

Bernardo Mafra Mendes

**Salários e Performance. É possível comprar
competitividade no futebol inglês?**

Brasília
2022

Bernardo Mafra Mendes

Salários e Performance. É possível comprar competitividade no futebol inglês?

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade de Brasília como requisito para a obtenção do título de Mestre em Economia.

Universidade de Brasília – UnB
Faculdade de Administração, Economia, Contabilidade e Gestão Pública
Programa de Mestrado em Economia

Orientador: Maria Eduarda Tannuri Pianto, PhD

Brasília
2022

Bernardo Mafra Mendes

Salários e Performance. É possível comprar competitividade no futebol inglês?

Dissertação apresentada ao Programa de Pós-Graduação em Economia da Universidade de Brasília como requisito para a obtenção do título de Mestre em Economia.

Maria Eduarda Tannuri Pianto, PhD
Orientador

Mauricio Soares Bugarin, PhD
Membro Interno

Regina Madalozzo, PhD
Membro Externo

Brasília
2022

Agradeço aos meus pais, Marcos e Ana, meu irmão, Gustavo, minha Avó Alba, minha namorada, Pietra, e meus amigos por todo o suporte. Sem eles nada disso seria possível.

Resumo

Este trabalho analisa a relação entre as despesas salariais dos times de futebol da primeira divisão inglesa e a performance esportiva dos clubes com a intenção de verificar se maiores gastos de fato são convertidos em melhor performance em campo e identificar qual a magnitude desse impacto. São adotadas 3 especificações diferentes do modelo, efeitos fixos, *System GMM* e *System GMM* com variáveis instrumentais exógenas para duas variáveis dependentes distintas (ranking da equipe e percentual de pontos obtidos). Ao final, consigo observar a relação positiva entre salário e performance além de identificar um comportamento interessante para a variável de tamanho do elenco. O modelo adotado nessa pesquisa apresenta um menor número de instrumentos em relação ao artigo que o motivou "*Aggregate wages of players and performance in Italian Serie A*", controlando assim para proliferação de instrumentos e *overfitting*.

Palavras-chave: Econometria do esporte. Futebol. Dados em painel.

Abstract

This work analyzes the link between wage expenditure among Premier League football clubs and their performance in the league. The main purpose is to check if higher wages imply on better performance on the pitch and to quantify this impact. I test three different econometric specifications for the two performance metrics chosen (team's ranking and points percentage). The regarded specifications are Fixed Effects, System GMM and Exogenous Instrumental Variable System GMM. In the end I am able to observe a positive relation between wages and performance and also check an interesting behavior for the roster variable. The models adopted in this research present a smaller group of instruments in comparison to "*Aggregate wages of players and performance in Italian Serie A*", which is the academic paper that motivated this research. The smaller group of instruments allow me to have a better control over instrument proliferation and overfitting.

Keywords: Econometrics of sports. Football. Panel data.

Sumário

1	Introdução	8
1.1	Panorama do futebol inglês	8
1.2	Revisão de literatura	16
2	Modelo Teórico	19
2.1	Introdução	19
2.2	Equações e solução do modelo	20
3	Modelo Empírico	29
3.1	Dados e Estatísticas descritivas	29
3.2	Variáveis	30
4	Especificação do modelo	33
4.1	Instrumentos Exógenos	35
5	Resultados	38
6	Conclusão	50
	REFERÊNCIAS	51
A	APÊNDICE A	53
B	APÊNDICE B	55

Lista de figuras

Figura 1 – Despesa salarial real anual dos clubes que disputam a primeira divisão inglesa - Milhões de libras	9
Figura 2 – Percentual do salário anual pago: Big 6 versus demais clubes	11
Figura 3 – Cotas de TV domésticas nominais e variação ano a ano - Premier League - £ Milhões	13
Figura 4 – Lucro/Prejuízo pre-tax dos clubes da Premier League por posição, 1993-2012	15
Figura 5 – Média das estatísticas descritivas	33
Figura 6 – Diferença salarial entre os Big6 campeões e os não campeões	46

Lista de tabelas

Tabela 1 – Campeões e Artilheiros da Premier League por temporada	12
Tabela 2 – Estatísticas Descritivas	30
Tabela 3 – Correlação de instrumentos com variáveis dependentes e endógenas	36
Tabela 4 – Validação dos Instrumentos	37
Tabela 5 – Segunda validação dos instrumentos	38
Tabela 6 – Resultados log(percentual_pontos)	40
Tabela 7 – Resultados log(rank)	41
Tabela 8 – Estudo da variável elenco	43
Tabela 9 – Resultados log(percentual_pontos) - modelo em log	47
Tabela 10 – Resultados log(rank) - modelo em log	48
Tabela A1 – Resultados log_rank sem real_GDHI	53
Tabela A2 – Resultados log_percentual_pontos sem real_GDHI	54

1 Introdução

1.1 Panorama do futebol inglês

A primeira divisão inglesa, *Premier League*, é uma das competições mais assistidas em todo o mundo e alcança 800 milhões de domicílios em 188 países. Em 2022, pela primeira vez a liga terá mais receita por acordos de transmissão internacionais que pelos contratos domésticos. Nos próximos três anos, espera-se que a liga receba £5,3 bilhões de libras pelos acordos internacionais, enquanto os contratos domésticos paguem £5,1 bilhões.¹

A internacionalização dos direitos de transmissão foi um ponto essencial para o desenvolvimento da indústria. Uma evidência disso é que as primeiras duas ligas a conseguir internacionalizar os direitos de transmissão e conseqüentemente arrecadar mais por esse mecanismo, *La Liga* (Espanha) e *Premier League* (Inglaterra), concentram o maior número de títulos de Liga dos Campeões (o mais importante torneio internacional de clubes) de toda a Europa.²

Em uma primeira análise, o tema do estudo parece trivial. Gastar mais permite adquirir mais talento e conseqüentemente montar um melhor time que ganha mais partidas. Mas se as coisas fossem estritamente lineares dessa forma, bastaria um gestor gastar mais que seus rivais e se tornaria campeão de tudo, lucrando com a operação e reinvestindo mais no próximo ano para ganhar novamente. Dessa forma, a indústria caminharia para um monopólio. Apesar disso, esse não é o formato que se observa.

Além de recorrentemente surgirem novos bilionários dispostos a despejar caminhões de dinheiro para formar equipes campeãs e competir com os atuais vencedores, discute-se a lucratividade da indústria. De fato compensa gastar mais para vencer mais? Existem exemplos na história que mostram o oposto de forma evidente e outros exemplos em que por mais que o time vença, os prêmios não compensam o volume gasto. Este estudo apresenta uma especificação teórica de Szymanski e Smith (1997), que modela como os clubes tentam balancear a rentabilidade e a performance no mercado de transferências e um modelo empírico que mensura o impacto percentual sobre a pontuação da equipe ao aumentar

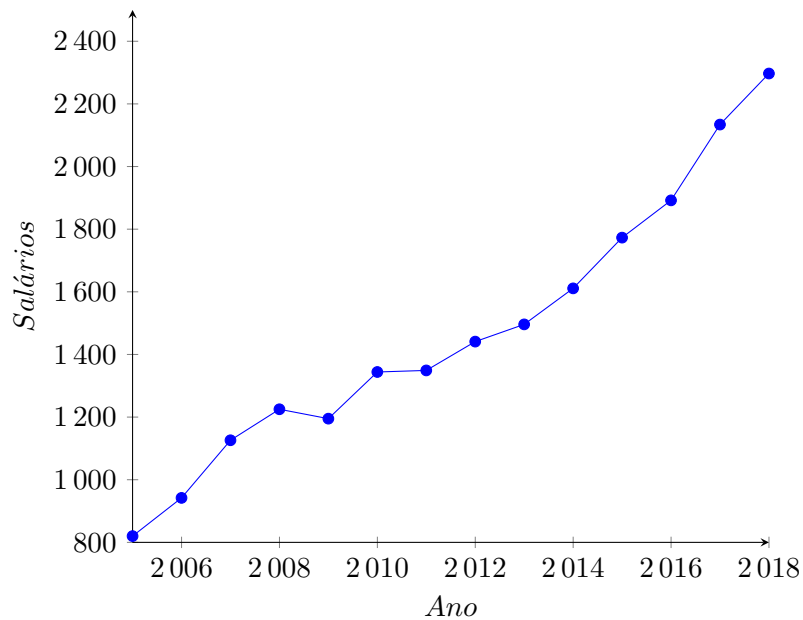
¹ <https://www.reuters.com/lifestyle/sports/premier-league-celebrates-30-year-rise-global-dominance-2022-08-16/:.text=Now%20broadcast%20to%20800%20million,billion%20followers%20on%20social%20media>.

² <https://www.uefa.com/uefachampionsleague/history/winners/>

em 1% o dispêndio com salários para o time.

A seguir, a figura 1, que indica a despesa anual real com salário de todos os times que disputam a primeira divisão inglesa ano a ano. Deve se notar que de 2005 até 2018 a despesa real dos clubes com salários aumentou mais de 180%. Simultaneamente, os prêmios, patrocínios e cotas de televisão acompanharam essa evolução, porém ainda é difícil obter dados oficiais sobre a evolução real dessas premiações. Diante disso, observa-se que o futebol inglês deixou de ser um pequeno entretenimento e passou para a escala de um evento global, com direitos de transmissão negociados em todos os continentes e clubes que gastam centenas de milhões de libras em salários anualmente.

Figura 1 – Despesa salarial real anual dos clubes que disputam a primeira divisão inglesa - Milhões de libras



Fonte: Governo do Reino Unido, *BBC Sport* e *The Guardian*

Isso leva a seguinte reflexão: os maiores gastos com salários, quando comparados à média dos times que competem na liga, levam a melhores performances esportivas? Qual o impacto no aumento marginal dos salários sobre a performance do clube em uma temporada de 38 partidas em que todos os times se enfrentam duas vezes? Este estudo busca recursos teóricos e empíricos para apresentar soluções a estes questionamentos. Além de se apoiar

na bibliografia existente para construir a argumentação e levantar questionamentos que virão ao longo deste estudo.

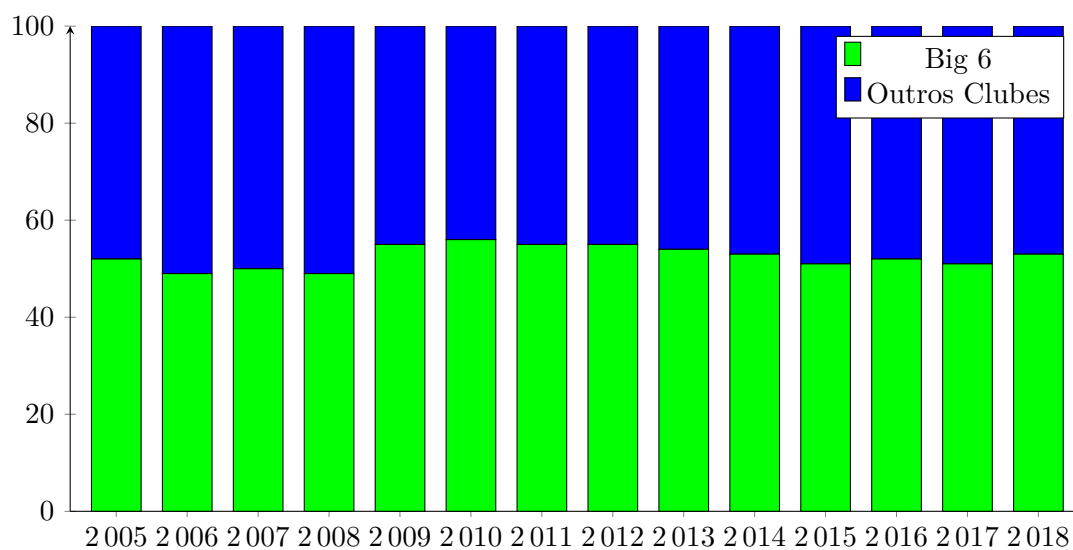
Com a evolução dos salários apresentada acima, é importante entender como esses salários são distribuídos entre os times que disputam a primeira divisão. Para fornecer essa visualização e motivar uma nova discussão, os clubes são separados em dois grupos: o *Big 6* e os demais times. O *Big 6* é um grupo formado pelos clubes de maior poderio financeiro entre os que disputam a liga. Estes são: Arsenal, Chelsea, Liverpool, Manchester United, Manchester City e Tottenham. Essa nomenclatura é recente e surge a partir da aquisição de três dos times por donos bilionários, enquanto outros três já possuíam forte tradição histórica na competição. Os recém-entrantes no *Big 6* são Chelsea, Manchester City e Tottenham. Os três foram inicialmente adquiridos por bilionários, porém o dono do Chelsea, que comprou o clube em 2003 e possuía negócios no setor de gás russo, teve de deixar o cargo por conta das sanções decorrentes dos conflitos entre a Rússia e Ucrânia em 2022 e o clube foi comprado por Todd Boehly, CEO da *Eldridge Industries*. O Manchester City foi adquirido em 2008 pelo xeque Mansour, membro da família real de Abu Dhabi, que pagou USD 212 milhões e até 2021 investiu USD 1,8 bilhão, segundo informações da Forbes para transformar um pequeno time em um multicampeão doméstico. Já o Tottenham, foi comprado em 2001 pelo *Enic group*, e um de seus acionistas majoritários, Daniel Levy, é hoje o presidente do clube. Os demais clubes que compõem o *Big Six* também são propriedade de bilionários. Entretanto, como mencionado, já possuíam uma tradição histórica e foram posteriormente adquiridos por empresários do setor esportivo. Esse é o caso do Arsenal, propriedade do bilionário Stan Kroenke, que possui atividades no setor petrolífero estadunidense e é dono dos Denver Nuggets (time de basquetebol da NBA) e dos Los Angeles Rams (time de futebol americano da NFL). O Manchester United foi adquirido pela família Glazer em 2003, com origem da fortuna no setor imobiliário estadunidense e donos dos Tampa Bay Buccaneers (time de futebol americano da NFL). O Liverpool é parte do grupo *Fenway Sports Group*, um tradicional investidor em clubes esportivos nos Estados Unidos.

Abaixo, a distribuição dos salários reais em milhões de libras ano a ano. Observe que na janela de tempo definida para esse estudo, o conjunto formado pelos seis clubes mais ricos

<https://www.forbes.com/sites/zakgarnerpurkis/2021/05/29/sheikh-mansours-biggest-win-manchester-citys-46-billion-increase-in-value/?sh=5dd6a8f66f2f>

(Big Six) paga, em média, 50% dos salários totais na liga. Isso é muito, haja vista que os 14 clubes restantes, aproximadamente 75% da liga, são responsáveis pelos outros 50%.

Figura 2 – Percentual do salário anual pago: Big 6 versus demais clubes



Fonte: Governo do Reino Unido, *BBC Sport* e *The Guardian*.

A análise se torna ainda mais interessante e motivadora para o estudo empírico que proponho quando se verifica que ao longo da janela de tempo estudada (2005 - 2018), em apenas uma temporada um clube não pertencente ao *Big 6* foi campeão da *Premier League* e em todas as temporadas o artilheiro da competição jogava por um dos clubes do *Big 6*, como podemos conferir abaixo:

Tabela 1 – Campeões e Artilheiros da Premier League por temporada

Temporada	Campeão	Artilheiro
2005	Chelsea	Thierry Henry (Arsenal)
2006	Manchester United	Didier Drogba (Chelsea)
2007	Manchester United	Cristiano Ronaldo (Manchester United)
2008	Manchester United	Nicolas Anelka (Chelsea)
2009	Chelsea	Didier Drogba (Chelsea)
2010	Manchester United	Berbatov (Manchester United)/Tévez (Manchester City)
2011	Manchester City	Robin van Persie (Arsenal)
2012	Manchester United	Robin van Persie (Manchester United)
2013	Manchester City	Luis Suárez (Liverpool)
2014	Chelsea	Sergio Agüero (Manchester City)
2015	Leicester City	Harry Kane (Tottenham)
2016	Chelsea	Harry Kane (Tottenham)
2017	Manchester City	Mohamed Salah (Liverpool)
2018	Manchester City	Mohamed Salah (Liverpool)

Fonte: Premier League

Essa tabela apresenta outra informação interessante. Antes de ser adquirido, o Manchester City havia conquistado a primeira divisão apenas duas vezes, em 1936 e 1967. Após a aquisição, foi campeão em 4 das 11 temporadas (dentro do horizonte do estudo) e em 6 das 14 temporadas até o fim da temporada 2021-2022. O Chelsea, antes de ser comprado, foi campeão apenas uma vez em 1954. Após a aquisição, o clube venceu 5 das 18 temporadas disputadas. É um aumento notável na performance após a injeção de um fluxo relevante de dinheiro no clube.

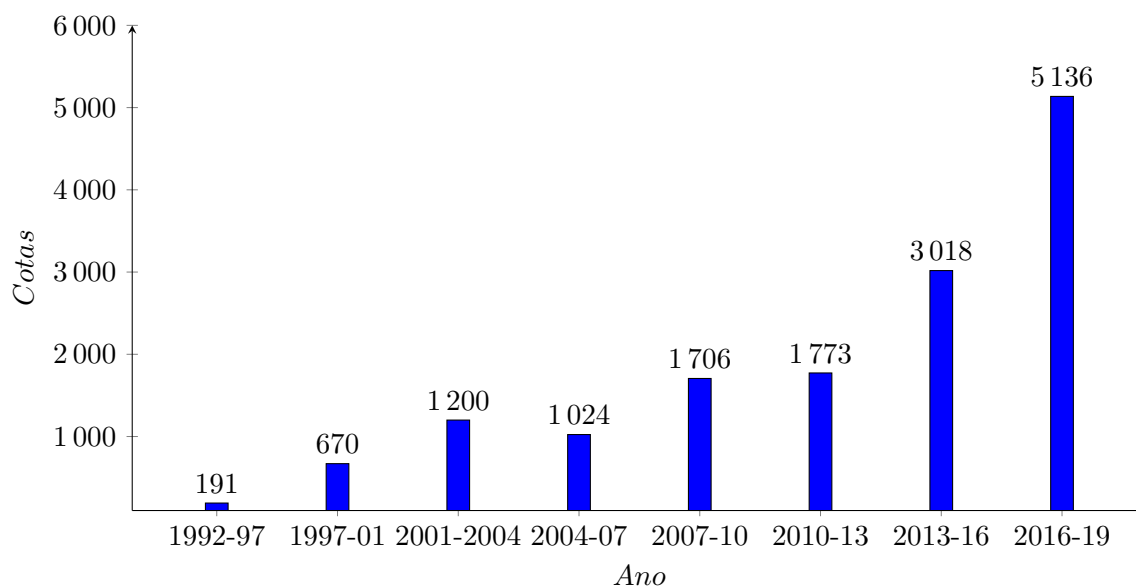
Por outro lado, clubes do *Big 6* como o Arsenal e Liverpool, tiveram uma piora significativa no número de triunfos na competição. O Arsenal não conquista um título desde 2003 e o Liverpool quebrou o jejum de títulos desde 1989 em 2019. Duas equipes de grande tradição e multicampeãs (terceiro e segundo maior campeão da competição, respectivamente) que possuíam uma estrutura financeira robusta, mas que não conseguiram suportar o grande

fluxo financeiro nos rivais recém-adquiridos que mudou o panorama da indústria e tomou talentos que poderiam ter o destino de um dos dois clubes.

Após a apresentação desses números, surge um questionamento: como surgiu tamanha atratividade para investimentos na indústria do futebol? Foram os retornos financeiros de patrocínios e cotas de transmissão de TV? Ou existem outros fatores envolvidos?

Essa discussão é iniciada abordando o ponto das cotas de televisão, uma das maiores fontes de receita dos clubes de futebol na Premier League. É inevitável que com o passar do tempo os valores evoluam, tanto pela inflação quanto pelo aumento de consumidores dessa forma de entretenimento. Entretanto, a velocidade com que os números aumentaram, reflete a agressividade dos lances que as emissoras de TV deram para obter o direito de transmissão da competição para capturar uma demanda crescente tanto dentro do Reino Unido quanto fora dele. Na figura abaixo, verifica-se a evolução dos valores em bilhões de libras e variação a cada renovação de contrato, com base em informações do *The Athletic*, um dos mais respeitados jornais esportivos do Reino Unido.

Figura 3 – Cotas de TV domésticas nominais e variação ano a ano - Premier League - £ Milhões



Fonte: *The Athletic*

Os direitos de transmissão internacionais são mais difíceis de mensurar, pois existe a necessidade de compilar dados de todos os países que transmitem a competição. Apesar disso, o *The Athletic* estima um montante de £780 milhões pago em 2016, além das cotas domésticas de transmissão que são divididos entre os 20 clubes participantes.

A divisão dos valores domésticos também ocorre de forma interessante. A proposta da divisão é não concentrar o pagamento em poucos clubes, mas tentar dividir de forma igualitária para que liga não perca sua competitividade. O formato atual propõe que 50% do valor total seja dividido de forma igualitária, 25% com base na colocação do time na tabela da competição (mérito da equipe) e os 25% restantes são definidos com base na quantidade de jogos da equipe transmitidos na TV, com uma remuneração de £1 milhão por transmissão e um piso de 10 partidas transmitidas por temporada. Em termos gerais, o campeão de 2016, Chelsea, obteve receitas totais de transmissão equivalentes a £150,8 milhões de libras e o último colocado, Sunderland, de £93,47 milhões, aproximadamente 61% a mais para o campeão. Pode parecer uma grande variação, mas é menor que o observado em outros países. Na Espanha, por exemplo, em 2013 o campeão foi o Atlético de Madri, que acumulou €42 milhões, já o Barcelona e o Real Madrid (segundo e terceiro colocados, respectivamente) acumularam €140 milhões e o lanterna, Almería, €18 milhões. No Brasil, o campeão de 2021, Atlético Mineiro, recebeu R\$ 66,1 milhões, enquanto o lanterna, Chapecoense, apenas R\$9,4 milhões, totalizando 600% de diferença entre o campeão e o lanterna. Diante dessa grande diferença para os times brasileiros, em 2022, a Rede Globo (principal emissora de TV responsável pela transmissão das partidas) reformulou o modelo de pagamento para um formato semelhante ao da *Premier League*, com os mesmos princípios de divisão, mas em um formato 40/30/30.

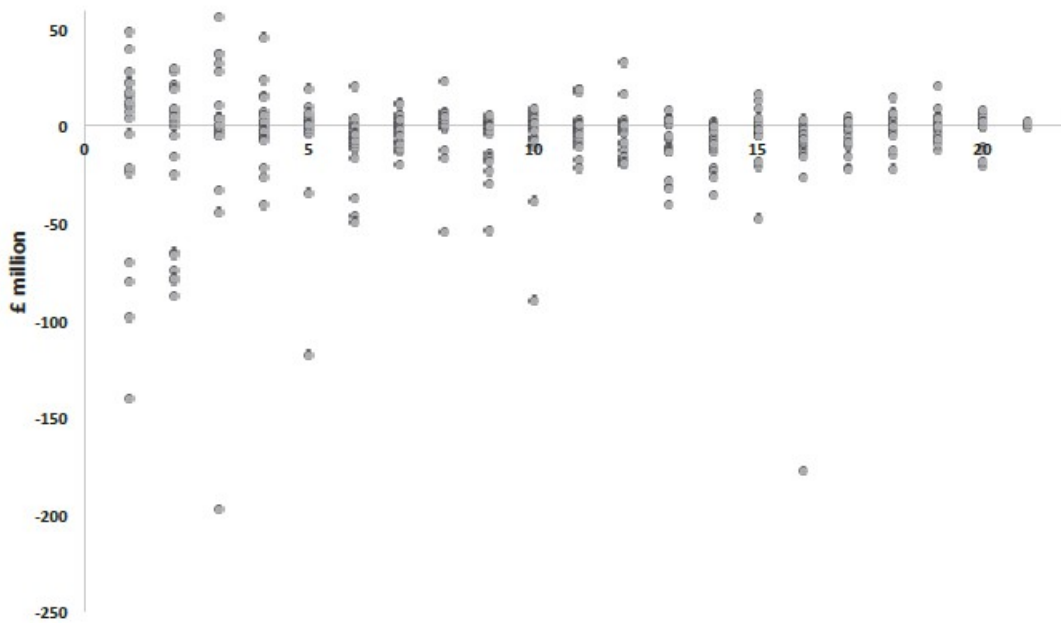
Com uma diferença tão grande no aporte que os donos realizam entre os clubes, considerar uma divisão com parâmetros para balancear os pagamentos é uma forma de ajudar os menores clubes e motivar novos empresários a entrarem como donos de equipes menores e igualar os fluxos exógenos de capital. Como no caso do Newcastle, recentemente adquirido pelo príncipe da Arábia Saudita Mohammed bin Salman, que realiza aportes a partir do Fundo Soberano do país.

<https://www.youtube.com/watch?v=a68Gw4rsV4>

<https://www.espn.com.br/futebol/artigo/id/10316945/liga-por-que-dinheiro-tv-ainda-divide-clubes-como-funciona-modelo-laliga-visto-como-ideal-maioria>

Porém, mesmo com esse expressivo número de receitas de transmissão, a conta parece não fechar. Segundo Szymanski e Kuper em *Soccernomics* (2010), poucas equipes são lucrativas e um dos motivos relevantes para a aquisição de clubes de futebol por bilionários é a proteção política e social que ser dono de um clube de futebol acrescenta à figura pública do proprietário. Abaixo, o leitor pode conferir um estudo apresentado no livro mencionado nesse parágrafo sobre o lucro/prejuízo *pre-tax* dos times da *Premier League* por posição de 1993-2012, que levam os autores a seguinte conclusão: quando empresários tentam gerir clubes de futebol como empresas, não só o futebol "sofre", como a empresa também. Mais do que isso, é interessante observar como esse comportamento se dá especialmente nas melhores posições da tabela (próximos ao eixo x), onde as apostas, medidas pelo resultado financeiro no eixo y, são maiores para tentar conquistar o campeonato, porém raramente recompensadas financeiramente.

Figura 4 – Lucro/Prejuízo pre-tax dos clubes da Premier League por posição, 1993-2012



Fonte: *Soccernomics*; eixo horizontal mede a colocação do time no campeonato

A divisão dos conteúdos desse trabalho ocorre em 6 seções. Esta que aborda a motivação da pesquisa, uma revisão de literatura sobre o tema e introduz os modelos e métodos

a serem abordados. A segunda seção, que traz um modelo teórico sobre contratações e características do mercado de transações de atletas. A seção 3, que apresenta as variáveis e a montagem da base de dados, a seção 4 que apresenta os modelos empíricos desenvolvidos, a seção 5 que aborda as variáveis incluídas, os resultados e as limitações de cada uma das abordagens estatísticas. Por fim, a seção 6 é uma síntese dos principais pontos do estudo e a conclusão da pesquisa.

Este trabalho analisa tópicos relacionados às características do mercado que os clubes se deparam ao tentar contratar novos atletas e a relação das contratações com a performance da equipe em sua liga doméstica, observando especificamente dados da primeira divisão inglesa, a *Premier League* e incluindo a performance na temporada anterior entre os regressores para verificar persistência entre as temporadas. O modelo empírico que motivou a pesquisa foi desenvolvido por Caruso et al. (2017), em um estudo sobre a relação dos salários pagos pelos clubes italianos e a performance das equipes na série A, a primeira divisão italiana.

1.2 Revisão de literatura

O aumento das cifras envolvidas no futebol naturalmente levou a uma maior profissionalização das áreas periféricas ao esporte. Departamentos jurídicos, médicos, táticos, financeiros e estatísticos passaram a apresentar uma relevância significativamente maior após a competitividade não envolver somente tradição, mas também volumosos prêmios em dinheiro, patrocínios e cotas de televisão que ajudam os clubes a financiar seus elencos milionários e, em alguns casos, bilionários. Nesse cenário, montar um time competitivo tornou-se extremamente caro. O dinheiro compra talento, e o talento é escasso. Os clubes calibram sua disposição a pagar a partir de um nível de qualidade técnica esperada e competem, tanto no mercado quanto em campo, com outras equipes de poderio financeiro e ambições distintas.

Para motivar o modelo empírico, este trabalho dedicará um capítulo ao modelo teórico desenvolvido por Szymanski e Smith (1997), apenas como forma de introduzir ao leitor uma perspectiva da literatura sobre o mercado de transferências e tecer observações. O modelo apresenta a dinâmica do mercado de transferências a partir de uma ótica semelhante àquela aplicada no modelo empírico apresentado neste estudo: os times competem financeiramente por talento e determinam sua disposição a pagar a partir de sua restrição orçamentária, sua

dotação inicial e pela ambição na temporada. Ademais, o modelo também aborda paralelamente uma discussão que não é o foco deste trabalho, mas é relevante para a literatura de econometria esportiva de forma geral, isto é: os clubes são maximizadores de lucro ou de performance? Em outras palavras, os proprietários dão maior relevância ao desempenho da equipe ou à saúde financeira? Esse questionamento pode ter desdobramentos complexos, haja vista a interação entre as duas variáveis e a observação da figura 4, onde é ressaltada a dificuldade de gerir clubes como empresas. Esse tópico é estudado com maior profundidade em Barrio e Szymanski (2009) . O propósito dessa seção não é trazer alterações ao modelo dos autores, mas sim apresentá-lo como referência bibliográfica ao estudo.

A contribuição inicial da literatura empírica vem de El Hodiri e Quirk (1971) e Scully (1974) que associam a performance das equipes à quantidade de talento que elas conseguem adquirir, assumindo a folha de pagamento como melhor proxy para os gastos com talento. Entretanto, mecanismos mais complexos entram em ação ao determinar a capacidade financeira e o sucesso das equipes. Hall et al. (2002) sugere que uma equipe vencedora aumenta sua receita pelas premiações, vendas de mercadorias licenciadas, maiores cotas de TV e ingressos. O aumento de receita permite uma maior aquisição de talentos e isso pode gerar dinastias nas ligas de futebol. Contudo, são raros os casos em que uma equipe domina completamente o cenário futebolístico doméstico e por isso existe o interesse em entender outros fatores, além da capacidade de geração de receita e de aquisição de talentos, que podem influenciar na performance.

Em uma literatura focada na parte empírica e econométrica, Szymanski (2004) considera a folha salarial e a posição final de 27 clubes da primeira e segunda divisões italianas entre 1987-2001. O autor percebe que em 62% dos casos um aumento da folha salarial impacta positivamente na posição final ocupada pela equipe. No mesmo livro, o autor compara os dados do futebol italiano com as duas primeiras divisões do futebol inglês, em uma janela de tempo de 1991-2001, e também verifica uma forte relação entre performance e pagamento de salários, constatando uma maior capacidade explicativa na liga italiana ($R^2 = 0,93$) do que na liga inglesa ($R^2 = 0,82$). Szymanski (2004) conclui essa pesquisa afirmando que os clubes italianos obtêm, em termos de resultado, aproximadamente o que eles pagaram. O autor, em outra publicação , apresenta dados da *Premier League* sobre as temporadas de 1993-2008, indicando uma forte correlação entre a performance dos clubes e os salários pagos. Além disso, Szymanski (2010) também reforça a perspectiva de que, apesar de

existir correlação entre o valor das transferências e os salários pagos, o segundo é uma melhor variável independente. Essa sugestão é recorrente na literatura e também é debatida no livro *Soccernomics* (2010) , onde Szymanski e Kuper (2010) também concluem que os salários agregados são uma melhor variável independente para estudos econométricos que o valor total gasto em transferências.

Por fim, a pesquisa mais recente, direcionada à parte empírica, foi realizada por Caruso et al. (2017) . No artigo, a partir de um modelo de efeitos fixos, e avançando para metodologias mais sofisticadas que visam interpretar a persistência do desempenho em temporadas anteriores como fator explicativo para a performance atual, os autores corrigem o viés dinâmico por um modelo de *Panel-Corrected Standard Errors (PSCE)* e incluem uma estrutura de autocorrelação AR1. Permitindo que os erros padrão sejam heterocedásticos e correlacionados entre os painéis, simultaneamente, corrigindo para autocorrelação de primeira ordem entre os painéis. Em sequência, os autores implementam um modelo *System-GMM* em primeiras diferenças, visando remover a autocorrelação e endogeneidade simultaneamente. Em todas as aplicações os autores encontram significância no coeficiente dos salários reais pagos, indicando que esses exercem influência sobre a performance das equipes na primeira divisão italiana. A abordagem empírica realizada neste estudo será pautada na metodologia que os autores aplicaram.

2 Modelo Teórico

2.1 Introdução

A proposta dessa seção é estudar a estrutura da indústria do futebol inglês pela ótica das transferências de jogadores e motivar a análise do modelo econométrico. Assim como no modelo empírico deste trabalho, os autores consideram o salário ofertado como o lance dado pelo jogador no mercado de transferências, analisam o comportamento dos times no mercado em busca de adquirir talento e estudam a formação dos equilíbrios no mercado. Será apresentado o modelo de Szymanski e Smith (1997) , como forma de revisão de literatura e tecer observações sobre as conclusões dos autores na publicação original.

A dinâmica do modelo assume as hipóteses a seguir: o talento é adquirido em um mercado competitivo de jogadores, onde o salário ajustado pela performance é determinado via equilíbrio de Nash pelos clubes. A quantidade de talento comprada determina diretamente a posição da equipe na liga doméstica e corresponde à função de produção da indústria que irei apresentar. A posição do time na liga determina a receita obtida pelo clube via bilheteria, cotas de televisão, patrocínios, vendas de mercadorias autenticadas, premiações e demais formas de receitas. Esse conjunto corresponde a função de demanda da indústria. Combinando a demanda estimada e a função de produção com a restrição orçamentária a ser imposta, obtemos o *trade-off* entre performance financeira e futebolística que o clube se depara. A função objetivo do proprietário do clube considera a saúde financeira do time e sua performance.

É importante mencionar que, para não termos uma trajetória explosiva em que apenas um clube com maiores dotações iniciais domina o cenário doméstico, os autores impõem que maiores gastos em atletas não são autofinanciados pela melhor performance e receita. Isso é compatível com o estudo de Barrio e Szymanski (2009) , onde os autores estudam o comportamento de clubes espanhóis e ingleses e verificam que as equipes gastam além do ótimo financeiro em aquisição de talentos e que não necessariamente isto é revertido em saúde financeira mesmo em caso de boa performance. Mesmo que os clubes individualmente se deparem com um *trade-off* negativo entre lucros e posição na tabela (isto é, maiores gastos não se financiam integralmente por performance), essa relação se dá de forma positiva entre eles, ou seja, melhores colocações pagam mais. Isso ocorre ao conside-

rar a restrição mencionada acima, que maiores gastos em atletas não são autofinanciados pela melhor performance, consequentemente maior receita. Apesar disso, eles competem entre si por melhores posições que levam a maiores prêmios, justificando o que é apresentado como *trade-off* positivo entre os times. Os clubes possuem diferentes dotações iniciais e capacidade de geração de receita. São essas diferenças que permitem uma distribuição não-idêntica de lucro e posição.

2.2 Equações e solução do modelo

No modelo apresentado por Barrio e Szymanski (2009) a folha salarial do clube é determinada como $W = wL$, onde w é o salário ajustado ao talento (definido via equilíbrio de Nash nas alocações individuais escolhidas por cada time), e L é a quantidade de talento empregada pelo clube. Assim como no mundo real, existe uma oferta limitada de talento e o mercado se equilibra (*market clears*) quando a despesa total de todos os clubes é igual à folha salarial total determinada pela oferta de talento, gerando w_t , medida de *salário/talento* no período t tal que:

$$w_t = \frac{\sum_{i=1}^N W_{i,t}}{\sum_{i=1}^N L_{i,t}} \quad (1)$$

Onde i = equipe, t = tempo e existem N equipes. Portanto, a despesa de cada clube depende da proporção do total de talento disponível que o clube demanda e do dispêndio total dos demais clubes. Isso ocorre, pois ao calcular W , os autores usam w_t que depende do comportamento dos demais clubes. Essa formulação assume, como já mencionado acima, a existência de um equilíbrio de Nash em salários. Ou seja: cada time propõe salários para contratar atletas conforme a maximização da sua função objetivo, que será discutida nesta sessão, assumindo que os demais clubes também agem da mesma forma, o que nos leva a um equilíbrio de melhor resposta, ou equilíbrio de Nash.

A partir disso, note que a parcela de talento que um clube contrata é igual a sua fração do total pago em salários por todos os clubes, isto é:

$$\frac{L_{i,t}}{\sum_{i=1}^N L_{i,t}} = \frac{w_t L_{i,t}}{w_t \sum_{i=1}^N L_{i,t}} = \frac{W_{i,t}}{\sum_{i=1}^N W_{i,t}} \quad (2)$$

Note que o primeiro termo da igualdade representa a fração do talento total da liga que o time contrata. Em seguida, é aplicada a igualdade $W = wL$ e a equação 1 para ir da primeira igualdade para a segunda e obter uma fração que representa o que foi exposto acima: a parcela de talento que um clube contrata é igual a sua fração do total pago em salários por todos os clubes.

Dando sequência à exposição do modelo, os autores caracterizam a tecnologia do clube por duas equações: uma função de produção e uma função demanda. A primeira define como a performance ($Q_{i,t}$) é produzida a partir dos salários ajustados ao talento pagos em um mercado competitivo, e de um intercepto individual c_i que capta as particularidades e será explorado no próximo parágrafo e de um coeficiente angular geral, que os autores assumem que na margem os clubes convertem salário em performance na mesma magnitude.

$$Q_{i,t} = c_i + dW_{i,t} \quad (3)$$

É evidente que os clubes diferem entre si em termos de performance por fatores como a idade dos jogadores, tipo de tática estabelecida pelo treinador e outros fatores intrínsecos a cada instituição. O intercepto c_i é responsável por captar tais particularidades. Ademais, os autores assumem uma conversão de salários em performance idêntica para todos os clubes da liga, captada por d que é invariante no tempo e entre os times.

Cada equipe se depara com uma restrição orçamentária que identifica a lucratividade do clube a partir de variáveis de decisão que serão modeladas. A restrição é apresentada como:

$$P_{i,t} = p_{i,t}G_{i,t} + O_{i,t} - T_{i,t} - N_{i,t} - W_{i,t} \quad (4)$$

Onde $P_{i,t}$ indica o lucro, $p_{i,t}$ o preço cobrado pelos ingressos, $G_{i,t}$ o público das partidas, $O_{i,t}$ as outras formas de receita (cotas de televisão, patrocínios, vendas de mercadorias autenticadas, premiações), $T_{i,t}$ é o saldo líquido de transferências e $N_{i,t}$ o conjunto de despesas não-salariais.

Cada um dos clubes se depara com uma função demanda de ingressos dos torcedores que, dada a performance da equipe $Q_{i,t}$, varia conforme os preços dos ingressos.

$$G_{i,t} = f(p_{i,t}, Q_{i,t}) \quad (5)$$

Para estimar essa demanda, os autores assumem que elas são lineares e separáveis em preço e performance. Podemos verificar que elas variam de acordo com cada clube e isso faz sentido sob uma ótica socio-econômica e geográfica. Cada região possui um montante diferente de habitantes, um histórico maior ou menor de investimento nos clubes locais e até uma maior possibilidade de internacionalização das marcas, como pode ser o caso de uma equipe nativa de Londres, uma cidade global. Barrio e Szymanski (2009) assumem que os clubes resolvem internamente uma maximização que determina o preço ótimo dos ingressos visando a maior receita líquida possível, ou seja, o preço dos ingressos é exógeno. Isso permite obter uma forma reduzida da função em que a receita de público depende exclusivamente da performance das equipes, ignorando um efeito *feedback* do público sobre a performance e seja uma limitação do modelo.

Considerando que o público das partidas ($G_{i,t}$) e as outras receitas especificadas acima ($O_{i,t}$) dependem da performance e são as únicas duas fontes de receita, com exceção do resultado financeiro das transações no mercado de transferências ($T_{i,t}$), define-se a função receita ($R_{i,t}$) diretamente relacionada à performance $Q_{i,t}$.

$$R_{i,t} = a_i + bQ_{i,t} \quad (6)$$

Note que, assim como a função $Q_{i,t}$, a função $R_{i,t}$ possui um coeficiente angular que na margem impacta as equipes na mesma intensidade (os clubes possuem a mesma capacidade de converter performance em receitas atreladas à performance). Entretanto, os autores assumem que as equipes possuem intercepto distinto a_i , captando a diferença na capacidade de geração de receita. Note que tanto a função $Q_{i,t}$, quanto $R_{i,t}$, possuem formato linear e de mesmo coeficiente angular entre os clubes, ou seja, clubes com intercepto maior nunca são alcançados pelos clubes com intercepto menor.

Os autores assumem que as despesas não atreladas diretamente ao salário ($N_{i,t}$) e o valor líquido das transferências ($T_{i,t}$) são relacionados aos custos salariais.

$$N_{i,t} + T_{i,t} = \tau W_{i,t} \quad (7)$$

Essa igualdade falha em captar características da indústria. Um dos evidentes problemas é o caso da contratação de jovens jogadores, que possuem um elevado valor de mercado por conta de seu potencial e durabilidade, contudo nem sempre recebem salários altos. Apesar disso, a equação 7 capta outros aspectos importantes relativos às finanças dos clubes. Times que pagam elevados salários usualmente possuem uma maior estrutura física e estão dispostos a arcar com maiores custos fixos, conseqüentemente é possível notar uma relação entre $N_{i,t}$ e $W_{i,t}$. Olhando para $T_{i,t}$, notamos que os clubes podem adquirir talento de duas formas: contratando no mercado competitivo que mencionamos, ou treinando atletas em suas categorias de base. Em ambos os casos estes atletas podem ser negociados com outras equipes, levando o clube vendedor a receber uma compensação financeira pelo atleta negociado. A partir disso, os clubes se deparam com uma folha salarial presente $W_{i,t}^A$ e buscam uma folha salarial ótima $W_{i,t}^*$ de acordo com sua alocação de performance obtida via equilíbrio de Nash no mercado por meio da equação 1. Essa adaptação ocorre por meio das transferências de atletas, de forma que $T_{i,t}$ pode ser negativo ou positivo.

$$T_{i,t} = v(W_{i,t}^* - W_{i,t}^A) \quad (8)$$

Portanto $T_{i,t}$ se torna uma função linear dos salários, como exposto acima. Equipes que desejam manter uma elevada folha salarial irão trazer atletas ($T_{i,t} > 0$) e equipes que desejam economizar irão vendê-los ($T_{i,t} < 0$). Em uma abordagem sequencial, é possível verificar a ascensão de equipes menores que se tornam especialistas em formar e vender atletas de sua categoria de base e começam a financiar um projeto de longo prazo para o clube, permitindo gradualmente adquirir mais talento e figurar entre as melhores equipes, como o Red Bull Bragantino no Brasil. Esse termo é responsável por dar uma dinâmica intertemporal entre as equipes, de forma que clubes que formam atletas vendidos por um maior valor conseguem disputar melhores prêmios nas temporadas seguintes.

Voltando ao desenvolvimento do modelo, a equação 4 é reescrita e define o lucro a partir das equações explicitadas acima. O passo a passo será disponibilizado no apêndice.

$$P_{i,t} = R_{i,t} - (1 + \tau)W_{i,t} \quad (9)$$

Portanto, o *tradeoff* entre performance e lucratividade com o qual o dono do clube se depara é:

$$P_{i,t} = \pi_i + \phi Q_{i,t} \quad (10)$$

onde

$$\pi_i = \frac{(a_i d + (1 + \tau)c_i)}{d} \quad (11)$$

e

$$\phi = b - \frac{(1 + \tau)}{d} \quad (12)$$

Os parâmetros das equações 11 e 12 serão explicados a seguir. Todos os clubes se deparam com o mesmo coeficiente angular, ou preço de aquisição do talento (ϕ), contudo diferem na dotação, determinada pelo intercepto (π_i), apresentado em 11. O π_i é composto pela capacidade do clube em transformar talento em performance (c_i), pela capacidade de gerar receita (a_i), pelo coeficiente angular da equação 3 (d), que relaciona performance e salários ofertados para contratar e por (τ), o parâmetro que modela a fração dos salários que é dedicada a custos não relacionados diretamente ao salário $N_{i,t}$ e $T_{i,t}$, o valor líquido das transferências, apresentados na equação 7.

Por meio da equação 11, identifica-se que uma maior capacidade de gerar receita a_i e uma maior capacidade de transformar talento em performance c_i multiplicado por $(1 + \tau)$, levam a um maior intercepto para a equação 10. A única parte não intuitiva dessa afirmação é como $(1 + \tau)$ entra de forma positiva. Os autores não exploram isso de forma clara. Na minha leitura, é uma forma de introduzir uma parcela positiva do impacto do talento sobre a lucratividade, uma vez que o coeficiente angular ϕ é negativo por hipótese. De forma direta, (π_i) é a capacidade individual de lucratividade de cada um dos times. Ele também pode ser observado como uma parte do efeito fixo que irei estudar no modelo empírico, uma vez que carrega condições particulares de cada um dos clubes.

No que lhe concerne, a equação 12 identifica o coeficiente angular da função. O parâmetro b é a taxa constante entre os times de conversão de performance em receita e o termo

negativo o parâmetro de gastos com transferências e custos não relacionados τ normalizado pela capacidade do clube transformar salários em performance d . Observe que isso pode ser interpretado como um preço da qualidade constante entre os times dessa economia.

Em adendo, se os lucros fossem uma função positiva da performance, o proprietário do time não teria problema a ser resolvido. A estratégia ótima seria gastar ao máximo em talentos para obter boas performances e automaticamente maior lucro. Este é o problema que a equação 12 aborda e apresentado no início da sessão quando a concentração de todo o talento da liga em apenas um clube é discutida. Logo, é importante equilibrar performance e lucratividade, não havendo solução de canto.

Passando para o estudo da função objetivo do proprietário, como mencionado anteriormente, o dono do clube leva em consideração tanto a performance $Q_{i,t}$ como os lucros $P_{i,t}$ do time, de forma que sua função objetivo (bem-estar) no período t pode ser especificada como:

$$W_t = P_{i,t} - \frac{\alpha}{2} P_{i,t}^2 + \beta Q_{i,t} - \frac{\gamma}{2} Q_{i,t}^2 \quad (13)$$

Na minha percepção, modelar a função objetivo do proprietário considerando a performance esportiva e financeira do clube é uma boa forma de entender o problema. Nenhum clube de futebol é imune a má gestão financeira. Aumentar as despesas excessivamente em um curto espaço de tempo para tentar ganhar competições pode ser uma receita para o desastre, até mesmo no caso em que o clube conquista títulos, como foi o caso do Cruzeiro no Brasil. Quando o nível de despesas se torna insustentável diante do montante de dívidas, o clube precisa dispensar atletas e passar vários anos distante da possibilidade de competir entre os melhores clubes do país, correndo até risco de ser extinto em determinados casos.

Por outro lado, nos últimos 10 anos tornou-se comum a aquisição de clubes de futebol por grandes bilionários, como Roman Abramovich (proprietário do Chelsea). Nesse caso, o comprador busca prestígio e proteção social por ter um clube de futebol e dificilmente enfrenta uma restrição orçamentária. Apesar disso, a FIFA introduziu um mecanismo de *Fair Play* financeiro que previne despesas exacerbadas e limita os gastos nas janelas de transferência com base nas receitas originadas pelo próprio operacional do clube.

Ao resolver a maximização, disponível no apêndice, resolvendo para $Q_{i,t}$, a variável de performance, obtemos:

$$Q_{i,t}^* = \frac{(\beta + \phi(1 - \alpha\pi_i))}{(\gamma + \alpha\phi^2)} \quad (14)$$

Para $P_{i,t}$, a variável de lucro, a seguinte expressão determina se o clube terá lucro ou prejuízo:

$$P_{i,t}^* = \frac{(\pi_i\gamma + \phi(1 + \phi\beta))}{\gamma + \alpha\phi^2} \quad (15)$$

Realizando uma análise de sensibilidade dos valores ótimos de $Q_{i,t}^*$ e $P_{i,t}^*$ verifica-se como essas variáveis se comportam diante de variações nos parâmetros que as determinam. Inicialmente, note que o vetor $[\beta, \phi, \alpha, \gamma]$ é composto por constantes, logo, sob a especificação que proponho aqui, não devem variar entre as equipes nem entre períodos. Entretanto, π_i varia entre equipes e suas variações ocorrem consoante o comportamento de a_i e c_i , que correspondem ao intercepto da função receita e da função performance respectivamente.

Derivando ambas as expressões em relação a π_i obtemos:

$$\frac{\partial Q_{i,t}^*}{\partial \pi_i} = \frac{-\phi\alpha}{\alpha\phi^2 + \gamma} > 0$$

$$\frac{\partial P_{i,t}^*}{\partial \pi_i} = \frac{\gamma}{\alpha\phi^2 + \gamma} > 0$$

(16)

Note que quando definimos o *tradeoff* entre performance e lucratividade pela equação 10, o coeficiente ϕ passa a ser negativo, mesmo que representado com um sinal positivo a sua frente. As demais constantes são, por hipótese, positivas.

Pelas expressões obtidas para $Q_{i,t}^*$ e $P_{i,t}^*$ e suas derivadas parciais em relação a π_i , é notável que as duas variáveis estudadas variam positivamente em relação a π_i . Além disso, parâmetros que determinam essa variação positiva e inferir o que eles indicam sobre o modelo. Observe que em 11, π_i , a capacidade individual de lucratividade de cada um

dos times, é determinado pela capacidade de gerar receita do time a_i e pela capacidade de transformar talento em performance c_i . Intuitivamente, times com maiores valores para essas variáveis conseguem melhores performances financeiras e esportivas. Ademais, quando as derivadas parciais dos valores ótimos em relação ao parâmetro π_i são analisadas, fica mais claro o impacto dos pesos relativos da equação 13, γ e α , para $P_{i,t}$ e $Q_{i,t}$. Um maior γ indica um menor peso relativo à performance $Q_{i,t}$ e permite uma maior lucratividade, indicado pela derivada parcial de $P_{i,t}$. De forma similar, um maior α indica um menor peso relativo à lucratividade e leva a uma decisão de maior investimento em performance do elenco, indicada pela derivada parcial em relação a $Q_{i,t}$.

Após estudar o modelo dos autores, elenco forças e limitações na especificação. Inicialmente é relevante mencionar a existência de uma literatura inicial em termos de modelagem teórica no mercado de transferências de futebol. A percepção do equilíbrio de Nash como forma de alocar talento, sujeito a uma restrição orçamentária, modela razoavelmente o comportamento das equipes. O *tradeoff* entre performance e lucratividade também é uma força do modelo ao se aproximar da realidade. Porém, diante das aquisições de clubes por bilionários, a restrição orçamentária passa a ser imposta pela FIFA e não pela performance esportiva da equipe ou pela capacidade de geração de receita do clube. É um consenso entre executivos e empresários que a transferência do Neymar para o *Paris Saint-Germain* em 2017 por €222 milhões representa bem essa transição em que os clubes mais ricos demandam de forma inelástica o talento e não estão sujeitos a restrições financeiras internas. Vale pensar se a forma com que é modelado o comportamento dos times mais ricos deve diferir dos menores clubes que precisam se comportar como empresas para sobreviver na primeira divisão e se consolidar gradualmente.

Ainda sobre os pontos pouco aprofundados no modelo em que novas pesquisas podem trazer melhoramentos, percebo que a formação de atletas na categoria de base deve ser percebido como uma parcela relevante da obtenção de receita dos clubes. Verifica-se que boa parte dos times brasileiros se financiam dessa forma atualmente, isto é, exportando jovens talentos. Na primeira divisão inglesa não é diferente. Os menores clubes recorrentemente vendem suas promessas para clubes maiores e demandantes inelásticos por talento.

Por fim, acredito que a introdução de um comportamento dinâmico e não estático no coeficiente angular das funções, tanto entre as equipes quanto para uma mesma equipe entre

temporadas, seria um avanço difícil de modelar, porém de grande valor para a literatura.

3 Modelo Empírico

3.1 Dados e Estatísticas descritivas

Para explorar o problema de performance e finanças no futebol inglês, construí uma base de dados com 280 observações de times da primeira divisão da Inglaterra em 14 temporadas diferentes. Essa é uma contribuição da minha pesquisa, uma vez que esses dados agrupados não estão disponíveis. A base é composta de informações sobre as temporadas disputadas entre 2005 e 2018 com dados para 39 equipes diferentes que disputaram a primeira divisão inglesa no período, resultando em um painel desbalanceado. As informações de elenco, média de idade, salário real, número de estrangeiros e performance (rank e percentual de pontos) foram obtidas por meio de *web scrapping* no site *Transfermarkt*, uma base de dados alemã que dispõe de informações sobre clubes de futebol do mundo inteiro. Já os dados de salários reais foram compilados por meio dos portais de notícias *BBC Sport*, *TalkSport*, *Express UK* e *The Guardian*, que produzem acompanhamentos dessas informações. Idealmente, esses dados deveriam ser divulgados por meio de apenas um portal da *Football Association*, o órgão inglês que regula o futebol inglês, em outros termos, a CBF inglesa. Os dados não são abertos ao público pela associação, o que implicou em compilar os dados para a série histórica a partir dos portais de notícias. Para obter os valores reais dos salários, obtive os dados de inflação no Reino Unido a partir do portal *Office for National Statistics UK* e transformei os salários em milhões de libras por ano. Após compilar os dados, apliquei transformações logarítmicas em todas as variáveis para testar diferentes especificações dos modelos e apresentar as que julguei mais fáceis de interpretar. É importante ressaltar que dois clubes não apresentaram dados para salários pagos ao longo de temporadas específicas, rebaixando o número de observações para 278, como podemos conferir na tabela 2.

O motivo pela seleção das variáveis apresentadas é discutido abaixo, onde cada uma delas é abordada individualmente. De antemão, apenas a variável de salário real será incluída em log para análise de elasticidade. As demais variáveis serão estudadas em nível para uma interpretação mais clara. Para fins de comparação com o trabalho de Carusto et al. (2017), apresentarei os resultados do modelo com todas as variáveis em log ao final da seção de resultados.

Tabela 2 – Estatísticas Descritivas

	Média	DP	Min	Max	N
Ano	-	-	2.005	2.018	280
Elenco	35,40	5,50	21,00	54,00	280
Idade	24,78	1,14	21,30	28,10	280
Salário Real (£ mm)	74,26	47,78	12,11	244,58	278
Estrangeiros	22,34	5,47	8,00	41,00	280
Rank	2,78	4,58	0,05	20,00	280
Percentual de pontos	0,46	0,15	0,10	0,88	280

A janela de estudo é motivada por um período onde a liga inglesa recebeu fortes aportes de capital estrangeiro via investidores e patrocinadores e evita usar informações referentes ao período da pandemia, quando o mundo se deparou com situações extraordinárias.

3.2 Variáveis

Como mencionado nas seções anteriores, a intenção deste trabalho é verificar se maiores salários pagos aos atletas implicam maior sucesso na liga doméstica inglesa, a *Premier League*. Para testar essa hipótese empiricamente é necessário definir uma métrica objetiva de sucesso na competição, que será a variável dependente dos modelos apresentados a seguir.

Para captar a performance dos clubes foram selecionadas duas variáveis dependentes distintas para trazer maior robustez aos resultados dessa pesquisa. A primeira delas é sugerida por Szymanski e Smith (1997) e implementada por Szymanski (2004) e Caruso et al. (2017) para medir o sucesso dos clubes no futebol italiano. Ela é computada a cada temporada disputada, varia segundo a posição do clube ao final do campeonato e enfatiza a diferença entre clubes posicionados no topo ou ao final da tabela da competição em relação aos clubes que terminam na metade da tabela. A fórmula que define a variável é apresentada como:

$$rank_{i,t} = -\ln\left[\frac{P_{i,t}}{N_t + 1 - P_{i,t}}\right] \quad (17)$$

Onde $P_{i,t}$ representa a posição na tabela do i -ésimo clube na temporada t e N_t indica o número de clubes disputando a competição, fixado em 20 no painel de dados. Para es-

clarecer a formulação, maiores valores de $rank_{i,t}$ representam uma melhor performance. Portanto, o primeiro colocado terá uma fração menor em módulo do que o último colocado, porém o sinal negativo que antecede a fração corrige esse ordenamento e permite interpretações unidirecionais.

É importante notar que essa variável tem um comportamento diferente do percentual de pontos em termos de progressão na tabela de classificação. Uma equipe que ocupa a última posição possui um valor de 20 para o rank, enquanto uma equipe que ocupa a penúltima colocação possui um valor de 9,5. Essa diferença decresce ao subir a colocação na tabela e impõe uma grande variação entre o valor do primeiro e último colocado, alterando a escala dos coeficientes associados aos regressores em relação ao percentual de pontos.

A segunda abordagem escolhida para medir sucesso na liga doméstica é proposta por Franck e Nüesch, (2010) e apresenta o percentual de pontos obtidos por cada equipe i em cada temporada t :

$$percentual_pontos_{i,t} = \frac{pontos_{i,t}}{pontos_t} \quad (18)$$

O numerador representa o número de pontos obtidos pela equipe i na temporada t e o denominador o máximo de pontos a serem obtidos durante o torneio.

A escolha das variáveis independentes é inspirada na pesquisa de Caruso et al. (2017), como mencionado no primeiro capítulo. A primeira, e mais importante variável independente do estudo, é o salário real em milhões de libras pago ao elenco das equipes em valores anuais. A variável é especificada no modelo como $salreal$ e está em milhões de libras.

A segunda variável independente incluída no modelo é a média de idade dos atletas que compõem o elenco, denominada $idade$. A inclusão dessa variável capta se a média de idade possui impacto direto na performance do clube na competição e determinar se esse impacto é positivo, isto é: jogadores mais velhos tendem a trazer mais pontos para o clube, ou se é negativo, indicando que uma maior juventude no elenco permite maior rendimento. A média de idade também é incluída com um termo quadrático $idade^2$ que permite

identificar um ponto ótimo em termos de média de idade do plantel para obter mais pontos. Além de Caruso et al. (2017), Buccioli et al. (2014) e Yamamura (2013) também incluem média de idade como um dos regressores para controlar o impacto dessa variável sobre a performance.

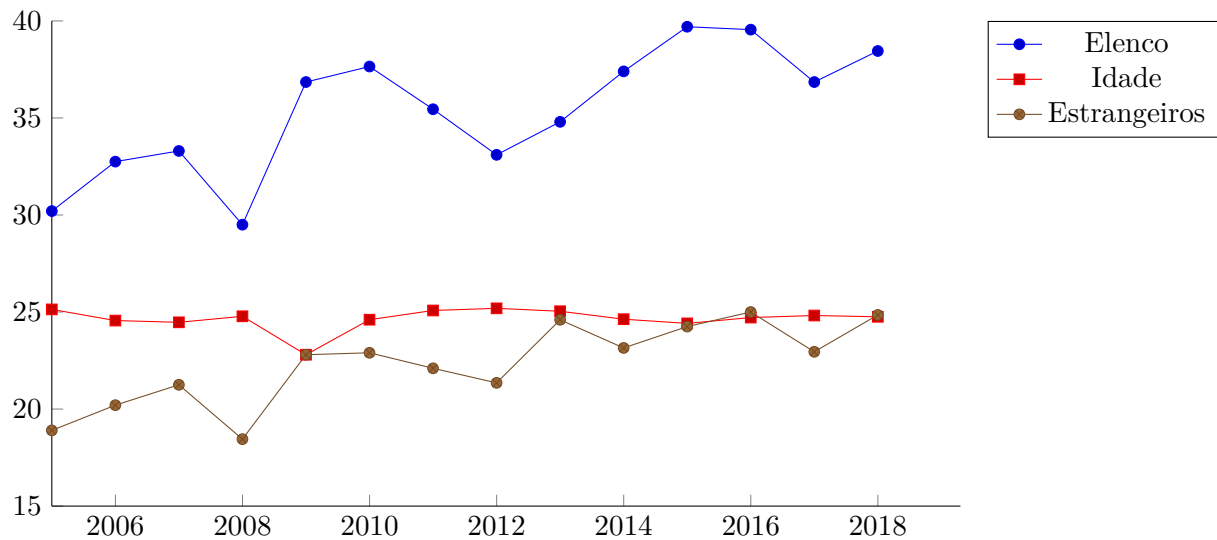
Com um intuito semelhante à variável de idade, o tamanho do elenco é incluído como regressor do modelo. Aqui, a motivação para incluir a variável independente (elenco) surge pela discussão se um maior elenco permite uma maior obtenção de pontos por disponibilizar ao treinador mais peças de reposição em casos de lesões e suspensões e também se mais jogadores que possuem características distintas podem ajudar o time em situações diferentes. Por outro lado, mais atletas pode dificultar a obtenção de uma escalação ideal, uma vez que o treinador teria mais combinações de jogadores a serem testadas em comparação ao elenco mais enxuto. Para testar esse comportamento e verificar se existe um ótimo no número de jogadores no elenco, a variável elenco quadrado é incluída. Essa variável é apresentada no trabalho de Franck e Nüesche (2010) e amplamente adotada na literatura.

Uma das hipóteses do estudo nas seções anteriores é de que o mercado de transferência de atletas pode ser considerado eficiente. Contudo, existem imperfeições que merecem atenção e induzem a inclusão do regressor estrangeiros, que tenta captar o efeito dessas imperfeições. Os clubes estão dispostos a pagar maiores salários para os jogadores quando tem uma percepção sobre seu nível de talento, contudo, nem todos os clubes têm acesso aos mesmos atletas. As equipes possuem profissionais de *scout* que analisam atletas em outros países e continentes para tentar adquirir talentos mal precificados. Assim, é justificada a inclusão da variável estrangeiros, que contabiliza o número total de estrangeiros no elenco. Uma crítica a inclusão dessa variável é que os atletas estrangeiros nem sempre são desconhecidos e podem ser estrelas transferidas de outras ligas com elevados salários. Exemplos recentes são a transferência do Haaland (norueguês) ou Darwin Núñez (uruguaio) na última janela de transferências para Manchester City e Liverpool, respectivamente. Outra forma de analisar essa variável é como um teste se atletas ingleses performam, em média, melhor ou pior que os estrangeiros. A inclusão dessa variável é sugestão de Migali et al, (2014).

O gráfico a seguir apresenta o comportamento médio das variáveis independentes ao longo do estudo. Observa-se uma baixa amplitude na média de idade dos elencos e no número de estrangeiros, apesar de uma tendência crescente na segunda variável. Já o tamanho do

elenco cresce significativamente entre o começo e o fim do horizonte de estudo. A hipótese levantada é que a evolução das categorias de base leva os clubes a registrarem jovens jogadores (categorias de base) no time principal para compor o elenco e utilizá-los apenas em situações esporádicas como forma de teste. Durante o restante da temporada os atletas jogam nos times sub-23, que funcionam como um último passo até o futebol profissional. A *Premier League* permite que isso seja feito.

Figura 5 – Média das estatísticas descritivas



4 Especificação do modelo

Para identificar o impacto das variáveis independentes elencadas acima sobre as métricas de sucesso propostas, dois métodos estatísticos serão apresentados para trazer maior robustez aos resultados desse trabalho e tratar problemas referentes aos regressores. Inicialmente uma especificação em efeitos fixos, que busca captar o efeito das variáveis independentes por meio de uma abordagem OLS e considera um termo de efeitos fixos para cada time (μ_t), para controlar os fatores invariantes ao longo do tempo e intrínsecos a cada um dos clubes analisados no painel. Além disso, essa especificação elimina a correlação entre a

<https://theathletic.com/3370929/2022/06/21/premier-league-transfers-homegrown/>

heterogeneidade não observada particular aos times e as variáveis explicativas.

O modelo enunciado acima possui a seguinte especificação:

$$\ln(y_{i,t}) = \alpha + \beta_1 \ln(\text{salreal}_{i,t}) + \beta_2 (\text{idade}_{i,t}) + \beta_3 [(\text{idade}_{i,t})]^2 + \beta_4 [\text{estrangeiros}_{i,t}] + \beta_5 (\text{elenco}_{i,t}) + \beta_6 [(\text{elenco}_{i,t})]^2 + \mu_i + \epsilon_{i,t} \quad (19)$$

Onde $\epsilon_{i,t}$ é o termo de erro estocástico e μ_i é o termo específico para cada equipe e invariante no tempo, portanto, a variável responsável por captar os efeitos fixos.

A decisão pelo modelo de efeitos fixos em relação ao modelo de efeitos aleatórios foi resultado da interpretação dos resultados fornecidos por ambas especificações e da observação de melhor ajustamento a essa especificação. Ademais, a literatura permite uma inclinação à escolha de um modelo de efeitos fixos pela crença de que os clubes possuem particularidades que não são distribuídas de forma aleatória, especialmente em um recorte temporal de 14 temporadas que carrega informações de anos anteriores, as quais interferem nas dotações iniciais de riqueza, torcida, talento, acesso a patrocínios, tipo de governança e outros fatores.

Na segunda especificação serão tratados dois problemas comuns na literatura de salários e performance esportiva: a endogeneidade e a potencial causalidade simultânea das variáveis, como apontado por Hall et al. (2002). Os autores notam que existe endogeneidade nos salários reais pagos, uma vez que o sucesso da equipe em uma temporada tende a afetar a folha salarial do clube no próximo ano. Dessa forma, ao não incluir πy_{t-1} na especificação, o termo de erro passaria a captar efeitos dessa variável e violaríamos a hipótese de exogeneidade estrita dos regressores, uma vez que $E[\mathbf{X}\epsilon_{i,t}] \neq 0$. Portanto, a folha salarial em "t" deve ser considerada endógena em relação às realizações passadas da variável que mede performance da equipe.

Outro problema é a persistência da performance entre temporadas: o trabalho realizado no ano anterior não é completamente zerado quando um novo ano inicia. Diversos jogadores permanecem em seus clubes (não buscam transferência) e os treinadores também

podem permanecer, o que gera continuidade do trabalho e alguma dependência temporal de $y_{i,t}$ em relação à realização de $y_{i,t-1}$. Dada essa hipótese, faz-se necessário apresentar um modelo que considere esse comportamento dinâmico da variável dependente e possa instrumentalizar esse regressor a partir de realizações passadas.

Ademais, se por hipótese assumo que as realizações passadas de y_i podem afetar as realizações presentes de $salreal$, o modelo incorre em uma violação da exogeneidade estrita. Essa situação introduz um viés dinâmico no modelo que pode ser resolvido a partir da instrumentação com defasagens e instrumentos exógenos.

$$\Delta \ln(y_{i,t}) = \pi \Delta \ln(y_{i,t-1}) + \beta_1 \Delta [\ln(salreal_{i,t})] + \beta_2 \Delta (idade_{i,t}) + \beta_3 \Delta [(idade_{i,t})]^2 + \beta_4 \Delta (estrangeiros_{i,t}) + \beta_5 \Delta (elenco_{i,t}) + \beta_6 \Delta [(elenco_{i,t})]^2 + \Delta \epsilon_{i,t} \quad (20)$$

As duas questões relatadas acima são apresentadas em Caruso et al., (2017) e tornam o problema a ser estudado um caso tratável a partir da especificação System GMM, como observado em Arellano e Bond (1991), Blundell e Bond (1998), Arellano e Bover (1995) e Roodman (2009). A análise dos dados por uma abordagem System GMM permite expurgar a endogeneidade da variável dependente defasada que é incorporada entre os regressores, $y_{i,t-1}$, e dos salários reais ($salreal$) que por hipótese é considerado endógeno. Após obter um modelo em primeira diferença, o tratamento da endogeneidade dessas variáveis ocorre pela instrumentalização por defasagens e de instrumentos exógenos que serão discutidos a seguir.

4.1 Instrumentos Exógenos

Para reforçar o tratamento dos regressores entendidos como endógenos, duas variáveis instrumentais excluídas são selecionadas. Estas denominadas assim por não pertencer à equação estrutural. Essa abordagem também foi adotada por Caruso et. al. (2017) utilizada para dar maior robustez à instrumentalização dos regressores entendidos como endógenos, além de capturar fatores socioeconômicos não compreendidos pelas variáveis já incluídas no modelo que podem afetar a folha salarial. A folha de pagamento dos clubes sofre influência histórica das condições socio-econômicas e ambientais de onde o clube é

estabelecido, uma vez que a instituição está sujeita a maiores ou menores investimentos decorrentes das condições da localização de estabelecimento. É comum observar esse tipo de análise quando se discute cultura e crescimento econômico ou demais tópicos de persistência.

A partir disso, é selecionada uma medida de riqueza amplamente utilizada na literatura econômica, o PIB per Capita Real (GDHI - Gross Domestic Household Income) de cada uma das regiões as quais os clubes pertencem para ser um dos instrumentos do salário real. O segundo instrumento selecionado foi a altitude da cidade onde o clube foi estabelecido. A argumentação em que baseia a seleção da segunda variável é que por muito tempo as regiões mais ricas da Inglaterra foram regiões localizadas ao nível do mar, por conta das trocas marítimas. Com isso, serão apresentados os testes de exogeneidade e relevância dos instrumentos selecionados, validando via forma reduzida da equação e inspecionando a matriz de correlação das variáveis relevantes.

O primeiro passo para a validação dos instrumentos selecionados é verificar se esses são correlacionados com a variável que ele irá instrumentar, pois a correlação entre o instrumento e as variáveis dependentes deve ser exclusivamente via variável endógena. Abaixo é possível ver o resultado do teste de correlação:

Tabela 3 – Correlação de instrumentos com variáveis dependentes e endógenas

<i>Variável Instrumental</i>	<i>log_rank</i>	<i>log_percentual_pontos</i>	<i>log_salreal</i>
<i>PIB per capita real (real_GDHI)</i>	0.1294*	0.1397*	0.1842*
<i>Altitude (altitude)</i>	-0.1291*	-0.1451*	-0.1735*

*Nota: * indica 5% de significância estatística*

Além disso, é necessário estimar um modelo MQO incluindo as variáveis instrumentais entre os regressores. Caso haja significância estatística dos coeficientes associados a esses instrumentos, é um argumento contra a validade dos instrumentos escolhidos, uma vez que o modelo incorreria em viés de variável omitida ao não considerá-las entre os regressores (Wooldridge 2002) . A tabela abaixo indica o resultado da regressão realizada para testar a validade dos instrumentos considerando cada uma das variáveis dependentes como $y_{i,t}$:

Tabela 4 – Validação dos Instrumentos

<i>Var. Dep.</i>	<i>log_rank</i>		<i>log_percentual_pontos</i>	
<i>Var. Instrumental</i>	<i>altitude</i>	<i>real_GDHI/1000</i>	<i>altitude</i>	<i>real_GDHI/1000</i>
<i>Instrumento</i>	-0.0019 (0.0014)	0.0105 (0.0174)	-0.0004 (0.0003)	0.0033 (0.0037)
<i>Lag Var. Dep.</i>	0.7881*** (0.0608)	0.7967*** (0.0605)	0.8569*** (0.0601)	0.8656*** (0.0597)
<i>idade</i>	0.2473 (1.7250)	0.2577 (1.7456)	0.0934 (0.3744)	0.1100 (0.3787)
<i>idade_quadrado</i>	-0.0087 (0.0348)	-0.0092 (0.0352)	-0.0025 (0.0075)	-0.0029 (0.0076)
<i>estrangeiro</i>	0.0127 (0.0169)	0.0100 (0.0169)	0.0024 (0.0036)	0.0017 (0.0036)
<i>elenco</i>	-0.3024** (0.1189)	-0.3016** (0.1193)	-0.4826* (0.0258)	-0.0478* (0.0259)
<i>elenco_quadrado</i>	0.0037** (0.0016)	0.0037** (0.0016)	0.0005* (0.0003)	0.0005 (0.0003)
<i>cons</i>	4.8698 (21.4750)	4.7067 (21.7542)	0.0499 (4.6624)	-0.1705 (4.7214)
<i>N</i>	221	221	221	221
<i>Adj Overall R-Sq</i>	0.5337	0.5308	0.5810	0.5786

Outro modo de testar a validade dos instrumentos selecionados é testar uma regressão da variável considerada endógena contra os demais regressores incluídos na equação principal e realizar um teste F para ver se os instrumentos podem ter valor igual a zero. Nesse caso, a capacidade dos instrumentos em explicar o regressor entendido como endógeno seria nula e inviabilizaria a utilização conjunta das variáveis como instrumentos na equação principal. Abaixo é realizado o teste proposto.

Tabela 5 – Segunda validação dos instrumentos

<i>Var. Dep.</i>	<i>log_salreal</i>
<i>elenco</i>	-0.0218 (0.0559)
<i>elenco_quadrado</i>	0.0003 (0.0008)
<i>estrangeiros</i>	0.0286*** (0.0074)
<i>idade</i>	0.2739 (0.8412)
<i>idade_quadrado</i>	-0.0082 (0.0169)
<i>altitude</i>	-0.0019** (0.0008)
<i>real_GDHI/1000</i>	0.0003 (0.0000)
<i>_cons</i>	2.1649 (10.4786)
<i>N</i>	278
<i>Adj Overall R-Sq</i>	0.2152
<i>F(2, 270)</i>	4.0600
<i>Prob ≥ F</i>	0.0183

A partir dos resultados obtidos, verifico que os instrumentos são válidos conjuntamente a pesar do baixo impacto do PIB per capita real, diferente de zero, porém de valor muito pequeno. A estatística $P > |t|$ associada ao coeficiente do PIB per capita real é 0,968. Apesar disso, a exclusão do instrumento nas estimações futuras não altera os resultados, o que nos leva a continuar com ele nas especificações futuras, já que conjuntamente com a altitude, é significativo. No apêndice o leitor pode verificar as estimações sem o instrumento *real_GDHI*.

5 Resultados

A partir das especificações mencionadas ao longo da seção 4, a seção 5 é dedicada à análise dos resultados obtidos a partir das especificações entendidas como melhor ajustadas à

análise do problema de performance e pagamento de salários.

No modelo de efeitos fixos definido pela equação 19, a intenção é controlar para os efeitos fixos relativos a cada uma das equipes que as diferenciam. Após filtrar por esse impacto, o modelo mensura o impacto do salário real sobre as medidas de performance esportiva determinadas.

O modelo definido pela equação 20 prevê uma especificação mais elaborada e possui algumas particularidades. É notável que como o arcabouço do *System GMM* opera em primeiras diferenças, inicialmente, algumas observações se perdem, nos restando 220 observações no painel desbalanceado. Além disso, para aprimorar as estimações do modelo, a proliferação de instrumentos é controlada e evita que distorça todo o processo de remoção da endogeneidade dos regressores.

Tabela 6 – Resultados log(percentual_pontos)

Model	FE	GMM1	GMM2
	(1)	(2)	(3)
<i>log_percentual_pontos_L1</i>		0.3622 (0.2194)	0.3751* (0.1926)
<i>elenco</i>	-0.0519** (0.0237)	-0.0772** (0.0262)	-0.0791** (0.0246)
<i>elenco_quadrado</i>	0.0006* (0.0003)	0.0009** (0.0004)	0.0009** (0.0004)
<i>idade</i>	-0.2382 (0.3898)	-0.0482 (0.5521)	-0.0562 (0.5460)
<i>idade_quadrado</i>	0.0049 (0.0079)	-0.0002 (0.0115)	0.0000 (0.0114)
<i>estrangeiros</i>	0.0025 (0.0047)	0.0018 (0.0069)	0.0016 (0.0064)
<i>log_salreal</i>	0.1669*** (0.0440)	0.3232** (0.1471)	0.3300** (0.1531)
<i>_cons</i>	2.4083 (4.8165)	0.9839 (6.4070)	1.0964 (6.3026)
<i>N° Observações:</i>	278	220	220
<i>N° Grupos:</i>	39	35	35
<i>N° Instrumentos:</i>		20	22
<i>R-sq Overall:</i>	0.4595		
<i>AR(1) Pr ≥ z:</i>		0.02	0.018
<i>AR(2) Pr ≥ z:</i>		0.525	0.514
<i>Hansen Pr ≥ Chi2:</i>		0.042	0.081

Nota: GMM1 é o modelo sem variáveis instrumentais exógenas e GMM2 o modelo que inclui realGDHI e altitude como instrumentos exógenos. A significância a 10% é representada por *, enquanto 5% é representado por ** e 1% é ***. Teste AR(1) e AR(2) referentes a autocorrelação serial e teste de Hansen para validação dos instrumentos.

Tabela 7 – Resultados log(rank)

Model	FE	GMM1	GMM2
	(1)	(2)	(3)
<i>log_rank_L1</i>		0.2810 (0.1886)	0.2812 (0.1794)
<i>elenco</i>	-0.2469** (0.1069)	-0.4250*** (0.1149)	-0.4194*** (0.1056)
<i>elenco_quadrado</i>	0.0027* (0.0014)	0.0047*** (0.0016)	0.0046*** (0.0015)
<i>idade</i>	-1.7864 (1.7623)	-0.5635 (2.7133)	-0.6438 (2.8858)
<i>idade_quadrado</i>	0.0362 (0.0355)	0.0062 (0.0566)	0.0080 (0.0602)
<i>estrangeiros</i>	0.0156 (0.0211)	0.0107 (0.0333)	0.0115 (0.0311)
<i>log_salreal</i>	0.7007*** (0.1989)	1.9068*** (0.3301)	1.9027*** (0.3241)
<i>_cons</i>	23.9929 (21.7748)	10.9146 (31.8516)	11.6519 (33.8836)
<i>N° Observações:</i>	278	220	220
<i>N° Grupos:</i>	39	35	35
<i>N° Instrumentos:</i>		20	22
<i>R-sq Overall:</i>	0.4503		
<i>AR(1) Pr ≥ z:</i>		0.018	0.018
<i>AR(2) Pr ≥ z:</i>		0.097	0.094
<i>Hansen Pr ≥ Chi2:</i>		0.027	0.053

Nota: GMM1 é o modelo sem variáveis instrumentais exógenas e GMM2 o modelo que inclui realGDHI e altitude como instrumentos exógenos. A significância a 10% é representada por *, enquanto 5% é representado por ** e 1% é ***. Teste AR(1) e AR(2) referentes a autocorrelação serial e teste de Hansen para validação dos instrumentos.

As tabelas 6 e 7 em que as três especificações do modelo são apresentadas (efeitos fixos, *System GMM* sem instrumentos exógenos e *System GMM* com instrumentos exógenos). A divisão em duas tabelas ocorre, pois os modelos são testados para variáveis dependentes distintas, $\ln(\text{percentualpontos})$ e $\ln(\text{rank})$, para prover maior robustez aos resultados e per-

mitir a identificação da questão debatida a partir de mais de uma ótica.

Analisando os resultados obtidos, verifico que em todas as estimações apresentadas a variável de interesse do modelo, o logaritmo dos salários reais, é no mínimo significativa a 5% e apresenta sinal positivo. Isso corrobora o propósito da pesquisa de que maiores salários anuais são bons medidores da performance esperada da equipe. Isto é, os clubes de fato calibram bem seu pagamento de salário aos jogadores para captar o impacto dos atletas na obtenção de pontos ao longo do campeonato, ou seja, a performance do time.

Passando para a análise individual dos modelos, iniciarei a discussão pelo modelo de efeitos fixos.

Nessa especificação, 3 regressores são significantes. Com um nível de 10%, o tamanho do elenco ao quadrado, a 5% o tamanho do elenco e os salários reais ($\log(\text{salreal})$) a 1%. Observando o resultado para as variáveis de tamanho do elenco e tamanho do elenco ao quadrado, o modelo indica que mais jogadores disponíveis piora a performance do time. Isso pode ser entendido de duas formas: a primeira análise é que um grande número de jogadores dificulta a escolha ótima dos 11 iniciais para o time, o que faz com que o treinador tenha problemas ao encontrar uma escalação ótima e conseqüentemente obtenha menos pontos. Uma segunda é que os clubes reconhecidamente menores contratem vários jogadores por um baixo salário e fazem testes para verificar quais possuem talento suficiente para permanecer no time e quais serão descartados na próxima janela. Por tratar-se de atletas com menor valor de transferência e com salários menores, tornam-se ativos facilmente descartáveis, gerando um incentivo a esse tipo de estratégia. Ademais, esse comportamento permite que o time com menor poder financeiro possa revelar atletas e negociá-los a um maior preço no mercado de transferências, podendo escalar de patamar após negociações de sucesso e contratar melhores jogadores para competir em um nível mais alto, como é discutido ao longo do modelo teórico. De forma resumida, essa variável capta uma preferência por parte dos times com menor poder financeiro em montar elencos extensos e baratos.

Por outro lado, o coeficiente da variável de tamanho de elenco ao quadrado, observa-se que apesar de um ter baixo impacto, ela possui sinal positivo, existindo um ponto de inflexão onde ter mais jogadores representa uma potencial melhora na performance da equipe. Para

entender o comportamento da variável, o ponto em que a variável passa a ter impacto positivo é 44 atletas. Ao investigar a base de dados com 220 observações utilizada para a estimação dos modelos GMM, verifica-se que usualmente times que possuem mais de 44 atletas são integrantes do Big Six, justificando essa melhora na performance a partir desse limiar de 44 atletas.

Uma explicação para esse fenômeno é que os clubes mais ricos oferecem contratos profissionais a jovens atletas e tentam os emprestar para obterem experiência em outras equipes antes de jogar pelo time que de fato o contratou. Entretanto, os empréstimos nem sempre se concretizam e o clube fica com um alto número de atletas que sequer são utilizados em partidas oficiais. Abaixo é possível observar os casos em que clubes possuem mais de 44 atletas em seu elenco. E as respectivas pontuações na campanha de referência.

Tabela 8 – Estudo da variável elenco

Ano	Tam Elenco	Posição	Time
<i>2009</i>	<i>45</i>	<i>3^o</i>	<i>FC Arsenal</i>
<i>2010</i>	<i>45</i>	<i>3^o</i>	<i>Manchester City FC</i>
<i>2011</i>	<i>49</i>	<i>4^o</i>	<i>Tottenham Hotspur</i>
<i>2013</i>	<i>48</i>	<i>1^o</i>	<i>Chelsea FC</i>
<i>2015</i>	<i>50</i>	<i>7^o</i>	<i>West Ham United</i>
<i>2016</i>	<i>49</i>	<i>1^o</i>	<i>Chelsea FC</i>
<i>2016</i>	<i>45</i>	<i>14^o</i>	<i>Crystal Palace FC</i>
<i>2016</i>	<i>46</i>	<i>17^o</i>	<i>FC Watford</i>
<i>2017</i>	<i>46</i>	<i>5^o</i>	<i>Chelsea FC</i>
<i>2017</i>	<i>45</i>	<i>6^o</i>	<i>FC Arsenal</i>
<i>2017</i>	<i>48</i>	<i>8^o</i>	<i>FC Everton</i>
<i>2018</i>	<i>54</i>	<i>3^o</i>	<i>Chelsea FC</i>

Por fim, é relevante verificar a alta significância estatística da variável que representa os salários reais, para induzir o teste dessa variável em outras especificações controlando para endogeneidade, conforme discutido na seção 4.

Começando a abordagem dos modelos *System GMM*, serão inicialmente discutidos os resultados obtidos sem a inclusão de variáveis instrumentais exógenas. Nessa especificação obtêm-se significância estatística, como definido acima, apenas para o tamanho do elenco (elenco) e para os salários reais ($\ln(\text{salreal})$). Isso reforça a explicação apresentada acima sobre o impacto negativo dos elencos de tamanho maior e o canal de influência que identifiquei nessa variável. Na análise sobre os salários reais, a significância estatística do coeficiente corrobora a proposta da pesquisa.

Vale ressaltar que a performance esportiva na temporada anterior ($L1.\ln(\text{percentual_pontos})$) fica muito próxima de ser significativa a 10%, o que me leva a considerar que a hipótese levantada ao longo do trabalho, de que existe persistência entre uma temporada e outra, é válida para a discussão econométrica entre salários e performance esportiva e que estudar essa questão sob a ótica dos modelos *System GMM* é interessante, pois permite explicitar essa variável como um dos regressores responsáveis pela obtenção de pontos ao longo do campeonato.

Na análise do modelo *System GMM* com variáveis instrumentais exógenas, a significância dos mesmos regressores apresentados nos parágrafos anteriores é verificada. Contudo, a grande diferença é a obtenção de significância estatística para o regressor ($L1.\ln(\text{percentual_pontos})$) a 10%. Isso reforça a hipótese de persistência do trabalho da equipe entre temporadas, o que é um dos objetivos do trabalho.

Analisando o resultado dos testes de Hansen, ou testes de robustez para os instrumentos, o modelo sem as variáveis instrumentais exógenas não me permite aceitar o conjunto de instrumentos a um nível de significância de 5%. Entretanto, o modelo com os instrumentos exógenos me permite aceitar este conjunto de instrumentos a um nível de 5% com significância estatística para os regressores mais relevantes para minha pesquisa, o que torna esse formato mais interessante do ponto de vista de resultados. Para verificar estes resultados basta se remeter ao teste Hansen na tabela de resultados. Vale ressaltar que em todas as especificações apresentadas, o número de instrumentos é inferior ao número de times, evidenciando o cuidado em não *overfit* o modelo. Outra estatística relevante a ser observada é o resultado do teste $AR(2)$, que em ambas especificações indica que o modelo não sofre de autocorrelação serial de segunda ordem. Em ambos os casos o p-valor do

teste está bem acima de 0.05. Vale ressaltar que o teste AR(1) não é relevante para os resultados, pois a especificação GMM é realizada em primeiras diferenças.

Passando para a análise do segundo grupo de regressões em que eu utilizo $\ln(\text{rank})$ como variável dependente, vou seguir a mesma ordem de explicação, isto é, iniciar pelo modelo de efeitos fixos, passando pelo modelo *System GMM* sem instrumentos exógenos e por fim comentando os aspectos mais relevantes do modelo *System GMM* com instrumentos exógenos. É importante ressaltar que, como pode ser evidenciado na tabela de estatísticas descritivas, a escala das variáveis *percentual_pontos* e *rank* é consideravelmente diferente, explicando a diferença na magnitude dos coeficientes da regressão. Nesta parte a exposição será mais breve pela similaridade entre os resultados encontrados. Abaixo, é possível conferir a tabela com as estimações.

Nos modelos GMM, a grande diferença está na dificuldade em apresentar significância para o regressor de *lag* na performance da temporada passada. Em ambas especificações ele fica próximo aos 10% de significância, porém não consegue ser incluída no grupo de coeficientes significantes. Apesar disso, o salário real é significativo nessas especificações.

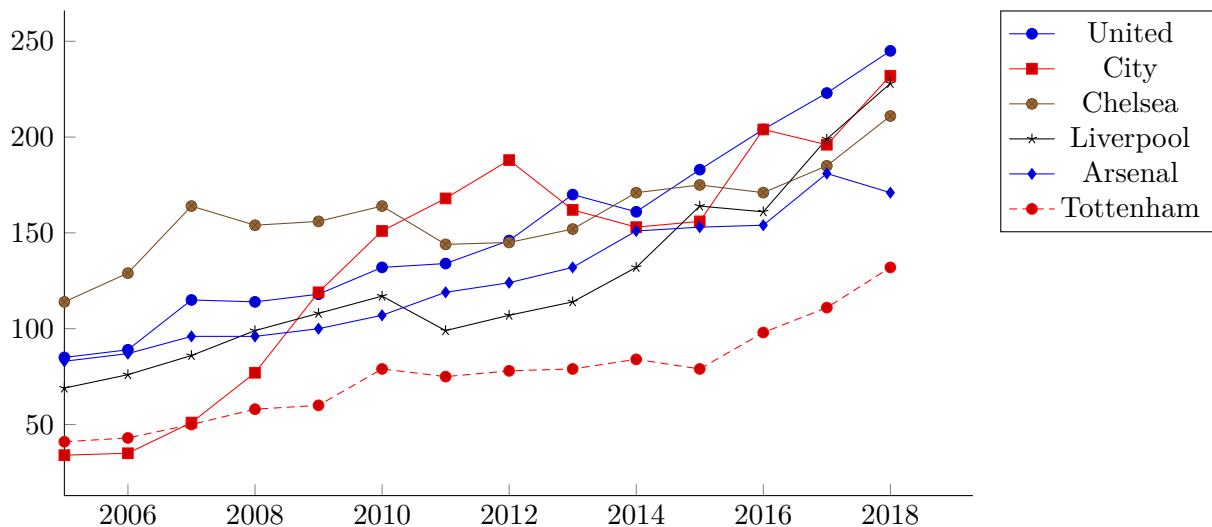
Abordando os testes relevantes para os modelos GMM, o conjunto de instrumentos sem as variáveis instrumentais exógenas não é sólido, indicando baixa relevância dos instrumentos na primeira especificação. Contudo, na especificação com instrumentos exógenos, o conjunto de instrumentos é relevante. No estudo de autocorrelação de segunda ordem, os dois modelos estão bem especificados.

Interpretando individualmente as variáveis estatisticamente significativas do modelo, verifica-se que nos modelos GMM, um aumento de 1% no valor dos salários reais das equipes gera aproximadamente 0,3% de aumento no percentual da pontuação das equipes, conforme a tabela 6 coluna 3. No modelo que considera a posição na tabela como variável dependente, verifica-se que um aumento de 1% nos salários reais da equipe promove uma melhora de aproximadamente 1,9% no ranking (tabela 7 coluna 3) obtido consoante aos modelos GMM.

Observando a base de dados, os times que obtiveram maior número de títulos ao longo dos anos: Chelsea, Manchester United e Manchester City, que concentram 13 conquistas nas 14 temporadas estudadas, verifica-se um distanciamento em relação a Arsenal e Tot-

tenham, que não ganharam nenhum título neste período. A partir do gráfico, levanta-se duas questões interessantes. A primeira delas é que os clubes se veem em uma situação semelhante ao dilema dos prisioneiros. Se todos os big six gastam, eu preciso gastar também. Isso gera um ciclo de elevados dispêndios salariais no topo. O segundo ponto é a presença do Liverpool entre os times que mais gastam apesar de não conquistar nenhum título levanta um questionamento interessantes: basta gastar ou é necessário gastar bem? Onde é melhor gastar? Trazer novos jogadores ou garantir a permanência dos já consolidados a partir de maiores salários? Os treinadores fazem a diferença? Todos esses tópicos podem ser abordados em futuras pesquisas para entender a receita para um time campeão.

Figura 6 – Diferença salarial entre os Big6 campeões e os não campeões



Inevitavelmente faz-se necessária a comparação dos resultados aqui obtidos com aqueles apresentados pelos autores do artigo base, *Aggregate Wages of players and performance in Italian Serie A 2017*. Uma diferença imediatamente notada é o número de instrumentos que os autores usam para provar a significância das variáveis selecionadas. Dispondo de apenas 212 observações e 33 grupos, os autores utilizam 183 instrumentos, gerando um notável problema de *overfitting* via instrumentos e promove uma significância "artificial" aos regressores, incluindo o *lag* de uma temporada para a performance esportiva.

Abaixo, o leitor pode comparar os resultados obtidos no meu estudo (usando um formato

semelhante à especificação dos autores) para a primeira divisão da Inglaterra e os resultados obtidos pelos autores na primeira divisão italiana. A comparação se dá pela utilização das variáveis em log, assim como os autores as utilizam.

Tabela 9 – Resultados log(percentual_pontos) - modelo em log

Model	FE	GMM1	GMM2
	(1)	(2)	(3)
<i>log_percentual_pontos_L1</i>		0.3654 (0.2203)	0.3739* (0.2195)
<i>log_elenco</i>	-5.0587* (2.8384)	-6.2717* (3.643)	-6.6182* (3.6135)
<i>log_elenco_quadrado</i>	0.6712* (0.3981)	0.8099 (0.5208)	0.8552 (0.5208)
<i>log_idade</i>	-18.3446 (30.4789)	1.936 (39.7374)	1.131 (50.0215)
<i>log_idade_quadrado</i>	2.8784 (4.7549)	-0.5168 (6.2734)	-0.3877 (7.8953)
<i>log_estrangeiros</i>	0.0125 (0.1002)	0.0544 (0.1434)	0.0536 (0.1461)
<i>log_salreal</i>	0.1674*** (0.0438)	0.321** (0.146)	0.3295** (0.1574)
<i>_cons</i>	37.1409 (48.7393)	9.1089 (62.2698)	10.9939 (78.4357)
<i>N° Observações:</i>	278	220	220
<i>N° Grupos:</i>	39	35	35
<i>N° Instrumentos:</i>		20	22
<i>R-sq Overall:</i>	0.4543		
<i>AR(1) Pr ≥ z:</i>		0.018	0.022
<i>AR(2) Pr >≥ z:</i>		0.484	0.495
<i>Hansen Pr ≥ Chi2:</i>		0.042	0.079
<i>Prob ≥ F =</i>	0.0011	0.000	0.000

Nota: GMM1 é o modelo sem variáveis instrumentais exógenas e GMM2 o modelo que inclui realGDHI e altitude como instrumentos exógenos. A significância a 10% é representada por *, enquanto 5% é representado por ** e 1% é ***. Teste AR(1) e AR(2) referentes a autocorrelação serial e teste de Hansen para validação dos instrumentos.

Tabela 10 – Resultados log(rank) - modelo em log

Model	FE	GMM1	GMM2
	(1)	(2)	(3)
<i>log_rank_L1</i>		0.2746 (0.1915)	0.2788 (.1849)
<i>log_elenco</i>	-22.4746* (12.8439)	-32.5276* (17.1136)	-32.2713* (16.3962)
<i>log_elenco_quadrado</i>	29.267 (1.8016)	4.1328 (2.4507)	4.1107* (2.3398)
<i>log_idade</i>	-133.5149 (137.9141)	-14.2817 (210.9504)	-27.4413 (257.8609)
<i>log_idade_quadrado</i>	20.8804 (21.5158)	1.2357 (33.2713)	3.323069 (40.6775)
<i>log_estrangeiro</i>	0.2356 (0.4535)	0.3035 (0.71068)	0.2981 (0.6874)
<i>log_salreal</i>	0.6941*** (0.1985)	1.8988*** (0.3375)	1.8958*** (0.3336)
<i>_cons</i>	252.6417 (220.5408)	87.4372 (332.8006)	107.5625 (406.423)
<i>N° Observações:</i>	278	220	220
<i>N° Grupos:</i>	39	35	35
<i>N° Instrumentos:</i>		20	22
<i>R-sq Overall:</i>	0.4471		
<i>AR(1) Pr ≥ z:</i>		0.021	0.021
<i>AR(2) Pr ≥ z:</i>		0.103	0.102
<i>Hansen Pr ≥ Chi2:</i>		0.026	0.054
<i>Prob ≥ F =</i>	0.000	0.000	0.000

Nota: GMM1 é o modelo sem variáveis instrumentais exógenas e GMM2 o modelo que inclui realGDHI e altitude como instrumentos exógenos. A significância a 10% é representada por *, enquanto 5% é representado por ** e 1% é ***. Teste AR(1) e AR(2) referentes a autocorrelação serial e teste de Hansen para validação dos instrumentos.

De forma geral, os valores para os coeficientes são semelhantes entre os estudos. A diferença mais relevante está na magnitude dos coeficientes de salários reais, onde encontro significância para um coeficiente de aproximadamente 0,3 e os autores de 0,2 quando a variável

dependente é percentual de pontos obtidos. Os autores também encontram significância para média de idade e média de idade ao quadrado do elenco. Acredito que isso se deve ao *overfitting* gerado pela proliferação de instrumentos. Outros detalhes são a inclusão de uma variável *dummy* que controla para temporadas com 18 times, o que ocorre na Itália, porém não ocorre na Inglaterra. Caruso também inclui uma variável independente de aristocracia, que testa se os times que tem mais participações na primeira divisão tendem a carregar um maior "peso histórico" e se isso influencia na performance.

6 Conclusão

O estudo chega ao fim com seu propósito inicial completo: clubes que pagam maiores salários de fato conseguem melhores resultados na primeira divisão inglesa. O modelo teórico construiu uma narrativa sobre a dinâmica do mercado de transferências e a abordagem de cada um dos clubes quando se depara com as possibilidades de reforçar seu elenco. Essa discussão motiva a apresentação do modelo empírico, que a partir de seis diferentes especificações, consegue resultados significativos para cumprir com a proposta do trabalho.

As maiores contribuições são: a criação de uma base de dados original que pode ser aproveitada e aprimorada para novos estudos, aprimorar um modelo inovador ao reduzir o número de instrumentos necessários para obter significância das variáveis e aplicar ao universo de dados da liga mais rica do mundo, verificando se a hipótese de salários e performance se aplica no extremo positivo da distribuição de salários internacionais entre as ligas.

A pesquisa pode evoluir em novas direções a partir de uma maior disponibilidade de dados. A primeira sugestão é verificar se a presença de uma grade estrela traz um impacto individual no resultado. Outra sugestão nessa mesma linha de pesquisa é estudar se a distribuição dos salários no elenco afeta o desempenho, ou seja, é melhor concentrar mais em grandes jogadores ou ter um elenco nivelado?

Por fim, pensando na implicação prática do que essa pesquisa traz ao leitor, é possível identificar que o futebol não é um jogo limitado ao universo das quatro linhas. Boas gestões, grandes patrocínios e maiores premiações, tornaram o esporte uma indústria em que os clubes competem fora de campo pelos pontos obtidos dentro de campo. Na Europa, essa narrativa faz parte do presente. No Brasil, o país do futebol, temos um longo caminho a trilhar, porém, existem movimentações de bancos, com uma área dedicada ao estudo de gestão e aprimoramento financeiro das equipes da América do Sul. Algo impensável nas últimas décadas de gestões amadoras.

Referências

- ROODMAN, David. How to do xtabond2: An introduction to difference and system gmm in stata. **STATA JOURNAL**, v. 9, n. 1, p. 86-136, 2009
- ARELLANO, Manuel; BOND, Stephen. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. **Review of Economic Studies**, v. 58, n. 2, p. 277–297, 1991.
- ARELLANO, Manuel; BOVER, Olympia. Another look at the instrumental variable estimation of error-components models. **Journal of Econometrics**, v. 68, n. 1, p. 29–51, 1995.
- BLUNDELL, Richard; BOND, Stephen Bond. Initial conditions and moment restrictions in dynamic panel data models. **Journal of Econometrics**, v. 87, n. 1, p. 115–143, 1998.
- BUCCIOL, Alessandro et al. Pay dispersion and performance in teams. **PloS ONE**, v. 9 n. 1 p. 112-131, 2014.
- CARUSO, Raul et al. Aggregate wages of players and performance in Italian Serie A. **Economia Politica**, v. 34, n. 3, p. 515–531, 2017.
- EL-HODIRI, Mohamed; QUIRK, James. An economic model of a professional sports league. **Journal of Political Economy**, v. 79, n. 6, p.1302–1319, 1971.
- FRANCK, Egon; NUESCH, Stephan. The effect of wage dispersion on team outcome and the way team outcome is produced. **Applied Economics**, v. 43, n. 11, p. 3037–3049, 2010.
- BARRIO, Pedro Garcia-del, SZYMANSKI, Stefan. Goal! Profit Maximization Versus Win Maximization in Soccer. **Review of Industrial Organization**, v. 34, n.1, 45–68, 2009.
- HALL, Stephen et al. Testing Causality Between Team Performance and Payroll: The Cases of Major League Baseball and English Soccer. **Journal of Sports Economics**, v. 3, n. 2, p. 149-168, 2002.
- LAGO, Umberto et al. Il business del calcio: successi sportivi e rovesci finanziari **Egea**, v.1, n.1, p. 1-313, 2004.
- SCULLY, Gerald. Pay and performance in major league baseball. **American Economic Review**, v. 64, n. 6, p. 915–930, 1974.
- SZYMANSKI, Szymanski. Wages, transfers, and the variation of team performance in the English premier league. **In The Econometrics of Sport**, p. 53–62, 2013.
- SZYMANSKI, Stefan; KUPER, Simon, Soccernomics, **Versione esclusiva per edizione brasiliana**, p. 310-311, 2009
- SZYMANSKI, Stefan; SMITH, Ron. The English Football Industry: profit, performance and industrial structure. **International Review of Applied Economics**, v. 11, n. 1, p 135–153, 1997.

WOOLDRIDGE, Jeffrey M. Instructional Stata datasets for econometrics,
Boston College Department of Economics, November 2002.

YAMAMURA, Eiji. Wage disparity and team performance in the process of industry development: Evidence from japan's professional football league. **Journal of Sports Economics**, v. 16, n. 02, p. 214-223, 2013.

A Apêndice A

Tabela A1 – Resultados log_rank sem real_GDHI

<i>Var. Dep.</i>	<i>log_rank</i>
<i>log_rank L1.</i>	0.2746 (0.1803)
<i>elenco</i>	-0.4263*** (0.1094)
<i>elenco_quadrado</i>	0.0046*** (0.0015)
<i>idade</i>	0.5811 (2.8218)
<i>idade_quadrado</i>	0.0066 (0.0588)
<i>estrangeiros</i>	0.0122 (0.0312)
<i>log_salreal</i>	1.9172*** (0.3285)
<i>_cons</i>	11.03472 (33.1945)
<hr/>	
<i>N° Observações: 220</i>	
<i>N° Grupos: 35</i>	
<i>N° Instrumentos: 21</i>	
<i>AR(1) Pr ≥ z: 0.018</i>	
<i>AR(2) Pr ≥ z: 0.097</i>	
<i>Hansen Pr ≥ Chi2: 0.038</i>	

Tabela A2 – Resultados log_percentual_pontos sem real_GDHI

<i>Var. Dep.</i>	<i>log_percentual_pontos</i>
<i>log_percentual_pontos L1.</i>	0.3742* (0.1872)
<i>elenco</i>	-0.0764*** (0.0261)
<i>elenco_quadrado</i>	0.0008** (0.0003)
<i>idade</i>	0.0477 (0.5466)
<i>idade_quadrado</i>	0.0001 (0.0114)
<i>estrangeiros</i>	0.0014 (0.0064)
<i>log_salreal</i>	0.3265** (0.1472)
<i>_cons</i>	0.9500 (6.311)
<hr/>	
<i>N° Observações: 220</i>	
<i>N° Grupos: 35</i>	
<i>N° Instrumentos: 21</i>	
<i>AR(1) Pr ≥ z: 0.017</i>	
<i>AR(2) Pr ≥ z: 0.510</i>	
<i>Hansen Pr ≥ Chi2: 0.061</i>	

B Apêndice B

Para obter (9):

Os lucros são definidos, por hipótese do modelo, como a diferença entre receitas e custos. Como as receitas são definidas por $R_{i,t}$ e as despesas estão relacionadas diretamente a $W_{i,t}$, ou indiretamente, como verificamos na equação 7 é imediato a dedução de 9:

$$P_{i,t} = R_{i,t} - (1 + \tau)W_{i,t}$$

Para obter 10:

Partimos de:

$$P_{i,t} = R_{i,t} - (1 + \tau)W_{i,t}$$

Podemos reescrever 3 como:

$$\frac{Q_{i,t} - c_i}{d} = W_{i,t}$$

Substituindo o novo formato de 3 e 6 em 9, temos:

$$P_{i,t} = a_i + bQ_{i,t} - (1 + \tau)\left[\frac{Q_{i,t} - c_i}{d}\right]$$

$$P_{i,t} = \frac{[a_id + (1 + \tau)c_{i,t}]}{d} + \frac{bdQ_{i,t}}{d} - \frac{(1 + \tau)Q_{i,t}}{d}$$

Note que o primeiro termo do lado direito da igualdade é igual ao que definimos como π .

Reescrevendo, temos:

$$P_{i,t} = \pi_{i,t} + Q_{i,t}\left[b - \frac{(1 + \tau)}{d}\right]$$

O termo que multiplica $Q_{i,t}$ é exatamente o que definimos como ϕ , portanto:

$$P_{i,t} = \pi_i + \phi Q_{i,t}$$

que é (2.10).

Otimização de 13:

$$W_t = P_{i,t} - \frac{\alpha}{2}P_{i,t}^2 + \beta Q_{i,t} - \frac{\gamma}{2}Q_{i,t}^2$$

Substituindo $P_{i,t}$ como em 10:

$$W_t = \pi_i + \phi Q_{i,t} - \frac{\alpha}{2}(\pi_i + \phi Q_{i,t})^2 + \beta Q_{i,t} - \frac{\gamma}{2}Q_{i,t}^2$$

Tomando a derivada parcial em relação a $Q_{i,t}$:

$$\frac{\partial W_t}{\partial Q_{i,t}} = \phi - \frac{\alpha}{2}2(\pi_i + \phi Q_{i,t})\phi + \beta - \gamma Q_{i,t} = 0$$

$$\frac{\partial W_t}{\partial Q_{i,t}} = \phi - \alpha\phi\pi_i - \alpha\phi^2 Q_{i,t} + \beta - \gamma Q_{i,t} = 0$$

$$\frac{\partial W_t}{\partial Q_{i,t}} = \phi(1 - \alpha\pi_i) + \beta = Q_{i,t}(\alpha\phi^2 + \gamma)$$

$$Q_{i,t}^* = \frac{[\phi(1 - \alpha\pi_i) + \beta]}{[\alpha\phi^2 + \gamma]}$$

Substituindo $Q_{i,t}^*$ em 10:

$$P_{i,t}^* = \pi_i + \phi \frac{[\phi(1 - \alpha\pi_i) + \beta]}{[\alpha\phi^2 + \gamma]}$$

$$P_{i,t}^* = \pi_i \frac{[\alpha\phi^2 + \gamma]}{[\alpha\phi^2 + \gamma]} + \frac{\phi^2 - \alpha\phi^2\pi_i + \phi\beta}{\alpha\phi^2 + \gamma}$$

$$P_{i,t}^* = \frac{\pi_i\gamma + \phi^2 + \phi\beta}{[\alpha\phi^2 + \gamma]}$$

$$P_{i,t}^* = \frac{\pi_i\gamma + \phi(\phi + \beta)}{[\alpha\phi^2 + \gamma]}$$