

**NOVA**

**IMS**

Information  
Management  
School

# MEGI

**Mestrado em  
Estatística e Gestão de Informação**

**ESTIMAÇÃO DOS DETERMINANTES DO DESEMPENHO DAS  
EQUIPAS DE FUTEBOL NA PRIMEIRA DIVISÃO PORTUGUESA**

Rui Alexandre Amaro Figueiras

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em  
Estatística e Gestão de Informação

**NOVA Information Management School**  
**Instituto Superior de Estatística e Gestão de Informação**  
Universidade NOVA de Lisboa



**NOVA Information Management School**  
**Instituto Superior de Estatística e Gestão de Informação**  
Universidade NOVA de Lisboa

# **ESTIMAÇÃO DOS DETERMINANTES DO DESEMPENHO DAS EQUIPAS DE FUTEBOL NA PRIMEIRA DIVISÃO PORTUGUESA**

por

Rui Alexandre Amaro Figueiras

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre em Estatística e Gestão de Informação, com especialização em Análise e Gestão de Risco

**Orientada por:**

Bruno Damásio, Professor Auxiliar, NOVA University Lisbon

Novembro, 2024

## STATEMENT OF INTEGRITY

Eu, Rui Figueiras, declaro ter realizado o presente trabalho académico com integridade. Confirmando que não recorri à prática de plágio ou de qualquer outra forma de utilização indevida de informação ou de falsificação de resultados durante o processo de elaboração deste trabalho. Declaro ainda que tenho conhecimento das Regras de Conduta e do Código de Honra da NOVA Information Management School.

Lisboa, 0 de novembro de 2024

## **ESTIMAÇÃO DOS DETERMINANTES DO DESEMPENHO DAS EQUIPAS DE FUTEBOL NA PRIMEIRA DIVISÃO PORTUGUESA**

Copyright © Rui Alexandre Amaro Figueiras, Instituto Superior de Estatística e Gestão de Informação, Universidade NOVA de Lisboa.

A Instituto Superior de Estatística e Gestão de Informação e a Universidade NOVA de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.

---

Este documento foi gerado com o processador (pdf/Xe/Lua) $\LaTeX$  e o modelo [NOVAthesis](#) (v7.2.4) (Lourenço, 2021).

”

*«Recomeça. . . se poderes, sem angústia e sem pressa e os passos que deres, nesse caminho duro do futuro, dá-os em liberdade, enquanto não alcances não descanses, de nenhum fruto queiras só metade. (Torga, M.)»*

---

# Resumo

---

O futebol é o desporto mais popular na Europa, onde os jogadores assumem um papel fundamental. Nos últimos anos, temos assistido a uma inflação generalizada do preço dos jogadores, o que pode representar um aumento da visibilidade deste desporto. Por conseguinte, tem-se tornado cada vez mais importante compreender quais os fatores que influenciam o desempenho dos jogadores de futebol de forma a entender e a interpretar estas variações no seu valor de mercado, bem como na própria performance das equipas.

Em Portugal, embora a visibilidade atual deste desporto seja inquestionável, a dificuldade em aceder a dados económicos e financeiros dos clubes profissionais revela-se como uma potencial motivação para adiamento de um estudo científico mais alargado deste tema. É neste sentido que surge esta dissertação, visando colmatar uma parte desta lacuna na literatura, estudando o impacto das medidas de performance nos jogadores da primeira liga portuguesa e, consequentemente, nos resultados das equipas. O principal objetivo passou por estabelecer uma relação entre a percentagem dos pontos por época de cada equipa como o respetivo posicionamento na liga (variáveis dependentes) e um conjunto de variáveis independentes, escolhido com base na prática em estudos feitos para outros países, balanceando esta seleção com a informação de acesso livre. Neste sentido, o conjunto de variáveis explicativas escolhido passou pelos salários dos jogadores, percentagem de lotação nos estádios, idade média do plantel, número de jogadores por equipa, o saldo de transferências e a percentagem de jogadores estrangeiros. Para o efeito, optou-se por recorrer aos dados das últimas dez épocas (de desde a época 2013/14 a 2022/23), incluindo um global de 31 equipas, cuja informação se recolheu de fontes públicas e gratuitas online, via *web scraping*, distinguindo este trabalho dos demais existentes pela quantidade mais significativa de dados utilizada. Como ponto de partida, fez-se um estudo descritivo das variáveis, não só para avaliar possíveis falhas de informação, bem como ganhar uma perceção mais palpável dos dados. Numa análise posterior, já numa perspetiva de modelação e estimação, utilizaram-se três técnicas: o Estimador de Efeitos Fixos (EEF), Método

dos Momentos Generalizados (MMG) e Método dos Mínimos Quadrados (MMQ). Nas três opções os resultados corroboraram-se mutuamente, revelando que uma melhor assistência nos seus estádios e, sobretudo, despesas salariais mais elevadas melhoram o desempenho das equipas ao longo da temporada.

**Palavras-chave:** Futebol, Performance desportiva, Portugal, Dados em painel

### Sustainable Development Goals (SDG):



---

# Abstract

---

Football is the most popular sport in Europe, where the players assume a fundamental role. In recent years, we have seen a generalised inflation in the price of players, which may represent an increase in the visibility of this sport. It has therefore become increasingly important to learn about the factors that influence the performance of football players in order to understand and interpret these variations in their market value, as well as the performance of the teams themselves.

In Portugal, although the current visibility of this sport is unquestionable, the difficulty in accessing economic and financial data from professional clubs is a potential motivation for postponing a broader scientific study of this subject. This dissertation aims to fill a gap in the literature by studying the impact of performance measures on players in the Portuguese top league and, consequently, on team results. The main objective was to establish a relationship between each team's percentage of points per season and its position in the league (dependent variables) and a set of independent variables, chosen on the basis of practice in studies carried out in other countries, balancing this selection with freely available information. The set of explanatory variables chosen included players' salaries, the percentage of stadium attendances, the average age of the squad, the number of players per team, the transfer balance and the percentage of foreign players. For this purpose, we chose to use data from the last ten seasons (from 2013/14 to 2022/23), including a total of 31 teams, whose information was collected from public and free online sources using web scraping, distinguishing this work from others because of the more significant amount of data used. As a starting point, a descriptive study of the variables was carried out, not only to analyse possible gaps in information, but also to gain a more accurate perception of the data. In a further analysis, from a modelling and estimation perspective, three techniques were used: the Fixed Effects Estimator, the Generalised Method of Moments and the Least Squares Method. In all three options, the results supported each other, revealing that better attendance at their stadiums and, above all, higher wage bills improve the performance of the teams throughout the season.

**Keywords:** Football, Sports performance, Portugal, Panel data,

---

# Índice

---

<b>Índice de Figuras</b>	<b>x</b>
<b>Lista de Siglas e Abreviaturas</b>	<b>xi</b>
<b>1 Introdução</b>	<b>1</b>
1.1 O futebol em Portugal . . . . .	1
1.2 Enquadramento do tema e relevância da investigação . . . . .	2
1.3 Formulação do problema e objetivos da investigação . . . . .	2
<b>2 Revisão de Literatura</b>	<b>4</b>
<b>3 Análise</b>	<b>8</b>
3.1 Dados . . . . .	8
3.1.1 Variáveis Dependentes . . . . .	8
3.1.2 Variáveis Explicativas . . . . .	9
3.1.3 Relação entre salários e performance . . . . .	10
3.2 Modelo . . . . .	11
3.3 Controlo de Endogeneidade . . . . .	12
<b>4 Resultados</b>	<b>15</b>
4.1 Percentagem de pontos enquanto variável dependente . . . . .	15
4.1.1 Método dos Mínimos Quadrados . . . . .	15
4.1.2 Estimador dos Efeitos Fixos . . . . .	16
4.1.3 Método dos Momentos Generalizados . . . . .	17
4.2 Posição na liga enquanto variável dependente . . . . .	17
4.2.1 Método dos Mínimos Quadrados . . . . .	17
4.2.2 Estimador dos Efeitos Fixos . . . . .	18
4.2.3 Método dos Momentos Generalizados . . . . .	19

<b>5 Conclusão</b>	<b>20</b>
<b>Bibliografia</b>	<b>22</b>
<b>Apêndices</b>	
<b>A Apêndice</b>	<b>25</b>

---

# Índice de Figuras

---

3.1	Relação entre Salários e Percentagem de pontos . . . . .	11
3.2	Relação entre Salários e Posição na liga . . . . .	11
A.1	Matriz de correlação com a Percentagem de pontos . . . . .	25
A.2	Matriz de correlação com a Posição na liga . . . . .	26

---

# Lista de Siglas e Abreviaturas

---

- EEA** Estimador de Efeitos Aleatórios (*pp. 11–13*)  
**EEF** Estimador de Efeitos Fixos (*pp. iv, 11–13, 15, 16, 19*)
- MMG** Método dos Momentos Generalizados (*pp. iv, 13–15, 17, 19*)  
**MMQ** Método dos Mínimos Quadrados (*pp. v, 12, 13, 15, 16, 18*)  
**MQ2E** Mínimos Quadrados de Dois Estágios (*p. 13*)

---

# Capítulo 1

## Introdução

---

### 1.1 O futebol em Portugal

O futebol chegou a Portugal nas últimas décadas do século XIX, trazido por alunos portugueses vindos de Inglaterra. Guilherme Pinto Basto foi o pioneiro na sua implementação. Ele teve a iniciativa de realizar uma demonstração deste novo jogo em outubro de 1888, e também foi o responsável pela organização da primeira partida, em janeiro do ano seguinte. Este jogo, ocorrido onde atualmente está localizada a praça de touros do Campo Pequeno, envolveu uma equipa composta por portugueses e outra equipa formada por ingleses. A equipa das quinas triunfou por 2-1.

No início do século XX, tentaram organizar competições. No entanto, o primeiro campeonato nacional só se realizaria em 1921-22.

O escalão de séniores, encontra-se organizado por três ligas profissionais, uma semi-profissional e abaixo disputam-se provas de âmbito distrital (amadoras). A *Liga Portugal Betclit*<sup>1</sup>, a principal divisão do futebol profissional português, congrega 18 equipas que se enfrentam em 34 jogos. O vencedor está assegurado na *UEFA Champions League*<sup>2</sup>, enquanto o segundo colocado enfrenta a terceira pré-eliminatória de acesso ao torneio milionário. O terceiro e o quarto colocados qualificam-se, respetivamente, para as segunda e terceira pré-eliminatórias da *UEFA Europa Conference League*, enquanto as duas equipas com menor pontuação caem para a *Liga Portugal SABSEG*. O 16º colocado

---

<sup>1</sup>A liga portuguesa de futebol foi sofrendo mudanças na sua denominação uma vez que a atribuição do nome da liga refere-se ao seu principal patrocinador e estes vão-se alterando ao longo dos anos. Sendo que, este é o nome atual da liga.

<sup>2</sup>É uma competição anual de futebol a nível continental, organizada pela União das Associações Europeias de Futebol (UEFA) e disputada por clubes da Europa. É uma das competições mais prestigiadas do mundo e a de clubes mais prestigiada no futebol europeu, disputada pelas equipas melhor classificadas nas respetivas ligas nacionais na temporada anterior (Wikipedia, s.d.).

jogará um Playoff de Manutenção com o terceiro colocado da *Liga Portugal SABSEG*. («Apresentação da Liga Portugal», 2024)

O futebol português é frequentemente considerado como uma rampa de lançamento para aceder, no futuro, às principais ligas europeias, daí que os clubes portugueses perceberam a importância de uma boa gestão. Efetivamente, tentam atrair jovens talentos ou jogadores estrangeiros menos conhecidos, mantendo-os na sua maioria por 2/3 anos e, depois, vendendo-os por valores avultados de modo a ter um grande retorno financeiro.

### 1.2 Enquadramento do tema e relevância da investigação

É indiscutível a presença e o impacto do futebol na sociedade moderna. No entanto, a sua beleza e a glorificação dos atletas pela sua importância já quase nada significam, passando de um desporto popular a um setor global com um enorme impacto económico e social. Isto ocorre devido à sua contínua comercialização (direitos de transmissão televisiva, patrocínios, publicidade...) e à expansão global inédita do mercado de trabalho no setor desportivo.

Na sociedade portuguesa, este desporto tem um enorme impacto social, económico e cultural. Um estudo publicado pela *Liga Portugal* e pela *EY*, relativo à época 2021-22, a *Liga Portugal Bwin* (primeira divisão) e a *Liga Portugal SABSEG* (segunda divisão) contribuíram com 617 milhões (0,29%) para o PIB português, mais 12,22% do que na época anterior (EY, 2022).

Este acréscimo é justificado pelo regresso do público aos estádios (depois de terem sido proibidos os jogos com público devido à COVID-19), permitindo um aumento das receitas de bilheteira e das participações das equipas portuguesas na Liga dos Campeões na época 2021-22.

Outro aspeto relevante é que os clubes de futebol aplicam uma parte substancial das suas receitas ao pagamento de salários de jogadores profissionais, na expectativa de obter retorno financeiro e desportivo mediante o seu desempenho em campo. Vários estudos foram realizadas para estabelecer empiricamente a relação causal entre salários e desempenho desportivo, seja no âmbito do clube ou do atleta. No entanto, o cenário português tem recebido pouca atenção neste aspeto.

Neste sentido, considerando a importância da *Primeira Liga* não apenas no futebol português, mas também na sociedade portuguesa, o foco deste estudo recairá sobre os clubes que a integram.

### 1.3 Formulação do problema e objetivos da investigação

A principal motivação na realização deste estudo resulta do reconhecimento por parte dos vários agentes desportivos (e.g. atletas, treinadores, etc), da existência dos

vários fatores que impactam a performance dos atletas e, por consequência, os resultados globais dos próprios clubes.

Portanto, com todos estes elementos, surge uma questão: quais as variáveis que têm uma influência significativa na performance de uma equipa de futebol? Alguns estudos já tentaram responder a esta questão no passado noutras competições ou até mesmo referente a outros desportos. Desde o início do século XXI, a literatura sobre o assunto tem crescido rapidamente. No entanto, nenhuma investigação se centrou em Portugal. Assim, propus-me a colmatar esta lacuna, contribuindo com esta dissertação, com especial atenção ao mais alto nível do futebol em Portugal.

Para realizar esta investigação, foram recolhidos dados de um painel de 31 equipas que participaram pelo menos uma vez na primeira divisão portuguesa, entre as épocas 2013/14 até 2022/23.

Uma vez que os estudos empíricos anteriores utilizaram, principalmente, uma variável independente nos seus modelos, vamos recorrer a várias variáveis independentes no sucesso da equipa com o objetivo de controlar o enviesamento e identificar as variáveis mais relevantes. Consequentemente, recorrer-se-á de modelos lineares para analisar dados de painel estáticos e dinâmicos.

Esta dissertação está organizada da seguinte forma: o capítulo 2 apresenta a literatura atualmente existente relacionada com este tema; o capítulo 3 é descrita a proveniência dos dados, enquanto a última parte é dedicada às estatísticas descritivas; os métodos escolhidos para desenvolver o modelo e as análises empíricas são explorados no capítulo 4; o capítulo 5 apresenta os resultados obtidos com as respetivas interpretações e verificações de robustez; por fim, o capítulo 6 encerra esta dissertação com algumas sugestões e uma conclusão.

Assim, este estudo, tal como outros na mesma área de investigação, visam contribuir para um entendimento mais profundo dos fatores que restringem ou potencializam a performance desportiva dos atletas durante a competição. Isso permite aos atletas aprimorar o seu desempenho competitivo com o apoio dos seus treinadores, que devem estar sempre atualizados com os estudos divulgados pelos investigadores da área do desporto.

---

## Capítulo 2

# Revisão de Literatura

---

A avaliação do desempenho das equipas desportivas não é inédita, mesmo que o interesse pelo futebol seja recente. Os economistas sempre se interessaram pela economia do desporto.

O final do século XX marcou finalmente a chegada da análise relacionada com o futebol com Szymanski e Smith (1997). Estes autores apresentaram um resultado significativo: as equipas que podem pagar mais dinheiro para comprar novos jogadores ou cobrir os seus salários têm mais vitórias do que as equipas com menos dinheiro. Depois disso, com mais dados disponíveis, a literatura sobre o desempenho no futebol aumentou rapidamente, com mais transparência por parte dos clubes e mais recursos fiáveis na Internet. Depois, as investigações sobre os clubes também aumentaram. "Os economistas queriam determinar, por exemplo, o efeito da distribuição da massa salarial no desempenho das equipas ((Franck & Nüesch, 2011); (Buccioli et al., 2014); (Caruso et al., 2016); (Yamamura, 2015)), o efeito dos jogadores estrangeiros no desempenho das equipas (Bachan et al., 2014), (Gásquez & Royuela, 2016) e outras variáveis independentes como, contribuição do treinador (Muehlheusser et al., 2018) ou disputar diferentes competições (Picazo-Tadeo & González-Gómez, 2010)." (Piccoli, 2020, p.11).

Szymanski e Kuypers (1999), apresentam duas hipóteses económicas principais que poderão ser determinantes para o sucesso de um clube de futebol. Sendo a primeira a performance de uma equipa, de forma a verificar se há uma correlação positiva entre o posicionamento na tabela classificativa e os custos salariais e, a segunda, o facto das receitas do clube estarem positivamente correlacionadas com a performance da equipa.

Num estudo desenvolvido por Szymanski e Kuypers (1999), 48 clubes da liga inglesa foram examinados ao longo de 20 anos. Foi concluído que os gastos dos clubes em salários eram responsáveis por 95% das variações na performance desportiva. Forrest e Simmons (2002) usaram os salários como uma variável explicativa da performance,

esta última avaliada através do rácio de pontos obtidos sobre o total possível de pontos, chegaram a resultados semelhantes na liga italiana.

Por seu turno, o futebol inglês foi analisado entre 1974 e 1999 por um estudo desenvolvido por Hall et al. (2002), os autores defenderam que a correlação entre salários e desempenho flui do primeiro para o segundo, ou seja, quanto maior o salário, melhor o desempenho registado.

A análise da relação das receitas totais obtidas pelos clubes com a classificação final na liga norueguesa de Kringstad e Olsen (2016), obteve resultados díspares, embora não tenha tido grande impacto os clubes da metade inferior da tabela classificativa, as receitas desempenham um papel significativo na classificação dos clubes da metade superior. No entanto, com base na amostra de toda a liga, os resultados mostraram que o fator financeiro tem um impacto significativo nos resultados de um jogo.

Numa análise entre a relação da performance económico-financeira e a performance desportiva dos clubes de futebol no caso dos três grandes, Rocha (2016) concluiu que, no caso destes clubes, existe uma significativa correlação positiva, o que refletiu uma clara relação entre a performance desportiva e os resultados financeiros de uma equipa, permitindo-lhe concluir a importância da definição dos objetivos desportivos dos clubes tem para a sua sustentabilidade financeira e, consequentemente, desportiva.

Ao analisar a relação *Pay-Performance*, no futebol português, Carvalho (2019) estudou a forma como os clubes de futebol investem as suas receitas em salários de jogadores, esperando obter um retorno financeiro e desportivo através da performance desportiva dos mesmos. Ele comprovou que: 1) existe correlação positiva entre os salários pagos a jogadores e a sua performance desportiva; 2) o custo salarial é a variável financeira que melhor explica a performance desportiva.

Piccoli (2020), num estudo relativo aos determinantes da performance dos clubes da liga Belga, analisou o comportamento das equipas que passaram na primeira divisão durante 10 temporadas. Começou por descobrir que os gastos em salários são positivamente correlacionados com o sucesso desportivo, sendo esta a principal variável que determina a performance de uma equipa. Demonstrou que o efeito de outras variáveis, como o total de receitas ou valores de transferências, é insignificante. Verificou ainda, como expectável, que a assistência nos estádios tem uma grande importância no sucesso desportivo e portanto os clubes têm de maximizar a fidelidade dos apoiantes.

O mercado de futebol é caracterizado por um elevado nível de assimetria de informação, onde cada clube desconhece a avaliação exata do índice global de capacidades e falhas de um determinado jogador e não pode ter a certeza do desempenho dos mesmos (Oprean & Oprisor, 2014). Além disso, é difícil avaliar se o valor de mercado dos jogadores de futebol está ajustado ao desporto e aos benefícios financeiros que o jogador de futebol trará ao clube que o contratou.

De acordo com Ante (2019), o valor de mercado de um jogador de futebol pode ser explicado por três indicadores: características dos jogadores, indicadores de desempenho

individual e coletivo e popularidade dos jogadores. Nos últimos anos, alguns estudos identificaram os fatores que influenciam o valor de transferência, os valores de mercado das mesmas e os salários. De acordo com Bryson et al. (2013), os salários dos jogadores são influenciados pelos mesmos fatores que influenciam o valor da transferência.

Por sua vez, Barajas e Rodríguez (2010), ao analisarem o futebol espanhol, afirmaram que o custo real de um plantel é calculado pela soma dos gastos com transferências de passes dos atletas (como amortizações do plantel) e gastos em salários. Eles argumentaram que essa variável tem maior relação com a performance desportiva do que apenas os gastos com salários.

Na procura de identificar que variáveis influenciavam a assistência nos estádios em Portugal, Martins e Cró (2018) verificaram que este fator é positivamente correlacionado com a performance de uma equipa, dérbis e com jogos em que a equipa visitante seja *SL Benfica*, *Sporting CP* ou *FC Porto*. Os resultados também mostraram que as pessoas preferem ir a jogos no início e no fim da temporada, e durante os fins de semana. Observaram também que há um aumento de atendimento nos estádios quando a equipa da casa tem hipóteses de ganhar o campeonato ou de se qualificar para a *UEFA Champions League*.

Na mesma linha de pensamento, Addesa e Bond (2021), ao examinarem a assistência nos estádios da Liga Italiana, durante 7 épocas (2012/13-2018/19), notaram que os adeptos que apoiam as equipas que deverão terminar nas sete primeiras posições parecem ser particularmente atraídos por jogos mais equilibrados, enquanto os adeptos das outras equipas são atraídos por jogos mais desequilibrados. Ainda concluíram que os adeptos italianos têm preferência por jogos contra adversários de maior qualidade.

Outra última variável incluída na análise de Tillou e Liarte (2012) foi a importância atribuída ao efeito da idade. Na realidade, pretendiam verificar o efeito da experiência no desempenho. Descobriram que as equipas com apenas jogadores velhos não são uma prova de sucesso. O melhor cenário é ter uma maioria de jogadores jovens enquadrados com uma minoria de jogadores velhos. Esta combinação permite que o grupo tenha a experiência dos mais velhos e a frescura dos mais novos.

Concluindo, a evidência mais importante que os investigadores descobriram é a influência dos salários. De facto, a maioria dos trabalhos focou-se em determinar se compensaria o gasto de milhares de milhões de euros em novos jogadores em vez de aumentar a massa salarial do plantel. A segunda opção é, em teoria, a eficaz. No seu livro, Kuper e Szymanski (2018) confirmam esta teoria através da análise do campeonato em Inglaterra. Como observado: "Quanto maior for a massa salarial total, melhor o plantel, mais alta será a classificação da equipa no campeonato. [...] Os clubes de futebol precisam de fazer menos transferências".

---

## Capítulo 3

# Análise

---

Foi analisada a relação entre a performance desportiva na Liga Portuguesa (mais concretamente, a classificação na liga em cada ano e a percentagem de pontos também de cada ano) e variáveis explicativas identificadas pelos investigadores em estudos anteriores. Variáveis estas que serão enunciadas na tabela 3.1.

### 3.1 Dados

O processo de extração, transformação e carregamento dos dados foi o primeiro passo deste projeto. O conjunto de dados considerado é constituído pelos clubes que participam na primeira divisão portuguesa durante as épocas 2013/14 a 2022/23 . A sua extração e recolha foi possível através de três websites: *Transfermarkt.com*, *Fbref.com* e *Carpology.com*.

#### 3.1.1 Variáveis Dependentes

A posição na liga e a percentagem de pontos são variáveis tipicamente utilizadas para mensurar o sucesso de uma equipa numa temporada e, conseqüentemente, são consideradas as variáveis dependentes. Quanto maior o seu valor, melhor será a performance da equipa em questão. Ambas as variáveis foram retiradas do *Transfermarkt*. Diversos investigadores utilizam os valores produzidos por esta plataforma em vários estudos sobre futebol. A página *Transfermarkt* é o *website* líder em bancos de dados transferências de futebol e outro tipo de informações relacionadas com resultados, notícias, rumores e valores de mercado dos jogadores (Müller et al., 2017).

### 3.1.2 Variáveis Explicativas

A escolha de variáveis explicativas foi feita com base na revisão de literatura, uma vez que, refletem um impacto real na performance dos jogadores (por exemplo a lotação nos estádios pode incentivar os mesmos dando-lhes apoio moral ou até os salários que refletem a capacidade financeira dos clubes para atrair e reter talento). Outro fator importante na escolha destas variáveis foi a disponibilidade de informação.

Maior parte destas variáveis foram retiradas do *Transfermarkt* e do *Capology*. A página *Capology* contém informação sobre os salários dos jogadores de várias ligas de futebol a nível mundial, onde 30% do total salários dos jogadores (em 16 ligas) são verificados diretamente através dos clubes, agências de jogadores e jornalistas, enquanto os restantes salários dos jogadores são calculados através de modelos que se baseiam nos salários oficiais e têm uma taxa de erro inferior a 5% (Capology, 2020).

Variáveis	Definição	Fonte
pontos_perc	Percentagem de pontos. Calculado da seguinte forma: $pontos\_perc_{i,s} = \frac{pontos_{i,s}}{pontos_s}$	Transfermarkt
liga_pos	Posição na liga. Calculado da seguinte forma: $liga\_pos_{i,s}^a = -\log\left(\frac{pos_{i,s}}{31^b - pos_{i,s}}\right)$	Transfermarkt
n_jogadores	Número total de jogadores no plantel.	Transfermarkt
idade_media	Idade média da equipa.	Transfermarkt
salarios_rel	Salários relativos. Calculado da seguinte forma: $salarios\_rel_{i,s} = \frac{salarios_{i,s}}{salarios\_media_s}$	Capology
transf_desp	Valor gasto em transferências.	Transfermarkt
transf_rec	Valor recebido em transferências.	Transfermarkt
transf_saldo	Diferença entre o valor recebido em transferências e o valor gasto.	Transfermarkt
perc_lotacao	Percentagem de lotação nos estádios. Calculado da seguinte forma: $perc\_lotacao_{i,s} = \frac{lotacao\_media_{i,s}}{capacidade\_estadio_{i,s}}$	Transfermarkt
perc_estrangeiros	Percentagem de estrangeiros no plantel. Calculado da seguinte forma: $perc\_estrangeiros_{i,s} = \frac{jogadores\_estrangeiros_{i,s}}{n\_jogadores_{i,s}}$	Transfermarkt

*a*: *i* indica a *i*-ésima equipa e *s* indica a *s*-ésima época

*b*: número total de equipas na amostra

Tabela 3.1: Descrição das variáveis e suas fontes

A Tabela 3.1 mostra o resumo estatística de todas as variáveis. Cada equipa é tratada como um indivíduo ao longo do período. De notar que na maioria das variáveis o número de observações é 178 à exceção de *perc\_lotacao*, devido à época de pandemia (2020/2021) em que não houve público nos estádios e também ao facto do clube Belenenses SAD não ter sido considerado para esta variável uma vez que não tinha um estádio fixo.

O quadro da estatística resumida dá uma análise preliminar interessante. A liga portuguesa é considerada um campeonato para jovens jogadores em ascensão e depende principalmente de jogadores estrangeiros para se manter competitiva contra equipas nacionais ou internacionais. De facto, uma equipa nacional, nas últimas 10 temporadas, tem uma média de idades de 25,77 e uma proporção de jogadores estrangeiros de 59%. Em relação a transferências, devido ao facto do saldo de transferências apresentar uma média positiva, isso prova que os clubes portugueses tentam comprar jogadores e vendê-los mais tarde a uma valor mais elevado de forma a obter lucro. É possível assumir que esses jogadores vêm para a liga portuguesa, vindos de ligas europeias consideradas inferiores ou mesmo de ligas fora da Europa, e portanto, o salário não será tão elevado, assim como o valor de compra dos mesmos. Para além disso, o lotação nos estádios é relativamente baixa, 45%, ainda que possam chegar a valores acima dos 90% quando a equipa joga contra uma das favoritas a vencer o campeonato ou quando se trata de um jogo de competições europeias. De realçar também que o número de jogadores numa equipa atinge o máximo de 47 jogadores, ou seja, mais de quatro vezes o número de jogadores que começa um jogo.

Variáveis	N	Média	Desvio Padrão	CV (%)	Mediana	IQR	Min	Max	Skewness	Kurtosis
pontos_perc	178	0.06	0.02	33.33	0.05	0.02	0.02	0.11	0.93	-0.12
liga_pos	178	0.47	0.44	93.62	0.41	0.62	-0.11	1.49	0.74	-0.25
n_jogadores	178	36.01	4.70	13.05	35.50	6.75	26.00	47.00	0.44	-0.55
idade_media	178	25.77	0.97	3.77	25.85	1.20	22.80	28.30	-0.37	0.13
salarios	178	1.00	1.28	128.00	0.40	0.47	0.04	4.67	1.86	1.91
transf_desp	178	8,053,123.60	18,376,225.36	228.20	262,500.00	3,992,500.00	0.00	115,000,000.00	3.24	12.46
transf_rec	178	19,707,803.37	40,185,030.21	203.90	3,525,000.00	13,759,000.00	0.00	276,220,000.00	3.36	14.04
transf_saldo	178	11,654,679.78	26,631,717.96	228.48	2,500,000.00	7,620,000.00	-33,280,000.00	175,290,000.00	3.40	14.38
perc_lotacao	157	0.45	0.20	44.44	0.41	0.27	0.09	0.90	0.47	-0.67
perc_estrangeiros	178	0.59	0.10	16.95	0.59	0.17	0.23	0.82	-0.40	-0.10

Tabela 3.2: Estatísticas descritivas das variáveis

### 3.1.3 Relação entre salários e performance

As figuras 3.1 e 3.2 mostram a relação entre as variáveis dependentes e os salários. É possível observar uma relação linear positiva entre os salários e o sucesso durante a temporada. Apesar de alguns valores isolados, os resultados parecem estar de acordo com a literatura existente.

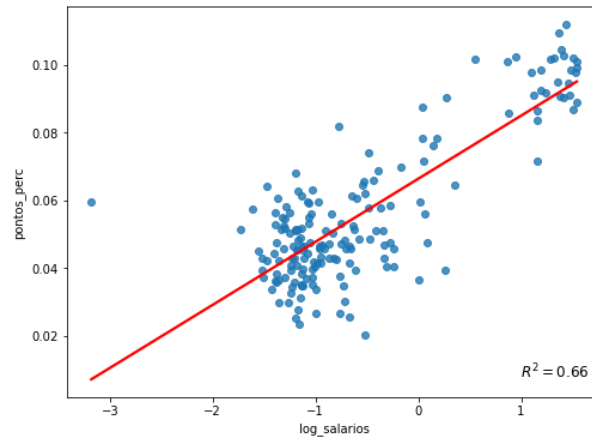


Figura 3.1: Relação entre Salários e Percentagem de pontos

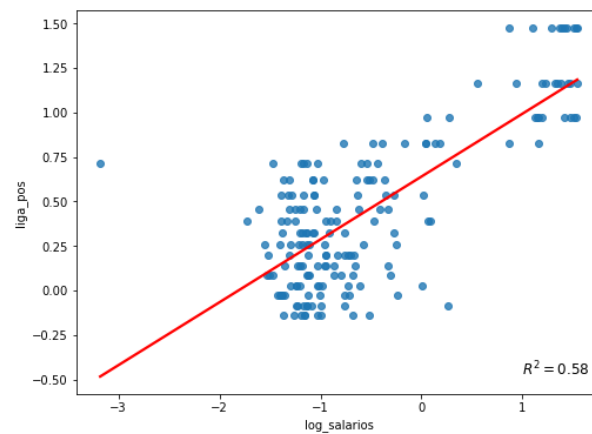


Figura 3.2: Relação entre Salários e Posição na liga

## 3.2 Modelo

Como já foi dito no início deste capítulo, esta base de dados inclui informação sobre 31 equipas durante 10 épocas. E portanto, temos um painel não equilibrado com 178 observações. Assim, permite usar-se uma análise de dados em painel, que providencia estimadores mais eficientes e consistentes do que um simples cruzamento de dados. Inclusive, resolve o problema de endogeneidade (Wooldridge, 2020).

Existem dois métodos muito comuns quando se trabalha com dados em painel: EEF e Estimador de Efeitos Aleatórios (EEA). O EEF é muito útil porque remove os efeitos fixos no tempo, no entanto, removem regressores que não variam ao longo do tempo. O EEA não aplica este mecanismo e tende a ser mais atrativo mas baseia-se em dois pressupostos: as variáveis explicativas são retiradas aleatoriamente de uma população e não estão correlacionadas com fatores não observados (Baltagi, 2021), o que é um pressuposto muito forte. Wooldridge (2010) discute que nos modelos de efeitos fixos, os efeitos específicos de cada entidade (neste caso, clubes) já capturam uma constante para cada grupo, substituindo a necessidade de uma constante global. Assim, uma

constante global não é necessária, pois cada grupo já tem a sua própria implícita, que é representada pelos efeitos fixos de cada entidade.

Ao analisar o problema em questão, este apresenta características semelhantes às do modelo de efeitos fixos porque supomos que estamos na presença de efeitos específicos individuais. De facto, cada equipa tem características individuais que não são possíveis de observações ou estudo, como por exemplo a pressão exercida pelos fãs, a sua própria história, a tendência para promover jovens jogadores das academias, etc. Estes fatores podem influenciar as variáveis utilizadas para este estudo, logo, é importante controlá-las para diminuir o viés da omissão (adicionando equipas fictícias). Da mesma forma que, as variáveis não observadas no tempo também serão controladas através de variáveis fictícias sazonais. Basicamente, é controlado as alterações que possam ocorrer ao longo da época que impacta a classificação final ou a gestão financeira de cada clube.

O modelo a estimar, onde índice "i" representa a i-ésima equipa ( $i = 1, \dots, 31$ ) e o índice "s" representa a s-ésima época ( $s=1, \dots, 10$ ), é apresentado da seguinte forma:

$$\begin{aligned}
 Y_{i,s} = & \beta_0 + \beta_1 n\_jogadores_{i,s} + \beta_2 idade\_media_{i,s} + \\
 & + \beta_3 salarios_{i,s} + \beta_4 transf\_desp_{i,s} + transf\_rec_{i,s} + \beta_6 transf\_saldo_{i,s} + \\
 & + \beta_7 perc\_lotacao_{i,s} + \beta_8 perc\_estrangeiros_{i,s} + \\
 & + \left( \delta_1 \sum_{i=1}^{31} equipa_i \right) + \delta_2 \left( \sum_{s=9}^{18} época_s + \sum_{s=1}^{16} época_s \right) + \\
 & + \alpha_i + \gamma_s + \epsilon_{i,s}
 \end{aligned}$$

Onde,  $Y_{i,s}$  equivale às variáveis dependentes,  $\beta_0$  é a constante,  $\alpha_i$  e  $\gamma_s$  são, respetivamente, as condições aleatórias das equipas e de cada época que não são calculáveis (como dito acima), e  $\epsilon_{i,s}$  corresponde ao termo de erro.

Inicialmente, será utilizado o MMQ. Este estimador é muito útil pois multiplica o número de indivíduos pelo período de tempo e, assim, aumenta o tamanho da amostra e o poder estatístico (Wooldridge, 2020).

De seguida, para saber qual o melhor estimador a ser implementado (EEF ou EEA), fazer-se-á um teste de Hausman. Este teste verifica se os erros únicos estão correlacionados com os regressores.

### 3.3 Controlo de Endogeneidade

Até agora, para obter estimadores não enviesados e consistentes, assumiu-se a hipótese de exogeneidade estrita <sup>1</sup>. No entanto, este pressuposto é forte porque exige que os acontecimentos passados, presentes e futuros não tenham qualquer influência na situação atual, sendo que, para este estudo, isto torna-se irrealista. É legítimo suspeitar

---

<sup>1</sup>De acordo com Mayer (2016), exogeneidade estrita consiste no facto do termo de erro não estar correlacionado com a variável x.

que o desempenho passado determine, por exemplo, os salários ou as receitas atuais. Mostrando um caso mais prático, se uma equipa tiver um bom desempenho durante a temporada, é provável que os clubes tentem renegociar o contrato dos jogadores para o ano seguinte. Isto porque jogaram bem e para evitar uma transferência para adversários. É também normal, que na hora de negociação de contrato, os jogadores pedirem mais do que aquilo que já recebem. Além disso, quando uma equipa está nas primeiras posições da sua liga, haverá uma maior *hype* à volta dos seus jogadores e, qualificando-se para as competições europeias, envolverá rendimentos maiores. Portanto, as variáveis explicativas do modelo podem estar correlacionadas com variáveis que não foram observadas. Uma solução para controlar esta questão é encontrar uma variável instrumental (se houver disponibilidade de dados) e esta variável deve ser exógena e correlacionada com a variável endógena e, em seguida, estimar por regressão dos Mínimos Quadrados de Dois Estágios (MQ2E) ou adicionar uma variável desfasada. Nesta situação, a segunda opção é selecionada devido à impossibilidade de encontrar uma boa variável instrumental.

Contudo, o problema quando se aplica este método é que passamos de um painel de dados estático para um painel de dados dinâmico (o que significa interação de realizações passadas) e os estimadores padrão, como o MMQ, o EEF ou o EEA, deixam de ser válidos porque assentam no pressuposto discutido acima (Baltagi, 2021). Portanto, serão necessários outros métodos de estimação. (Arellano & Bond, 1991) e (Blundell & Bond, 1998) propuseram o MMG como estimador para resolver este problema. Este estimador foi concebido para que, na presença de um painel dinâmico, os regressores possam ser endógenos, o período T seja menor que os indivíduos N, haja presença de efeitos fixos e heteroscedasticidade (Roodman, 2009).

Na verdade, existem dois estimadores MMG, o MMG de Diferença (Arellano & Bond, 1991) e o MMG de Sistema (Blundell & Bond, 1998). O estimador de diferença subtrai a última observação do período para o atual, pelo que  $\Delta y = y_1 - y_2$ . O problema deste método é que estamos na presença de um painel desequilibrado pois, uma vez que a diferença entre as observações pode aumentar e, assim, diminuir o poder estatístico. Além disso, a variável desfasada pode continuar a estar correlacionada com o termo de erro pois, com este método, precisamos sempre de observações anteriores da mesma. O estimador de sistema, no entanto, subtrai a observação da variável atual pela média das observações futuras (possíveis até  $n - 1$ ), pelo que,  $\Delta y = y_1 - \bar{y}_{futuro}$ . Adicionalmente, ao remover os efeitos fixos constantes, a perda de dados é minimizada e, no cálculo, não é necessária nenhuma variável desfasada e o termo constante pode ser mantido. E por isso, o MMG de sistema é mais desejável.

O novo modelo a estimar (com a variável dependente desfasada), onde índice "i" representa a i-ésima equipa ( $i = 1, \dots, 31$ ) e o índice "s" representa a s-ésima época ( $s = 1, \dots, 10$ ), é apresentado da seguinte forma:

$$\begin{aligned}\Delta Y_{i,s} = & \beta_0 + \beta_1 \Delta n\_jogadores_{i,s} + \beta_2 \Delta idade\_media_{i,s} + \\ & + \beta_3 \Delta salarios_{i,s} + \beta_4 \Delta transf\_desp_{i,s} + \beta_5 \Delta transf\_rec_{i,s} + \\ & + \beta_6 \Delta transf\_saldo_{i,s} + \beta_7 \Delta perc\_lotacao_{i,s} + \beta_8 \Delta perc\_estrangeiros_{i,s} + \\ & + \beta_9 \Delta desfasado Y_{i,s} + \epsilon_{i,s}\end{aligned}$$

Onde, novamente,  $Y_{i,s}$  equivale às duas variáveis dependentes,  $\beta_0$  é a constante e  $\epsilon_{i,s}$  corresponde à margem de erro ( $\alpha_i$  e  $\gamma_i$  desapareceram devido ao processo de ortogonalidade).

Para estimar esta equação, será utilizado o MMG de Sistema.

---

## Capítulo 4

# Resultados

---

Neste capítulo será apresentado o resultado de todos os modelos testados. Vamos obter os fatores que serão considerados determinantes para o triunfo na liga portuguesa. Vão ser utilizados três tipos de estimadores: MMQ, EEF e MMG. Para todos os testes será utilizado um nível de significância de 5%. Para testar estes modelos foram alterados os valores da época 2020/21 da variável `perc_lotacao` para 0 de forma a incluir as 31 equipas de igual forma para todas as variáveis ao longo das 10 épocas. Também foi testado sem esta temporada, para efeitos de comparação, e os resultados permanecem idênticos.

### 4.1 Percentagem de pontos enquanto variável dependente

#### 4.1.1 Método dos Mínimos Quadrados

A tabela 4.1 apresenta os resultados do estimador dos MMQ.

Numa análise preliminar, é possível observar que o modelo aparenta explicar razoavelmente esta variável dependente, uma vez que o  $R^2$  é relativamente alto (71.82%).

Ao analisar as variáveis explicativas, os salários são a única variável que  $p\text{-value} = 0$  e, portanto, é estatisticamente significativa para o modelo. Isso sugere que os salários têm um impacto positivo claro sobre a performance de uma equipa.

Outra variável significativa para o modelo é a percentagem de estrangeiros num plantel, pois o seu  $p\text{-value}$  é inferior ao nível de significância. Esta variável tem um impacto negativo neste modelo e indica que um plantel com maior percentagem de jogadores estrangeiros consigam menos pontos.

Existem outras duas variáveis que apesar do  $p\text{-value}$  ser ligeiramente superior a 0.05, são consideradas ligeiramente significativas. Tanto a variável idade média como o

número de jogadores numa equipa são negativamente correlacionadas com a variável dependente, o que mostra que planteis com jogadores mais velhos e mais extensos têm uma performance inferior.

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	0.1044	0.0281	3.7191	0.0003	0.0490	0.1598
n_jogadores	-0.0003	0.0002	-1.7126	0.0886	-0.0007	5.163e-05
idade_media	-0.0016	0.0010	-1.7304	0.0854	-0.0035	0.0002
salarios	0.0132	0.0014	9.3395	0.0000	0.0104	0.0160
transf_desp	4.453e-11	1.043e-10	0.4271	0.6699	-1.613e-10	2.504e-10
transf_rec	8.174e-12	4.368e-11	0.1871	0.8518	-7.805e-11	9.44e-11
perc_lotacao	0.0063	0.0043	1.4574	0.1468	-0.0022	0.0147
perc_estrangeiros	-0.0167	0.0074	-2.2539	0.0255	-0.0314	-0.0021

$R^2 = 0.7182$

Tabela 4.1: MMQ com pontos\_perc

#### 4.1.2 Estimador dos Efeitos Fixos

A tabela 4.2 apresenta os resultados do EEF.

Neste estimador, o  $R^2 = 0.9611$ , o que indica que aproximadamente 96.11% da variação na variável dependente (pontos\_perc) é explicada pelas variáveis independentes incluídas no modelo. Esse valor é bastante alto, sugerindo que o modelo se deverá ajustar muito bem aos dados.

A idade média dos jogadores tem um efeito positivo e é estatisticamente significativo para o modelo. Isso sugere que as equipas com uma idade média mais alta tendem a ter uma percentagem de pontos menor. Esta variável segue um sentido inverso neste modelo ao que seguia no MMQ. Com o mesmo racional, também estão os salários, com um coeficiente positivo e altamente significativo. Sendo que ambas estas variáveis têm um  $p - value = 0$ .

A percentagem de lotação nos estádios também é uma variável interessante de se considerar neste modelo, apesar de ter um  $p - value$  ligeiramente acima do nível de significância, apresenta um valor positivo e portanto tem uma correlação positiva com a variável dependente.

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
n_jogadores	3.396e-05	0.0002	0.1923	0.8477	-0.0003	0.0004
idade_media	0.0017	0.0003	6.0446	0.0000	0.0012	0.0023
salarios	0.0131	0.0015	8.9188	0.0000	0.0102	0.0160
transf_desp	2.013e-11	1.079e-10	0.1866	0.8522	-1.928e-10	2.331e-10
transf_rec	1.179e-11	4.528e-11	0.2605	0.7948	-7.758e-11	1.012e-10
perc_lotacao	0.0076	0.0044	1.7111	0.0889	-0.0012	0.0163
perc_estrangeiros	-0.0111	0.0075	-1.4744	0.1422	-0.0260	0.0038

$R^2 = 0.9611$

Tabela 4.2: EEF com pontos\_perc

### 4.1.3 Método dos Momentos Generalizados

A tabela 4.3 apresenta os resultados do MMG.

O coeficiente de determinação ajustado do modelo é de 0.7339, indicando que aproximadamente 73.39% da variação na percentagem de pontos conquistados pode ser explicada pelas variáveis incluídas no modelo.

Entre as variáveis analisadas, os salários desfasados apresentam um coeficiente positivo (0.0138) e altamente significativo ( $p$ -value = 0), o que indica que salários mais altos estão fortemente associados a um aumento na percentagem de pontos conquistados.

Por outro lado, a percentagem de jogadores estrangeiros desfasada apresenta um coeficiente negativo (-0.0163) e também é estatisticamente significativa ( $p$ -value = 0.0442). Isso sugere que uma maior percentagem de jogadores estrangeiros na temporada anterior está associada a uma redução na percentagem de pontos conquistados.

Também o coeficiente desfasado da idade média dos jogadores é negativo (-0.0159) e ligeiramente significativo com um  $p$ -value de 0.0921. Isso sugere que um aumento na idade média dos jogadores na temporada anterior tende a reduzir a percentagem de pontos conquistados na temporada atual, embora o efeito seja pequeno.

Resumindo, os resultados sugerem que, enquanto salários mais altos podem estar associados a um melhor desempenho dos clubes, uma maior idade média dos jogadores e uma maior percentagem de estrangeiros parecem estar relacionados a um desempenho inferior.

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	0.0855	0.0306	2.7912	0.0053	0.0255	0.1455
n_jogadores_lag	0.0003	0.0002	1.2835	0.1993	-0.0001	0.0007
idade_media_lag	-0.0017	0.0010	-1.6844	0.0921	-0.0037	0.0003
salarios_lag	0.0138	0.0013	10.401	0.0000	0.0112	0.0164
transf_desp_lag	-4.772e-11	9.08e-11	-0.5255	0.5992	-2.257e-10	1.302e-10
transf_rec_lag	4.63e-11	4.037e-11	1.1469	0.2514	-3.283e-11	1.254e-10
perc_lotacao_lag	-0.0010	0.0046	-0.2119	0.8322	-0.0101	0.0081
perc_estrangeiros_lag	-0.0159	0.0081	-1.9643	0.0495	-0.0317	-3.5e-05

$R^2$  Ajustado = 0.7339

Tabela 4.3: MMG com pontos\_perc

## 4.2 Posição na liga enquanto variável dependente

### 4.2.1 Método dos Mínimos Quadrados

Este modelo explica 63.59% da variação na posição da liga, conforme indicado pelo  $R^2$ .

A idade média dos jogadores também apresenta um coeficiente negativo (-0.0417), sendo ligeiramente significativo ao nível de 5% de significância. Este resultado sugere que as equipas com jogadores mais velhos tendem a ter uma pior classificação na liga.

Por outro lado, os salários têm um coeficiente positivo (0.2176) e são estatisticamente significativos, o que indica que os clubes com salários mais elevados tendem a ficar mais acima na tabela. Esse resultado é intuitivo, pois equipas que investem mais nos seus jogadores geralmente têm um melhor desempenho.

A relação entre percentagem de lotação do estádio e desempenho, embora não estatisticamente significativa, sugere que maiores lotações podem estar associadas a um desempenho ligeiramente melhor, uma observação que poderia ser explorada em estudos futuros.

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	1.6534	0.6497	2.5450	0.0118	0.3709	2.9358
n_jogadores	-0.0077	0.0046	-1.6802	0.0948	-0.0167	0.0013
idade_media	-0.0417	0.0220	-1.8934	0.0600	-0.0852	0.0018
salarios	0.2287	0.0328	6.9678	0.0000	0.1639	0.2935
transf_desp	1.224e-09	2.413e-09	0.5074	0.6125	-3.539e-09	5.987e-09
transf_rec	8.178e-10	1.011e-09	0.8091	0.4196	-1.177e-09	2.813e-09
perc_lotacao	0.1473	0.0993	1.4830	0.1399	-0.0488	0.3434
perc_estrangeiros	-0.2848	0.1717	-1.6587	0.0990	-0.6237	0.0541

$R^2 = 0.6359$

Tabela 4.4: MMQ com liga\_pos

#### 4.2.2 Estimador dos Efeitos Fixos

O modelo de efeitos fixos foi utilizado para explicar a variação na posição das equipas na liga ao longo do tempo, com um bom desempenho geral. O coeficiente de determinação ( $R^2$ ) indica que o modelo explica cerca de 81.51% da variação na classificação das equipas.

Ao analisar os coeficientes, observamos que os salários têm um impacto positivo e significativo na posição das equipas, sugerindo que equipas com salários mais elevados tendem a se classificar melhor. Este resultado é consistente com a ideia de que maiores investimentos em jogadores podem melhorar o desempenho da equipa.

A percentagem de lotação nos estádios apresenta um coeficiente bastante alto e é ligeiramente significativo no modelo com um p-value um pouco acima do nível de significância. O que mostra que equipas com maior percentagem de lotação nos jogos em casa tendem a ter uma posição na liga melhor.

A idade média, apesar de também estar ligeiramente acima do nível de significância também é considerada ligeiramente significativa para o modelo, indicando um valor positivo. O que significa que planteis com jogadores mais velhos, tendem a ficar numa melhor posição na tabela classificativa.

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
n_jogadores	-0.0018	0.0040	-0.4460	0.6561	-0.0097	0.0061
idade_media	0.0119	0.0065	1.8266	0.0695	-0.0010	0.0248
salarios	0.2265	0.0333	6.7970	0.0000	0.1607	0.2923
transf_desp	8.378e-10	2.446e-09	0.3425	0.7324	-3.991e-09	5.667e-09
transf_rec	8.751e-10	1.027e-09	0.8525	0.3952	-1.151e-09	2.901e-09
perc_lotacao	0.1684	0.1005	1.6748	0.0958	-0.0301	0.3669
perc_estrangeiros	-0.1958	0.1708	-1.1467	0.2531	-0.5329	0.1413

$R^2 = 0.8151$

Tabela 4.5: EEF com liga\_pos

### 4.2.3 Método dos Momentos Generalizados

Os resultados deste modelo mostram que o  $R^2$  ajustado é de 0.6449, indicando que cerca de 64.49% da variação na posição das equipas pode ser explicada pelas variáveis incluídas no modelo. O salário desfasado é a variável mais significativa, com um coeficiente de 0.2826 e um p-value = 0, o que indica que equipas que pagaram salários mais altos no ano anterior tendem a melhorar significativamente sua posição na liga. Esse resultado aponta para uma forte correlação entre altos salários e o desempenho das equipas.

Todas as outras variáveis não demonstram uma influência significativa no modelo, sugerindo que, isoladamente, elas não são bons indicadores do sucesso das equipas na liga.

	Parameter	Std. Err.	T-stat	P-value	Lower CI	Upper CI
const	0.8775	0.6296	1.3939	0.1634	-0.3564	2.1115
n_jogadores_lag	0.0053	0.0047	1.1272	0.2597	-0.0039	0.0146
idade_media_lag	-0.0303	0.0209	-1.4499	0.1471	-0.0713	0.0107
salarios_lag	0.2826	0.0314	9.0084	0.0000	0.2211	0.3440
transf_desp_lag	-1.942e-09	2.053e-09	-0.9461	0.3441	-5.966e-09	2.081e-09
transf_rec_lag	5.771e-10	1.031e-09	0.5597	0.5757	-1.444e-09	2.598e-09
perc_lotacao_lag	-0.0014	0.1084	-0.0128	0.9898	-0.2139	0.2112
perc_estrangeiros_lag	-0.1973	0.1867	-1.0568	0.2906	-0.5632	0.1686

$R^2$  Ajustado = 0.6449

Tabela 4.6: MMG com liga\_pos

---

## Capítulo 5

# Conclusão

---

Em geral, esta dissertação procurou determinar os fatores determinantes do desempenho dos clubes que participam no campeonato português de futebol profissional. Para tal, foi construído um painel de dados não equilibrado, que agrupa 178 observações ao longo de dez épocas para trinta e uma equipas. Graças a diferentes técnicas de estimação dos dados, foi possível controlar as potenciais características não observadas das equipas e a endogeneidade. Além disso, para apoiar a investigação, foram aplicadas duas variáveis de referência que avaliam o desempenho das equipas: a posição na liga e a percentagem de pontos.

Para começar, descobriu-se que o efeito causado pelas despesas salariais no sucesso desportivo português está positivamente correlacionado e verifica a literatura existente para outros campeonatos na Europa ((Szymanski & Smith, 1997); (Forrest & Simmons, 2002); (Hall et al., 2002); (Kuper & Szymanski, 2018)). Os efeitos de outras variáveis financeiras, como o valor recebido e gasto em transferências, são insignificantes. Os salários dos jogadores refletem supostamente o seu verdadeiro valor, enquanto a compra de jogadores depende mais da oferta e da procura no mercado do futebol, que pode sobrevalorizar (ou subvalorizar) o seu valor real.

Depois, foram encontradas evidências contrárias relativamente à idade média e à percentagem de estrangeiros num plantel. O que significa que equipas com jogadores mais velhos e menos jogadores portugueses, tendem a ter uma performance pior. Isto vai ao encontro ao facto das grandes equipas possuírem academias para a formação de jovens jogadores.

Depois disso, como era de esperar, uma assistência significativa durante a época está associada a melhores prestações. Os clubes devem, portanto, tentar maximizar a fidelidade dos seus adeptos.

Conclui-se assim que, para ganhar campeonatos, os salários elevados são cruciais

para atrair os melhores jogadores, mas para permanecer na primeira divisão, gerar receitas é vital para alargar as suas despesas de modo a poder fazer crescer o clube.

No entanto, esta investigação também tem as suas próprias limitações a salientar. A maior delas diz respeito aos dados utilizados. Infelizmente, estes não foram recolhidos diretamente junto das equipas ou da liga, mas sim de fontes externas. Assim sendo, é possível que o conjunto de dados não corresponda a 100% dos valores reais, especialmente no que se refere aos salários e rendimentos, mesmo que se assuma uma proximidade razoável. Tendo em conta este facto, alguns coeficientes estimados poderão sofrer alterações. Além disso, esta análise foi efetuada especificamente para a liga portuguesa. As estimativas de outras ligas podem variar.

Por último, este estudo constitui uma das primeiras abordagens empíricas da realidade portuguesa. Para reforçar esta análise, devem ser realizados projetos semelhantes, tendo presente porém que será essencial a acessibilidade a dados mais rigorosos e/ou oficiais, de forma a reforçar novos resultados. Outras análises poderiam, por exemplo, recolher dados relativos a um período de tempo mais alargado de modo a explorar se as mesmas variáveis se mantinham significantes para o modelo. Na mesma ideia, o alargamento dos dados a mais divisões em Portugal faz sentido para representar tanto quanto possível a realidade.

Além disso, a inclusão de uma variável de substituição para medir o impacto do treinador ou da disparidade salarial no triunfo poderia ser interessante para a liga portuguesa, tal como vários autores fizeram para outras ligas ((Franck & Nüesch, 2011); (Yamamura, 2015); (Muehlheusser et al., 2018)). Paralelamente, ter em conta o efeito do desenvolvimento dos jovens pode ser interessante. Um jogador vindo da academia, que conhece bem a cultura do clube e não tem nada a perder, tende a ter um bom desempenho e a melhorar a qualidade da equipa.

---

# Bibliografia

---

- Addesa, F., & Bond, A. J. (2021). Determinants of stadium attendance in Italian Serie A: New evidence based on fan expectations. *PLoS One*, 16(12), e0261419 (ver p. 6).
- Ante, L. (2019). *Determinants of transfers fees: Evidence from the five major European football leagues*. (Ver p. 5).
- Apresentação da Liga Portugal [Acessado em: 1 de fevereiro de 2024]. (2024). (Ver p. 2).
- Arellano, M., & Bond, S. (1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *Rev. Econ. Stud.*, 58(2), 277 (ver p. 13).
- Bachan, R., Reilly, B., & Witt, R. (2014). Team performance and race: evidence from the English and French national soccer teams. *Appl. Econ.*, 46(13), 1535–1546 (ver p. 4).
- Baltagi, B. H. (2021). *Econometric analysis of panel data* (6<sup>a</sup> ed.). Springer Nature. (Ver pp. 11, 13).
- Barajas, A., & Rodríguez, P. (2010). Spanish football clubs finances: Crisis and player salaries. *Int. J. Sports Financ.*, 5(1), 52–66 (ver p. 6).
- Blundell, R., & Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *J. Econom.*, 87(1), 115–143 (ver p. 13).
- Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent. *J. Sports Econom.*, 14(6), 606–628 (ver p. 6).
- Buccioli, A., Foss, N. J., & Piovesan, M. (2014). Pay dispersion and performance in teams. *PLoS One*, 9(11), e112631 (ver p. 4).
- Capology. (2020). <https://twitter.com/CapologyDB/status/1330906457668276233> (ver p. 9).
- Caruso, R., Carlo, B. P., & Marco, D. (2016). Does diversity in the payroll affect soccer teams' performance? Evidence from the Italian Serie A (ver p. 4).
- Carvalho, R. M. (2019). *A relação pay-performance no futebol português* [tese de mestrado, ISEG]. (Ver p. 5).

- EY. (2022). *Contributo do futebol profissional para o pib português aumentou para 617 milhões de euros na época 2021-22*. [https://www.ey.com/pt\\_pt/news/2022/12/anuario-do-futebol-profissional-portugues-2021-2022](https://www.ey.com/pt_pt/news/2022/12/anuario-do-futebol-profissional-portugues-2021-2022) (ver p. 2).
- Forrest, D., & Simmons, R. (2002). Team salaries and playing success in sports: A comparative perspective. Em *Sportökonomie* (pp. 221–238). Gabler Verlag. (Ver pp. 4, 20).
- Franck, E., & Nüesch, S. (2011). The effect of wage dispersion on team outcome and the way team outcome is produced. *Appl. Econ.*, 43(23), 3037–3049 (ver pp. 4, 21).
- Gásquez, R., & Royuela, V. (2016). The determinants of international football success: A panel data analysis of the Elo rating. *Soc. Sci. Q.*, 97(2), 125–141 (ver p. 4).
- Hall, S., Szymanski, S., & Zimbalist, A. S. (2002). Testing causality between team performance and payroll. *J. Sports Econom.*, 3(2), 149–168 (ver pp. 5, 20).
- Kringstad, M., & Olsen, T.-E. (2016). Can sporting success in Norwegian football be predicted from budgeted revenues? *Eur. Sport Manag. Q.*, 16(1), 20–37 (ver p. 5).
- Kuper, S., & Szymanski, S. (2018). *Soccernomics (2018 world cup edition)* (2018 world cup edition). Nation Books. (Ver pp. 7, 20).
- Lourenço, J. M. (2021). *The NOVAthesis L<sup>A</sup>T<sub>E</sub>X Template User's Manual*. NOVA University Lisbon. <https://github.com/joaomlourenco/novathesis/raw/main/template.pdf> (ver p. ii).
- Martins, A. M., & Cró, S. (2018). The demand for football in Portugal. *J. Sports Econom.*, 19(4), 473–497 (ver p. 6).
- Mayer, A. (2016). Testing for strict exogeneity in fixed-effects panel models. *SSRN Electron. J.* (ver p. 12).
- Muehlheusser, G., Schneemann, S., Sliwka, D., & Wallmeier, N. (2018). The contribution of managers to organizational success. *J. Sports Econom.*, 19(6), 786–819 (ver pp. 4, 21).
- Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *Eur. J. Oper. Res.*, 263(2), 611–624 (ver p. 8).
- Oprean, V.-B., & Oprisor, T. (2014). Accounting for soccer players: Capitalization paradigm vs. Expenditure. *Procedia Econ. Finance*, 15, 1647–1654 (ver p. 5).
- Picazo-Tadeo, A. J., & González-Gómez, F. (2010). Does playing several competitions influence a team's league performance? Evidence from Spanish professional football. *Cent. Eur. J. Oper. Res.*, 18(3), 413–432 (ver p. 4).
- Piccoli, A. (2020). *Evaluation of determinants of performance of the clubs participating in the Belgian football championship* [tese de mestrado, University of Liège]. (Ver pp. 4, 5).
- Rocha, R. M. G. (2016). *A correlação entre a performance económico-financeira e a performance desportiva dos clubes de futebol: O caso dos três grandes em Portugal* [tese de mestrado, ISCTE]. (Ver p. 5).

- Roodman, D. (2009). How to do Xtabond2: An introduction to difference and system GMM in Stata. *Stata J.*, 9(1), 86–136 (ver p. 13).
- Szymanski, S., & Kuypers, T. (1999). *Winners and losers*. Viking. (Ver p. 4).
- Szymanski, S., & Smith, R. (1997). The English Football Industry: profit, performance and industrial structure. *Int. Rev. Appl. Econ.*, 11(1), 135–153 (ver pp. 4, 20).
- Tillou, C., & Liarte, S. (2012). L'impact de l'expérience sur la performance du groupe. Le cas de l'équipe nationale de football du Brésil de 1954 à 2010. *Manag. Int.*, 16(2), 113–128 (ver p. 6).
- Wikipedia. (s.d.). Liga dos Campeões da UEFA — Wikipédia, a enciclopédia livre [[Acessado em: 27-12-2024]]. [https://pt.wikipedia.org/wiki/Liga\\_dos\\_Campe%C3%B5es\\_da\\_UEFA](https://pt.wikipedia.org/wiki/Liga_dos_Campe%C3%B5es_da_UEFA) (ver p. 1).
- Wooldridge, J. (2010). *Econometric Analysis of Cross Section and Panel Data* (2<sup>a</sup> ed.). MIT Press. (Ver p. 11).
- Wooldridge, J. (2020). *Introductory econometrics* (7<sup>a</sup> ed.). South-Western College Publishing. (Ver pp. 11, 12).
- Yamamura, E. (2015). Wage disparity and team performance in the process of industry development. *J. Sports Econom.*, 16(2), 214–223 (ver pp. 4, 21).

---

## Apêndice A

# Apêndice

---

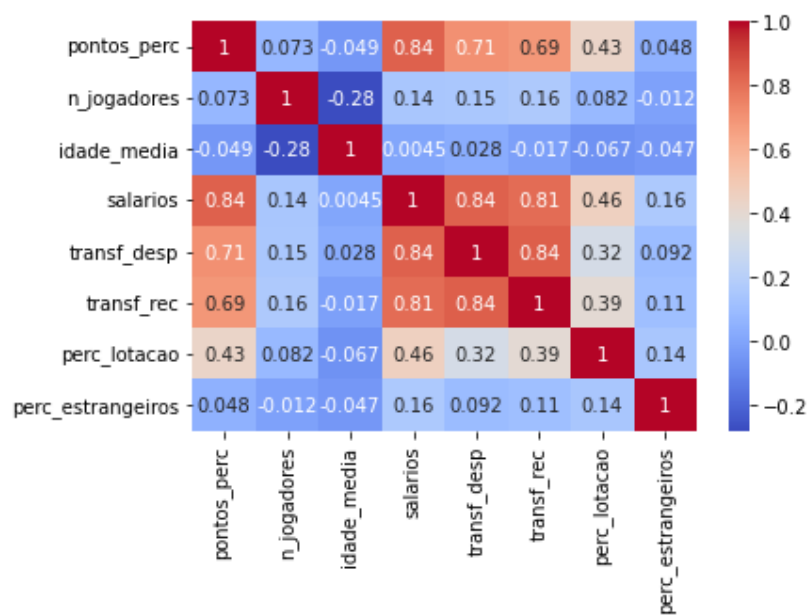


Figura A.1: Matriz de correlação com a Percentagem de pontos

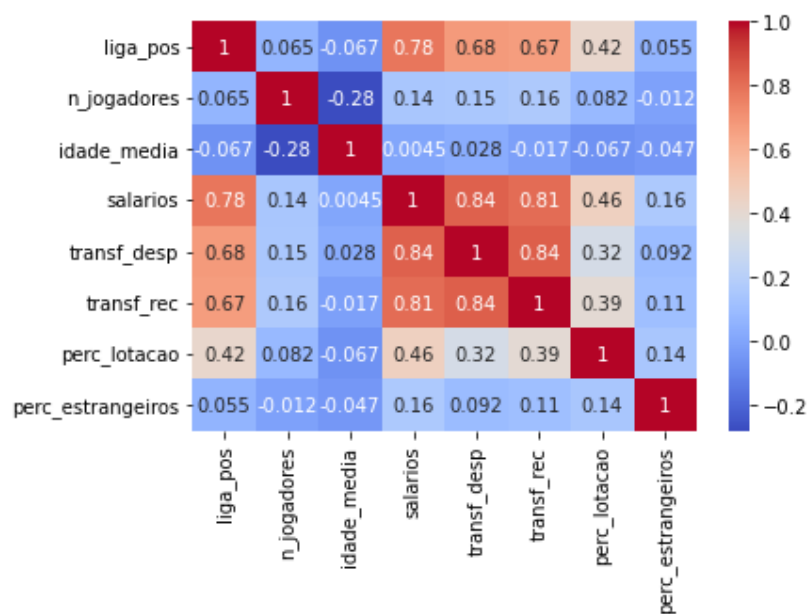


Figura A.2: Matriz de correlação com a Posição na liga



**NOVA Information Management School**  
**Instituto Superior de Estatística e Gestão de Informação**

Universidade NOVA de Lisboa