

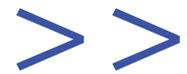
OBEGEF – Observatório de Economia e Gestão de Fraude

WORKING PAPERS

#59

A fraude no futebol:
Viciação de resultados.
O caso Calciocaos

Jorge Saraiva



OBEGEF
Observatório de Economia
e Gestão de Fraude

>> FICHA TÉCNICA**A FRAUDE NO FUTEBOL: VICIAÇÃO DE RESULTADOS.
O CASO CALCIOCAOS**

WORKING PAPERS Nº 59 / 2018

OBEGEF – Observatório de Economia e Gestão de Fraude

Autores: Jorge Saraiva¹

Editor: Edições Húmus

1ª Edição: Julho de 2018

ISBN: 978-989-755-355-4

Localização web: <http://www.gestaodefraude.eu>

Preço: gratuito na edição electrónica, acesso por download.

Solicitação ao leitor: Transmita-nos a sua opinião sobre este trabalho.

©: É permitida a cópia de partes deste documento, sem qualquer modificação, para utilização individual. A reprodução de partes do seu conteúdo é permitida exclusivamente em documentos científicos, com indicação expressa da fonte.

Não é permitida qualquer utilização comercial. Não é permitida a sua disponibilização através de rede electrónica ou qualquer forma de partilha electrónica.

Em caso de dúvida ou pedido de autorização, contactar directamente o OBEGEF (obegef@fep.up.pt).

©: Permission to copy parts of this document, without modification, for individual use. The reproduction of parts of the text only is permitted in scientific papers, with bibliographic information of the source.

No commercial use is allowed. Not allowed put it in any network or in any form of electronic sharing.

In case of doubt or request authorization, contact directly the OBEGEF (obegef@fep.up.pt).

¹ Faculdade de Economia da Universidade do Porto. Email: jorgesaraiva@gestaodefraude.eu

>> ÍNDICE

I. INTRODUÇÃO	6
II. ENQUADRAMENTO TEÓRICO	10
RACIONALIDADE E MOTIVAÇÕES	10
CONTRIBUTOS TEÓRICOS DA ECONOMIA	12
Modelo de Procura e Oferta	12
Teoria dos Jogos	13
III. ANÁLISE EMPÍRICA – O CASO CALCIOCAOS	15
O CASO CALCIOCAOS	15
DETALHE DA BASE DE DADOS	16
ANÁLISE ESTATÍSTICA	17
Golos	18
Minutos dos golos	19
Infrações	19
Tabelas de contingência	20
MODELO ECONOMETRICO	21
Especificação do modelo	21
IV. CONCLUSÕES	34
REFERÊNCIAS BIBLIOGRÁFICAS	36
ANEXO 1	37
MATRIZ DE P-VALUES DAS VARIÁVEIS RELATIVAS A GOLOS	37
ANEXO 2	38
MATRIZ DE P-VALUES DAS VARIÁVEIS RELATIVAS A MINUTOS DOS GOLOS	38
ANEXO 3	39
MATRIZ DE P-VALUES DAS VARIÁVEIS RELATIVAS A INFRAÇÕES	39
ANEXO 4	40
TABELAS DE CONTINGÊNCIA DE ALGUMAS VARIÁVEIS RELATIVAMENTE À VARIÁVEL FIX	40

>> RESUMO

O presente artigo pretende analisar os fatores que influenciam a existência de viciação de resultados num jogo de futebol. O ponto de partida é a contextualização abrangente da fraude desportiva e, mais concretamente, da viciação de resultados, em termos teóricos. Já o aspeto central do artigo é o estudo do Caso Calciocaos, quer numa vertente de análise estatística, quer numa vertente econométrica, onde são estimadas quais as contribuições – para a probabilidade de existência de fraude – de cada uma das variáveis consideradas. Por fim, o modelo econométrico é ainda transformado de forma a possibilitar a sua utilização com diferentes amostras onde se suspeita que existe viciação de resultados do tipo arrangement fix perpetrada por árbitros.

Palavras-chave: viciação de resultados, fraude desportiva, futebol, Calciocaos

>> ABSTRACT

The following article intends to analyse the factors that drive the presence of match fixing in football matches. The starting point is a wide theoretical framework of sports' fraud and, more precisely, of match fixing. The main matter of the article is the study of the Calciocaos Case both in statistical and econometrical analysis. The econometric model estimates the contributions, for the probability of fraud's existence, of each one of the considered variables. Lastly, the econometric model is then transformed in order to enable its' use with different samples given suspicions of arrangement fixes made by referees.

Keywords: match fixing, sports' fraud, football, Calciocaos

>> AGRADECIMENTOS

Em primeiro lugar, quero agradecer ao professor de Seminário de Economia por toda a confiança, paciência e empenho que colocou quer em mim, quer neste trabalho, e cuja colaboração espero poder contar, de novo, no futuro. Quero também deixar algumas palavras ao professor Mendes de Oliveira pela brilhante discussão que promoveu e por todo o tempo que me despendeu. Sem dúvida que o seu contributo foi fulcral para o aperfeiçoamento da estimação econométrica, o que gerou melhorias preditivas significativas.

Finalmente, agradeço à Francisca pelas várias sugestões e leituras prévias do trabalho, assim como todo o apoio quando ele é necessário.

>> I. INTRODUÇÃO

A profissionalização da atividade desportiva, a transformação de muitos clubes desportivos em empresas capitalistas que procuram essencialmente o lucro, o envolvimento de elevados recursos públicos – nomeadamente em grandes empreendimentos –, a multiplicidade de atividades económicas satélite, o volume de recursos financeiros envolvidos e a intangibilidade de muitos dos objetos de negócio dão particular relevância ao sector desportivo na sociedade contemporânea e contribuem para o aumento significativo do risco de fraude.

Contudo, a importância do desporto ultrapassa em muito a dimensão económica. Esta importância comporta também uma dimensão lúdica, identitária e simbólica, cultural e social, jurídica, gestonária e política. Assim, as fraudes desportivas têm um conjunto de características próprias conferidas pela multiplicidade de facetas a elas associadas.

As fraudes no desporto podem ser de dois tipos diferentes. Por um lado, as instituições desportivas estão sujeitas a fraudes e praticam fraudes como qualquer outra atividade económico-social, num determinado contexto social. As empresas desportivas ou delas dependentes (empresas de apostas, empresas de organização de eventos, agentes desportivos, *media*, as próprias instituições desportivas de gestão e de coordenação como, por exemplo, as associações e as federações) podem ser alvo de fraudes por parte dos seus funcionários, podem fazer manipulação contabilística, associar às suas atividades empresas fantasma e operações secretas em paraísos fiscais. O mesmo se pode dizer dos próprios atletas.

A corrupção ativa de instituições, pessoas e responsáveis políticos, as burlas, a informação privilegiada e o tráfico de influência, a fraude fiscal e o branqueamento de capitais são algumas das fraudes que, suscetíveis de serem praticadas em e por qualquer sector de atividade, podem assumir nas atividades desportivas um conjunto de características especiais. De facto, a projeção social e popular do desporto, a impossibilidade de estimar antecipadamente o justo valor de muitas das suas transações, o elevado valor de muitas das operações financeiras e o significado que tem para as populações podem dar a estas fraudes características e dimensões próprias. Ainda que constituam fraudes no setor desportivo, estas não podem ser categorizadas como fraudes no desporto, uma vez que as fraudes no desporto são sim “um desvio das expectativas públicas de que o desporto será jogado e adminis-

trado de uma forma honesta”¹, como refere Masters (2015, p. 113).

Este tipo de fraude pode assumir diversas formas, desde o atleta que simula uma falta, o dirigente que paga aos adversários para perderem uma partida ou mesmo um atleta que toma substâncias químicas para melhorar o seu rendimento desportivo. Embora todas estas fraudes sejam importantes, não é errado assumir que a viciação de resultados é uma das mais relevantes. Não se trata apenas do seu carácter enganoso, presente em qualquer fraude, em relação aos espectadores, mas também do que representa para o crime organizado, em particular para as máfias e para as suas operações de lavagem de dinheiro Hill² (2011).

A propósito da importância da viciação de resultados, Hill (2013) afirma que o tema é um *santo graal* da sociologia, uma vez que representa um desvio universal, isto é, em todas as culturas a viciação de resultados foi sempre vista como um comportamento desviante. Tal como Hill (2013) explica, a existência de desvios universais não é comum, uma vez que grande parte das vezes os temas dividem culturas, seja em termos históricos ou geográficos, não sendo, portanto, universais. Um exemplo disto está presente na obra de Sir Arthur Conan Doyle, onde Sherlock Holmes era consumidor de cocaína, situação encarada com naturalidade pela sociedade do século XIX e que atualmente é reprovada.

O objeto deste estudo será uma fraude desportiva em particular – a viciação de resultados. Em termos simples, este tipo de fraude trata-se de um conjunto de ações praticadas, de forma intencional, com o intuito de alcançar um resultado pré-determinado e que vai contra o fator de aleatoriedade normal do desporto. Assim, a viciação de resultados enquadra-se na definição de fraude aqui adotada.

Será igualmente importante distinguir os dois tipos de viciação de resultados que podem existir (Boeri & Severgnini, 2013, pp. 105-108). Neste âmbito, serão utilizados os termos anglo-saxónicos por conferirem mais clareza ao texto do que os termos traduzidos dariam. Assim, o termo a utilizar quando a viciação tiver como objetivo principal a obtenção de resultados desportivos favoráveis corresponde a *arrangement fix*³. Já no caso em que

¹ Tradução livre — “the deviation from public expectations that sport will be played and administered in an honest manner”.

² Declan Hill é um jornalista e académico canadiano com vasta experiência na área da viciação de resultados e corrupção no desporto internacional. As suas principais obras — “The Fix” e “The Insider’s Guide to Match Fixing in Football” — serão alvo de análise mais detalhada no decorrer do texto.

³ De notar que um *arrangement fix* não está limitado ao caso em que uma equipa A paga a uma equipa B para que B perca o jogo contra A. Uma situação que também é possível é o caso em que uma equipa A paga a uma equipa B para que B perca o jogo contra C, de tal forma que A fica favorecida desportivamente por este resultado.

o objetivo principal é o ganho financeiro nos mercados de apostas o termo é *gambling fix*.

Para além dos tipos de viciação, torna-se de igual forma importante definir quem são os agentes corrompidos nesta fraude, ou seja, quem de facto pratica a fraude. Assim, embora o agente corruptor possa ser bastante diversificado – desde um dirigente desportivo a um membro da máfia –, a figura de agente corrompido é algo limitada. Ora, na análise do caso do futebol surgem dois tipos de agentes: atletas e árbitros. De notar que também é concebível considerar os treinadores como agentes corrompidos, no entanto, dado que em última instância apenas os atletas poderão ter uma influência direta no jogo – ainda que possam agir exclusivamente por força das decisões dos treinadores – será considerado, para fins de sistematização, que os treinadores estão incluídos na categoria de atletas¹.

Na tabela 1 é feita uma sistematização tendo em conta o tipo de vício e o agente corrompido, expondo casos concretos para cada uma das situações.

Tabela 1

		Agente corrompido	
		Atleta	Árbitro
Vício	Arrangement	Totonero (1980)	Calciocaos (2006)
	Gambling	BBS (1964)	Hoyzer (2005)

Observação: BBS = British Betting Scandal.

Muito embora esta tipificação seja possível, não devemos descurar a possibilidade de duas situações ocorrerem em simultâneo, isto é, é possível imaginar uma situação em que recorrendo a um único agente, se verifiquem ambos os vícios. Para além disto, também é plausível o caso em que são utilizados ambos os agentes e só ocorre um vício. Naturalmente, podemos também conjecturar um cenário onde ambos os vícios são alcançados por recurso a ambos os agentes. Assim, é importante salientar que os casos expostos na tabela 1 são casos onde é possível fazer a distinção clara entre tipo de vício e agente corrompido.

Podem também ser feitas duas analogias relativamente ao significado económico que ocorre em cada um dos tipos de vício. No caso dos *gambling fixes*, um jogo viciado está para o corruptor como a *inside information* está

¹ De facto, é incomum o treinador ser o único agente corrompido.

para um investidor na bolsa. No caso dos *arrangement fixes*, um jogo viciado funciona como um seguro para o agente corruptor.

Uma vez dado o mote para o estudo da viciação de resultados no futebol, é importante essencial qual será o rumo do restante artigo. Em primeiro lugar serão abordadas as questões ligadas às motivações dos agentes envolvidos; de seguida, será apresentada a revisão de literatura relativa à viciação de resultados; por último será analisado um caso empírico – o caso Calcioaos –, recorrendo a técnicas estatísticas e econométricas, sendo os resultados obtidos no estudo empírico generalizados para permitir a sua utilização em análises posteriores.

>> II. ENQUADRAMENTO TEÓRICO

Racionalidade e Motivações

Nesta fase é fundamental abordar o modelo desenvolvido por Cressey (1953), o famoso triângulo da fraude. Na sua obra, o autor postulou que a fraude teria sempre por base três fatores: pressão, oportunidade e racionalização. Na situação mais comum estes fatores atuam de forma algo sequencial, isto é, um indivíduo idóneo sofre uma pressão (podendo ser de ordem financeira ou não) que não é passível de ser partilhada com uma terceira pessoa. Na sequência desta pressão, o indivíduo apercebe-se de uma oportunidade que lhe permite cessar a pressão inicial por meio da concretização de uma fraude. Por último, o indivíduo acaba por, inconscientemente, racionalizar uma razão para cometer a fraude de tal forma que a sua consciência fica tranquila após realização da fraude. No modelo de base, este triângulo era aplicado a fraudes nas empresas cometidas por trabalhadores, mas, naturalmente, pode ser expandido a outros âmbitos.

Embora existam diversas teorias no âmbito social, a abordagem irá assentar na conceitualização do *homo economicus*, ou seja, na premissa de que a ação dos agentes é determinada pela comparação entre ganhos¹ esperados e custos² potenciais conforme proposto por Forrest (2013, p. 180).

Assim, um agente neutro ao risco apenas irá cometer fraude se os ganhos forem superiores às perdas esperadas. No entanto, um agente que seja amante do risco irá cometer fraude mesmo se os custos potenciais não forem muito superiores aos ganhos esperados. Neste ponto, Forrest (2013, p. 180) afirma que do grupo de atletas profissionais deve constar uma significativa percentagem de amantes do risco comparativamente à população geral. Partindo do princípio que existe evidência que a predisposição para participar em estratégias arriscadas é uma característica do sucesso desportivo, o autor conclui que o grupo de atletas que conseguiram chegar a profissionais deve incluir vários agentes amantes do risco.

Para além disto, também as motivações dos agentes fraudulentos são relevantes, uma vez que também estas podem influenciar a sua decisão. Neste ponto note-se que o estudo recai nas motivações dos agentes corrompidos

¹ O autor estipula que o ganho esperado é igual ao valor recebido ponderado pela probabilidade de deteção da fraude.

² Relativamente aos custos o autor estipula que dependem da probabilidade de deteção da fraude, das consequências da deteção, da perda de valor intrínseco de um atleta (ou árbitro) por uma má exibição e do custo psicológico de viver com a consciência "pesada".

e não dos agentes corruptores. Assim, o foco não está em perceber o que leva um indivíduo a querer viciar um resultado – até porque, grosso modo, essas razões coincidem com o objetivo inerente ao tipo do vício, conforme detalhado na tabela 1 –, mas sim o que leva certos indivíduos – jogadores e árbitros – a aceitarem incorrer num risco para cometer uma fraude.

Embora possam existir circunstâncias especiais de caso para caso, é possível elencar algumas razões universais. A motivação financeira é a principal, seja por necessidade ou ganância, os ganhos monetários são, sem dúvida, os mais atraentes (Hill, 2011, p. 28). Com esta realidade em mente, é necessário ter em atenção que existem circunstâncias mais propícias à fraude, são elas: salários em atraso, salários baixos, existência de dívidas por parte dos agentes, vício no jogo por parte dos agentes e aproximação do final da carreira dos agentes. Estas razões aparentam ser comuns quer a jogadores, quer a árbitros, existindo, no entanto, existem mais duas categorias principais.

Principalmente no caso das apostas, existem ameaças ao agente e às suas famílias caso não cumpra o pretendido, por parte das máfias (Hill, 2011, pp. 195-196). Muitas das vezes, os agentes acabam mesmo por ceder e cometer a fraude sem receber qualquer valor monetário, procurando apenas evitar a concretização das ameaças. No caso dos árbitros, o prestígio pode também estar em causa, uma vez que os dirigentes podem exercer influência para que um árbitro que comete fraude tenha benefícios na sua carreira futura. O corolário da afirmação anterior é que, caso um árbitro não cometa uma fraude que lhe foi encomendada, os dirigentes poderão exercer influência em sentido oposto, ou seja, denegrindo o árbitro.

Associado a qualquer fraude é comum existir um problema de medição inerente, isto é, nunca é fácil medir nem o número de fraudes, nem os seus impactos (Brooks, Aleem, & Button, 2013, pp. 31-34). O primeiro aspeto não deve levantar dúvidas: a fraude é, por definição, um fenómeno oculto, logo é expectável que uma grande parte das fraudes não seja contabilizada por não ser descoberta. A medição dos impactos é igualmente complicada. Por um lado, não existe ainda um enquadramento legal que seja suficientemente forte e completo para enquadrar as fraudes descobertas. Por outro, é importante definir quem são as vítimas e quais as suas perdas, o que, na viciação de resultados, se apresenta como particularmente complicado, uma vez que as vítimas são muitas e que o cálculo das suas perdas não é linear (Brooks et al., 2013, p. 32). Em termos simples, os espectadores de um jogo viciado foram defraudados uma vez que não assistiram a um jogo “limpo”, mas será que as vítimas se limitam aos espectadores que pagaram o bilhete? Mesmo que admitamos que sim, será que são todos os espectadores ou

apenas os de uma das equipas? Este exemplo é suficiente para demonstrar o problema da vitimização gerado pela fraude desportiva e poderia facilmente ser aumentado para situações ainda mais complexas.

Contributos teóricos da Economia

Modelo de Procura e Oferta

O primeiro contributo analisado foi dado por Forrest (2013) e baseia-se na aplicação do modelo de procura e oferta de mercado. Em termos simples, neste modelo existem as duas curvas habituais do modelo de procura e oferta, com a particularidade de o bem em causa neste mercado serem os jogos viciados. Como em qualquer modelo deste género, é a interação entre procura e oferta que gera o par preço/quantidade. Neste caso, irá determinar quantos jogos viciados existirão e qual o seu preço.

De forma a compreender qual o funcionamento de ambas as curvas é importante ter em conta quais os seus determinantes e quais os agentes presentes em cada uma delas. Do lado da oferta, temos que os possíveis “fornecedores” são os atletas e os árbitros, isto é, os agentes corrompidos. Já do lado da procura existe uma maior diversidade. Para além dos indivíduos internos ao desporto – como dirigentes, treinadores, etc. –, também existem indivíduos externos.

Em termos da oferta de *fixes* é possível elencar alguns dos seus determinantes. Assim, a oferta será tanto maior quanto:

1. menor for o risco de deteção da fraude;
2. menor for a punição associada à fraude;
3. menor for a perda de reputação associada à descoberta da fraude;
4. menor for o sentimento de pertença ao clube/comunidade.

Já a procura modelizada por Forrest (2013, pp. 181-182) é sobretudo relacionada com *gambling fixes*, embora possa ser generalizada para ser independente do tipo de vício, sem perda de significado nas conclusões. Assim, o autor estipula que a procura irá depender positivamente da liquidez dos mercados de apostas e negativamente do risco de deteção da fraude.

Por fim, apenas do encontro entre a procura e a oferta é possível determinar as quantidades e preços que iriam vigorar. O autor chega mesmo a dar exemplos simples de casos extremos onde fica patente que nenhuma das curvas por si só consegue definir os valores de equilíbrio. Por fim, conclui também que o modelo tanto pode ser utilizado para uma análise *cross-section* como para uma análise temporal.

Na perspetiva *cross-section* postula que bastaria uma análise ao nível de salários dos jogadores – lado da procura – e da liquidez do mercado de apostas – lado da oferta. Nesta análise, a situação onde o risco de fraude seria maior consistiria numa realidade de baixos salários e grande liquidez do mercado de apostas.

Já sobre a análise temporal, a grande valia do modelo passa pela identificação de choques de oferta e procura e consequente explicação do novo equilíbrio. Forrest (2013, p. 183) termina a sua obra dando o exemplo das mudanças ocorridas em Inglaterra nos anos 60 – onde existiu um forte aumento de salários e onde os mercados de apostas passaram a ser legais e regulados. Afirma também que não existe razão aparente para uma mudança das condições da oferta, pelo que a explicação para a quantidade de viciação de resultados atual terá de passar pela procura, onde, na sua opinião, a grande diferença reside nos recentes e alargados desenvolvimentos dos mercados de apostas desportivas.

Teoria dos Jogos

Ainda no domínio dos modelos base da microeconomia temos o contributo de Hakeem (2013), que propõe uma análise da temática da fraude através da utilização da Teoria dos Jogos. A sua análise não é concretizada de um ponto de vista empírico, sendo que o autor apenas elenca as bases teóricas nas quais esta abordagem pode ser efetuada.

Numa primeira fase, o autor apresenta os conceitos de Equilíbrio de Nash¹ e de Dilema do Prisioneiro² como forma de introduzir as principais bases da Teoria dos Jogos. De seguida, Hakeem concebe cinco dimensões de análise do problema da viciação de resultados. Começa por considerar ambas as equipas presentes no jogo; apresenta a segunda dimensão como

¹ Em termos simples, um jogo está em Equilíbrio de Nash se ambos os jogadores estiverem satisfeitos com o resultado.

² O Dilema do Prisioneiro é um dos jogos mais conhecidos em Teoria dos Jogos e consiste num jogo em que existem dois prisioneiros, capturados pela polícia, embora não existam provas conclusivas caso não exista confissão. As ações possíveis de ambos os prisioneiros são confessar ou não confessar. Assim, dita a matriz de ganhos deste jogo que se um confessar e o outro não confessar, o primeiro é posto em liberdade e o último apanha pena máxima; se ambos confessarem, ambos apanham uma pena intermédia; se nenhum confessar, ambos apanham uma pena mínima. Fica assim evidente que a solução de cooperação (nenhum confessar) é o melhor resultado que podem obter, no entanto, existe um desvio a esta solução, pois caso apenas um mantenha o acordo e outro confesse, este último teria grande vantagem. Assim, este desvio leva a que, em equilíbrio, ambos confessem e a solução seja a pior possível para ambos. Existe evidência que caso este jogo seja repetido é de esperar que os jogadores cooperem na maioria das vezes como forma de não serem punidos em jogadas futuras, embora com o aproximar do fim das repetições do jogo a tendência seja a de retomar a solução de não cooperação.

sendo a interação entre o árbitro e agentes externos que o possam subornar; a terceira dimensão está relacionada com o papel que os espetadores têm no jogo; a quarta dimensão foca-se no papel que os apostadores têm na viciação de resultados e, por fim, a quinta dimensão está relacionada com a interação dos *media* no desporto.

Na conceção da matriz de ganhos¹ do jogo proposto pelo autor, existem duas classes de jogadores. De um lado existe a classe dos agentes da Lei, cujas ações possíveis são uma atuação proativa ou apenas reativa. De outro lado existe a classe dos agentes passíveis de cometer fraude, que coincidem com os vários agentes identificados nas cinco dimensões anteriores (jogadores, árbitros, casas de apostas, apostadores, espetadores e *media*) e cujas ações são praticar ou não fraude. É também pressuposto que uma ação proativa das autoridades faz com que exista uma menor propensão à prática da fraude.

Assim, a matriz de ganhos tem dimensão $2 \times 6 \times 3^2$ e o seu preenchimento é possível através de estimação com base em inquéritos. De facto, Hakeem apresenta alguns valores aproximados para as várias células da matriz, partindo do conhecimento comum das razões da participação de vários dos agentes. Por exemplo, o objetivo de uma casa de apostas passa somente pelo lucro, enquanto que um adepto apenas estará interessado com o seu entretenimento e satisfação. Por sua vez, um jogador poderá estar interessado no lucro assim como na sua satisfação.

Desta forma, segundo Hakeem (2013, p. 255), será a aplicação das várias técnicas de Teoria dos Jogos a esta matriz estimada que poderá dar indicações às autoridades competentes de como proceder corretamente de forma a combater a fraude desportiva.

¹ Ou mapa de valências cognitivas, como designada pelo autor.

² Duas jogadas possíveis das autoridades, seis tipos de agentes fraudulentos e três razões para participação no jogo concebido (lucros, entretenimento e satisfação).

>> III. ANÁLISE EMPÍRICA – O CASO CALCIOCAOS

O Caso Calciocaos

O Caso Calciocaos (Boeri & Severgnini, 2013, pp. 104-105) constitui dos maiores escândalos mundiais de manipulação de resultados, tendo mesmo provocado a retirada de um título de campeão e na descida de divisão de várias equipas. Este caso ocorreu na primeira liga italiana – Serie A –, também conhecida por Calcio, na época desportiva de 2004/05, embora apenas tenha vindo a ser descoberto em 2006 e com as consequências a terem lugar apenas a partir da época de 2006/07.

De forma muito resumida, a manipulação de resultados consistiu na escolha dos árbitros por parte de algumas das equipas para alguns dos seus jogos, através de alguns membros das administrações dos clubes. Os árbitros escolhidos eram sobretudo os de segunda linha, isto é, ainda sem estatuto internacional, mas com ambição de o alcançar num futuro próximo. No fundo, a escolha recaía em árbitros com experiência mediana, ou seja, nem árbitros já de topo, nem árbitros cuja carreira era ainda curta.

A principal punição resultante do julgamento deste caso tratou-se da descida de divisão da Juventus – que foi também a principal equipa a manipular os resultados – e da retirada do seu título de campeã em 2004/05, embora outros emblemas italianos tenham sofrido sanções idênticas, para além de multas e banimentos.

A análise empírica foi feita com base nos dados de todos os jogos da Serie A na época de 2004/05. A base de dados foi elaborada após serem reunidas várias fontes¹. Alguns dados já estavam disponíveis com formatação adequada ao tratamento informático, enquanto que outros foram criados de forma manual para o efeito – seja na forma, seja no conteúdo. Nos casos em que tal separação faz sentido, os valores globais foram desagregados em valores para a equipa da casa e visitante (número de golos da equipa da casa e número de golos da equipa visitante, por exemplo). Todos os testes de hipóteses que serão apresentados foram realizados através do programa estatístico R.

Em anexo estão disponíveis as tabelas referidas ao longo do texto. As tabelas em anexo foram já objeto de tratamento de dados de forma a serem de leitura mais fácil. Assim, apenas figuram os valores que teriam o maior in-

¹ Neste ponto, notar que a contribuição de Boeri e Severgnini (2008) foi fulcral na identificação dos vários jogos viciados.

teresse para análise. Valores esses que estão também ligados a um simples código de cor: na diagonal o valor estará a verde se H0 não foi rejeitada e a vermelho se foi rejeitada; nos restantes campos o valor estará a verde se H0 foi rejeitada, a vermelho se não foi rejeitada e a amarelo caso a situação seja dúbia. Também em anexo está disponível uma tabela com várias medidas de estatística descritiva acerca das diferentes variáveis.

Na restante análise está subjacente algum conhecimento da gíria futebolística como forma de compreensão da totalidade do conteúdo.

Detalhe da base de dados

- Fix – Variável binária que assume o valor 1 se o jogo tiver sido viciado e o valor 0 em caso contrário;
- HomeTeam – Variável categórica nominal que contém o nome da equipa da casa;
- AwayTeam – Variável categórica nominal que contém o nome da equipa visitante;
- HalfTimeGoals (HTG) – Variável numérica discreta que representa o número total de golos marcados no jogo até ao intervalo;
- FullTimeGoals (FTG) – Variável numérica discreta que representa o número total de golos marcados no jogo;
- 1.15 – Variável numérica discreta que representa o número total de golos marcados entre os minutos presentes no nome da variável;
- 16.30 – Variável numérica discreta que representa o número total de golos marcados entre os minutos presentes no nome da variável;
- 31.45 – Variável numérica discreta que representa o número total de golos marcados entre os minutos presentes no nome da variável;
- 45+ – Variável numérica discreta que representa o número total de golos marcados no período de descontos da primeira parte;
- 46.60 – Variável numérica discreta que representa o número total de golos marcados entre os minutos presentes no nome da variável;
- 61.75 – Variável numérica discreta que representa o número total de golos marcados entre os minutos presentes no nome da variável;
- 76.90 – Variável numérica discreta que representa o número total de golos marcados entre os minutos presentes no nome da variável;
- 90+ – Variável numérica discreta que representa o número total de golos no período de descontos da segunda parte;
- OwnGoals (OG) – Variável numérica discreta que representa o número total de autogolos marcados no jogo;

- Penáltis (PEN) – Variável numérica discreta que representa o número total de penáltis assinalados no jogo (não necessariamente convertidos em golo);
- YellowCards (YC) – Variável numérica discreta que representa o número total de cartões amarelos no jogo;
- RedCards (RC) – Variável numérica discreta que representa o número total de cartões vermelhos no jogo;
- HomeTeamPos (HPOS) – Variável categórica ordinal que representa a classificação da equipa da casa no começo da jornada;
- AwayTeamPos (APOS) – Variável categórica ordinal que representa a classificação da equipa visitante no começo da jornada.

Análise Estatística

A análise estatística foi dividida em quatro grupos: golos (Anexo I), minutos dos golos (Anexo II), infrações (Anexo III) e tabelas de contingência (Anexo IV).

As matrizes presentes nos anexos pretendem sistematizar dois tipos de testes de hipóteses. Na diagonal principal está inscrito o *p-value* associado a um teste de ajustamento ao Qui-Quadrado da variável em causa relativamente a uma *Poisson* com parâmetro igual à média amostral da variável. Em termos mais técnicos, temos as seguintes hipóteses¹

$$H_0: X \sim \text{Po}(\bar{X}) ; H_1: H_0 \text{ é falsa}$$

A realização do teste no R tem uma limitação: o valor crítico do teste é dado por Qui-Quadrado com $k-1$ graus de liberdade, no entanto, dado que o parâmetro foi estimado, os graus de liberdade deveriam ser reduzidos numa unidade. Assim, nos casos em que o *p-value* está suficientemente afastado de alfa não foi feita qualquer correção, já nos casos em que pudessem existir problemas o valor foi estimado manualmente. De notar que os *p-values* foram obtidos por aplicação da rotina de Monte Carlo nos casos em que os valores esperados de cada categoria eram inferiores a 5 pelo método normal.

¹ O objetivo de um teste desta natureza é a não rejeição de H_0 , pelo que o *p-value* deverá ser maior que alfa, caso contrário é rejeitada H_0 .

Na metade superior da matriz estão inscritos os *p-values* associados ao teste de razão entre as médias de duas variáveis *Poisson*¹. Em termos mais técnicos, temos as seguintes hipóteses²

$H_0: \frac{\bar{X}}{\bar{Y}} = 1$; $H_1: \frac{\bar{X}}{\bar{Y}} \neq 1$; sendo X a variável-linha e Y a variável-coluna, respetivamente

Na metade inferior da matriz está inscrito o sentido desta razão, ou seja, caso esteja inscrito o sinal “+” significa que $X > Y$, caso seja o sinal “-” significa o inverso. De notar que o sinal está na posição da matriz onde o p-value estaria caso a matriz fosse transposta.

Na sequência serão expostas as conclusões referentes à análise das várias tabelas de forma sintética. Quando for referido “em geral” pretende-se significar que está a ser tomada em consideração a totalidade da amostra, já quando for referido “em particular” pretende-se significar que apenas parte da amostra foi utilizada, no caso, ou apenas os jogos viciados ou apenas os jogos limpos.

Golos

Todas as variáveis ajustam a uma $Po(\bar{X})$ para $\alpha=0,01$.

Em geral (ou seja, na totalidade da amostra, incluindo jogos viciados e limpos) e nos jogos limpos, em particular, existe uma vantagem de jogar em casa. Isto é, o número de golos da equipa da casa é superior ao da equipa visitante com um p-value inferior a uma centésima. Nos jogos viciados esta vantagem desaparece (*p-value*=0,15).

Isto pode ser indicação de dois fenómenos: a probabilidade de um jogo ter sido viciado é maior caso a equipa da casa não ganhe; a viciação pode ocorrer mais vezes fora de casa do que em casa (à partida, caso a viciação fosse neutra ao fenómeno casa/fora, a vantagem de jogar em casa não deveria desvanecer de forma tão acentuada).

¹ Nos casos em que o teste de ajustamento concluir que o ajustamento a uma distribuição Poisson não se verifica, os valores na metade superior da matriz estão de acordo com o teste não paramétrico de Wilcoxon. Assim, caso que quer o teste paramétrico, quer o não paramétrico, levem à mesma conclusão, é apresentado o p-value do teste paramétrico.

² O objetivo de um teste desta natureza é a rejeição de H_0 , pelo que o valor de alfa deverá ser maior que o p-value, caso contrário H_0 não é rejeitada.

Minutos dos golos

Todas as variáveis ajustam a uma $Po(\bar{X})$ para qualquer α habitual.

O número de golos marcados entre o minuto 46 e 60 é maior nos jogos limpos do que nos viciados ($p\text{-value}=0,032$) pelo que é expectável que, num jogo com vários golos neste período, exista uma baixa probabilidade de ter sido viciado.

Infrações

À exceção de algumas variáveis relacionadas com os cartões amarelos, todas as outras ajustam a uma $Po(\bar{X})$ para qualquer alfa habitual.

Em nenhum caso o número de autogolos beneficia uma equipa de forma sistemática ($p\text{-values}$ entre 0,61 e 1).

No caso geral, existe vantagem de jogar em casa no número de penáltis a favor ($p\text{-value}=0,045$). No entanto, olhando aos casos particulares, esta vantagem desaparece ($p\text{-values}$ superiores a 0,1).

O número de cartões vermelhos também não apresenta nenhuma tendência para o favorecimento de uma das equipas ($p\text{-values}$ superiores a 0,15).

No caso geral e no caso particular dos jogos limpos, os $p\text{-values}$ são tais que o rácio entre número de penáltis e vermelhos não é significativamente diferente de um, o que pode ser visto como uma situação no jogo em que uma falta mais grave que dá origem ao penálti e ao respetivo vermelho. No entanto, no caso dos jogos viciados o $p\text{-value}$ é de 0,036, o que demonstra que o rácio é diferente de 1 (no caso, menor), significando que existem mais cartões vermelhos do que penáltis neste tipo de jogos. Assim sendo, e tratando-se de um caso de viciação de árbitros, verifica-se um agravamento relativo de cartões vermelhos nestes jogos, situação que acaba por ser muito vantajosa para a equipa que vicia o jogo e que acaba por passar mais despercebida do que, por exemplo, assinalar um penálti falso ou outra situação do género.

De facto, o número de penáltis tende mesmo a ser maior nos jogos não viciados do que nos jogos viciados (os $p\text{-values}$ não são ideais, mas apontam neste sentido), situação que pode dever-se precisamente a este receio de que um penálti seja demasiado óbvio. Assim, um jogo sem penáltis pode estar associado a uma probabilidade de viciação superior.

Por fim, o rácio de cartões amarelos à equipa da casa, relativamente à visitante, tende a ser menor que um em geral e nos jogos limpos (mais uma vantagem de jogar em casa), com $p\text{-values}$ inferiores a uma centésima. No

entanto, nos jogos viciados o *p-value* é de 0,21, o que mais uma vez desvanece a vantagem de jogar em casa.

Tabelas de contingência

Na tabela 2 estão presentes os *p-values* referentes ao Teste de Independência do Qui-Quadrado, ao qual as variáveis que figuram nas tabelas de contingência foram submetidas. Em termos mais técnicos, temos as seguintes hipóteses¹

H0: As variáveis são independentes ($p_{ij}=p_i.p_j, \forall i,j$); H1: H0 é falsa

Tabela 2 – P-values do teste de independência

Variável	<i>P-value</i>
T	0
REF	0
HOMETEAM	0,626
AWAYTEAM	0,074
HPOS	0,458
APOS	0,017

Observ: De notar que a Regra de Cochran apenas é verificada para a variável T, ainda assim, para os restantes casos apenas não é verificada por algumas diferenças ligeiras que, dada a dimensão de algumas das tabelas de contingência, acaba por ser razoável.

Desta forma, facilmente se verifica que, para um alfa de 10%², existe dependência estatisticamente significativa entre a presença ou não de viciação num jogo e as variáveis trimestre, ranking do árbitro, equipa adversária e classificação da equipa adversária no começo da jornada. O principal resultado a reter desta análise é a existência de dependência entre a existência de viciação num jogo e o ranking do árbitro³, uma vez que corrobora os acontecimentos do Caso Calciocaos.

¹ O objetivo de um teste desta natureza é a rejeição de H0, pelo que o valor de alfa deverá ser maior que o *p-value*, caso contrário H0 não é rejeitada.

² Ou mesmo mais exigente, como 2%, excetuando apenas uma variável neste caso.

³ O que equivale a dizer: entre existência de viciação num jogo e o árbitro nomeado.

Modelo Econométrico

Especificação do modelo

A primeira etapa da concetualização do modelo passa por escolher a forma funcional a adotar para a estimação. Será uma regressão linear a especificação mais adequada?

Uma vez que o modelo representa uma situação dicotómica, isto é, ou ocorreu fraude num jogo ou não, o que significa que os resultados estimados deverão estar situados entre zero e um, concluímos que o modelo linear não garante que as estimativas estejam compreendidas neste intervalo. Esta situação decorre do facto da variável dependente ser função linear das variáveis explicativas, não existindo, assim, qualquer restrição ao valor estimado, podendo, inclusive, ser negativo ou superior a um.

De igual forma, dada a dicotomia da variável dependente, existirá um problema de heteroscedasticidade¹. Embora este último problema pudesse ser resolvido recorrendo a técnicas econométricas comuns, continuaríamos a estar na presença de outros problemas, principalmente a questão da não restrição dos valores estimados, pelo que a melhor opção passa pela utilização de um modelo diferente — um modelo *probit*.

Neste ponto é importante explicitar quais as variáveis que serão utilizadas no modelo. Assim, cada observação é composta por oito variáveis explicativas e pela variável dependente e diz respeito a um jogo. As variáveis incluídas no modelo são

- FIX – Variável dependente. Igual à unidade caso o jogo tenha sido viciado e de valor nulo caso o jogo seja limpo;
- DAY – Número da jornada do jogo;
- ELO – Somatório do *rating* ELO² das duas equipas no início da jornada³;
- APOS – Classificação da equipa visitante no começo da jornada;
- 31.45 – Número total de golos marcados entre os 31 e os 45 minutos do jogo;
- REF_INT – Variável binária igual a 1 caso o árbitro tenha estatuto internacional e nula caso contrário;
- APEN – Número total de penáltis assinalados, a favor da equipa visitante, no jogo;

¹ Heteroscedasticidade pode ser definida como ausência de homoscedasticidade, o que, em termos simples, equivale a dizer que a variância do erro aleatório do modelo não é constante.

² O *rating* ELO surgiu originalmente como força de medir a força relativa entre jogadores de xadrez, tendo sido adaptado para medir a força das equipas de futebol. Os valores foram obtidos em clubelo.com, onde pode também ser estudado com mais cuidado a forma como o *rating* é construído

³ Embora o somatório de ELO seja feito jornada a jornada, os valores de ELO de cada equipa foram obtidos tendo como referência o começo da temporada.

- REFAGE - Idade do árbitro do encontro¹;
- REF – *Ranking* do árbitro. Ordenação descendente tendo como critério o número total de jogos arbitrados.

Acerca da variável FIX, existem alguns aspetos que devem ser clarificados, por forma a definir de forma inequívoca o estudo que se pretende fazer. Os próximos parágrafos fazem esse aprofundamento.

Logo pela definição de FIX surgem duas grandes problemáticas: a eficácia da fraude e a fraude indireta. A questão da eficácia passa pela viciação ser ou não útil, isto é, se de facto a viciação foi crucial para definir o resultado do jogo ou se, pelo contrário, a equipa até acabou por perder o jogo. Já a questão da fraude indireta passa pela possibilidade de ser uma terceira entidade (outra equipa, por hipótese) a tomar o papel de agente corruptor e não nenhuma das duas equipas envolvidas no jogo. É também fundamental não perder de vista que o âmbito do presente modelo visa apenas estudar a realidade de jogos onde os árbitros são o agente corrompido e se trata de um *arrangement*.

Assim, FIX tomará valor unitário caso exista viciação, independentemente da sua eficácia. A ideia presente neste pressuposto é a de que, desde que tenha existido um efetivo suborno ao árbitro e que o mesmo tenha sido aceite e realizado em campo, independentemente de a equipa ter ganho (ou não), ou de ter ganho apenas por isso (ou não), ocorreu viciação e, por isso, FIX tomará o valor de um.

Relativamente à fraude indireta (ou cruzada), o seu tratamento passaria pela utilização de variáveis dicotómicas adicionais, de forma a compreender o seu funcionamento, isto é, uma variável W que indicasse se a fraude era incitada pela equipa visitante (ou não) e outra variável Z que indicasse se a fraude era incitada por uma equipa terceira (ou não). Infelizmente, a inclusão das ditas variáveis tem um problema imediato, sendo alguma delas unitária, também FIX terá de o ser, pelo que, neste caso, a previsão perde o valor.

Formalizando o modelo enquanto um modelo *probit* de variável latente — isto é, apenas observamos a ocorrência ou não de fraude, mas não o valor que resulta da especificação funcional das variáveis explicativas —, temos o seguinte,

¹ A idade do árbitro tem por referência o começo da temporada.

$$Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & Y_i^* \leq 0 \end{cases}$$

com

$$Y_i^* = X_i\beta + \mu_i, \mu_i \sim N(0,1)$$

e com

$$\begin{aligned} X_i\beta = & \beta_0 \\ & +\beta_1 DAY_i + \beta_2 DAY_i^2 \\ & +\beta_3 ELO_i + \beta_4 ELO_i^2 \\ & +\beta_5 APOS_i + \beta_6 APOS_i^2 \\ & +\beta_7 31.45_i + \beta_8 APEN_i + \beta_9 REF_INT_i \\ & +\beta_{10} REFAGE_i + \beta_{11} REFAGE_i^2 + \beta_{12} REFAGE_i^3 \\ & +\beta_{13} REF_i + \beta_{14} REF_i^2 + \beta_{15} REF_i^3 + \beta_{16} REF_i^4 + \beta_{17} REF_i^5 \end{aligned}$$

De seguida, é importante compreender que, no âmbito do modelo *probit*, o valor de cada coeficiente é algo irrelevante do ponto de vista da interpretação, ou seja, apenas o seu sinal terá interesse nesta análise. Para determinar o efeito concreto que cada variável tem, em termos probabilísticos, devem ser calculados os seus efeitos marginais.

Os efeitos marginais servem precisamente para explicar qual é, em média, a variação na probabilidade estimada dada uma variação infinitesimal de uma variável explicativa. Teoricamente, dado o carácter não linear do modelo, o cálculo dos efeitos marginais implica uma escolha de indexação, isto é, para calcular o efeito marginal de uma variável é necessário fixar todas as outras. A opção mais utilizada é fixar todas as variáveis nas suas médias amostrais, gerando, assim, os efeitos marginais na média. Outra opção, possível através dos *softwares* estatísticos atuais, passa por calcular uma média de efeitos marginais.

Porém, dado o tipo de variáveis presente neste modelo, esta abordagem tem de ser feita de forma cautelosa, uma vez que os efeitos marginais têm

por base variações infinitesimais e não variações discretas que as variáveis utilizadas implicam. Já no caso da variável REF_INT, o cálculo do seu efeito marginal é dado pela diferença entre o valor de Y* quando REF_INT é unitária e quando é nula.

De qualquer forma, segundo Greene (2003, p. 690), o cálculo de efeitos marginais para variáveis dicotômicas como se de variáveis contínuas se tratassem revela-se bastante acertado, pelo que a opção adotada para o cálculo dos efeitos marginais nestas circunstâncias foi a utilização da média dos efeitos marginais.

Importa também explicar a interpretação dos coeficientes e ainda como pode o modelo ser utilizado em termos de previsão. No geral, caso o coeficiente seja positivo, o aumento da variável induz um aumento da probabilidade estimada, caso seja negativo, o efeito é inverso. No entanto, a variação da probabilidade será sempre função de uma regra normal e não diretamente do valor do coeficiente, isto é, o valor estimado pela equação do probit é o argumento da função normal reduzida e será o valor que resulta da aplicação desta função que representa a probabilidade estimada.

Adicionalmente, o modelo pode ainda ser melhorado se considerarmos alguns termos de interação relacionados, sobretudo, com as características dos árbitros. Assim, as interações a incluir são

$$+\beta_{18}REF_iELO_i+\beta_{19}REF_iAPEN_i+\beta_{20}REF_i31.45_i$$

Finalmente, na tabela 3 estão presentes ambas as estimações relativas aos modelos até aqui detalhados, assim como os respetivos efeitos marginais¹ na tabela 4

Tabela 3 – Estimação do probit sem (1) e com (2) interações

	(1)		(2)	
	FIX		FIX	
DAY	0,284***	(5,42)	0,327***	(6,41)
DAY ²	-0,00759***	(-5,69)	-0,00851***	(-6,39)
ELO	-0,0997**	(-2,99)	-0,165***	(-4,61)
ELO ²	0,0000150**	(3,00)	0,0000243***	(4,57)
APOS	-0,249***	(-3,40)	-0,281***	(-3,78)
APOS ²	0,0123***	(3,63)	0,0136***	(4,04)

¹ Devido à singularidade da matriz de variâncias, não é possível calcular o *p-value* para efeito marginal da variável REFAGE. No entanto, dado o seu valor estar em linha com o do modelo (1), a expectativa é a de que a sua significância também seja semelhante.

	(1) FIX		(2) FIX	
31.45	0,366**	(2,81)	0,879***	(3,35)
APEN	-0,905*	(-2,30)	-3,229***	(-3,77)
REF_INT	-1,256***	(-5,27)	-1,442***	(-5,48)
REFAGE	-104,9***	(-4,41)	-116,8***	(-4,85)
REFAGE ²	2,118***	(4,45)	2,361***	(4,89)
REFAGE ³	-0,0142***	(-4,49)	-0,0159***	(-4,93)
REF	-4,791***	(-5,15)	-6,730***	(-5,13)
REF ²	1,151***	(5,24)	1,237***	(4,94)
REF ³	-0,116***	(-5,28)	-0,123***	(-4,98)
REF ⁴	0,00511***	(5,26)	0,00531***	(4,96)
REF ⁵	-0,0000808***	(-5,20)	-0,0000829***	(-4,89)
REF . ELO			0,000447**	(2,71)
REF . APEN			0,206***	(3,33)
REF . 31.45			-0,0672*	(-2,35)
Constante	1898,8***	(4,64)	2205,8***	(5,31)
Observações	380		380	
Pseudo R ²	0,430		0,483	

Observ: Estatísticas t entre parentesis. * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Tabela 4 – Efeitos marginais médios, por variável, associados ao modelo (1) e (2)

	(1) FIX		(2) FIX	
DAY	-0.00136**	(-2.69)	-0.00113*	(-2.17)
ELO	-0.000105	(-0.69)	-0.0000343	(-0.22)
APOS	0.000949	(0.27)	0.000227	(0.07)
31.45	0.0608**	(3.08)	0.0394	(1.85)
APEN	-0.150*	(-2.35)	-0.201***	(-3.35)
REF_INT	-0.209***	(-7.16)	-0.218***	(-13.38)
REFAGE	-0.0820***	(-6.88)	-0.0838	
REF	-0.0561***	(-3.87)	-0.0496***	(-4.73)

Observ: Estatísticas t entre parentesis. * $p < 0,05$; » $p < 0,01$; *** $p < 0,001$.

Em primeiro lugar, constatamos que todas as variáveis são estatisticamente significativas. Para além disto, os vários testes de significância conjunta (ou seja, testando se os coeficientes de, por exemplo, DAY e DAY² são

conjuntamente diferentes de zero) resultam sempre na rejeição da hipótese nula¹ para o valor habitual de $p=0.05$.

Por último, relativamente ainda a questões de significância, verificamos que os efeitos marginais das variáveis ELO, APOS e 31.45 não são significativos, ainda que os coeficientes relativos às variáveis e aos seus quadrados o sejam.

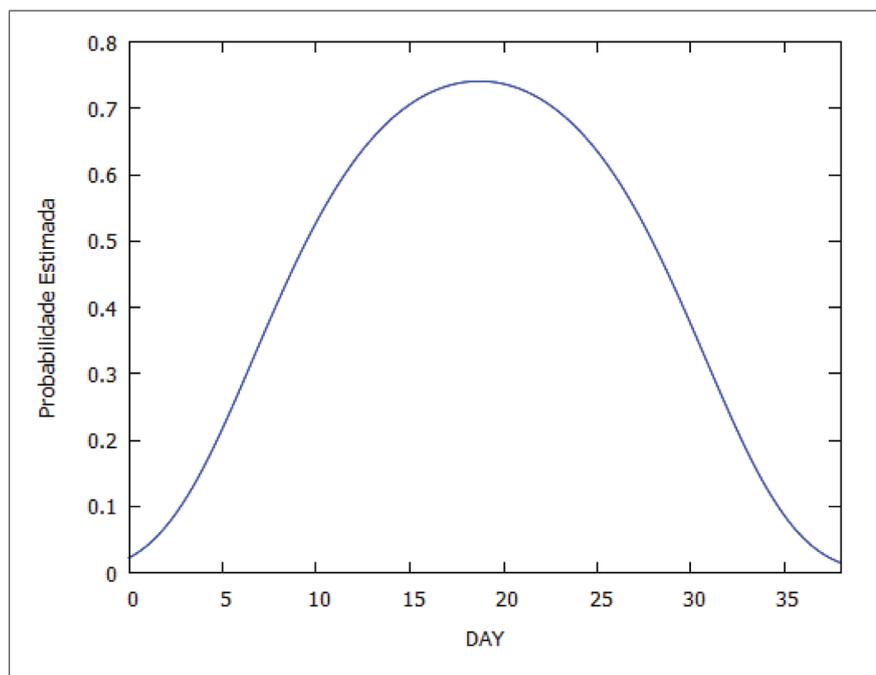
Esta aparente contradição deve-se ao facto dos coeficientes do modelo e dos efeitos marginais medirem diferentes aspetos. Os coeficientes do modelo identificam a contribuição de determinada variável (ou transformações dela) para a predição da probabilidade latente, ou seja, Y_i^* . Por sua vez, os efeitos marginais medem o impacto de cada variável em Y_i . Para além da diferença que é imediatamente apercebida (Y_i^* contra Y_i), é também fundamental não esquecer que os coeficientes são estimados componente a componente, mas os efeitos marginais são estimados variável a variável, daí decorre que o número de efeitos marginais (j) num modelo com k coeficientes obedecerá sempre a $j \leq k$.

Em suma, esta diferença nas significâncias pode ser resumida ao seguinte: as variáveis ELO, APOS e 31.45 são importantes para determinar o valor de Y_i^* e, assim, determinar se \hat{Y} deve ser nulo ou unitário aquando de um *cutoff*; no entanto, essas mesmas variáveis deixam de ser importantes para determinar o valor de Y_i , ou seja, o valor concreto de \hat{Y} .

Relativamente aos coeficientes, um dos primeiros pontos a notar é que nenhum dos sinais se altera entre o modelo 1 e o modelo 2. Seguidamente importa perceber qual é a configuração das várias parábolas inerentes ao modelo. Assim, temos que à variável DAY está associada uma parábola cuja concavidade está voltada para baixo, o que significa que a probabilidade de fraude aumenta com o decorrer da época até um determinado ponto a partir do qual essa probabilidade começa a decrescer. A explicação para isto passa pela tentativa de ocultação da fraude ao longo da época desportiva, sendo as equipas avessas a concentrarem no final da época muitos fenómenos de fraude, uma vez que tal situação seria mais suspeita, embora, em termos de pontos ganhos, o resultado seja perfeitamente equivalente. Esta evidência vai de encontro ao também demonstrado por Boeri e Severgnini (2008, p. 23).

¹ Isto é, rejeita-se que, conjuntamente, os coeficientes sejam nulos.

Figura 1 - Gráfico que relaciona a probabilidade estimada de fraude, segundo o modelo 1, com a variável DAY, ceteris paribus.

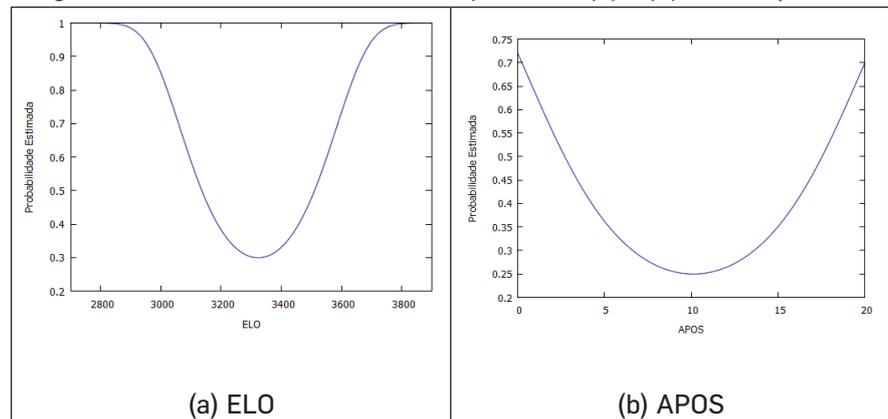


Observações: Todas as restantes variáveis estão fixadas na sua média amostral.

Quanto às variáveis ELO e APOS, o seu comportamento é bastante semelhante já que ambas apresentam parábolas com a concavidade voltada para cima, o que significa que a probabilidade de fraude é menor quando APOS está perto dos seus valores mais centrais e quando a variável ELO associada ao jogo tende, também, para a sua média. Caso contrário, isto é, se estas variáveis estiverem a tender para os seus valores mais extremos, a probabilidade de fraude aumenta (Figura 2).

Através desta descrição deve também ser óbvia outra razão para o efeito marginal destas duas variáveis não ser significativo. Uma vez que estão a ser calculados efeitos marginais médios e que estas variáveis estão modeladas enquanto parábolas, dada a sua simetria, é natural que a média dos seus efeitos marginais seja nula, fazendo com que o seu efeito marginal não seja significativo. Já no caso da variável DAY, embora também seja modelada como uma parábola, existe uma assimetria relativa — devido à aversão à prática de fraude no final da época — que faz com que o seu efeito marginal médio seja negativo e significativo, embora bastante perto de zero.

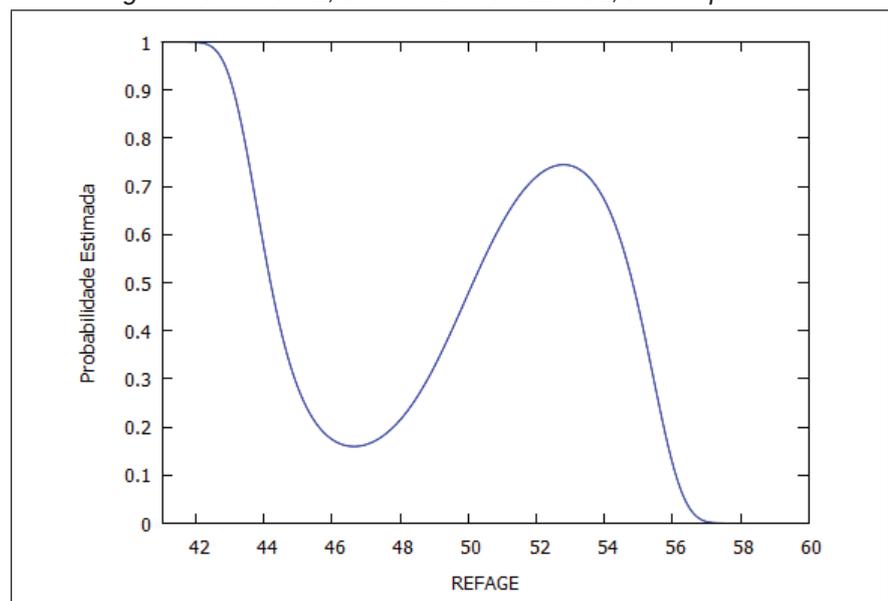
Figura 2 - Gráfico que relaciona a probabilidade estimada de fraude, segundo o modelo 1, com a variável referida em (a) e (b), ceteris paribus.



Observ: Todas as restantes variáveis estão fixadas na sua média amostral.

A configuração da curva associada à variável REFAGE é um pouco mais complexa. Nesta situação, inicialmente a probabilidade de fraude tende a diminuir, observando-se depois duas inflexões, voltando a decrescer à medida que se aproxima do seu valor máximo. Uma conclusão imediata a retirar é a de que os árbitros mais jovens e os árbitros perto do final de carreira constituem grupos de risco, segundo os resultados do modelo.

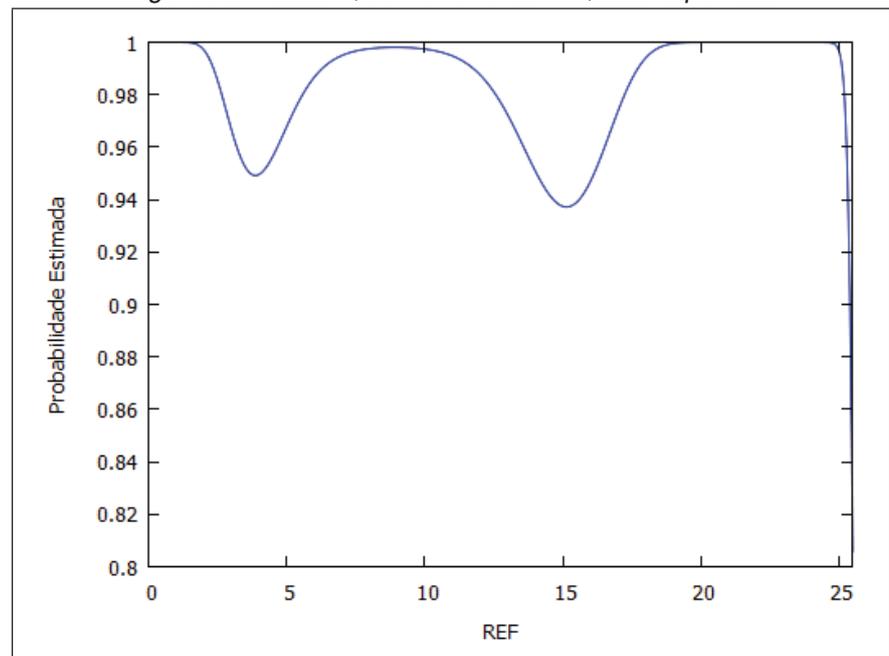
Figura 3 - Gráfico que relaciona a probabilidade estimada de fraude, segundo o modelo 1, com a variável REFAGE, ceteris paribus.



Observ; Todas as restantes variáveis estão fixadas na sua média amostral

Já o comportamento da variável REF tem algumas semelhanças ao da variável REFAGE, embora seja bastante peculiar. O que se verifica é que, a variável REF, apresenta três zonas onde a probabilidade estimada de fraude é elevada e outras tantas onde a probabilidade de fraude é relativamente baixa. Por um lado, a probabilidade é alta para árbitros cujo ranking é elevado — uma vez que estes árbitros podem ser prejudicados, em termos internacionais, caso não cooperem —, para os árbitros cujo ranking é médio — e aos quais são prometidas vantagens nas futuras nomeações internacionais — e aos árbitros que estarão quer no início, quer no final, de carreira e, portanto, fazem poucos jogos (tendo, assim, um ranking baixo) — já que poderão estar mais vulneráveis a subornos. Já os árbitros cuja probabilidade de fraude é relativamente baixa são os árbitros com ranking consolidado, embora ainda não de topo absoluto; os árbitros que começam a entrar em fases da carreira onde podem pensar almejar nomeações internacionais e os árbitros que arbitram poucos jogos por época e, por isso, acabam por não conseguir ter influência nos resultados de forma significativa¹.

Figura 4 - Gráfico que relaciona a probabilidade estimada de fraude, segundo o modelo 1, com a variável REF, ceteris paribus.



Observ: Todas as restantes variáveis estão fixadas na sua média amostral.

¹ Boeri e Severgnini (2008) exploram mais a fundo a relação entre a probabilidade de fraude e os árbitros.

Por fim, a interpretação das variáveis 31.45, APEN e REF_INT é mais imediata. No caso de 31.45, uma variação positiva no seu valor induz uma variação positiva na probabilidade de fraude, já no caso de APEN a conclusão é a simétrica. No caso da variável dicotómica REF_INT, se a variável for unitária, a probabilidade de fraude é menor. Analisando os termos de interação no modelo (2), é necessário ter em atenção que existem agora variáveis que não podem ser tomadas de forma independente (REF, ELO, APEN 31.45).

O poder explicativo de ambos os modelos pode ainda ser atestado através da comparação entre os valores reais da variável FIX e os valores por eles estimados — resultados patentes nas tabelas 5 e 6.

Tabela 5 – Quadro de predição do modelo (1)

Corte de sucesso: $C = 0,5$

Equação estimada	Dep=0	Dep=1	Total
$P(\text{Dep}=1) \leq C$	285	29	314
$P(\text{Dep}=1) > C$	15	51	66
Total	300	80	380
Corretos	285	51	336
% Corretos	95,00	63,75	88,42
% Incorretos	5,00	36,25	11,58

Tabela 6 – Quadro de predição do modelo (2)

Corte de sucesso: $C = 0,5$

Equação estimada	Dep=0	Dep=1	Total
$P(\text{Dep}=1) \leq C$	288	23	311
$P(\text{Dep}=1) > C$	12	57	69
Total	300	80	380
Corretos	288	57	345
% Corretos	96,00	71,25	90,79
% Incorretos	4,00	28,75	9,21

Dado que será também interessante aplicar os coeficientes estimados no modelo a outros casos, um pouco como medida de *benchmark* de possível manipulação de resultados, podemos estimar o mesmo modelo, mas utilizando agora variáveis *standardizadas*¹. De notar que este novo modelo, embora naturalmente apresente diferentes valores estimados para os coeficientes, mantém todas as propriedades do anterior, em particular tem o mesmo pseudo- R^2 e a mesma taxa de acerto².

De notar que a utilização deste modelo com uma finalidade preditiva é um exercício teórico, que possui limitações e cujos resultados não devem ser encarados como verdades absolutas, embora, observando o pressuposto principal de que a existir viciação de resultados se trata de um *arrangement* perpetrado por árbitros, o seu valor possa ser encarado como um indicador³ da existência, ou não, de fraude.

Finalmente, o modelo na sua versão *standardizada* tem os seus resultados presentes na tabela 7, assim como os respetivos efeitos marginais na tabela 8.

Tabela 7 – Probit *standardizado* sem (3) e com (4) interações

	(1)		(2)	
	FIX		FIX	
s_DAY	-0,131	(-0,98)	-0,0551	(-0,40)
s_DAY ²	-0,915***	(-5,69)	-1,026***	(-6,39)
s_ELO	-0,102	(-0,96)	-0,0284	(-0,26)
s_ELO ²	0,251**	(3,00)	0,408***	(4,57)
s_APOS	0,0340	(0,35)	0,000794	(0,01)
s_APOS ²	0,403***	(3,63)	0,445***	(4,04)
s_31.45	0,245**	(2,81)	0,147	(1,45)
s_APEN	-0,304*	(-2,30)	-0,405**	(-3,03)
REF_INT	-1,256***	(-5,27)	-1,442***	(-5,48)
s_REFAGE	1,217**	(3,14)	1,394***	(3,54)
s_REFAGE ²	-0,938***	(-5,03)	-1,110***	(-5,37)
s_REFAGE ³	-1,101***	(-4,49)	-1,229***	(-4,93)
s_REF	-1,720***	(-4,31)	-1,492***	(-3,64)
s_REF ²	-4,325***	(-4,95)	-4,623***	(-4,69)
s_REF ³	2,284***	(3,79)	2,114***	(3,57)

¹ Isto é, as variáveis transformadas segundo a regra $x_i^{\text{Standard}} = \frac{x_i - \bar{X}}{\sigma_x}$

² É assim desnecessário apresentar as tabelas de predição do modelo (3) e (4), uma vez que são exatamente iguais às do modelo (1) e (2), respetivamente.

³ Caso queiramos utilizar este modelo para efeitos de previsão, será até razoável corrigir a probabilidade estimada que obtivermos pela taxa de acerto do modelo aqui estimado, ou seja, 90,79% . $\overline{\text{FIX}}$.

	(1) FIX		(2) FIX	
s_REF ⁴	2,914***	(5,15)	3,169***	(4,89)
s_REF ⁵	-1,468***	(-5,20)	-1,506***	(-4,89)
s_REF . ELO			0,412**	(2,71)
s_REF . APEN			0,493***	(3,33)
s_REF . 31.45			-0,319*	(-2,35)
Constante	0,807*	(2,34)	0,887*	(2,40)
Observações	380		380	
Pseudo R ²	0,430		0,483	

Observ: Estatísticas *t* entre parentesis. * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Tabela 8 – Efeitos marginais médios, por variável, associados ao modelo (3) e (4)

	(3) FIX		(4) FIX	
s_DAY	-0,0150**	(-2,67)	-0,0124*	(-2,08)
s_ELO	-0,0136	(-0,69)	-0,00444	(-0,22)
s_APOS	0,00544	(0,27)	0,00130	(0,07)
s_31.45	0,0406**	(3,03)	0,0263	(1,80)
s_APEN	-0,0505*	(-2,33)	-0,0676**	(-3,09)
REF_INT	-0,209***	(-6,58)	-0,218***	(-7,10)
s_REFAGE	-0,350***	(-5,90)	-0,358***	(-6,34)
s_REF	-0,399***	(-3,61)	-0,353***	(-3,50)
Observações	380		380	

Observ: Estatísticas *t* entre parentesis. * $p < 0,05$; ** $p < 0,01$; *** $p < 0,001$.

Relativamente aos modelos (3) e (4) toda a interpretação anterior se aplica. Apenas se levanta uma ressalva acerca da não significância dos termos s_DAY, s_ELO e s_APOS. Dada a transformação das variáveis e a forma funcional das parábolas, é natural a perda de significância. Ora, dado que uma parábola tem a forma geral $y = a + bx + cx^2$ — onde *b* determina um deslocamento horizontal da função — e que as variáveis foram subtraídas da sua média — tendo assim média nula —, o termo *b* será nulo — pois não é necessário qualquer deslocamento horizontal —, perdendo assim significância.

Para terminar, fica apenas uma nota relativamente a algumas variáveis que não foram incluídas no modelo, ainda que, à partida, pudessem ser tidas em conta como significativas. O primeiro caso são os autogolos, de facto, a

inclusão desta variável no modelo é sempre estatisticamente insignificante e a explicação é simples: tratando-se do estudo de *arrangements*, onde são os árbitros os atores da fraude, é natural que os autogolos não sejam importantes. No entanto, caso se tratasse de um cenário onde os agentes corrompidos fossem os jogadores, esta variável iria, provavelmente, assumir importância.

Para além disto, e à exceção da variável 31.45 incluída, nenhuma outra variável relacionada com número de golos é significativa. Para explicar isto podem existir várias razões. A primeira é imediata: para ganhar um jogo basta marcar um golo e se uma equipa corrompe um árbitro é porque, à partida, está à espera de um jogo difícil, logo, será normal se vencer apenas pela margem mínima. Isto podia levar-nos a pensar que, então, um baixo número de golos indicia fraude, porém, no futebol, não é muito comum existirem resultados avultados, pelo que a média de golos marcados no total de uma competição é relativamente reduzida.

>> IV. CONCLUSÕES

A análise estatística ao Caso Calciopoli permite estabelecer algumas conclusões que são relevantes na concetualização do modelo econométrico. Assim, verificámos que existe vantagem de jogar em casa¹, embora esta vantagem desapareça nos jogos viciados, o que indica que as equipas que pretendem viciar os resultados fazem-no, sobretudo, nos jogos fora. Esta é a razão pela qual no modelo econométrico são utilizadas as variáveis APOS em vez de APOS e HPOS; e de APEN em vez de PEN.

O número de cartões vermelhos não se revelou significativo no modelo econométrico, pelo que não foi incluído. Este facto vai de encontro às conclusões de Hill (2013). Na mesma obra, Hill mostra também que, ao contrário do que muitas vezes se pensa, não existe uma concentração anormal de golos marcados no final do jogo quando existe viciação. Esta é a razão pela qual a variável 31.45 foi incluída no modelo econométrico. De qualquer forma, é importante ressaltar que esta será a variável que tem a base teórica mais frágil.

Para além disto, é de notar que a variável OWNG não se revelou, como esperado, significativa. Uma vez que estamos na presença de viciação por parte de árbitros, nada indicaria que esta variável apresentasse algum comportamento de interesse.

Por fim, ainda relativamente à análise estatística, é de notar a dependência significativa entre a existência de fraude e as variáveis T, REF e APOS. Assim, estes três factos foram incluídos² na especificação do modelo econométrico, revelando-se, também, significativos e encontrando bases teóricas nos já referidos trabalhos de Boeri e Severgnini (2008) e de Nevill e Holder (1999).

Relativamente ao modelo econométrico, as conclusões relativas à forma funcional variável DAY são consistentes com os resultados de Boeri e Severgnini (2008). Já as conclusões relativas às variáveis APOS e ELO são muito semelhantes e, em termos gerais, significam que a probabilidade estimada de fraude é tanto maior quanto mais perto dos extremos for a classificação da equipa visitante. De forma equivalente, se a variável ELO estiver nos seus extremos – sendo, assim, um jogo entre equipas ou ambas fortes ou ambas fracas – a probabilidade estimada de fraude também será maior. Em suma, quer isto dizer que a probabilidade de fraude será maior nos casos

¹ Um fenómeno comum a bastantes desportos e alvo de diversos estudos. Por exemplo, Nevill e Holder (1999) tentam identificar os principais fatores que influenciam esta vantagem.

² De notar que a variável T foi substituída pela variável DAY no modelo econométrico.

em que as equipas presentes nos jogos estão a disputar algo decisivo (seja o título, o acesso às competições europeias ou a manutenção de divisão), enquanto que as equipas do meio da tabela – que não têm grandes ambições em termos internacionais, mas também não estão muito pressionadas pela descida de divisão – serão menos propensas a praticar fraude.

Já as conclusões acerca das variáveis APEN e 31.45 estão de acordo com os resultados de Hill (2013). Especialmente, o coeficiente negativo estimado pelo modelo para a variável APEN corrobora a ideia de que os árbitros tendem a utilizar outros instrumentos que não os cartões vermelhos e os penáltis para concretizar a viciação de resultados.

Finalmente, relativamente às características específicas dos árbitros, verificamos que os grupos de maior risco são os árbitros mais jovens e os árbitros em fim de carreira. Para além disto, e de acordo com o que Boeri e Severgnini (2008) também referem, são também os árbitros sem estatuto internacional que apresentam uma maior probabilidade estimada de fraude – devido ao coeficiente negativo estimado para a variável REF_INT. Já no que toca ao *ranking* dos árbitros – avaliado pelo número de jogos arbitrado na época – as conclusões são mais complexas (principalmente se considerarmos o modelo com interações, situação na qual não é possível fazer uma análise *ceteris paribus* à variável REF sem considerar as variáveis que com ela interagem), embora o principal ponto a reter seja precisamente o de serem os árbitros com um *ranking* que poderá abrir portas a futuras carreiras internacionais os principais visados pelos corruptores.

Concluindo, dado que os resultados modelo econométrico encontram eco na literatura, será adequado utilizar o modelo na sua versão *standardizada* para o estudo da viciação de resultados, por parte dos árbitros, em *arrangements* em trabalhos futuros, embora tendo sempre em atenção as limitações inerentes a modelos econométricos, no geral, e ainda mais a modelos que tentam modelar a fraude, em particular.

Referências bibliográficas

- Boeri, T., & Severgnini, B. (2008). The italian job: Match rigging, career concerns and media concentration in 'Serie A'.
- Boeri, T., & Severgnini, B. (2013). Match rigging in Italian professional soccer: the economic determinants of corruption *Match-fixing in international sports* (pp. 101-112): Springer.
- Brooks, G., Aleem, A., & Button, M. (2013). *Fraud, corruption and sport*: Springer.
- Cressey, D. R. (1953). Other people's money; a study of the social psychology of embezzlement.
- Forrest, D. (2013). Match fixing: An economics perspective *Match-Fixing in International Sports* (pp. 177-197): Springer.
- Greene, W. H. (2003). *Econometric analysis*: Pearson Education India.
- Hakeem, F. B. (2013). Sports-related crime: A game theory approach *Match-Fixing in International Sports* (pp. 247-260): Springer.
- Hill, D. (2011). *Mafia no futebol: Saída de Emergência*.
- Hill, D. (2013). *The Insider's Guide to Match-fixing in Football*: Anne McDermid & Associates Limited.
- Masters, A. (2015). Corruption in sport: From the playing field to the field of policy. *Policy and Society*, 34(2), 111-123.
- Nevill, A. M., & Holder, R. L. (1999). Home advantage in sport. *Sports Medicine*, 28(4), 221-236.

>> ANEXO 4

Tabelas de contingência de algumas variáveis relativamente à variável FIX

T	T1	T2	T3	T4
0	80	82	92	46
1	10	38	30	2

REF	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35
0	25	14	24	23	20	6	22	13	17	8	15	16	7	15	12	11	10	6	4	3	2	2	4	3	3	3	3	2	1	1	1	1	1	1	1
1	2	13	1	1	4	17	0	8	3	9	2	0	9	0	0	1	0	1	2	3	3	1	0	0	0	0	0	0	0	0	0	0	0	0	0

HOME TEAM	Atalanta	Bologna	Brescia	Cagliari	Chievo	Florentina	Inter	Juventus	Lazio	Lecce	Livorno	Messina	Milan	Palermo	Parma	Reggina	Roma	Sampdoria	Sienna	Udinese
0	14	17	17	16	16	15	14	15	14	16	17	12	15	16	16	14	14	17	11	14
1	5	2	2	3	3	4	5	4	5	3	2	7	4	3	3	5	5	2	8	5

AWAY TEAM	Atalanta	Bologna	Brescia	Cagliari	Chievo	Florentina	Inter	Juventus	Lazio	Lecce	Livorno	Messina	Milan	Palermo	Parma	Reggina	Roma	Sampdoria	Sienna	Udinese
0	15	14	14	17	16	13	16	9	18	15	18	13	13	17	14	16	15	14	15	18
1	4	5	5	2	3	6	3	10	1	4	1	6	6	2	5	3	4	5	4	1

HPOS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	14	16	15	11	16	14	15	18	18	11	15	16	18	12	11	16	18	13	17	16
1	4	3	4	7	3	4	5	1	3	4	4	6	0	7	5	2	4	5	4	5

AROS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	10	13	16	19	13	19	16	16	11	21	17	14	17	14	19	12	13	15	13	12
1	10	6	3	1	6	1	2	3	6	2	4	1	2	5	3	8	3	5	4	5