	5.71	5.81	5.81	6.89	6.67	5.91
Week 27	6.40	6.31	6.38	6.36	6.58	7.61	6.67	6.20
Week 28	4.33	4.21	4.43	4.39	4.04	4.20	6.67	6.28
Week 29	5.78	5.80	5.88	5.84	5.90	5.06	6.67	6.65
Week 30	5.34	5.33	5.30	5.37	5.19	5.45	6.67	6.24
Week 31	6.49	6.51	6.42	6.47	6.70	6.88	6.67	7.06
Week 32	3.87	3.89	3.99	3.63	3.89	5.24	6.67	5.79
Subtotal (J20-32)	**74.42**	**73.94**	**74.23**	**74.00**	**74.35**	**79.63**	**86.67**	**82.18**
Week 33	5.34	5.27	5.35	5.43	5.11	5.37	6.67	6.28
Week 34	5.45	5.36	5.34	5.35	5.12	4.69	6.67	5.87
Week 35	5.42	5.29	5.45	5.44	4.69	5.52	6.67	6.61
Week 36	9.79	9.84	9.67	9.49	9.46	9.89	6.67	7.14
Week 37	9.11	9.07	8.77	8.82	8.85	9.68	6.67	6.32
Week 38	5.57	5.42	5.38	5.62	4.80	6.48	6.67	6.24
Subtotal (J33-J38)	**40.67**	**40.25**	**39.97**	**40.14**	**38.03**	**41.63**	**40.00**	**38.46**
Total	**115.10**	**114.19**	**114.20**	**114.14**	**112.38**	**121.27**	**126.67**	**120.64**

3.6.4 Model Ranking.

Now, we can rank the models according to their short-term prediction performance. Weighted OLS provides the most accurate predictions according to De Finetti measure (0.567 per match on average). BHM 2 ranks in second place (0.569 per match on average), followed by OLS and BHM 1 both with 0.57.

An alternative approach, only valid for bayesian model comparisons, consist on computing the Deviance Information criterion (DIC). DIC is a widely used hierarchical modelling generalization of the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). The idea is that models with smaller DIC should be preferred to models with larger DIC. DIC is easily calculated from the samples generated by a MCMC simulation. In our case, according to this criterion, the BHM 2 model (DIC=6112) is better than the BHM 1 model (DIC=6576), in the sense that it fits the data better. This results are also consistent with what was observed before, i.e, BHM 2 presents better short-term prediction capabilities.

In view of the results, the differences between models are small. There is slight evidence that Assumption 5.ii (parameters change over the season) seems more reasonable than Assumption 5.i. (parameters are constant), because Weighted OLS is better than OLS and BHM 2 is also better than BHM 1.

3.6.5 Tournament simulation (long-term forecast):

The forecasting algorithms proposed in this paper can also be applied to tournament simulation. At any date, we can simulate the rest of the season, obtaining predicted points at the end of the season for each team together with a set of simulated variables, such as position, games won, goals scored, probability of winning the league, chances of being relegated, etc.

The simulation procedure is described as follows. Firstly, we collect information of the first half of the season[26] (Betfair Exchange data) in order to estimate model parameters for each team (home attack, home defense, away attack and away defense). To do this, we focus on OLS Model. Secondly, we use these estimates to calculate the intensity[27] of the goal scoring processes for the remaining matches until the end of the season. To do this, we use the methodology described in Section 3.6.2. Finally, we run Monte Carlo simulations [28], sampling from these distributions. In each simulated league, we have to sample from 380 different Poisson distributions, given there exist two goal scoring processes per match. During a simulation, we apply the Liga official rules for tie breaking[29]. We repeat this process n times[30] and we get a simulated distribution for each variable.

[26] This exercise can be implemented at any point of the season.

[27] $(\hat{\boldsymbol{\lambda}}^H_{i,j},\hat{\boldsymbol{\lambda}}^A_{i,j})$, i=j=1,...,20

[28] Simulations are implemented in Matlab.

[29] In case of tie between six or more teams, we just break the tie randomly, because they are unlikely and it is really hard to code such situations.

[30] In our case, we consider that n=100,000 is enough.

Table 3.7 presents a summary of the simulation results[31]. Note that we are simulating from Week 19, approximately five months before the end of the season. Firstly, the medians[32] of the simulated distributions of several variables are presented. The list of variables includes: matches won, matches drawn, matches lost, goals scored (GF), goals conceded (GC) and goal difference (GD). In the Appendix (figures 3.A15, 3.A16 and 3.A17), we compare the actual and predicted values of GF, GC and GD . Secondly, we pay special attention to points, so that in addition to the median, it is also included the 5th percentile and the 95th percentile of the simulated distribution. Finally, some probabilities are presented: probability of winning La Liga, probability of qualifying for Champions League (UCL)[33], probability of qualifying for Europa League (UEL) and probability of being relegated to second division[34].

Table 3.7: Simulated Season (LIGA 2013-2014)

		End-of-season median values							Points			End-of-season probabilities			
Pos.	**Team**	**Played**	**Won**	**Drawn**	**Lost**	**GF**	**GC**	**GD**	**Median**	**Pctl 5**	**Pctl 95**	**Win Liga**	**UCL**	**UEL**	**Relegated**
1	**Barcelona**	38	30.9	4.46	2.7	103.8	24.7	79.1	97.0	90	103	66.6%	100%	0%	0%
2	**Real Madrid**	38	29.4	4.67	3.9	102.6	35.1	67.5	92.9	85	100	23.2%	100%	0%	0%
3	**Atlético**	38	28.2	5.89	3.9	81.5	24.3	57.1	90.4	82	98	10.2%	100%	0%	0%
4	**Athletic**	38	18.5	7.92	11.6	55.0	46.1	9.0	63.4	55	72	0%	36.3%	47%	0%
5	**Villarreal**	38	18.2	8.59	11.2	61.9	42.3	19.6	63.1	55	72	0%	34.3%	48.3%	0%
6	**Sevilla**	38	16.4	10.7	10.9	62.0	51.1	10.9	59.8	51	69	0%	16.0%	46.7%	0%
7	**Real Sociedad**	38	16.4	9.76	11.9	59.6	51.1	8.6	58.9	50	68	0%	12.0%	42.2%	0%
8	**Valencia**	38	15.4	6.55	16.1	52.2	53.2	-1.0	52.7	44	61	0%	1.4%	13.3%	0.1%
9	**Espanyol**	38	12.2	8.66	17.1	42.0	50.2	-8.2	45.4	37	54	0%	0%	0.9%	3.0%
10	**Granada**	38	12.2	6.65	19.1	36.9	52.6	-15.7	43.3	35	52	0%	0%	0.3%	6.8%
11	**Málaga**	38	11.3	9.83	16.9	39.5	49.4	-9.9	43.6	35	52	0%	0%	0.5%	6.9%
12	**Getafe**	38	12.2	6.93	18.9	37.6	57.8	-20.3	43.4	35	52	0%	0%	0.4%	7.5%
13	**Levante**	38	10.7	9.65	17.6	34.9	55.8	-20.9	41.8	34	50	0%	0%	0.1%	11.8%
14	**Osasuna**	38	11.1	7.80	19.1	34.7	57.1	-22.4	41.0	33	50	0%	0%	0.1%	15.4%
15	**Celta**	38	10.2	8.76	19.0	41.7	59.2	-17.5	39.4	31	48	0%	0%	0.1%	24.7%
16	**Almería**	38	9.91	8.60	19.5	39.0	67.8	-28.8	38.3	30	47	0%	0%	0%	31.5%
17	**Elche**	38	8.86	10.6	18.6	33.9	56.4	-22.5	37.2	29	45	0%	0%	0%	39.0%
18	**Valladolid**	38	8.18	11.9	17.9	38.1	58.9	-20.8	36.4	28	45	0%	0%	0%	44.2%
19	**Rayo Vallecano**	38	10.2	5.58	22.2	38.2	74.8	-36.6	36.2	28	45	0%	0%	0%	47.3%
20	**Betis**	38	8.06	9.90	20.0	36.4	63.7	-27.3	34.1	26	43	0%	0%	0%	61.9%

[31] Actual table of LIGA 2013-2014 is presented in the Appendix (Table 3.A8)
[32] The mean gives similar results.
[33] The top four teams are qualified to Champions League
[34] The bottom four teams are relegated.

Figure 3.9 shows the 90% interval for the points at the end of the season. The actual points fall within the bands for 18 out of the 20 teams. Only Barcelona (models over-estimate its performance) and Celta (models infraestimate its performance) are outside the interval. At the same time, **Figure 3.10** compares the actual points and the median predicted points, observing a good fit in general terms.

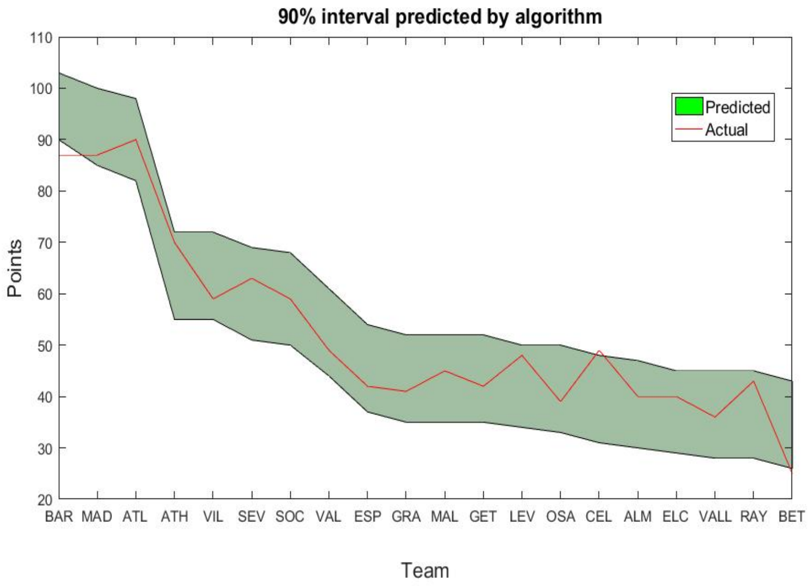


Figure 3.9: 90% confident interval for points.

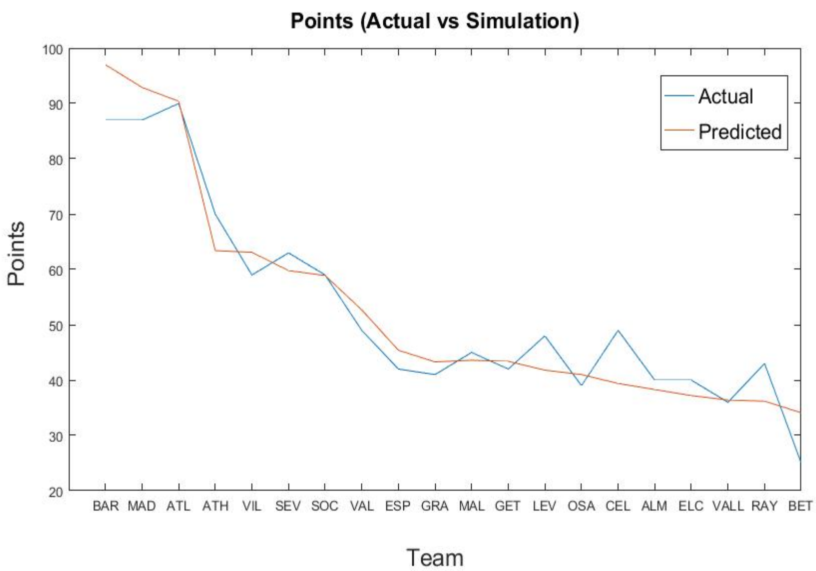


Figure 3.10: Points: Actual vs Simulation.

As an example, Figure 3.11 shows a probability distribution of points for Atlético de Madrid. The simulation was made considering only information from the first half of the season. At that point, 91 points was the most likely outcome with a probability of 8.5%, followed by 90 and 92 points, with a probability of 8% and 7.8%, respectively. Atlético de Madrid finally got 90 points at the end of the season.

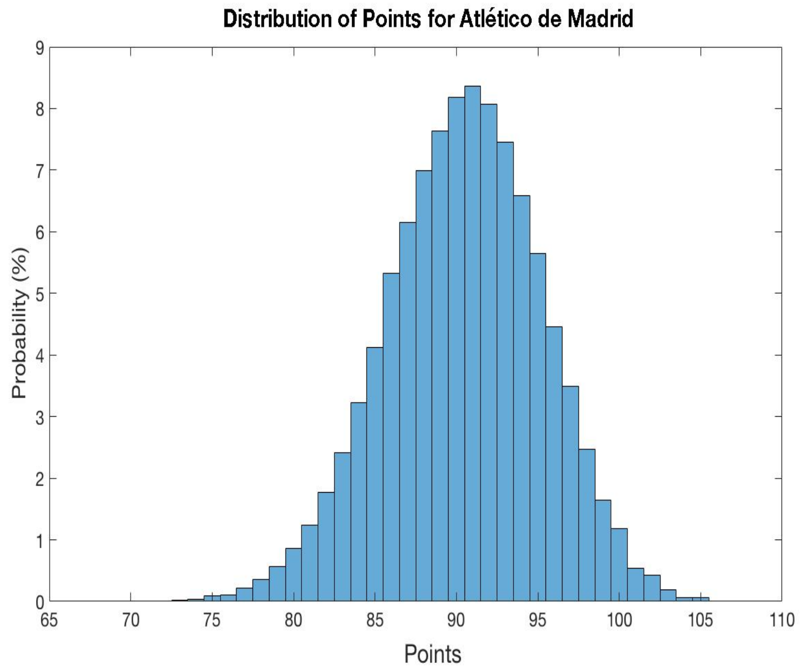


Figure 3.11: Simulated distribution of points for Atlético de Madrid at the end of the season.

3.6.6 Past performance vs Prediction markets data (long-term).

We define the "prediction error" as the difference between the actual points and the median value of the simulated distribution. In order to obtain a measure of long-term relative performance, we simulate the entire season using both the model based on past results and the model based on prediction markets with information until Week 4. Later we repeat the simulation but this time using information until Week 5 and so on until Week 37. Now we compute the sum of squared errors for the model based on past results and for the model based on prediction markets.

Figure 3.12 shows this information, i.e., the sum of squared errors of prediction (SSE) week by week. In the first weeks the gap is very large and as the season is advancing the difference is reduced, but the market-based model always perform better than the models based in past performance.

If we consider long term predictions the marked-based algorithms also generate better predictions than models based on past performance. Moreover, they need less amount of information to generate accurate predictions.

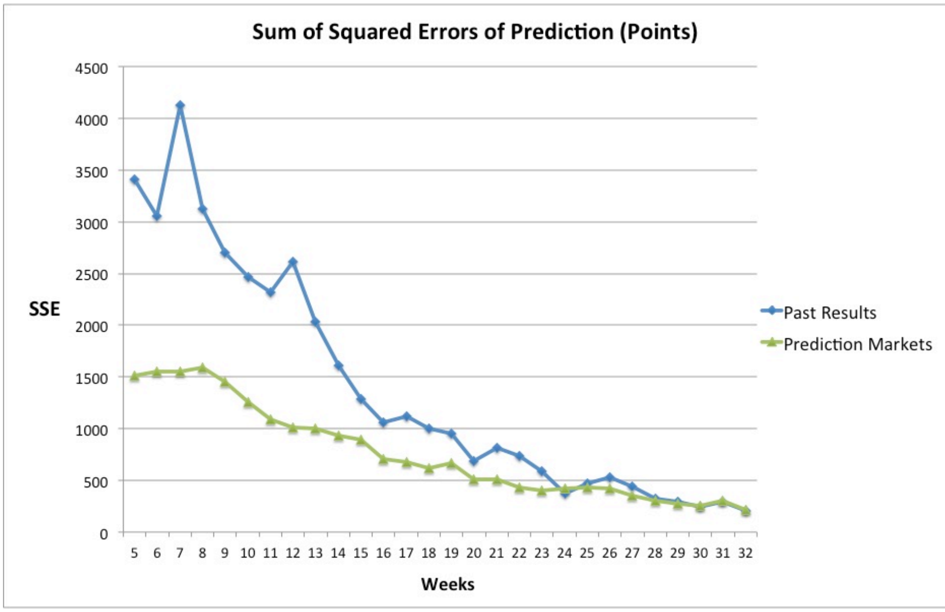


Figure 3.12: Sum of Squared Errors of Prediction (SSE).

3.7 Conclusions

In this article a market-based algorithm for predicting the outcome of events is proposed and illustrated for the 2013-2014 Liga. The main methodological contribution is the use of high-frequency data from betting exchanges as primary information for the estimation of predictive models. Until now, the models presented in the literature have been based solely on historical results or rankings, but have never used information from competitive markets. When modelling the stochastic processes that determine the goals scored by the different teams, we believe that markets can offer us better quality information that is more up-to-date than the final scores that, after all, are a simple realisation of each stochastic process.

First, in order to facilitate the process of inference, several assumptions are introduced. On the one hand, it is assumed that the goals scored by a certain team follow a time-homogeneous Poisson process. On the other hand, it is considered that the processes that determine the goals of the local and visiting team are statistically independent. These assumptions are widely used in the literature and the tests performed to ascertain their validity find no evidence against them.

Next, predictive market information is used to infer the parameters of the stochastic processes that determine the goals scored by each team. For each second of each match of the season, a system of equations is proposed that uses as inputs the probabilities provided by the markets and outputs the estimated intensity of the Poisson processes for each of the teams. In this way, for each match we have a whole probability distribution for each of the possible final scores instead of a single result.

Subsequently, the additive property of the Poisson distributions is used to estimate a total of four parameters per team: home attack, home defense, away attack and away defense. The intensities of the Poisson processes previously estimated are used as primary information. Several estimation methods are proposed: (1) (weighted) ordinary least squares, (2) hierarchical bayesian model with normal random shocks and (3) hierarchical bayesian model with autoregressive shocks. Now there is a mechanism by which any two teams can be paired and the outcome of their future matches predicted.

The methodology described above is applied to predict the probability of each Liga 2013-2014 match, calculated week by week, starting on Matchday 5 and ending on Matchday 38. The prediction capability of the proposed models is tested using "De Finetti" measure, which is defined as the Euclidean distance between the prediction and the realisation. The results obtained indicate that the models have good short-term predictive properties, being able to improve market predictions, except in the last weeks of the season. Moreover, they also generate better predictions than models based on past performance.

Finally, we verify the predictive ability of proposed algorithms when simulating a league. In this exercise, information from the betting markets (first half of the season) is used for estimating the structural parameters and then the rest of the season is simulated. Our models have also good long-term predictive properties, since the predictions derived from the simulation closely resemble what is observed in reality.

These models can be easily extended to other sports just changing the structural model and the distributions. This methodology can also be applied to other prediction markets, such as stock markets, political markets...

As a final remark, it can be concluded that the use of information from betting markets can lead to more accurate estimates of the probability distribution of future events.

3.8 References

Antoniou, C., & Mavis, C. P. (2016). Bayesian Reasoning: Evidence from the Field. ISO 690

Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. Journal of Applied Statistics, 37(2), 253-264.

Best, N., Mason, A., & Li, P. (2011). Introduction to Bayesian analysis using WinBUGS.

Brillinger, D. R. (2008). Modelling game outcomes of the Brazilian 2006 series a championship as ordinal-valued. Brazilian Journal of Probability and Statistics, 89-104.

Casella, G., & George, E. I. (1992). Explaining the Gibbs sampler. The American Statistician, 46(3), 167-174..

Croxson, K., & James Reade, J. (2014). Information and efficiency: Goal arrival in soccer betting. The Economic Journal, 124(575), 62-91.

Dyte, D., & Clarke, S. R. (2000). A ratings based Poisson model for World Cup soccer simulation. Journal of the Operational Research society, 51(8), 993-998.

De Finetti, B. (1972). Probability, induction, and statistics.

Dixon, M. J., & Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. Journal of the Royal Statistical Society: Series C (Applied Statistics), 46(2), 265-280.

Dixon, M., & Robinson, M. (1998). A birth process model for association football matches. Journal of the Royal Statistical Society: Series D (The Statistician), 47(3), 523-538.

Everson, P., & Goldsmith-Pinkham, P. S. (2008). Composite Poisson models for goal scoring. J Quant Anal Sports, 4(2), 13.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2014). Bayesian data analysis (Vol. 2). Boca Raton, FL, USA: Chapman & Hall/CRC.

Hernández, V. (2017b). On the aggregation of information in centralized prediction markets. Working paper.

Huang, X., Knottenbelt, W., & Bradley, J. (2011). Inferring tennis match progress from in-play betting odds. Final year project), Imperial College London, South Kensington Campus, London, SW7 2AZ.

Karlis, D., & Ntzoufras, I. (2000). On modelling soccer data. Student, 3, 229-245.

Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. Journal of the Royal Statistical Society: Series D (The Statistician), 52(3), 381-393.

Karlis, D., & Ntzoufras, I. (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. IMA Journal of Management Mathematics, 20(2), 133-145.

Keller, J. B. (1994). A characterization of the Poisson distribution and the probability of winning a game. The American Statistician, 48(4), 294-298.

Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. Applied Economics, 32(11), 1353-1363.

Langseth, H. (2013) Beating the bookie: A look at statistical models for prediction of football matches. In SCAI (pp. 165-174). ISO 690

Maher, M. J (1982). Modelling association football scores. Statistica Neerlandica, 36(3), 109-118.

Merritt, S., & Clauset, A. (2014). Scoring dynamics across professional team sports: tempo, balance and predictability. EPJ Data Science, 3(1), 1-21.

Mller, J., & Srensen, M. (1994). Statistical analysis of a spatial birth-and-death process model with a view to modelling linear dune fields. Scandinavian journal of statistics, 1-19.

Ntzoufras, I. (2011). Bayesian modeling using WinBUGS (Vol. 698). John Wiley & Sons.

Ridder, G., Cramer, J. S., & Hopstaken, P. (1994). Down to ten: Estimating the effect of a red card in soccer. Journal of the American Statistical Association, 89(427), 1124-1127.

Rue, H., & Martino, S. (2007). Approximate Bayesian inference for hierarchical Gaussian Markov random field models. Journal of statistical planning and inference, 137(10), 3177-3192.

Simonoff, J. S. (2013). Analyzing categorical data. Springer Science & Business Media.

Suzuki, A. K., Salasar, L. E. B., Leite, J. G., & Louzada-Neto, F. (2010). A Bayesian approach for predicting match outcomes: the 2006 (Association) Football World Cup. Journal of the Operational Research Society, 61(10), 1530-1539.

Volf, P. (2009). A random point process model for the score in sport matches. IMA Journal of Management Mathematics, 20(2), 121-131.

Wolfers, J., & Zitzewitz, E. (2006). nterpreting prediction market prices as probabilities (No. w12200). National Bureau of Economic Research.

Wang, H., Feng, C., Tu, X., & Kowalski, J. (2012). A comparison of two test statistics for Poisson overdispersion/underdispersion.

3.9 Statistical Appendix

A. Tables and figures

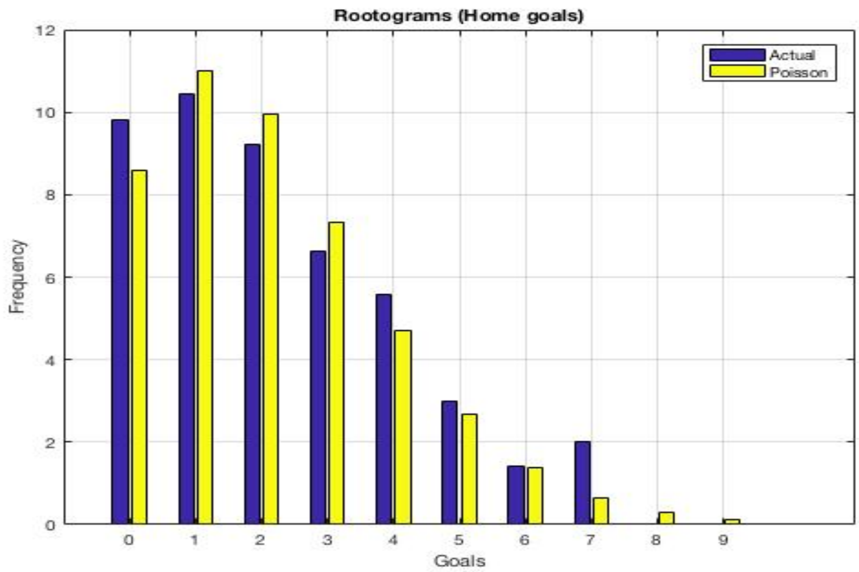


Figure 3.A13: Rootogram: Home goals

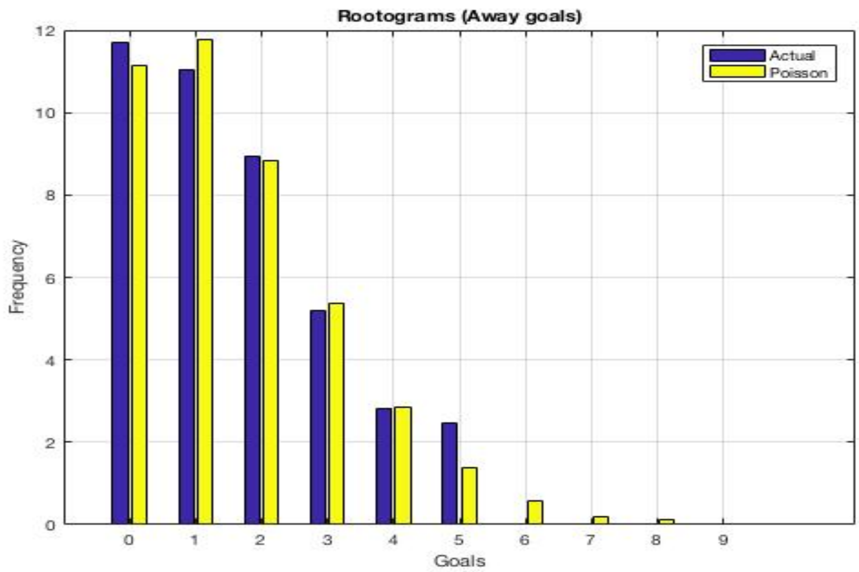


Figure 3.A14: Rootogram: Away goals

Table 3.A8: Final Standing (LIGA 2013-2014)

Pos.	Team	Played	Won	Drawn	Lost	GF	GC	GD	Points	Win Liga	UEFA CL	UEFA EL	Relegated
1	**Atletico**	38	28	6	4	77	26	51	90	Yes	Yes	No	No
2	**Barcelona**	38	27	6	5	100	33	67	87	No	Yes	No	No
3	**Real Madrid**	38	27	6	5	104	38	66	87	No	Yes	No	No
4	**Athletic**	38	20	10	8	66	39	27	70	No	Yes	No	No
5	**Sevilla**	38	18	9	11	69	52	17	63	No	No	Yes	No
6	**Villarreal**	38	17	8	13	60	44	16	59	No	No	Yes	No
7	**Real Sociedad**	38	16	11	11	62	55	7	59	No	No	No	No
8	**Valencia**	38	13	10	15	51	53	-2	49	No	No	No	No
9	**Celta**	38	14	7	17	49	54	-5	49	No	No	No	No
10	**Levante**	38	12	12	14	35	43	-8	48	No	No	No	No
11	**Malaga**	38	12	9	17	39	46	-7	45	No	No	No	No
12	**Rayo Vallecano**	38	13	4	21	46	80	-34	43	No	No	No	No
13	**Getafe**	38	11	9	18	35	54	-19	42	No	No	No	No
14	**Espanyol**	38	11	9	18	41	51	-10	42	No	No	No	No
15	**Granada**	38	12	5	21	32	56	-24	41	No	No	No	No
16	**Elche**	38	9	13	16	30	50	-20	40	No	No	No	No
17	**Almeria**	38	11	7	20	43	71	-28	40	No	No	No	No
18	**Osasuna**	38	10	9	19	32	62	-30	39	No	No	No	Yes
19	**Valladolid**	38	7	15	16	38	60	-22	36	No	No	No	Yes
20	**Betis**	38	6	7	25	36	78	-42	25	No	No	No	Yes

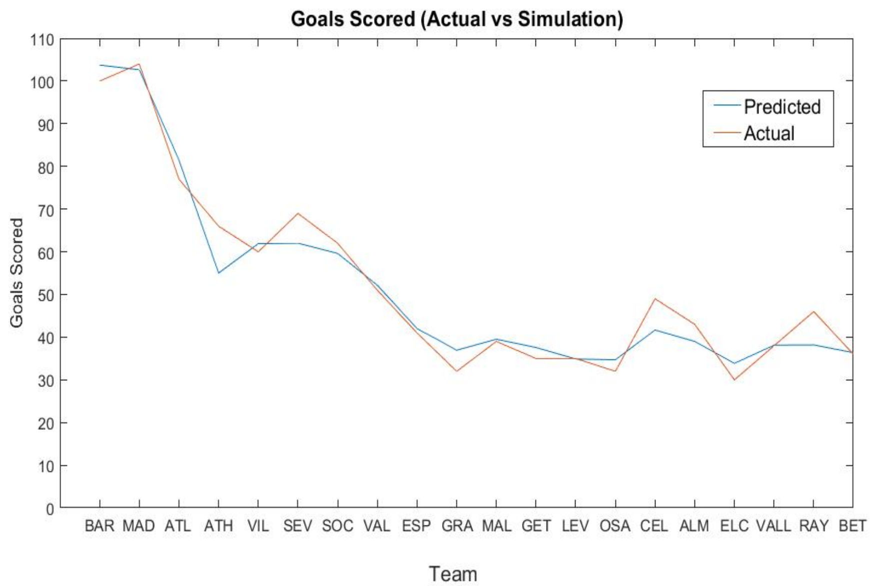


Figure 3.A15: Goals Scored (Actual vs Simulation).

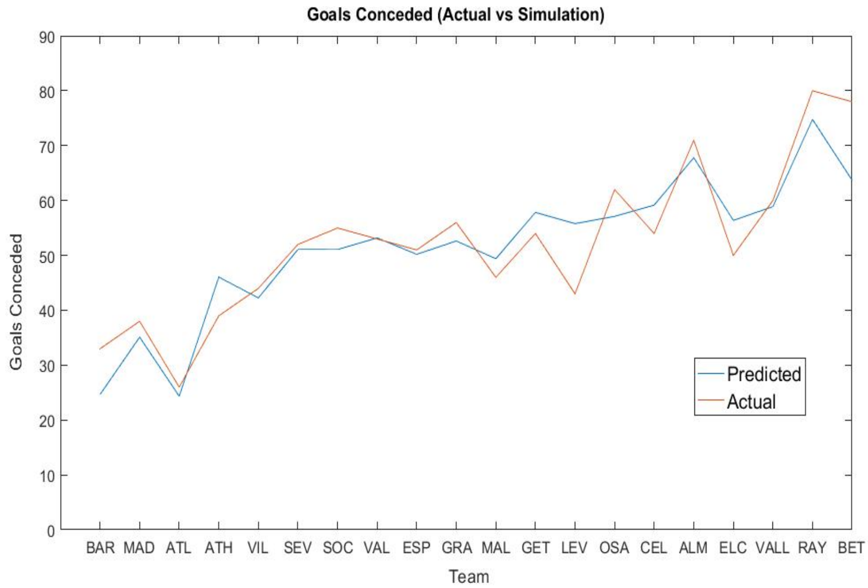


Figure 3.A16: Goals Conceded (Actual vs Simulation).

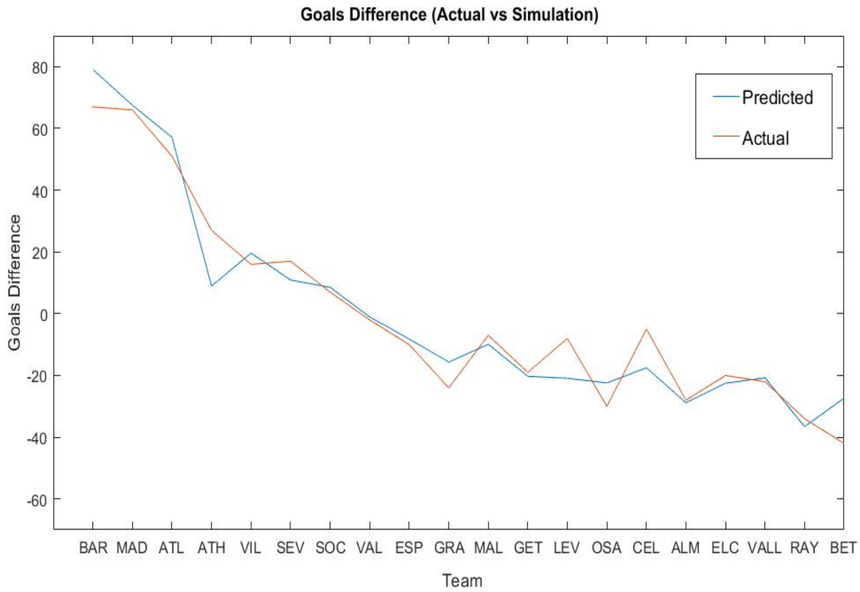


Figure 3.A17: Goals Difference (Actual vs Simulation).

B. Alternative models

Hierarchical Poisson-Bayesian Model with Gamma random effects priors and Normal-Gamma hyperpriors distribution.

Level 1 (Structural model). It contains the central part of the model.

$$X_{i,j}^{H} \mid \theta_i^H, \delta_j^A \sim Poisson((\theta_i^H + \delta_j^A) * F(t)) \qquad i = j = 1, 2, ..., 20; \qquad i \neq j \tag{3.57}$$

$$X_{i,j}^{A} \mid \theta_j^A, \delta_i^H \sim Poisson((\theta_j^A + \delta_i^H) * F(t)) \qquad i = j = 1, 2, ..., 20; \qquad i \neq j \tag{3.58}$$

$$X_{i,j}^{H} \mid \theta_i^H, \delta_j^A \perp X_{i,j}^{A} \mid \theta_j^A, \delta_i^H \qquad i = j = 1, 2, ..., 20; \qquad i \neq j \tag{3.59}$$

Level 2 (Hierarchical Normal Random Effect priors). In this level, the parameters of the estructural model (θ's and δ's) are modelled. We are assuming Gamma Random Effect priors for these parameters. Doing this, we are using the "assumption Parameters are constant over the season" (Section 4.5.1).

$$\theta_i^H \sim Gamma(\mu_1, \sigma_1) \qquad i = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.60}$$

$$\theta_j^A \sim Gamma(\mu_2, \sigma_2) \qquad j = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.61}$$

$$\delta_i^H \sim Gamma(\mu_3, \sigma_3) \qquad i = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.62}$$

$$\delta_j^A \sim Gamma(\mu_4, \sigma_4) \qquad i = 1, 2, ..., 20 \qquad k = 1, 2, ..., 38 \tag{3.63}$$

Hyperpriors distributions. Now we are defining the priors (also known as hyperpriors) for the parameters of the level 2. Following the literature (Best et al., 2011), we use non-informative Normal-Gamma hyperpriors.

$$\mu_1 \sim Normal(0, 0.0001) \tag{3.64}$$

$$\mu_2 \sim Normal(0, 0.0001) \tag{3.65}$$

$$\mu_3 \sim Normal(0, 0.0001) \tag{3.66}$$

$$\mu_4 \sim Normal(0, 0.0001) \tag{3.67}$$

$$\sigma_1 \sim Gamma(0.001, 0.001) \tag{3.68}$$

$$\sigma_2 \sim Gamma(0.001, 0.001) \tag{3.69}$$

$$\sigma_3 \sim Gamma(0.001, 0.001) \tag{3.70}$$

$$\sigma_4 \sim Gamma(0.001, 0.001) \tag{3.71}$$

On the Aggregation of Information in Centralized Prediction Markets.

Víctor Hernández García

November 2017

Resumen en Castellano

El principal objetivo de este trabajo es estudiar el proceso de agregación de información en los mercados predictivos. En particular, este artículo utiliza datos procedentes de Betfair Exchange. En primer lugar, se comprueba si los precios ofrecidos por el mercado estan bien calibrados. Tanto el test Hosmer-Lemeshow como las técnicas bootstrap indican que hay evidencia de falta de calibración. En segundo lugar, se realizan una serie de regresiones con el fin de investigar algunos de los sesgos más habituales en la literatura. Hay evidencias de "favourite long-shot bias", aunque su importancia es mucho menor debido probablemente a las particularidades del mercado predictivo analizado. A continuación, aplicamos el modelo propuesto en el segundo capitulo, tras añadirle una serie de refinamientos, con el objetivo de: (1) modelar "in-play prices", (2) evaluar la capacidad de los mercados para agregar información y actualizar los precios competitivos y (3) examinar la existencia de oportunidades de inversión. Los mercados predictivos en ocasiones tienen dificultades para agregar la información, en especial cuando gran cantidad de información es revelada en poco tiempo. El modelo planteado para modelar "in-play prices" presenta un mejor ajuste, medido por la distancia De Finetti, que los precios del mercado, a pesar de disponer de una menor cantidad de información. Por último, el algoritmo de inversión identifica estrategias rentables bajo ciertas circunstancias.

Palabras clave: Predicción, Mercados Predictivos, Distribución de Poisson
JEL classification: C5 C6 D4 D8

On the Aggregation of Information in Centralized Prediction Markets *

Víctor Hernández García†

November 2017

Abstract

The main goal of this paper is to study the process of aggregation of information in prediction markets. In particular, this article focus on high-frequency data from Betfair Exchange. Firstly, we analise whether market prices are well-calibrated. Both Hosmer-Lemeshow test and bootstrap methods indicate that there is evidence of miscalibration. Secondly, a number of regressions are conducted in order to identify potential biases in competitive prices. There is some evidence of "Favourite Longshot bias", although its magnitude is significantly smaller than that observed in the literature, probably due to the characteristics of the market. Thirdly, we apply the model developed in the second chapter, after adding some refinements, in order to: (1) model in-play soccer probabilities, (2) evaluate the prediction markets' ability to aggregate information and to update competitive prices and (3) examine the existence of investment opportunities. Markets predictions are very good at the beginning of the match and worsen as new information (goals) is incorporated. At the same time, the investment algorithm is able to detect opportunities for investment under certain circumstances. In conclusion, prediction markets, despite having a large amount of information, sometimes have problems aggregating information.

Keywords: Forecasting, Prediction Markets, Poisson distribution
JEL classification: C5 C6 D4 D8

*I thank José Ignacio Conde Ruiz and Marco Celentani for their help and support. I gratefully acknowledge the support from the Spanish Ministry of Science and Technology from grants ECO2011-30323-C03-01 & BES-2012-061597..

†E-mail: victorhzgarcia@gmail.com

4.1 Introduction

Prediction markets[1] are created for the purpose of trading the outcome of future events, so they can provide useful information about the probability distribution of the different outcomes. Although prediction markets exist from long time ago (Rhode & Strumpf, 2004), they have drawn considerable attention in the last years, both in the mass media and in the research community. In election time, media follow the evolution of political prediction markets, because they are often the most accurate predictors of election results. In the world of sports, prediction markets also attract media attention and sports enthusiasts get useful information from betting odds about the probability distribution of future sport events. At the same time, from a theoretical point of view, there is a growing literature about the functioning of these markets, with special attention to centralized prediction markets. Furthermore, prediction markets are an useful testing ground to check the validity of new models and statistical tools.

Although there is strong evidence that prediction markets can be a good information aggregation mechanism under certain assumptions (Efficient Market Hypothesis), it is necessary to point out some of their limitations. Firstly, prediction markets accuracy may depend on market liquidity (Tetlock, 2008 and Christiansen, 2007.). Secondly, competitive equilibrium is not very informative when public information is selective, inaccurate or misleading (Wolfers & Zitzewitz, 2004a). Thirdly, there is strong evidence of time horizon limitations of prediction markets (Antweiler, 2012). It is really hard to create a prediction market for events with a long time horizon because opportunity costs discourage potential investors. Finally, participants may have incentive to manipulate market prices in order to influence the resulting decision, although these manipulation attempts are unlikely to be successful even in the short term (Hanson & Oprea, 2005).

The main goal of this article is to study the process of aggregation of information and beliefs in prediction markets, in other words, how prediction markets incorporate new information and update competitive prices. Additionally, the following issues related to prediction markets are addressed. Firstly, one might wonder whether prediction markets produce well calibrated probability forecasts or not. Throughout this article, we will use data from a big-scale prediction market (Betfair Exchange) to answer this question. We will also check if prediction markets information can be systematically used to increase the accuracy of forecasting models. To to this, we will propose a prediction market-based non-homogenous independent Poisson model. Moreover, we are also interested in market efficiency, so the existence of investment opportunities will be examined.

Our initial hypothesis is that markets are not able to aggregate information correctly under certain circumstances, such as the presence of biases, the lack of liquidity or when bettors face very unlikely events, etc. In these cases prices do not reflect all the available information and Efficient Market Hypothesis does not apply.

[1] Also known as information markets, virtual markets or idea futures.

The remainder of this paper is organised as follows. Firstly, in Section 4.2, a brief review of the literature on "prediction markets" is presented. Then, in Section 4.3, we describe the contain of our database and we also introduce some descriptive analysis. In Section 4.4, we analise whether market prices are well-calibrated, using Hosmer-Lemeshow test and bootstrap methods. In Section 4.5, a number of regressions are carried out in order to identify potential biases in prediction markets, such as favourite long-shot bias, home-team bias, etc. In Sections 4.6 and 4.7 we propose a forecasting algorithm based on high-frequency markets data in order to get more accurate predictions. We assume that goal scoring can be modelled using a non-homogeneous independent Poisson model. In Section 4.8, the predictive power of the model is tested using De Finetti measure. At this point, we use market predictions as a reference to evaluate the performance of the model. At the same time, a high-frequency investment algorithm is designed in order to detect opportunities for investment. Finally, in Section 4.9, conclusions are presented.

4.2 Related Literature

Our work is related to the literature on prediction markets, which has grown significantly over the last years taking advantage of the availability of big-scale databases. Tziralis & Tatsiopoulo (2012) analyse the evolution of research on prediction markets (PM), providing an extended literature review that consists of more than one hundred and fifty articles, all of them published between 1990 and 2006.

Servan Schreiber et al. (2004) evaluates the relative accuracy of two different types of prediction markets: markets based on real money and markets based on play-money. Results show that the play-money markets perform as well as the real money markets. The authors speculate that this strange finding may reflect two opposing forces. On the one hand, real money markets generate incentives to discover new information. On the other hand, play-money markets may aggregate information in a more efficient way.

Hanson & Oprea (2005) adapts a Kyle-style market microstructure model to check whether prediction markets accuracy is affected by manipulators. It was found that manipulators are not able to influence on competitive prices and typically they lose money. At the same time, the rewards for informed trading and the accuracy of prediction market prices increase due to the rise in the volume of transactions.

Wolfers & Zitzewitz (2006) finds prediction markets prices typically provide useful (albeit sometimes biased) estimates of average beliefs about the probability of future events.

Applying the "Efficient Market Hypothesis", prediction markets have been exploited to forecast all types of events. Election future markets have been used to predict election outcomes (Wolfers & Zitzewitz, 2004b, Berg et al., 2008, Berg & Rietz, 2014). In the public sector, the Pentagon[2] has used prediction markets data sets in order to forecast geopolitical risk (Hanson & Oprea, 2005). In the health sector, prediction markets have been applied

[2]FutureMAP project

to forecast infectious diseases outbreaks (Polgreen et al., 2007). Recently, Cowgill & Zitzewitz (2015) analyzes the efficiency and performance of corporate prediction markets, including markets that forecast economic activity, demand, corporate sales, success of new products, etc.

Our work is also related to the literature on sport economics and more specifically to the literature on modelling soccer data. A number of studies have tried to model soccer goal scoring. Most of the articles use past performance as prior information to estimate the models (Maher, 1982, Dixon & Coles, 1997, Karlis & Ntzoufras, 2003, Everson & Goldsmith-Pinkham, 2008). Bueno et al. (2010) proposes a soccer model based on past performance and also on experts' opinions.

Some of the articles on sport economics are closely linked to the search for arbitrage opportunities. Dixon & Robinson (1998) propose a two-dimensional birth process to investigate setting prices in the spread betting market. This study finds evidence of inaccuracies in spread betting prices. Taking a different approach, Kuypers (2000) introduces a model of bookmaker behaviour and look for profitable opportunities in fixed odds betting markets using an ordered Probit. It provides evidence against the efficient market hypothesis and offers a betting rule to make money beating the bookie. Vlastakis et al. (2009) uses a Poisson count model and a multinomial Logit model to investigate profitable arbitrage opportunities. It also finds that the fixed-odds betting market are not completely compatible with weak-form market efficiency.

4.3 Data

In this article, we use the same database as in Hernandez (2017a). The dataset contain Betfair Exchange data for Spanish Football League (Liga 2013-2014)[3]. Betfair Exchange is an online centralized marketplace where agents can place back bets or they can play the role of traditional bookmakers laying a selection. Betfair acts as the intermediary and generates revenue charging commissions on net winnings[4]. It was impossible to choose a more recent league, as in November 2014 Betfair updated the application for downloading data[5], so the Liga 2013-2014 is the latest available.

The database consist of fully time-stamped data and includes pre-play and in-play information for several markets: Match Odds, Correct Score, Over-Under betting, Half Time Score, Next Goal... Our raw data set contain approximately 210 millions of observations and information for a total of 2680 markets[6], matching more than 1000 millions of euros in total. The frequency is one observation per second and the following information is included for each selection: best three back odds and its volume, best three lay odds and its volume, last price matched, total matched, market status and timestamp in deciseconds. Table 4.1 presents some descriptive statistics regarding the database:

[3]The data was downloaded from www.fracsoft.com

[4]The top rate is set at 5%.

[5]Betfair API 6.0

[6]On average, 7 markets per match

Table 4.1: Descriptive statistics (Betfair Exchange)

	Match Odds			Correct score			Other markets [a]		
	Total	**Mean**	**Median**	**Total**	**Mean**	**Median**	**Total**	**Mean**	**Median**
Observations (Total) [b]	22.8	0.6	0.4	136.8	0.36	0.16	50	0.13	0.08
Observations (Pre-play)	15.5	0.4	0.31	93	0.24	0.18	34	0.09	0.07
Observations (In-play)	7.3	0.2	0.15	43.8	0.12	0.10	16	0.04	0.03
Observations (Active)	20.1	0.53	0.52	126	0.33	0.33	45	0.11	0.12
Total Matched [c]	802	2.1	0.5	106	0.28	0.19	76	0.2	0.2
No. of matches	380			380			380		

[a] They include: half time correct score, over/under markets, correct score away, correct score home, next goal and both teams to score.

[b] The data are expressed in millions of observations.

[c] The data are expressed in millions of euros.

Figure 4.1 shows a density histogram of probabilities for Match Odds markets. The probabilities are obtained computing the inverse of the match odds and normalizing. The observations are then divided into 100 equal size groups. Match odds markets have three possible outcomes, so the mean value is 1/3. This explains why the distribution is right-skewed. It can be seen that probabilities tend to be concentrated around zero and around the mean. Note that about 13% of the probabilities are between 0% and 1% and 0% of the market probabilities are between 99% and 100%. The latter is explained by the fact that people are not willing to back at less than 1.01.

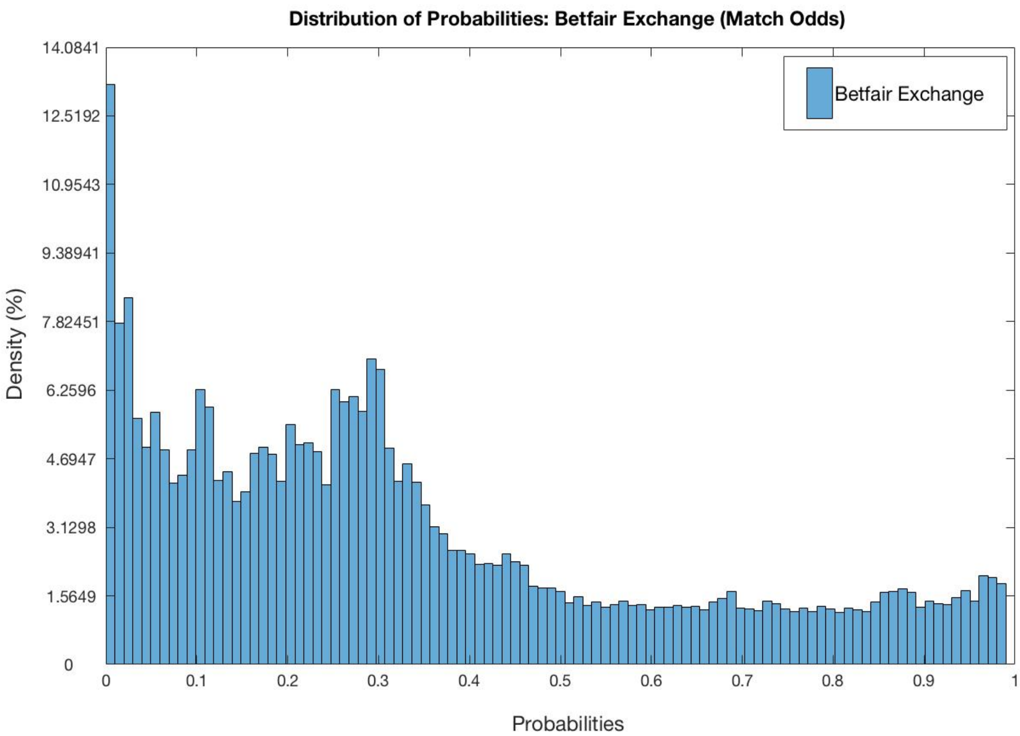


Figure 4.1: Distribution of market probabilities (Match odds)

4.4 Calibration.

Several studies have addressed the calibration of prediction market prices. Page & Clemen (2012) uses data from political and sport prediction markets in order to find evidence of (mis)calibration. They found that prediction markets are well calibrated in general when time to expiration is relatively short, but miscalibration is observed for events with a long time horizon. Christiansen (2007) discusses calibration in small-scale prediction markets (rowing). The study concludes that small markets are very well calibrated and determines a potential minimum threshold of participation to guarantee well-calibrated probabilities.

In this section, we study the calibration of the Betfair Exchange implied probabilities, after normalizing. We focus on the Match Odds markets because is the most liquid. Firstly, we discretize observations (market probabilities) into categories. Then observations are divided into 10 groups with equal range[7]. Secondly, we compute the mean for each group. Thirdly, we look at the actual outcomes (0 or 1) and compute the observed mean for each group. Finally, Figure 4.2 presents the calibration plot for our sample.

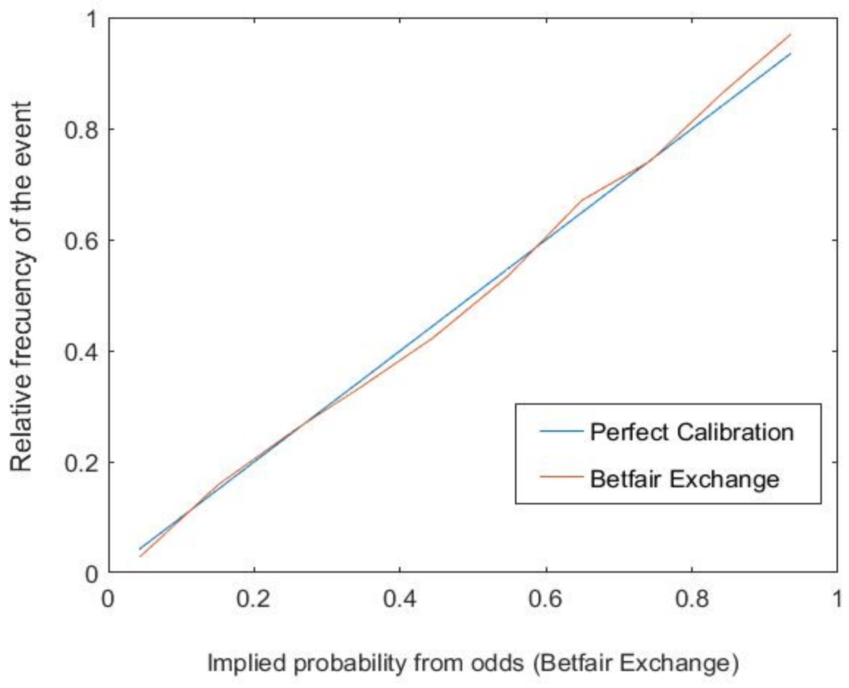


Figure 4.2: Calibration plot (Betfair Exchange)

Small deviations from the 90-degree line (perfect calibration) can be observed, mainly in the upper area. But, graphically we can not evaluate whether or not deviations are statistically significant. Then further analysis is needed.

[7] [(0, 0.1), (0.1, 0.2), (0.2, 0.3), (0.3, 0.4), (0.4, 0.5), (0.5, 0.6), (0.6, 0.7), (0.7, 0.8), (0.8, 0.9), (0.9, 1)]

Testing Calibration: Some tests are carried out to check whether our data is well calibrated.

i) **Hosmer-Lemeshow test**. The Hosmer-Lemeshow test is widely used to test calibration[8], in the context of the logistic regression models. The test evaluate whether the expected event rate match observed event rates in subgroups of the model population. Then this framework is perfectly applicable to our case.

In this section, we focus only on in-play probabilities[9], so we have about ten millions of observations. Firstly, before performing the test, we fit the market probabilities using a standard logistic regression, where the independent variable is the market probability and the dependent variable is the actual outcome for each observation (0 or 1). The curve obtained is almost a straight line. Secondly, we create ten groups[10] for the fitted probabilities as in the previous case. Finally, we carry out a Hosmer-leshow test using Stata software, setting ten groups. Table 4.2 presents the Stata output of the Hosmer-Lemeshow test:

Table 4.2: Hosmer-Lemeshow test for calibration

	Hosmer-Lemeshow Test
Number of observations	10,650,300
Number of groups	10
Hosmer-Lemeshow χ^2_8	1665
Prob $>\chi^2$	0.0001

Hosmer-Lemeshow test indicates that there is strong evidence of miscalibration, because the p-value is less than 0.05. These results should not be surprising since with such a big sample any deviation is statistically significant. We can conclude that Betfair Exchange implied probabilities are not well-calibrated. Note that Hosmer-Lemeshow test is for overall calibration. Then it is not valid to assess in which range of probabilities there are (mis)calibration. To answer this question we will use bootstrap methods in next section.

[8]For a comprehensive review of the test, see Lemeshow & Hosmer(1982) and for an application, see Kramer & Zimmerman (2007).

[9]In fact, pre-play probabilities are almost constant if we looking at a specific match, so they do not have much interest.

[10]Results remain constant if we change the number of groups.

ii) **Bootstrap Methods**. Page & Clemen (2012) propose an alternative approach to check whether prediction markets produce well-calibrated probabilities: ”a non-parametric estimation of the calibration curve using bootstrap[11]”. We also apply empirical bootstrap methods in order to compute a confident interval for our calibration curve. The bootstrap proceeds by resampling with replacement from our empirical distribution. Following the suggestion in Orloff & Bloom (2013), the resample size equals sample size. Figure 4.3 shows a 95% confident interval for our calibration curve[12]. Results suggest the presence of miscalibration when probability is greater than 80%, or in other words, when match odds are less than 1.25. For instance, a market probability of 95% is associated with a actual probability of 96.5%. The results are ambiguous with respect to the existence of favourite long-shot bias. On the one hand, favourites are slightly undervalued, so this is consistent with favourite long-shot bias. On the other hand, there is no evidence that bettors overvalue underdogs. Then further analysis about the existence of favourite long-shot bias is required. In the next section, several regression will be carried out to check the presence of this bias.

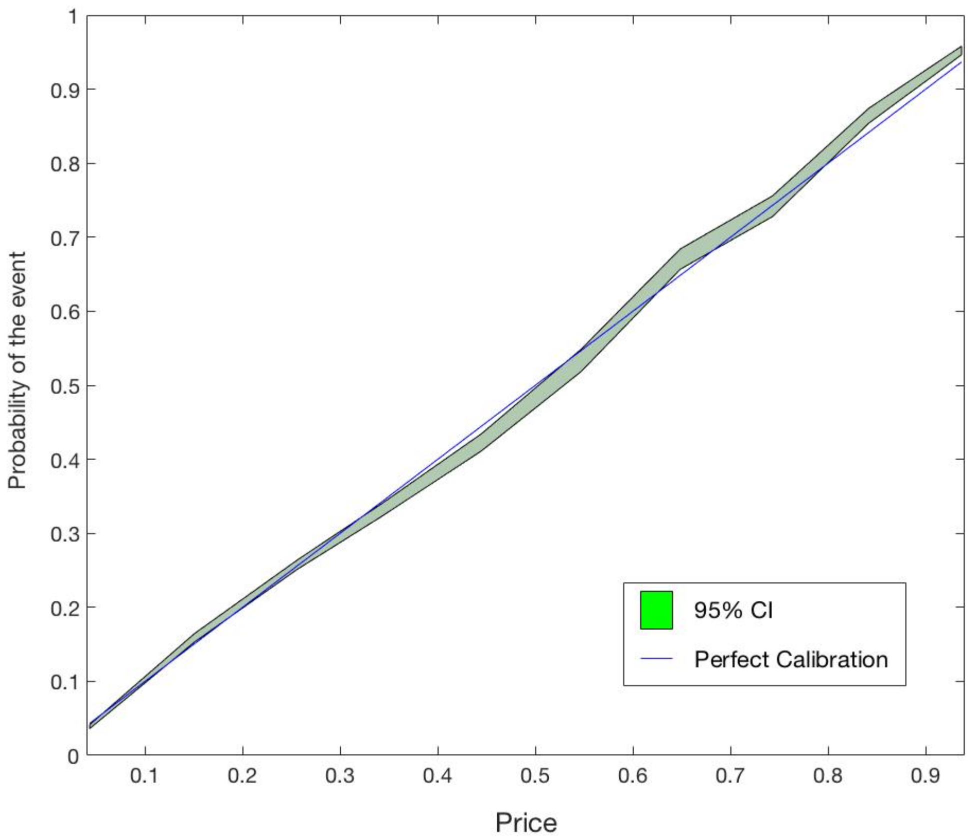


Figure 4.3: 95% interval for the calibration curve

[11]For a review of bootstrap methods, see Hardle (1990) and Davison (1997).
[12]We use the software Matlab to carry out bootstrap.

4.5 Odds biases.

In the betting literature, the favourite-longshot bias is probably the most analized and documented bias. It is probably the main source of sport betting markets inefficiency. This bias refers to the tendency of bettors to undervalue favourites and overvalue long-shots. Several empirical studies have found evidence of this bias in a variety of fixed-odds betting markets.

Two different explanations have been proposed to explain the existence of favourite long-shot bias: risk love models and misperceptions models (Snowberg & Wolfers, 2010):

The neoclassical approach focus on rational punters with (al least locally) risk-loving utility functions. According to this theory, on the demand side, gamblers (on average) are willing to accept lower expected returns betting on long-shots, i.e, the riskiest investment. At the same time, on the supply side, risk neutral profit-maximizing bookmakers know this fact and implement a strategy to take advantage of gambler's preferences, setting unfavourable odds on underdogs.

On the other hand, competing behavioural theories stress the role of misperceptions of probabilities. According to this explanation, cognitive errors explain the existence of the favourite long-shot bias. Several studies in the field of psychology have shown that humans are systematically poor at discerning between small and tiny probabilities. Sobel & Raines (2003) design an information model and suppose that Bayesian bettors start with a non informative prior (equal probabilities). They update their prior distributions according to private information. As a result, long-shots are overestimated and favourites underestimated.

From a empirical point of view, several studies have tried to identify the favourite long-shot bias in prediction markets. In general, the presence of favourite long-shot bias is widely accepted in the literature. Page & Clemen (2012) identify favourite long-shot bias in political and sport prediction markets using a non-parametric estimation of the calibration of market prices. At the same time, the magnitude of the bias is greater in long-term markets. Deschamps & Gergaud (2012) study the efficiency of soccer betting in England (from 2002 to 2006). As a conclusion, they found a positive favourite long-shot bias in both home and away odds, but surprisingly results indicate a negative favourite long-shot bias for draw odds. Constantinou & Fenton (2013) detect strong presence of favourite long-shot bias examining prediction markets based on 14 European football leagues. Vlastakis et al. (2009) confirm the existence of favourite long-shot bias and also find profitable betting opportunities.

Another bias that appears in the literature of sport economics is the home bias, defined as the tendency of bettors to overvalue home chances. Graham & Stott (2008) and Constantinou & Fenton (2013) identify a clear home-bias using football data.

Finally, a less common bias is the "immediate strike back bias". It refers to the tendency of bettors to overvalue comeback chances. Dixon & Robinson (1998) finds no evidence for the immediate strike back bias.

Figure 4.4 shows expected returns for different intervals of market probabilities. As in the literature, it can be observed that low probabilities (high odds) are associated with low expected returns, while the opposite occur with respect to high probabilities (low odds). Specifically, in Betfair Exchange bets with an associated probability less than 10% and 20% make a expected return of -4.2% and -1.8%, respectively. Note that the average return in Betfair exchange is about -2%.

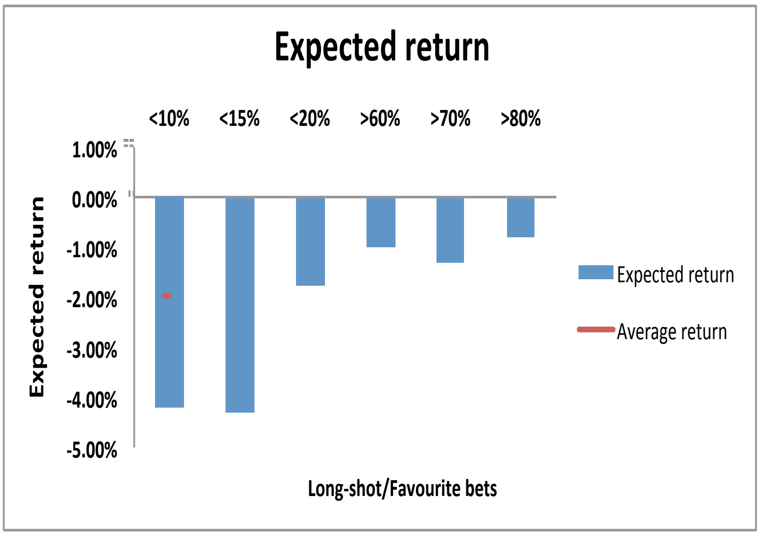


Figure 4.4: Favourite Long-shot bias

However, in terms of magnitude, these results contrast with the literature, because they are conclusive about the existence of the bias. For instance, Constantinou & Fenton (2013) find that bets with an associated probability less than 10% and 20% generate a expected return of -40% and -30%, respectively (the average return in those markets is about -8%). Andrew (2012) analizes tennis data and finds that odds greater than 10 (probabilities less than 10%) offer a expected return of -69%.

It should also be taken into consideration that literature usually focus on fixed-odds markets and pre-play bets, while we are considering a Betting Exchange market and in-play bets. Fixed-odds markets charge higher commissions than Betting Exchanges and the prices are not generated by competitive equilibrium, but they are fixed by the bookmakers. Theoretical models suggest that favourite long-shot bias is sometimes associated with bookmaker strategic behaviour. Bookmakers do not proportionally share the commission between all the possible events because it is not the strategy that maximizes their profits. Bookmakers seeks a balanced book and take into account bettors' behaviour while setting odds.

At the same time, in our market (Betfair Exchange) prices are competitively determined and bookmakers are only intermediaries, so the context is totally different. The characteristics of the market may explain the sharp reduction in the favourite long-shot bias.

Now, following the methodology described in Stanek (2016), we propose a regression to identify both favourite long-shot and home bias. The functional form of the econometric model is as follows:

$$Outcome_i = \beta_0 + \beta_1 * log(Odds_i - 1) + \beta_2 * Home_i + u_i \tag{4.1}$$

where Outcome is a binary variable that has a value of 1 if the bet was successful and 0 otherwise, Odds are the market odds and home is a dummy value taking the value 1 if the bet is on the victory of local team and 0 otherwise.

Interpretation of the coefficients: We expect $\beta_1 < -1$ in presence of favourite long-shot bias, i.e., bets on favourites are more profitable than bets on long-shots. Likewise, $\beta_1 > -1$ in case of negative favourite long-shot bias. The negative value of β_2 indicate the presence of home bias.

Table 4.3 shows the regression output. Column 1 presents the result for the entire sample (8,607,330 observations). In order to analyze whether the results depend on liquidity, we also divide the sample into two groups according to the level of liquidity, using the median value as threshold. Column 2 and 3 show the regression output for moments of low and high volume, respectively. P-values are reported in brackets. The reported p-values for the variable FLSbias consider the null hypothesis $\beta_1 = -1$ instead of $\beta_1 = 0$.

Table 4.3: Reggresion Analysis: Favourite Long-shot bias & Home bias

		(1)	(2)	(3)
EQUATION	**VARIABLES**	**Total**	**Low Volume**	**High Volume**
Outcome	**FLS_bias**	-1.020***	-1.051***	-0.992
		(0.0082)	(0.000)	(0.4407)
	Home_bias	0.0853***	-0.0248	0.191***
		(0.000)	(0.347)	(0.000)
	Constant	-0.0141	0.0594***	-0.0865***
		(0.189)	(0.000)	(0.0151)
	Observations	8,607,330	4,311,720	4,295,610
	Robust standard errors in parentheses	***p <0.01; **p <0.05; *p <0.1		

When we consider all the observations, there is evidence of the favourite long-shot bias. Moreover, it can be observed that the bias becomes more intense in times of low liquidity and disappear when the volume of transactions is high, then liquidity matters. However, the favourite long-shot bias is less intense than in Stanek (2016), since in regression $\beta_1 < -1.15$. At the same time, home bias is statistically significant in Regression 1 and 3, but when we compute the marginal effects at the mean, the magnitude is not economically relevant.

4.6 Model Assumptions.

Some assumptions are required in order to develop our time-inhomogeneous independent Poisson model. Before that, we introduce some notation:

In a match between teams indexed i (home team) and j (away team), we define the following variables:

- Let $X^H_{i,j}$ represents the goals scored by team i against team j, or equivalently, the goals conceded by team j against team i.
- Let $X^A_{i,j}$ represents the goals conceded by team i against team j, or equivalently, the goals scored by team j against team i.

4.6.1 The Poisson Assumption:

Assuming Poisson distributions for home and away goals is a classical assumption in the literature for modelling soccer probabilities (See Everson & Goldsmith-Pinkham, 2008, Bueno et al., 2010)

In Hernández (2017a) several test are carried out in order to check whether Poisson distribution fits soccer data (Liga 2013-2014). Results find no evidence against the Poisson assumption. That is the reason why we assume for the rest of the article that home-team and away-team goals follow a Poisson distribution with parameters $\lambda^H_{i,j}$ and $\lambda^A_{i,j}$, respectively.

$$\text{Assumption 1:} \quad X^H_{i,j} \sim Poisson(\lambda^H_{j,i} * t) \quad ; \quad X^A_{i,j} \sim Poisson(\lambda^A_{j,i} * t) \tag{4.2}$$

In the above expression t stands for "match time". This variable measure the percentage of the time[13] elapsed from the beginning of the match and ranges from 0 (the start of the match) to 1 (the end of the match). Then in the middle of a match, t equals 0.5.

4.6.2 The Independence Assumption:

In the literature, the decision of including the assumption of independence when modelling soccer data is widely debated. On the one hand, there is some evidence that the scores of the two competing teams are not independent (Maher, 1982, Dixon & Coles, 1997; Karlis & Ntzoufras; Hernandez, 2017a). On the other hand, empirical studies show that in practice the correlation between the goals of the two competing teams is small, usually less than 0.05 (Karlis & Ntzoufras, 2003).

In view of the results and for simplicity, from now on we assume the Independence Assumption, defined as follows:

$$\text{Assumption 2:} \quad X^H_{i,j}(t) \quad \perp \quad X^A_{j,i}(t) \quad i \neq j \tag{4.3}$$

[13] Given as a fraction of unity.

4.6.3 Time-inhomogeneous Poisson Process:

There is strong evidence that the probability of a goal being scored steadily increases over the course of the match, perhaps because of tiredness of players. (See Dixon & Robinson (1998)). Figure 4.5 show the sample goal distribution by minute (Liga 2013-2014). It can be observed (adjusted line) an increase in the frequency of goals scored as a match progresses. Figure 4.6 shows the sample cumulative distribution of goals (Liga 2013-2014).

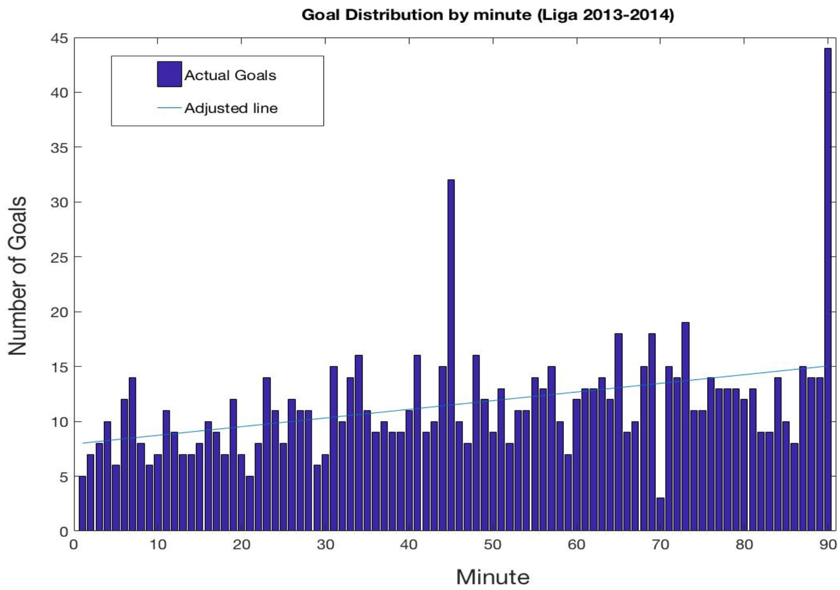


Figure 4.5: Goal distribution by minute

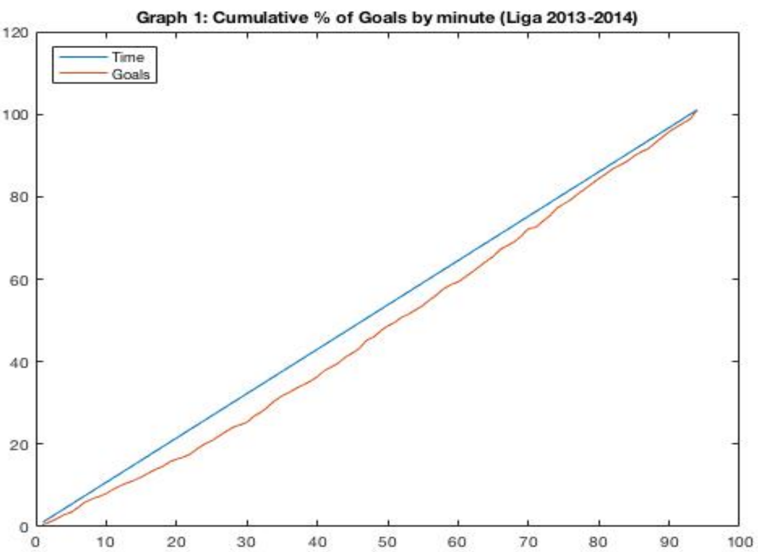


Figure 4.6: Cumulative (%) of goals by minute (LIGA 2013-2014)

In the light of this results, it is necessary to allow the intensity of the poisson process to vary over the course of the game. We need a time-inhomogeneous Poisson process, so we have to incorporate some refinements to the model proposed in Hernández (2017a). To do this, we use historical goal distribution data and we assume the same kind of adjustment for all teams. We also assume, for simplicity, that the intensity does not depend on the current score. We propose two possible methods to model time-inhomogeneous Poisson processes:

i) Method 1: Adjusting Match Time: One possibility is to take into account historical (cumulative) goal distribution to adjust match time according it. Note that this approach is analogous in some sense to "capital per unit of effective labour", used in some Macroeconomics Growth Models, simply that in this case we consider "effective unit of time". In order to apply this method, we introduced the function F (historical cumulative goal distribution). This function converts standard % of match time elapsed (t) into effective % of time elapsed (t^E), using the historical cumulative goal distribution.

$$\text{Assumption 4:} \qquad \text{F} : [0,1] \to [0,1]; \qquad t^E = F(t) \tag{4.4}$$

The following example tries to illustrate how this method works. We focus on minute 45, so we are exactly in the middle of the match. If we assume that the poisson process is time-invariant, we just consider that 50% of the time remains. But, as we know that the probability of a goal being scored steadily increases over the course of the match, we apply the F function. In this case, we obtain F(0.5)=0.43 (43%), i.e., only 43% of the time has been played, since looking at the cumulative distribution of goals we can observe that in the first half only 43% of the goals have been scored. In this way, we have used the cumulative goal distribution to adjust the model.

From now on, we assume goals in a football match follow the following time-inhomogeneous Poisson distributions:

$$\text{Home team:} \quad X^H_{i,j}(t) \sim Po(\lambda_{i,j} * F(t)) \qquad \text{i=j=1,...,20} \tag{4.5}$$

$$\text{Away team:} \quad X^A_{i,j}(t) \sim Po(\lambda_{i,j} * F(t)) \qquad \text{i=j=1,...,20} \tag{4.6}$$

where F uses the cumulative goal distribution to convert match time (minutes) into "effective percentage of time elapsed".

ii) Method 2: Estimating a parametric time-variant Poisson process: We assume that the Poisson parameters can be modelled as a linear function of time ($\lambda^H_{i,j}(t)$, $\lambda^A_{i,j}(t)$). Specifically, we assume: $\lambda^H_{i,j}(t) = \bar{\boldsymbol{\lambda}}^H_{i,j} * (a + b * t)$ and $\lambda^H_{i,j}(t) = \bar{\boldsymbol{\lambda}}^H_{i,j} * (a + b * t)$. Then, $\bar{\boldsymbol{\lambda}}^H_{i,j}$ and $\bar{\boldsymbol{\lambda}}^A_{i,j}$ can be interpreted as the average intensity over the match. In this case, goals in a football match follow the following time-inhomogeneous Poisson distributions:

$$\text{Home team:} \quad X^H_{i,j}(t) \sim Po(\bar{\boldsymbol{\lambda}}^H_{i,j} * (a + b * t) * t) \qquad \text{i=j=1,...,20} \tag{4.7}$$

$$\text{Away team:} \quad X^A_{i,j}(t) \sim Po(\bar{\boldsymbol{\lambda}}^A_{i,j} * (a + b * t) * t) \qquad \text{i=j=1,...,20} \tag{4.8}$$

We can use the historical goal distribution in order to estimate the parameters a and b. We expect $\hat{b} > 0$, since given $t_1 < t_2$, we expect that goal intensity will be greater in the second case. In addition, $\hat{a}$ must be greater than zero, because Poisson distribution does not admit negative values.

In this article, we will use "Method 1" mainly for two reasons: it is more intuitive and easy to implement.

4.7 Model for predicting in-play probabilities

In this section, we propose a time inhomogeneous independent Poisson model to forecast in-play prices (probabilities). In order to model in-play probabilities two steps are required. Firstly, we estimate the goal scoring intensities at the beginning of the match ($\lambda's$) for the two competing teams. In this part, we make use of: Assumption 1 (Poisson distribution), Assumption 2 (independence), Assumption 3 (Time-inhomogeneous Poisson distribution) and the PMF of a Poisson distribution. Additionally, we also use data from betting markets (Betting Exchange), in particular we look at the odds at the beginning of the match. In this phase, we estimate the goal scoring intensities for all league matches. Secondly, we use the previous estimates ($\lambda's$) and our time inhomogeneous independent Poisson model to forecast systematically in-play probabilities.

4.7.1 Step 1: Obtaining goal scoring intensities ($\hat{\boldsymbol{\lambda}}^H_{i,j,t}$, $\hat{\boldsymbol{\lambda}}^A_{i,j,t}$)

Firstly, it is necessary to calculate the inverse of the betting odds in order to obtain a probability measure. It is easy to check that for any match the sum of the odds of the three possible outcomes (home win, draw or away win) is more than 1, indicating that odds are not fair [14]. Next, we proceed to normalise the odds, so that the sum equals 1.

Secondly, we will also need the Probability Mass Function (PMF)[15] for a Poisson distribution, defined as:

$$\text{P(k events in the interval)} = \frac{e^{-\lambda}\lambda^k}{k!} \tag{4.9}$$

where λ is the average number of events per interval and k takes values 0, 1, 2...

[14] In fact, sample mean is about 1.018

[15] Discrete equivalent of a pdf.

If we apply this formula to the soccer context, the probability of observing a particular final score (k, m) is defined by the following joint mass probability function:

$$\text{P(team i scores k goals, team j scores m goals)} = \frac{e^{-(X_{i,j}^{H})}(X_{i,j}^{H})^{k}}{k!} \times \frac{e^{-(X_{i,j}^{A})}(X_{i,j}^{A})^{m}}{m!} \tag{4.10}$$

where $X_{i,j}^{H}$ and $X_{i,j}^{H}$ are the goal scoring parameters of team i and j, respectively, when playing at team i home. k takes values 0, 1, 2... and m takes values 0, 1, 2...

At this point some notation is required. In every moment of a match, let w be the current home team score and q the current away team score. Let t be match time (% of total time elapsed). Let Θ_H^t be the set of possible final home team scores conditioned to the current score (at time t). Analogously, let Θ_A^t be the set of possible final home team scores conditioned to the current score (at time t). For example, if the score at time t is (w=1, q=2), $\Theta_H^t = [1, 2, ..., \infty]$ and $\Theta_A^t = [2, 3, ..., \infty]$. Let $\hat{\boldsymbol{\lambda}}_{i,j,t}^{H}$ be the goal scoring parameter of team i against team j in time t. Let $\hat{\boldsymbol{\lambda}}_{i,j,t}^{A}$ be the goal scoring parameter of team j against team i at time t. Let $a_{i,j}^{t}$ and $b_{i,j}^{t}$ be the Betfair Exchange implicit probability of draw and local victory at time t, respectively.

Using the implicit probabilities of Betfair Exchange and the PMF formula, we can solve the following implicit system of equations[16] for every observation[17] of every match:

<u>Equation 1:</u>

$$0 = \sum_{k \subseteq \Theta_A^t} \sum_{\substack{m=k \\ m \subseteq \Theta_H^t}} \frac{e^{-(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^{H}}{1-F(t)})}(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^{H}}{1-F(t)})^{(k-w)}}{(k-w)!} \times \frac{e^{-(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^{A}}{1-F(t)})}(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^{A}}{1-F(t)})^{(m-q)}}{(m-q)!} - a_{i,j}^{1} \qquad \substack{t \subseteq [0,1] \\ i=j=1,...,20} \tag{4.11}$$

<u>Equation 2:</u>

$$0 = \sum_{k \subseteq \Theta_A^t} \sum_{\substack{m<k \\ k \subseteq \Theta_A^t}} \frac{e^{-(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^{H}}{1-F(t)})}(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^{H}}{1-F(t)})^{(k-w)}}{(k-w)!} \times \frac{e^{-(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^{H}}{1-F(t)})}(\frac{\hat{\boldsymbol{\lambda}}_{i,j,t}^{H}}{1-F(t)})^{(m-q)}}{(m-q)!} - b_{i,j}^{1} \qquad \substack{t \subseteq [0,1] \\ i=j=1,...,20} \tag{4.12}$$

The system is perfectly identified, so every pair ($a_{i,j}^{t}$ and $b_{i,j}^{t}$) identify one and only one pair ($\hat{\boldsymbol{\lambda}}_{i,j,t}^{H}$, $\hat{\boldsymbol{\lambda}}_{i,j,t}^{A}$). Solving all the systems of equations, we find the pair of unknown ($\hat{\boldsymbol{\lambda}}_{i,j,t}^{H}$, $\hat{\boldsymbol{\lambda}}_{i,j,t}^{A}$) for every second of every match, so we have a entire distribution for every match and every team (ranging from 5600 to 5800 depending of the match duration).

[16]We solve the implicit systems of equations using FSOLVE command in Matlab.
[17]The frequency is one observation per second.

Figure 4.7 shows a particular example, the match between Getafe and Valladolid, corresponding to Week 22 (Liga 2013-2014). Using the time-inhomogeneous independent Poisson model, goal scoring intensities are estimated for each second of the match. At the beginning of the match, Getafe's goal scoring intensity is 1.3[18] and Valladolid's goal scoring intensity is 0.8. Note that Getafe is playing at home and markets reflect this home advantage.

Once the match begins, bettors receive a lot of information about the match (goals, teams' performance, motivation, injuries, tiredness, etc), so they update their beliefs about the quality and chances of both teams. That is the reason why goal intensity estimates are continually evolving as it is shown in Figure 4.7.

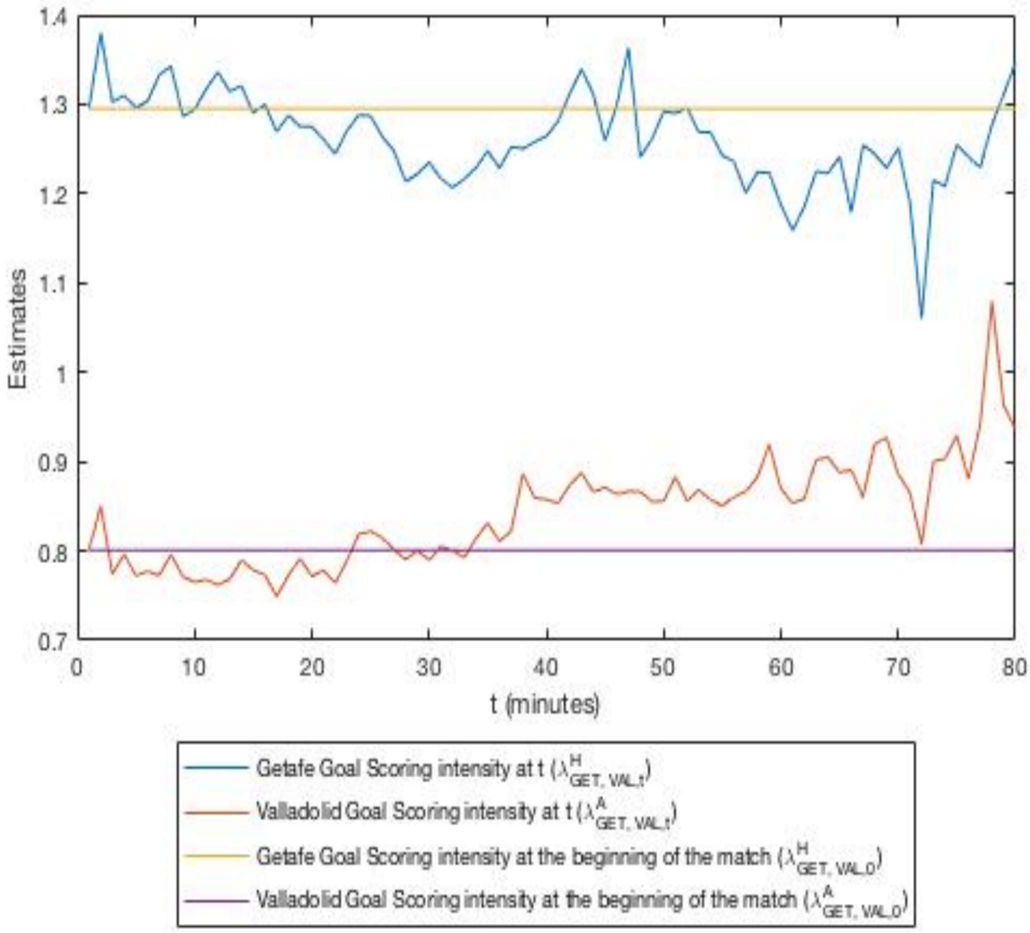


Figure 4.7: Goal scoring intensities over the match: Getafe-Valladolid (Week 22)

[18]Remember that goal scoring intensity can be interpreted as expected goals scored.

4.7.2 Step 2: Obtaining in-play probabilities

In order to forecast in-play probabilities we need the goal scoring intensities estimated in the previous section. In particular, we are going to use goal scoring intensities at the beginning of the match ($\hat{\boldsymbol{\lambda}}^H_{i,j,0}$, $\hat{\boldsymbol{\lambda}}^A_{i,j,0}$; $i \neq j$). We apply assumptions 1-3, then we have that the stochastic process that determine the goals scored by team i against team j follows a time-inhomogeneous independent Poisson process with parameter: $\hat{\boldsymbol{\lambda}}^H_{i,j,0}$. Note that when computing in-play probabilities we assume that we know match time and the current score. At this point, we can match any pair of teams, so we can obtain match probabilities for any possible match.

As an illustrative example, Figure 4.8 shows in-play match probabilities from both Betfair Exchange and the time-inhomogeneous independent Poisson model (Valencia vs Atletico). It can be observed that market probabilities have a lot of noise while the in-play probabilities derived from the model evolve smoothly.

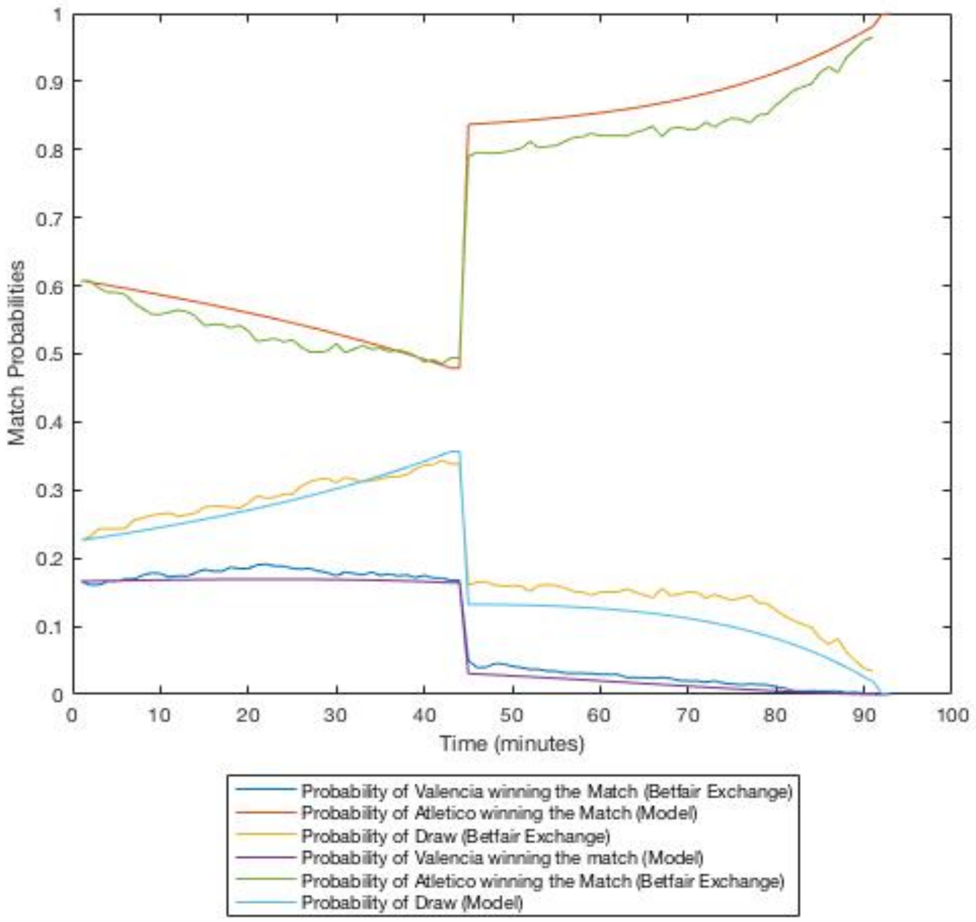


Figure 4.8: Match probabilities (Valencia vs Atlético)

4.8 Application: LIGA 2013-2014

Once we have developed a model to forecast in-play probabilities, we evaluate its performance. We are mainly interested in its predictive power, together with its ability to generate profitable investment opportunities. As an illustrative example, we use data from Liga 2013-2014 afresh.

4.8.1 Goodness of fit (De Finetti Measure):

We apply the time-inhomogeneous independent Poisson model proposed in the previous section in order to forecast in-play prices. As primary information, we just use the market odds at the beginning of the match, the match time and the goals scored by each team. Other information is not available for our model. Then we can apply this model to predict all the in-play match probabilities of the Liga 2013-2014, obtaining about 200 millions of in-play probabilities. At this point, we have two match probabilities for each second of each match: market probabilities (Betfair Exchange) and probabilities generated by our algorithm. The next step is to compare which are best.

To do this, we resort to De Finetti measure (De Finetti, 1972), which is defined as the Euclidean distance between the point correspondent to the outcome and the point correspondent to the prediction. For every second of every match of the Liga 2013-2014, we compute the De Finetti measure for both market predictions and model predictions. Note that a smaller De Finetti measure indicates better fit. As a way to test the robustness of the results, we repeat the previous exercise but considering this time Absolute Distance[19] instead of De Finetti measure.

Results are shown in Table 4.4. To facilitate their interpretation and to have a relative measure of performance (RP), results are reported in percentages, as indicated by the following formula:

$$RP = \frac{\text{Total sum of De Finetti measures (Model)}}{\text{Total sum of De Finetti measures (Betfair Exchange)}} \times 100. \tag{4.13}$$

Table 4.4: Relative Performance (Market vs Model)

	Relative Performance (Market vs Model)					
	De Finetti Measure			**Absolute Distance**		
	Total	**Liquidity**		**Total**	**Liquidity**	
		Low volume	**High volume**		**Low volume**	**High volume**
RP	99.8%	99%	100.9%	97.1%	96.9%	97.2%

[19] It reflects the distance on the real line between the actual outcome and the prediction.

According to Table 4.4, model's in-play probabilities are slightly better than market ones, given the De Finetti ratio is less than 100%. Table 4.4 also disaggregates the results according to the level of liquidity. When the liquidity is high, i.e, there are a lot of transactions, the predictive power of market predictions improve. So we can conclude that a certain level of liquidity is required to guarantee competitive prices are accurate.

Finally, Figure 4.9 compares De Finetti measures minute by minute. It can be observed that market performance worsen as match progresses. At the beginning of the match, model predictions and market predictions are more or less equally accurate. However, as match progresses and new information is revealed, it seems that prediction markets fails and the model presents better performance than the market. Finally, at the end of the match, differences become more intense.

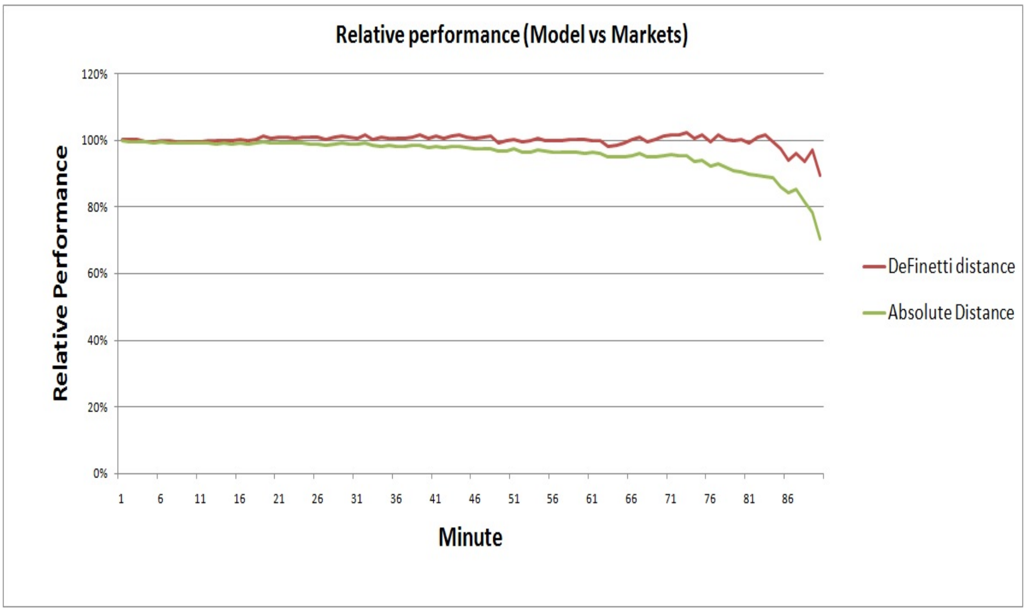


Figure 4.9: Relative Performance (minute by minute)

There is evidence that in-play probabilities estimated using our model are more accurate than markets probabilities, despite having much less information than the markets. Note that the market observe a lot of information: red cards, teams's performance, refereeing, tiredness, injuries, motivation, etc. It can be concluded that the prediction markets fail to aggregate the information in a proper way, at least when the level of liquidity is low. A simple model that uses odds at the beginning of the match, match time and goals is able to equalise or even to beat the market.

4.8.2 Beating the bookie:

In finance, a standard way to proof that markets do not aggregate information properly is to propose a profitable strategy to make money. (See Hogan et al., 2004; Draper& Fung, 2002)

In this part, a high-frequency investment algorithm is designed in order to detect opportunities for investment. For simplicity, we focus only on back bets, but it could be extended to lay betting. The algorithm works in real time using match time (t), current score and markets odds at the beginning of the match. The steps required are summarise below (base strategy):

-**Step 1**: At every second of every match, the algorithm receives information (input data) from the match: current score, market odds, volume of transactions, red cards, etc.

-**Step 2**: Using the model[20] and the input data mentioned before, the algorithm estimate in-play probabilities in every moment.

-**Step 3**: The algorithm compares both in-play probabilities. Rule of decision: If model's probability is higher than market probability for one selection (local victory, draw or away victory), algorithm decide to invest (one euro).

-**Step 4**: Average return is computed. Note that if we invest randomly, the expected return of this market (for back betting) is approximately -1.9%.

-**Additional requirements**: Extra conditions can be applied after Step 3. For instance: algorithm can take into account extra conditions, such as the level of liquidity, number of goals, red cards, etc. We propose the following three extra conditions:

- **Condition A[21]:** The level of liquidity is less than the median value.
- **Condition B[22]:** No player has been sent off.
- **Condition C[23]:** The current score is not 0-0.

Table 4.5 shows the average return for the base strategy (0.6%) and also when extra conditions are applied. It can be observed that profitable opportunities are detected, ranging from 0.6% to 13.5% depending on the strategy. This is further evidence that prediction markets are not fully efficient.

Table 4.5: Investment strategies

	Investment strategies					
	Random	**Base strategy**	**Base + A**	**Base + B**	**Base + C**	**Base + A + B + C**
Average Return	-1.93%	0.6%	2.55%	2.45%	2.25%	13.54%
% of observations	100%	43%	21%	40%	21%	8%

[20]Time-inhomogeneous independent Poisson model

[21]Markets' performance improves as liquidity increases, so we take advantage of this fact.

[22]Our model does not model red cards, so we decide not to invest if any player has been sent off.

[23]We know from previous section that markets do better when little amount of information has been revealed.

4.9 Conclusions

Throughout the present article we have focused on the prediction markets as a mechanism of aggregation of information. In particular we use information from Betfair Exchange, a centralized prediction market with high volume of transactions. We have chosen this market for several reason: prices are competitively determined unlike in fixed-odds markets, odds are almost fair (there are no commissions that distort the prices) and finally Betfair Exchange provides full information about the functioning of the market, including back and lay prices, volume, last price matched, etc.

Firstly, we have analized whether market prices are well-calibrated, i.e, are efficient. An overall calibration test, the Hosmer-Lemeshow test, indicates that there is evidence of miscalibration. Bootstrap methods help us to detect the origin of the miscalibration. Events with high probability, with probability greater than 80% are systematically overvalued. Despite these results, we can conclude that Betfair Exchange market presents a better calibration than other prediction markets analized in the literature, perhaps due to the characteristics of the market.

Secondly, we have checked the existence of biases. Following the literature, a regression is carried out in order to identify potential biases. There is evidence of "Favourite Long-shot bias", but its magnitude is very small comparing to the literature and disappear when the level of liquidity is high. At the same time, "Home bias" is identified but when we compute the marginal effects at the mean, we can conclude that it has no economic relevance.

Thirdly, we propose a time-inhomogeneous independent Poisson model in order to forecast in-play prices. As primary information, we just use the market odds at the beginning of the match, the match time and the goals scored by each team. Then we apply this model to predict all the in-play match probabilities of the Liga 2013-2014, about 200 millions of in-play probabilities. Later, we use De Finetti measure to compare the predictive power of the model. We obtain that the in-play probabilities estimated using the model are more accurate than markets probabilities, despite having much less information than the market. Note that the market observe a lot of information: red cards, teams's performance, refereeing, tiredness, injuries, motivation, etc. It can be concluded that the prediction markets fail to aggregate the information in a proper way. A simple model that uses odds at the beginning of the match, match time and goals is able to equalize or even to beat the market.

Finally, a high-frequency investment algorithm is designed in order to detect opportunities for investment. For simplicity, we focus only on back bets, but it could be extended to lay betting. Several simple strategies are implemented and profitable opportunities are detected under some circumstances, such as low level of liquidity, high-scoring match, no red cards, etc.

4.10 References

Antweiler, W. (2012). Long-term prediction markets. Journal of Prediction Markets, 6(3).

Baio, G., & Blangiardo, M. (2010). Bayesian hierarchical model for the prediction of football results. Journal of Applied Statistics, 37(2), 253-264.

Berg, J., Forsythe, R., Nelson, F., & Rietz, T (2008). Results from a dozen years of election futures markets research. Handbook of experimental economics results, 1, 742-751.

Berg, J. E., & Rietz, T. A. (2014). Market design, manipulation, and accuracy in political prediction markets: Lessons from the Iowa Electronic Markets. PS: Political Science & Politics, 47(02), 293-296.

Brillinger, D. R. (2008). Modelling game outcomes of the Brazilian 2006 series a championship as ordinal-valued. Brazilian Journal of Probability and Statistics, 89-104.

Christiansen, J.D. (2007). Prediction markets: practical experiments in small markets and behaviours observed, Journal of Prediction Markets, vol. 1(1), pp. 17–41.

Constantinou, A., & Fenton, N. (2013). Profiting from arbitrage and odds biases of the European football gambling market. Journal of Gambling Business and Economics.

Cowgill, B., & Zitzewitz, E. (2015). Corporate prediction markets: Evidence from google, ford, and firm x. The Review of Economic Studies, 82(4), 1309-1341.

Croxson, K., & James Reade, J. (2014). Information and efficiency: Goal arrival in soccer betting. The Economic Journal, 124(575), 62-91.

Davison, A. C., & Hinkley, D. V. (1997). Bootstrap methods and their application (Vol. 1). Cambridge university press.

Deschamps, B., & Gergaud, O. (2012). Efficiency in betting markets: evidence from English football. The Journal of Prediction Markets, 1(1), 61-73.

Dyte, D., & Clarke, S. R. (2000). A ratings based Poisson model for World Cup soccer simulation. Journal of the Operational Research society, 51(8), 993-998.

De Finetti, B. (1972). Probability, induction, and statistics.

Dixon, M. J., & Coles, S. G. (1997). Modelling association football scores and inefficiencies in the football betting market. Journal of the Royal Statistical Society: Series C (Applied Statistics), 46(2), 265-280.

Dixon, M., & Robinson, M. (1998). A birth process model for association football matches. Journal of the Royal Statistical Society: Series D (The Statistician), 47(3), 523-538.

Draper, P., & Fung, J. K. (2002). A Study of Arbitrage Efficiency Between the FTSE-100 Index Futures and Options Contracts. Journal of Futures Markets, 22(1), 31-58.

Everson, P., & Goldsmith-Pinkham, P. S. (2008). Composite Poisson models for goal scoring. J Quant Anal Sports, 4(2), 13.

Graham, I., & Stott, H. (2008). Predicting bookmaker odds and efficiency for UK football. Applied Economics, 40(1), 99-109.

Hanson, R., & Oprea, R. (2004). Manipulators increase information market accuracy. George Mason University.

Härdle, W. (1990). Applied nonparametric regression (No. 19). Cambridge university press.

Hernández, V, (2017a). A Market-based algorithm for predicting soccer outcomes. Working Paper.

Hogan, S., Jarrow, R., Teo, M., & Warachka, M., (2004). Testing market efficiency using statistical arbitrage with applications to momentum and value strategies. Journal of Financial economics, 73(3), 525-565.

Huang, X., Knottenbelt, W., & Bradley, J. (2011). Inferring tennis match progress from in-play betting odds. Final year project), Imperial College London, South Kensington Campus, London, SW7 2AZ.

Karlis, D., & Ntzoufras, I. (2000). On modelling soccer data. Student, 3, 229-245.

Karlis, D., & Ntzoufras, I. (2003). Analysis of sports data by using bivariate Poisson models. Journal of the Royal Statistical Society: Series D (The Statistician), 52(3), 381-393.

Karlis, D., & Ntzoufras, I. (2009). Bayesian modelling of football outcomes: using the Skellam's distribution for the goal difference. IMA Journal of Management Mathematics, 20(2), 133-145.

Keller, J. B. (1994). A characterization of the Poisson distribution and the probability of winning a game. The American Statistician, 48(4), 294-298.

Kramer, A. A., & Zimmerman, J. E. (2007). Assessing the calibration of mortality benchmarks in critical care: The Hosmer-Lemeshow test revisited. Critical care medicine, 35(9), 2052-2056.

Kuypers, T. (2000). Information and efficiency: an empirical study of a fixed odds betting market. Applied Economics, 32(11), 1353-1363.

Langseth, H. (2013) Beating the bookie: A look at statistical models for prediction of football matches. In SCAI (pp. 165-174). ISO 690

Lemeshow, S., & Hosmer Jr, D. W. (1982)A review of goodness of fit statistics for use in the development of logistic regression models. American journal of epidemiology, 115(1), 92-106.

Maher, M. J (1982). Modelling association football scores. Statistica Neerlandica, 36(3), 109-118.

Merritt, S., & Clauset, A. (2014). Scoring dynamics across professional team sports: tempo, balance and predictability. EPJ Data Science, 3(1), 1-21.

Orloff, J., & Bloom, J. (2013). 18.05 class 24, Bootstrap confidence intervals, Spring 2014

Page, L., & Clemen, R. T. (2013). Do Prediction Markets Produce Well-Calibrated Probability Forecasts?. The Economic Journal, 123(568), 491-513.

Polgreen, P. M., Nelson, F. D., Neumann, G. R., & Weinstein, R. A. (2007).Use of prediction markets to forecast infectious disease activity. Clinical Infectious Diseases, 44(2), 272-279.

Rhode, P. W., & Strumpf, K. S. (2004). Historical presidential betting markets. The Journal of Economic Perspectives, 18(2), 127-141.

Servan-Schreiber, E., Wolfers, J., Pennock, D. M., & Galebach, B. (2004). Prediction markets: Does money matter?. Electronic markets, 14(3), 243-251.

Simonoff, J. S. (2013). Analyzing categorical data. Springer Science & Business Media.

Snowberg, E., & Wolfers, J. (2010). Explaining the Favourite-Long Shot Bias: Is it Risk-Love or Misperceptions?. Journal of Political Economy, 118(4), 723-746.

Sobel, R. S., & Travis Raines, S. (2003). An examination of the empirical derivatives of the favourite-longshot bias in racetrack betting. Applied Economics, 35(4), 371-385.

Staněk, R. (2016). Home bias in sport betting: Evidence from Czech betting market. Judgment and Decision Making, 12(2), 168-172.

Suzuki, A. K., Salasar, L. E. B., Leite, J. G., & Louzada-Neto, F. (2010). A Bayesian approach for predicting match outcomes: the 2006 (Association) Football World Cup. Journal of the Operational Research Society, 61(10), 1530-1539.

Tetlock, P. C. (2008). Liquidity and prediction market efficiency.

Tziralis, G., & Tatsiopoulos, I. (2012). Prediction markets: An extended literature review. The journal of prediction markets, 1(1), 75-91.

Vlastakis, N., Dotsis, G., & Markellos, R. N. (2009). How efficient is the European football betting market? Evidence from arbitrage and trading strategies. Journal of Forecasting, 28(5), 426-444.

Volf, P. (2009). A random point process model for the score in sport matches. IMA Journal of Management Mathematics, 20(2), 121-131.

Wolfers, J., & Zitzewitz, E. (2004a). Prediction markets. The Journal of Economic Perspectives, 18(2), 107-126.

Wolfers, J., & Zitzewitz, E. (2004b). Experimental Political Betting Markets and the 2004 Election. The Economists' Voice, 1(2), 1-8.

Wolfers, J., & Zitzewitz, E. (2006). Interpreting prediction market prices as probabilities (No. w12200). National Bureau of Economic Research.

Zeelte, A. (2012). Explaining the Favourite-Long Shot Bias in Tennis:An Endogenous Expectations Approach. University of Amsterdam (2012).