Universitat de Barcelona

Facultat de Matemàtiques i Informàtica

Doble grau en Matemàtiques i Administració i Direcció d'Empreses

Treball final de grau

Avaluació de l'equilibri competitiu en lligues de futbol i hoquei: anàlisi basada en l'índex de gols

Roger Rovira Solé

Directors: Dr. Carles Rovira
Dr. Josep Vives

Realitzat a: Departament de Matemàtiques i Informàtica
Departament de Matemàtica Econòmica, Financera i Actuarial

Barcelona, 17 de gener de 2024

Abstract

This paper deals from a statistical and analytical point of view with how the five big leagues of European football are distributed in terms of competitiveness.

This study is focused on the use of a specific index, based on the goals scored in a match. Based on this index, a comparison with other theoretical measures is undertaken to test the efficiency of this index, and then the influence of this index in explaining the revenues of the football leagues is tested.

Finally, the efficiency of this index is tested with a completely different sport such as hockey, and from the results obtained, conclusions are drawn.

Resum

Aquest treball tracta des d'un punt de vista estadístic i analític com les cinc grans lligues del futbol europeu estan distribuïdes a nivell de competitivitat esportiva.

L'estudi es centre en l'ús d'un índex específic, el qual està basat en els gols que es marquen en un partit. A partir d'aquest índex, s'aborda una comparació amb altres mesures teòriques de balanç competitiu per tal de provar l'eficiència d'aquest índex i, posteriorment, es prova la influència d'aquest índex a l'hora d'explicar els ingressos de les lligues de futbol.

Per últim, es prova l'eficiència d'aquest índex amb un altre esport completament diferent com és l'hoquei, i a partir dels resultats obtinguts es treuen unes conclusions.

Agraïments

Abans d'iniciar el meu treball de fi de grau, voldria expressar el meu reconeixement i agraïment a totes aquelles persones que, amb la seva col·laboració han contribuït a la realització d'aquest treball.

En primer lloc vull agraïr l'ajuda i la dedicació durant el desenvolupament del meu treball als meus tutors Carles Rovira i Josep Vives.

També m'agradaria agraïr als familiars i amics tot el suport que m'han donat al llarg d'aquests anys estudiant el doble grau.

2020 Mathematics Subject Classification. 62-00, 62P20, 62J12, 62Q05

Índex

1 Introducció

1.1 Context

Actualment, el futbol és l'esport amb més influència d'espectadors de tot el món. Pel que fa a ingressos, tenim que les 5 grans lligues del futbol europeu (La Liga, la Premier League, la Serie A, la Ligue 1 i la Bundesliga) denominades 'Big Five' van generar uns ingressos totals de 17,2 bilions de dòlars. Aquestes 5 grans lligues són les que tenen més influència i generen la majoria d'ingressos en el futbol europeu. És per aquest motiu que estudiarem aquestes 5 grans lligues durant les temporades 2009/10 fins a la 2018/19.
Durant aquest període presentarem diferents mesures que intenten explicar el balanç competitiu d'una lliga de futbol, i les compararem amb el nostre índex basat en els gols per veure quina d'elles és més adient en el context del futbol europeu.

Abans de començar a presentar mesures per avaluar aquest balanç competitiu, necessitem definir en què consisteix aquest terme.
Per a definir aquest concepte, el primer que hem d'entendre és que l'economia dels esports funciona molt diferent de l'economia tradicional que tots entenem. Aquest fet és degut principalment a la *Paradoxa de Louis-Schmelling.* Aquesta paradoxa va ser introduïda per Walter C. Neale [18], i el que diu és bàsicament que a qualsevol empresa tradicional li interessa aconseguir una situació monopolística (mercat on només hi ha una empresa i, per tant, no té cap competidor), però en el cas de l'economia dels esports és al contrari. Walter C. Neale explica en el seu article el cas d'un boxejador de pesos pesants.
Aquest boxejador per tal d'obtenir més ingressos el que necessita no és guanyar tots els combats a la primera ronda i sense cap dificultat, perquè això el que provocarà serà que la gent ja no estigui tan interessada a veure la competició perquè ja saben quin serà el resultat. El que necessita aquest boxejador és aconseguir tenir un rival o diversos rivals molt exigents per tal que els combats estiguin molt més disputats, perquè això atraurà més espectadors i conseqüentment més ingressos pel boxejador.

Aquesta paradoxa no només funciona en la boxa, sinó que funciona exactament igual en la indústria del futbol que és en la que ens centrarem en aquest treball.

El concepte de balanç competitiu és una mesura que intenta mostrar com estan distribuïts els diferents jugadors amb talent entre tots els equips de la lliga. Rottenberg (1956) [4], va dur a terme un estudi sobre aquest concepte, i va arribar a la conclusió que en funció de com estan distribuïts aquests jugadors, existeix un grau d'incertesa en els resultats dels partits *(Uncertainty* *of Outcome Hypothesis).* I va relacionar el grau d'incertesa del resultat dels partits amb l'atracció del públic, és a dir, si en un partit el resultat és molt predictible, aleshores el públic no tindrà gaires incentius per a veure el partit.

Una altra definició totalment acceptada sobre el balanç competitiu és quan tots els equips tenen la mateixa probabilitat de guanyar a l'inici de la temporada. És per aquest motiu que no hi ha una única definició adient per a explicar el balanç competitiu.

Per estudiar aquest balanç competitiu s'han dut a terme molts estudis teòrics per veure quins són els més adequats per explicar la teoria econòmica del futbol. En aquest treball en veurem alguns i veurem quins són els més adients.

Hi ha diferents formes de mesurar correctament el balanç competitiu d'una lliga de futbol. Ja sigui pel nivell de partits, pel tipus de campionat, o per la del nivell d'una temporada. Nosaltres ens centrarem en aquest darrer, el d'una temporada, perquè així estudiarem tots els partits de les 5 grans lligues de futbol i podrem obtenir una gran comparació de les dades.

1.2 Mesures balanç competitiu

1.2.1 Desviació estàndard

El primer que podem pensar és en utilitzar la desviació estàndard, la qual és una mesura que mostra la variació de les dades que volem analitzar amb la mitjana d'aquestes.

Podríem usar la desviació estàndard per a comparar els percentatges reals de victòria dels equips amb els percentatges ideals, on tots els equips tenen la mateixa probabilitat de guanyar un partit. Aquesta mesura va ser estudiada intensament al llarg dels anys per Quirk and Fort (1997)[19], però va acabar sent rebutjada per diversos motius.

Els motius més importants són que no té en compte el percentatge que hi pot haver en cas d'empat, el qual és bastant elevat al futbol (aproximadament un 25%). Tampoc és adient per comparar diferents lligues perquè no té en compte la durada d'aquesta, ni tampoc la concentració de la lliga, el qual fa referència als equips amb una dominància respecte als altres equips ja sigui perquè tenen millors jugadors, o perquè tenen més pressupost...

1.2.2 Ràtio de Concentració

Aquesta mesura està inspirada en el concepte que van formular Hall i Tideman l'any 1967 [20], la qual quantifica la grandària que tenen les empreses més influents en una indústria qualsevol.

Si ho volem extrapolar al nostre cas, que és una lliga de futbol, el que podem estudiar és quin percentatge de punts obtenen els millors equips respecte a la resta d'equips. Així podem tenir una visió de quan pot estar equilibrada o no una lliga. En el nostre estudi, el que farem serà estudiar la concentració que tenen els 6 primers classificats de la lliga, ja que en les 5 grans lligues, els 6 primers classificats juguen les competicions europees. Dit això, es presenta l'índex C6 a continuació.

Definició 1.1 (Índex C6)**.** *Sigui una lliga on n és el número d'equips, denotarem per P_i el número de punts obtinguts per l'i-èsim equip al final de la temporada. La ponderació dels punts de l'i-èsim equip es defineix com $P_i^* = P_i / \sum_{i=1}^{n} P_i$. Ara sigui $P_{(j)}^*$ el j-èsim valor més gran del conjunt $\{P_1^*, P_2^*, ..., P_n^*\}$. Aleshores l'índex C6 es defineix com*

$$C6 = \frac{n}{6} \sum_{j=1}^{6} P_{(j)}^* \tag{1.1}$$

Aquest Índex C6 quantifica el desequilibri que hi ha entre els sis primers equips i la resta.

Observació 1.2. Aquest índex augmenta quan la desigualtat dels 6 equips amb la resta és més gran, i disminueix en cas contrari.
Pren el valor de $\frac{n}{6}$ quan la desigualtat és total, és a dir, que els primers 6 equips han aconseguit tots els punts possibles i els altres $n - 6$ equips no han aconseguit cap punt (cas impossible).
I aquest índex pren el valor d'1 quan tots els equips aconsegueixen els mateixos punts (cas també molt improbable però possible).
La fórmula que acabem de veure de l'índex C6 està condicionada al nombre d'equips de la lliga, com també a la ràtio de concentració, és per aquest motiu que J. James i Manasis (2011) [17] presenten aquest índex normalitzat, el qual definirem de la següent manera.

Definició 1.3 (Índex C6 Normalitzat)**.** *Sigui una lliga de n equips, i assumirem que cada victòria val 3 punts. Sigui $P_{(j)}^*$ el j-ésim valor del conjunt $\{P_1, P_2, ..., P_n\}$, on P_i fa referència al valor que s'ha definit a la definició anterior. Aleshores l'índex C6 normalitzat és*

$$NC6 = \frac{1}{18(n-6)} \left[\sum_{j=1}^{6} P_{(j)} - 18(n-1) \right]. \tag{1.2}$$

1.2.3 Índex Herfindahl del balanç competitiu

L'índex d'Herfindahl és una mesura molt usada en economia, que informa sobre la concentració econòmica d'un mercat qualsevol.
Si el valor d'aquesta mesura és elevat, vol dir que el mercat està molt concentrat i, per tant, no és gaire competitiu. Per contra, si el valor és baix, el que ens indica és que el mercat està poc concentrat i, per tant, hi ha una alta competència entre les empreses que conformen el mercat.
Si ho apliquem en l'àmbit del futbol, el calcularem en funció de la ponderació dels punts obtinguts durant la temporada.

Definició 1.4 (Índex Herfindahl)**.** *Sigui n el nombre d'equips de la lliga i sigui $P^*_{(j)}$ la ponderació esmentada a la* (1.1)*. Aleshores l'IHBC es defineix com*

$$IHBC = n \sum_{i=1}^{n} P_i^{*^2}. \tag{1.3}$$

Observació 1.5. Si tots els equips de la lliga obtenen els mateixos punts, aquest valor en valdrà 1. Altrament, aquest valor anirà augmentant, i com més elevat sigui aquest valor, voldrà dir que hi haurà més desigualtats a la lliga. Aquest índex no superarà mai el valor de n, ja que si l'índex val n, voldria dir que un equip ha aconseguit tots els punts i els altres cap, fet que en una lliga de futbol és impossible que passi.

Aquesta mesura de l'índex d'Herfindahl també està condicionada al nombre d'equips que conformen la lliga.
Antonio Avila Cano en la seva tesi doctoral [13] va proposar que el màxim valor de l'índex es produeix quan una lliga es distribueix en dos grans grups, els quals li direm $r(n)$ i $n - r(n)$, i en aquests dos grups el que ha de succeir perquè el valor de l'índex sigui el més gran, és que els del grup $r(n)$ guanyin els seus partits en cascada (els equips de dalt en la posició guanyen als de baix), i també guanyin a tots els altres equips del grup $n - r(n)$. I pel que fa al grup $n - r(n)$, el que ha de passar és que empatin entre ells, i perdin contra el grup $r(n)$.
En aquest valor màxim li direm $m(n)$, i clarament depèn del nombre d'equips de la lliga. Per tant, tenint en compte aquest estudi d'Avila Cano, a continuació es presenta l'Índex d'Herfindahl normalitzat.

Definició 1.6 (Índex Herfindahl Normalitzat)**.** *Sigui n el nombre d'equips de la lliga i sigui IHBC el valor calculat com a la Eq.*(1.3)*. Aleshores, l'IHBC normalitzat es defineix com*

$$IHNBC = \left(m(n) - \frac{1}{n}\right)^{-1} \left[\frac{IHBC}{n} - \frac{1}{n}\right]. \tag{1.4}$$

1.2.4 Mesura Normalitzada de la dinàmica

Per últim, analitzarem el concepte de mesura normalitzada de la dinàmica la qual va ser estudiada en profunditat per Haan el 2007 [23], i va proposar la següent definició.

Definició 1.7. *Considerem una lliga de n equips. Sigui* $C_{i,1}$ *la posició de l'i-èssim equip en una particular temporada, i sigui* $C_{i,0}$ *la posició a la lliga del mateix equip però la temporada anterior, aleshores diem que la mesura normalitzada de la dinàmica de la lliga per a la temporada actual és:*

$$MND = \frac{2}{n^2} \sum_{i=1}^{n} |C_{i,1} - C_{i,0}|. \tag{1.5}$$

2 Base Teòrica

En aquesta secció explicarem quin serà el marc teòric en el qual treballarem, per tal així de poder definir posteriorment l'índex basat en els gols, i comparar-la amb les mesures definides anteriorment.

Definició 2.1. *Sigui n el número d'equips a una lliga de futbol qualsevol on tots els equips juguen entre ells tant a casa com a fora, aleshores el número de partits que es juga en una lliga li direm N i és de $n(n-1)$ partits.*

Definició 2.2. *Denotarem per P_{ij1} els punts obtinguts per l'i-èssim equip quan el partit es juga a casa de l'i-èssim equip contra l'equip j-èssim, i respectivament denotarem per P_{ij0} els punts obtinguts com a visitant.*

Definició 2.3. *Denotarem per X_{ij1} els gols marcats per l'i-èssim equip al camp local i per X_{ij0} els gols marcats com a visitant respectivament.*

Definició 2.4. *Amb les dues definicions anteriors, podem definir els punts (P_i), els gols anotats (GA_i), els gols concedits (GC_i) i la diferència de gols totals (DG_i) per l'i-èssim equip al llarg de la temporada de la manera següent:*

$$\begin{aligned} P_i &= \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{k=0}^{1} P_{ijk}, \quad GA_i = \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{k=0}^{1} X_{ijk}. \\ GC_i &= \sum_{\substack{j=1 \\ j\neq i}}^{n} \sum_{k=0}^{1} X_{jik}, \quad DG_i = GA_i - GC_i \end{aligned} \tag{2.1}$$

Definició 2.5 (Clasificació lliga de futbol)**.** *La classificació final de qualsevol lliga de futbol professional ve determinada primer pel valor de (P_i) definit anteriorment. En cas d'empat es mira la diferència de gols que hi ha hagut entre els equips empatats, i si segueix havent-hi empats, aleshores es mira la diferència de gols totals (DG_i).*

Definició 2.6 (Lliga perfectament equilibrada)**.** *Direm que una lliga de futbol està perfectament equilibrada si qualsevol permutació dels equips de la lliga té la mateixa probabilitat d'acabar sent la posició final a la classificació.*

Nota 1. *A partir d'ara, tal com s'exposa en l'article* [1] *suposarem en tot moment que tots els gols (X_{ijk}) són independents i segueixen una distribució de Poisson de paràmetre λ_{ijk} per tot $i, j \in \{1, 2, 3, ..., n\}, k \in \{1, 0\}$.*

Teorema 2.7. *Suposant que estem en la hipòtesis de la Nota 1. Si $\lambda_{ijk} = \lambda > 0$ per a tot $i, j \in \{1, 2, 3, ..., n\}, k \in \{1, 0\}$, aleshores existeix un λ_0 tal que tots els partits tenen les mateixes possibilitats d'acabar en victòria, empat o derrota per a qualsevol equip.*

Demostració. Considerem un partit on l'i-èssim equip juga a casa contra el j'èssim equip. Per estalviar-nos notació, li direm L_1 a X_{ij1} (gols marcats de l'equip local) i L_2 a X_{ji0} (gols marcats per l'equip visitant) respectivament. Primer, considerem la probabilitat que el partit acabi en empat, és a dir, $\mathbb{P}(L_1 = L_2|\lambda)$. Aquesta expressió ens està dient la probabilitat que es marquin els mateixos gols condicionats a un valor de λ (mitjana de gols esperats en el partit) prèviament fixat.
Usant que les variables són iid's i la funció de massa de probabilitat de la Poisson, tenim que

$$\begin{aligned}\mathbb{P}(L_1 = L_2|\lambda) &= \sum_{k=0}^{\infty} \mathbb{P}(L_1 = k|\lambda)\mathbb{P}(L_2 = k|\lambda) \\ &= \sum_{k=0}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k!)} \frac{e^{-\lambda}\lambda^k}{(k!)} = \sum_{k=0}^{\infty} \frac{e^{-2\lambda}\lambda^{2k}}{(k!)^2}\end{aligned} \tag{2.2}$$

Ara podem rescriure l'equació (2.2), fent servir la funció de Bessel modificada de primer ordre $I_\alpha(x)$ on α és l'ordre de la funció de Bessel modificada i x l'argument. Aquesta funció es defineix de la següent manera:

$$I_\alpha(x) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(k+\alpha+1)} \left(\frac{x}{2}\right)^{2k+\alpha} \tag{2.3}$$

on Γ és la funció gamma.
Si prenem aquesta funció com a ordre 0 i argument 2λ tenim que

$$I_0(2\lambda) = \sum_{k=0}^{\infty} \frac{1}{k!\Gamma(k+1)} \left(\frac{2\lambda}{2}\right)^{2k} = \sum_{k=0}^{\infty} \frac{1}{(k!)^2} \lambda^{2k} \tag{2.4}$$

Aleshores usant (2.4) tenim que $\mathbb{P}(L_1 = L_2|\lambda) = e^{-2\lambda} I_0(2\lambda)$.
Quan $\lambda \to 0$, si fem el desenvolupament de la sèrie de Taylor de la funció de Bessel modificada d'ordre 0 i argument 2λ al voltant de 0, tenim la següent expressió:

$$I_0(2\lambda) \approx 1 + \lambda^2 + \frac{\lambda^4}{4^2} + \frac{\lambda^6}{6^2} + \frac{\lambda^8}{24^2} + O(\lambda^9) \tag{2.5}$$

Per tant, tenim que el límit que volem calcular serà el següent:

$$\lim_{\lambda \to 0} e^{-2\lambda} I_0(2\lambda) = \lim_{\lambda \to 0} \frac{1}{\Gamma(1)} = \lim_{\lambda \to 0} \frac{1}{1} = 1 \tag{2.6}$$

Ja que aquí hem usat la sèrie de Taylor de (2.5), on tots els termes són 0 excepte quan la $k = 0$, el qual el seu valor és 1.
Per tant, tenim que $\mathbb{P}(L_1 = L_2|\lambda) \to 1$, quan $\lambda \to 0$.
Ara, quan $\lambda \to \infty$, tenim que $I_0(2\lambda)$ és aproximadament

$$I_0(2\lambda) = \frac{e^{2\lambda}}{\sqrt{4\pi\lambda}} \left[1 + \frac{1}{16\lambda}\left(1 + \frac{9}{32\lambda}\left(1 + \frac{25}{48\lambda}(1 + ...)\right)\right)\right] \tag{2.7}$$

Aquesta aproximació es pot trobar a [15].
Per tant, usant aquesta aproximació tenim que $\mathbb{P}(L_1 = L_2|\lambda) \to 0$ quan $\lambda \to \infty$, ja que $\lim_{\lambda\to\infty} \frac{1}{\sqrt{4\pi\lambda}} = 0$.
Aleshores per continuïtat, tenim que existeix un λ_0, tal que la probabilitat d'empat és 1/3. Per últim, tenim que les variables són independents i idènticament distribuïdes, per tant, tenim que $\mathbb{P}(L_1 < L_2|\lambda_0) = \mathbb{P}(L_1 > L_2|\lambda_0) = 1/3$, on hem acabat la demostració.

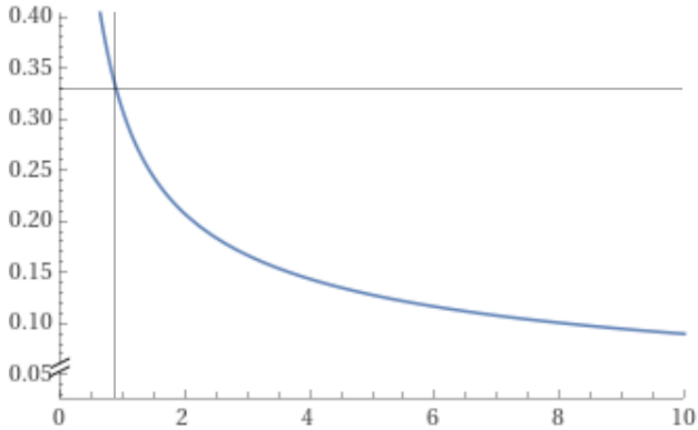
Corol·lari 2.8. *Si les variables X_{ijk} són iid Poisson de paràmetre λ_0 per a tot $i, j \in \{1, 2, 3, ..., n\}, k \in \{1, 0\}$, aleshores una lliga de futbol està equilibrada perfectament.*

Demostració. La demostració és simple usant el teorema demostrat anteriorment. Tenim que cada partit de la lliga segueix una distribució multinomial i tots els partits són iid's, amb la probabilitat que cada resultat possible és 1/3 (és a dir empat, victòria o derrota).
També sabem que el nombre de gols marcats i concedits per a cada equip són variables iid Poisson amb mitjana $2\lambda_0(n-1)$. Aquesta última afirmació ens assegura que en el supòsit que dos equips acaben amb els mateixos punts, també són equiprobables de tenir una diferència de gols positiva, més gols marcats i millor enfrontament cara a cara.

Observació 2.9. Per a calcular aquest valor de λ_0, el que farem és utilitzar el software *Wolfram Alpha* per a donar el valor que estem buscant de forma aproximada. Usarem aquest programa i li demanarem que ens resolgui l'equació $e^{-2\lambda} I_0(2\lambda) = \frac{1}{3}$ per tal de veure quin és el valor aproximat de λ_0 que estem buscant.
Escrivint el següent codi:

Solve[e(̂-2*lambda)*BesselI[0, 2*lambda] == 1/3, lambda]

el que obtenim és que el valor de $\lambda_0 \approx 0,88029$.
I també podem veure en la imatge que es mostra just a continuació com varia la probabilitat d'empat en funció del paràmetre λ.



Per tant, el que observem amb aquest gràfic és que a mesura que augmenta la λ, la probabilitat que hi hagi empat és cada vegada més baixa, i pel contrari, si la λ és propera a 0, la probabilitat d'empat serà més alta.

Ara, amb el següent Teorema, veurem que λ_0 no és l'únic valor en el qual fa que una lliga estigui perfectament equilibrada.

Teorema 2.10. *Tornem a estar sota les hipòtesis de la Nota 1 amb $\lambda_{ijk} = \lambda > 0$ per a tot $i, j \in \{1, 2, 3, ..., n\}, k \in \{1, 0\}$. Si els partits guanyats valen 3 punts, els perduts 0 i els empatats 1, aleshores una lliga de futbol està perfectament equilibrada.*

Demostració. Denotem per $J_{ij} = X_{ij1} - X_{ji0}$ la diferència dels gols marcats en un partit de la lliga qualsevol. Si $J_{ij} < 0$ tenim que el j-èssim equip ha guanyat el partit, en canvi si $J_{ij} > 0$ ens diu que l'i-èssim equip ha guanyat el partit, i per últim si $J_{ij} = 0$, vol dir que el partit ha acabat en empat.
Com que totes les variables X_{ijk} són iid's amb distribució de Poisson de paràmetre λ, les variables J_{ij} segueixen una distribució de Skellam amb paràmetres (λ,λ). Aquesta distribució té la funció de massa de probabilitat següent:

$$\mathbb{P}(J_{ij} = k) = e^{-2\lambda} \left(\frac{\lambda}{\lambda}\right)^{\frac{k}{2}} I_k(2\sqrt{\lambda\lambda}), \tag{2.8}$$

on $I_k(2\sqrt{\lambda\lambda})$ és la funció de Bessel modificada de primer ordre.
Aquesta funció de massa de probabilitat va ser introduïda per Irwin el 1937 [9], i es pot veure d'on surt en l'article, com també que si k és un nombre enter tenim que $I_k(z) = I_{|k|}(z)$.
Amb el mateix procediment que hem fet a la demostració del Teorema 2.7 la funció de massa de probabilitat de la J_{ij} es pot reescriure com

$$\mathbb{P}(J_{ij} = k) = e^{-2\lambda} I_{|k|}(2\lambda) \tag{2.9}$$

de la igualtat de (2.9) veiem que la distribució de la J_{ij} és simètrica, ja que $\mathbb{P}(J_{ij} = k) = \mathbb{P}(J_{ij} = -k)$.
Denotarem a la probabilitat d'empatar en un partit per a qualsevol equip $\big(\mathbb{P}(J_{ij} = 0) = e^{-2\lambda} I_0(2\lambda)\big)$ com a d_λ. I a la probabilitat de guanyar per a qualsevol equip en un partit serà doncs de $(1 - d_\lambda)/2$ que li direm g_λ.
Amb aquesta notació i recordant la de la Definició 2.4, tenim que la funció de massa de probabilitat de P_{ijk} (els punts obtinguts) és:

$$\begin{aligned} &\mathbb{P}(P_{ijk} = 0) = g_\lambda, \quad \mathbb{P}(P_{ijk} = 3) = g_\lambda, \\ &\mathbb{P}(P_{ijk} = 1) = d_\lambda \quad i \quad \mathbb{P}(P_{ijk} = t) = 0, t \neq 0, 1, 3 \end{aligned} \tag{2.10}$$

Sense pèrdua de generalitat, considerem dos equips de la lliga qualsevol, i definim com a S_1, S_2 el nombre total de punts aconseguits al final de la lliga

de l'equip 1 i l'equip 2 respectivament. Més concretament podem reescriure aquests S_1 i S_2 com $S_1 = S_1^{'} + S_{12}$, on $S_1^{'}$ són el nombre de punts obtinguts jugant contra tots els equips menys l'equip 2, i $S_{12} = S_{121} + S_{120}$, és a dir els punts obtinguts després de jugar contra l'equip 2 tant a casa com de visitant.
Fent el mateix amb l'equip 2, tenim $S_2 = S_2^{'} + S_{21}$.
Donats aquests supòsits tenim que
$\mathbb{P}(S_{12} = x, S_{21} = y)$

$$= \begin{cases} g_\lambda^2 & \text{si } x = 6, y = 0 \text{ } o \text{ } x = 0, y = 6 \\ 2g_\lambda d_\lambda & \text{si } x = 4, y = 1 \text{ } o \text{ } x = 1, b = 4 \\ 2g_\lambda^2 & \text{si } x = 3, y = 3 \\ d_\lambda^2 & \text{si } x = 2, y = 2 \\ 0 & \text{en cas contrari} \end{cases}$$

D'aquí podem veure que $\mathbb{P}(S_{12} > S_{21}) = \mathbb{P}(S_{12} < S_{21})$.
A continuació denotem el conjunt (S_{12}, S_{21}) com B $\subset A \times A$, on

$$\text{A} = \{0, 1, 2, 3, 4, 6\}.$$

Aleshores,

$$\begin{aligned} \mathbb{P}(S_1 > S_2) &= \mathbb{P}(S_1^{'} + S_{12} > S_2^{'} + S_{21}) = \mathbb{P}(S_1^{'} - S_2^{'} > S_{21} - S_{12}) \\ &= \sum_{(x,y)\in B} \mathbb{P}(S_1^{'} - S_2^{'} > y - x)\mathbb{P}(S_{12} = x, S_{21} = y) \end{aligned} \tag{2.11}$$

Les variables aleatòries $S_1^{'}$ i $S_2^{'}$ són iid's degut a la naturalesa de que cada partit és iid. Per tant, la $\mathbb{P}(S_1^{'} - S_2^{'} > -t) = \mathbb{P}(S_2^{'} - S_1^{'} > -t) = 1 - \mathbb{P}(S_1^{'} - S_2^{'} \leq t)$.
Aleshores, $\mathbb{P}(S_1^{'} - S_2^{'} > -t) + \mathbb{P}(S_1^{'} - S_2^{'} > t) = 1 - \mathbb{P}(S_1^{'} - S_2^{'} = t)$.
Ara calcularem el valor de l'equació (2.11) amb el que acabem de veure i tenim:

$$\begin{aligned} \mathbb{P}(S_1 > S_2) = g_\lambda^2(1 - \mathbb{P}(S_1^{'} > S_2^{'} = 6)) \\ + 2g_\lambda d_\lambda(1 - \mathbb{P}(S_1^{'} - S_2^{'} = 3)) \\ + (2g_\lambda^2 + d_\lambda^2)\mathbb{P}(S_1^{'} = S_2^{'}). \end{aligned} \tag{2.12}$$

Això implica clarament que $\mathbb{P}(S_1 > S_2) = \mathbb{P}(S_2 > S_1)$ ja que les variables $S_1^{'}$ i $S_2^{'}$ són iid's.

Per tant, hem vist que tots els equips tenen la mateixa probabilitat d'acabar amb els mateixos punts. Ara només ens falta veure si també es compleix amb els gols marcats i els gols concedits.

Siguin GA_1, GA_2, DG_1 i DG_2, les mateixes definicions de l'equació (2.1). Com abans, definim $GA_1 = GA_1^{'} + GA_{12}$ i $DG_1 = DG_1^{'} + DG_{12}$ i amb l'equip dos fem exactament el mateix.
Com que els gols anotats en un partit qualsevol són variables aleatòries iid amb paràmetre λ, aleshores les variables $GA_1^{'}$ i $GA_2^{'}$ són iid's, i segueixen una distribució de Poisson$\Big((2n-4)\lambda\Big)$ on aquest valor $(2n-4)$ és el número de partits que juga cada equip tal com està definit $GA_1^{'}$ ja que aquí no estem mirant els dos partits que juga l'equip 1 contra el 2.
Amb un raonament semblant, tenim que la distribució de GA_{12} i GA_{21} és una Poisson(2λ), i per tant les variables (diferència de gols) seguiran una distribució de Skellam de paràmetres (2λ,2λ).
Sabem que $DG_{21} = -DG_{12}$, ja que és la diferència de gols entre els dos partits que juguen l'equip 1 contra el 2, per tant,

$$\begin{aligned}\mathbb{P}(DG_1 > DG_2) &= \mathbb{P}(DG_1^{'} + DG_{12} > DG_2^{'} + DG_{21}) = \\ &= \mathbb{P}(DG_1^{'} - DG_2^{'} > DG_{21} - DG_{12}) = \mathbb{P}(R + 2DG_{12} > 0),\end{aligned} \tag{2.13}$$

on $R = DG_1^{'} - DG_2^{'}$ i aquesta variable R segueix una distribució de Skellam amb paràmetres $\Big((2n-4)\lambda, (2n-4)\lambda\Big)$.
Si ens fixem en les variables R i DG_{12}, veiem que totes dues són simètriques al voltant del 0, ja que són dues variables que segueixen una distribució de Skellam amb els dos paràmetres iguals respectivament, per tant, són simètriques al voltant del 0.
Per tant, tindrem que $\mathbb{P}(DG_1 > DG_2) = \mathbb{P}(DG_2 > DG_1)$. De la mateixa manera es calcula la probabilitat dels gols anotats, i arribem a la conclusió que $\mathbb{P}(GA_1 > GA_2) = \mathbb{P}(GA_2 > GA_1)$.
En resum, acabem de demostrar que la posició final dels dos equips és equiprobable que sigui (1,2) com (2,1), si aquesta posició està determinada en ordre per (S_1, S_2), (DG_1, DG_2), seguidament per (GA_1, GA_2) i per últim (S_{12}, S_{21}).

Corol·lari 2.11. *Si la distribució de probabilitat de (X_{ij1}, X_{ij0}) és tal que les distribucions de $J_{ij} = X_{ij1} - X_{ij0}$ (per tot i,j) són iid's aleshores una lliga de futbol està equilibrada perfectament.*

Demostració. En aquest cas no ho demostrarem, perquè es prova de manera similar al que hem fet servir a la demostració anterior, i no és el tema principal que volem estudiar.

3 Índex Basat en els gols

Abans de proposar el nostre índex basat en els gols, calcularem l'estimador de màxima versemblança, que serà clau per a poder definir el nostre índex.

Proposició 3.1. *Sota el supòsit que totes les variables X_{ijk} són iid's amb Poisson(λ) aleshores l'estimador de màxima versemblança és:*

$$\hat{\lambda} = \frac{1}{2N} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{k=0}^{1} X_{ijk}. \tag{3.1}$$

Demostració. Com que les variables són idd's amb Poisson(λ), aleshores podem calcular la funció de versemblança, que serà:

$$L(X_{ijk}|\lambda_{ijk} = \lambda; i, j \in \{1, 2, 3, ..., n\}, k \in \{1, 0\}) = \prod_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} \prod_{k=0}^{1} \frac{e^{-\lambda_{ijk}} \lambda_{ijk}^{X_{ijk}}}{X_{ijk}!}$$

$$= \prod_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} \prod_{k=0}^{1} \frac{e^{-\lambda} \lambda^{X_{ijk}}}{X_{ijk}!} = \frac{e^{-2N\lambda} \lambda^{\sum X_{ijk}}}{\prod X_{ijk}!} \tag{3.2}$$

Aleshores hem de maximitzar aquesta funció, que serà equivalent a maximitzar el numerador. Sigui Q=$e^{-2N\lambda} \lambda^{\sum X_{ijk}}$ el numerador de l'equació (3.2), aleshores prenent logaritmes tenim: $\ln Q = -2N\lambda + \sum X_{ijk} \ln \lambda$
Ara, si derivem aquesta funció tenim:

$$\frac{\partial \ln Q}{\partial \lambda} = -2N + \frac{\sum X_{ijk}}{\lambda} = 0 \Rightarrow \hat{\lambda}(x) = \frac{\sum X_{ijk}}{2N} \tag{3.3}$$

i és un màxim, ja que

$$\frac{\partial^2 \ln Q}{\partial \lambda^2} = -\frac{\sum X_{ijk}}{\lambda^2} < 0 \tag{3.4}$$

Per tant, tenim que $\hat{\lambda}$ correspon a la mitjana de gols marcats en una lliga de futbol.

En una distribució Poisson de paràmetre λ sabem que l'esperança matemàtica és igual a la variància, i prenen el valor de λ. En la Nota 1 hem suposat que les variables X_{ijk} eren independents, però no idènticament distribuïdes, ja que el valor del paràmetre λ_{ijk} pot dependre de diferents factors com ara el nivell dels jugadors en cada equip, el nombre d'aficionats que assisteixen a un camp, les condicions de la gespa, i d'aquests factors en trobaríem molts més.

Ara bé, si suposem que aquestes variables X_{ijk} són iid's aleshores les desviacions dels gols marcats (X_{ijk}) respecte de $\hat{\lambda}$ seran les que determinaran si una lliga difereix d'estar perfectament equilibrada o no. És per aquest motiu que a continuació plantegem el següent índex basat en els gols.

Definició 3.2 (Índex basat amb els gols). *Sigui X_{ij1} i X_{ij0} els gols tal com els hem definit a la Definició 2.3. Aleshores el nostre índex basat en els gols és el següent:*

$$IBG = \frac{1}{(2N-1)\hat{\lambda}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{k=0}^{1} (X_{ijk} - \hat{\lambda})^2. \tag{3.5}$$

Teorema 3.3. *Suposem que les variables aleatòries X_{ijk} són independents i distribuïdes segons lleis de Poisson. Aleshores si estem en els supòsits del Teorema 2.10, tenim que $(2N-1)IBG$ és asimptòticament equivalent a una variable aleatòria khi quadrat amb $2N-1$ graus de llibertat.*

Demostració. Tenim que les variables X_{ijk} són independents i segueixen una distribució de Poisson de paràmetre λ_{ijk} per a tot $i, j \in \{1, 2, 3, ..., n\}, k \in \{1, 0\}$. Ara construirem una prova de raó de versemblança de la manera següent:

$$H_0 : \lambda_{ijk} = \lambda$$
$$H_1 : \lambda_{ijk} \neq \lambda$$

Aquest test ens diu que la hipòtesi nul·la és un λ fixat per a tot $i, j \in \{1, 2, 3, ..., n\}, k \in \{1, 0\}$, mentre que la hipòtesi alternativa ens diu que les λ no són totes iguals.

La funció de versemblança de les X_{ijk} s'escriu de la següent manera:

$$L(X_{ijk}|\lambda_{ijk}) = \prod_{i=1}^{n} \prod_{\substack{j=1 \\ j \neq i}}^{n} \prod_{k=0}^{1} \frac{e^{-\lambda_{ijk}} \lambda_{ijk}^{X_{ijk}}}{X_{ijk}!} \tag{3.6}$$

Per a calcular la raó de versemblança d'aquest test, necessitem trobar l'estimador de màxima versemblança de les λ_{ijk} sota H_0, i també sota $H_0 \cup H_1$. Per a les λ_{ijk} sota H_0, tenim que l'EMV és el que hem demostrat a la Proposició 3.1, ja que estem en les mateixes condicions. Per tant, l'EMV és:

$$\hat{\lambda} = \frac{1}{2N} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{k=0}^{1} X_{ijk}. \tag{3.7}$$

Pel que fa a l'EMV de les λ_{ijk} sota $H_0 \cup H_1$, com que els paràmetres són diferents ho calcularem en funció de cada λ_{ijk}, de la següent forma:

Considerem el log de la funció de versemblança i ens centrem una altra vegada només amb el numerador, ja que és el que ens interessa maximitzar per tal de calcular la raó de versemblança. Dit això, l'expressió que ens queda un cop aplicant el logaritme al numerador de la funció de versemblança és la següent

$$\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{k=0}^{1}(-\lambda_{ijk}+X_{ijk}\log\lambda_{ijk}) \tag{3.8}$$

Ara si derivem aquesta expressió respecte al λ_{ijk} tenim que ens queda

$$-1+\frac{X_{ijk}}{\lambda_{ijk}} \tag{3.9}$$

i un cop ho igualem a 0, tenim que l'EMV sota $H_0 \cup H_1$ és X_{ijk}. I és un màxim perquè la segona derivada torna a ser negativa. Per tant, tenim que la raó de versemblança és:

$$\Lambda(x)=\prod_{i=1}^{n}\prod_{\substack{j=1\\j\neq i}}^{n}\prod_{k=0}^{1}\frac{e^{-\hat{\lambda}}\hat{\lambda}^{X_{ijk}}}{e^{-X_{ijk}}X_{ijk}^{X_{ijk}}}=\prod_{i=1}^{n}\prod_{\substack{j=1\\j\neq i}}^{n}\prod_{k=0}^{1}\frac{e^{X_{ijk}-\hat{\lambda}}\hat{\lambda}^{X_{ijk}}}{X_{ijk}^{X_{ijk}}} \tag{3.10}$$

Ara, si apliquem la teoria del comportament asimptòtic de la prova de versemblança, tenim que $-2\log\Lambda$ s'aproxima a una khi quadrat amb $2N-1$ graus de llibertat. Per tant hem de provar que $(2N-1)IBG$ és asimptòticament equivalent a $-2\log\Lambda$, i ja hauríem acabat la demostració.
Anem a calcular doncs aquest logaritme.

$$\begin{aligned}-2\log\Lambda &= -2\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{k=0}^{1}\left[(X_{ijk}-\hat{\lambda})+X_{ijk}\log\left(\frac{\hat{\lambda}}{X_{ijk}}\right)\right]=\\ &=2\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{k=0}^{1}X_{ijk}\log\left(\frac{X_{ijk}}{\hat{\lambda}}\right).\end{aligned} \tag{3.11}$$

ja que $\sum_{i=1}^{n}\sum_{\substack{j=1\\j\neq i}}^{n}\sum_{k=0}^{1}(X_{ijk}-\hat{\lambda})=0$.
Ara, considerem la funció $f(x)=x\log(x/\lambda)$ per a una λ fixada, i la sèrie de Taylor d'aquesta funció al voltant de λ serà la següent:

$$f(x)=f(\lambda)+f'(\lambda)(x-\lambda)+\frac{f''(\lambda)}{2!}(x-\lambda)^2+\frac{f'''(\lambda)}{3!}(x-\lambda)^3+\cdots \tag{3.12}$$

Si fem els càlculs tenim que $f'(x)=\log\left(\frac{x}{\lambda}\right)+1$ i, $f''(x)=\frac{1}{x}$ per tant, la sèrie de Taylor és

$$f(x)=(x-\lambda)+\frac{1}{2\lambda}(x-\lambda)^2+\epsilon, \tag{3.13}$$

on ϵ és negligible sota H_0 i per a una N prou gran, concretament $N = O(n^2)$ Per tant, si ho substituïm al resultat de l'Equació (3.11), tenim que

$$-2\log\Lambda = 2\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\sum_{k=0}^{1}\left[(X_{ijk}-\hat{\lambda}) + \frac{(X_{ijk}-\hat{\lambda})^2}{2\hat{\lambda}}\right] \tag{3.14}$$

I com que el primer terme és 0, veiem clarament que $-2\log\Lambda$ s'aproxima asimptòticament a $(2N-1)IBG$.

3.1 Altres índexs basats en els gols

En aquesta secció definirem dos índexs diferents també basats en els gols per tal de mostrar que se n'han proposat més, [21] i [22] respectivament. Però no els estudiarem en aquest treball. Suposem en els dos casos que les variables X_{ijk} tenen el mateix significat que en la Definició 2.3

- **Bivariant de Poisson-lognormal**
  Suposem que tenim (X_{ij1}, X_{ij0}) que segueixen una distribució bivariant de Poisson-lognormal amb paràmetres μ_{ij} i θ_{ij}.
  Considerem la prova de raó de versemblança següent:

  $$H_0 : \mu_{ijk} = \mu, \quad \theta_{ij} = \theta \text{ per a tot } i, j$$

  contra la hipòtesi alternativa que H_0 no és certa.
  Sigui BPL la raó de versemblança de la prova anterior, on li associem una distribució nul·la de $-2\log BPL \sim \chi^2_b$, aleshores l'índex basat en els gols d'aquesta distribució bivariant és:

  $$IBGBP = \frac{-2\log BPL}{b} \tag{3.15}$$

- **Distribució de Skellam**
  Suposem que tenim $(X_{ij1} - X_{ij0})$ que segueixen una distribució de Skellam amb paràmetres (λ_i, λ_j).
  Considerem la prova de raó de versemblança següent:

  $$H_0 : \lambda_i = \lambda_p, \quad \lambda_j = \lambda_k \text{ per a tot } i, j$$

  contra la hipòtesi alternativa que H_0 no sigui certa.
  Sigui DS la raó de versemblança de la prova que acabem de definir, tal que sota H_0, $-2\log DS \sim \chi^2_s$. Aleshores l'índex basat en la diferència dels gols és:

  $$IDG = \frac{-2\log DS}{s}. \tag{3.16}$$

Observació 3.4. Les variables definides en els dos índexs anteriors són iid's sota la hipòtesis nul·la, i per tant pel Corol·lari 2.11, podem afirmar que en els dos casos estem en la noció de lliga perfectament equilibrada .

4 Aplicació i Càlculs dels mètodes

En tota aquesta secció calcularem i analitzarem els diferents tipus de mètodes definits anteriorment i veurem quin d'ells és el més adequat per tal de veure si una lliga està perfectament equilibrada o no.
Tots els càlculs que farem involucraran les 5 grans lligues de futbol que com bé sabem són la Premier League, La Liga, la Bundeslliga, la sèrie A, i la Ligue 1. Analitzarem un total de deu temporades, perquè així podrem obtenir suficient informació per a poder treure conclusions, i les temporades que estudiarem seran de la 2009/10 fins a la 2018/19.

4.1 Càlculs Mitjana gols i probabilitats empat

En el context teòric explicat a les seccions anteriors, el primer que farem serà calcular la mitjana de gols marcats en una lliga ($\hat{\lambda}$) per a cada una de les temporades i per a les 5 lligues. El càlcul de $\hat{\lambda}$ fent servir les dades que ens proporciona la pàgina web *Datahub* són molt senzills, ja que ens ofereix tota la informació que necessitem, que no són més que tots els gols marcats en una temporada.
Seguidament calcularem la probabilitat teòrica d'empat, i buscarem a qualsevol pàgina web on hi hagi la classificació d'aquestes lligues, quina ha sigut la proporció real d'aquests empats per tal de veure si difereixen molt o no de la nostra probabilitat teòrica.
A la probabilitat teòrica d'empat la denotarem $d_{\hat{\lambda}}$ i a la probabilitat real d'empat la denotarem $\hat{E}$.
Per a calcular el $d_{\hat{\lambda}}$ farem servir l'expressió que hem vist a la demostració del Teorema 2.7, tenint en compte ara que la Poisson és de paràmetre $\hat{\lambda}$.
Per tant, l'expressió que hem de calcular és la següent: $\mathbb{P}(L_1 = L_2|\hat{\lambda}) = e^{-2\hat{\lambda}} I_0(2\hat{\lambda})$
Aquests càlculs els he fet fent servir l'aplicació *wolframalpha*, per tal d'evitar càlculs molt carregosos. Ho he fet usant la funció $BesselI[0, 2 * \hat{\lambda}]$, la qual és la funció de Bessel modificada d'ordre 0, que és la que nosaltres volem calcular.
Amb aquests dos càlculs i mirant per a cada lliga quina ha sigut la probabilitat real d'empat, a la taula 1 que es mostra a continuació es poden veure quins han sigut tots els resultats.

Lligues	Var.	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015-16	2016-17	2017-18	2018-19
Serie A	$\hat{\lambda}$	1,31	1,26	1,28	1,32	1,36	1,35	1.29	1.48	1.34	1.34
	$d_{\hat{\lambda}}$	0,26	0,27	0,27	0,26	0,26	0,26	0.27	0.25	0.26	0.26
	$\hat{E}$	0,27	0,26	0,29	0,25	0,24	0,32	0.25	0.21	0.22	0.28
La Liga	$\hat{\lambda}$	1,36	1,37	1,38	1,44	1,38	1,33	1.37	1.47	1.35	1.29
	$d_{\hat{\lambda}}$	0,26	0,26	0,25	0,25	0,26	0,26	0.26	0.25	0.26	0.26
	$\hat{E}$	0,25	0,21	0,25	0,22	0,23	0,24	0.24	0.23	0.23	0.29
Bundeslliga	$\hat{\lambda}$	1,42	1,46	1,43	1,47	1,58	1,38	1.42	1.43	1.40	1.59
	$d_{\hat{\lambda}}$	0,25	0,25	0,25	0,25	0,24	0,26	0.25	0.25	0.25	0.24
	$\hat{E}$	0,28	0,21	0,26	0,26	0,21	0,27	0.23	0.24	0.27	0.24
Premier League	$\hat{\lambda}$	1,39	1,40	1,40	1,40	1,38	1,28	1.35	1.40	1.34	1.41
	$d_{\hat{\lambda}}$	0,25	0,25	0,25	0,25	0,25	0,27	0.26	0.25	0.26	0.25
	$\hat{E}$	0,25	0,29	0,24	0,28	0,20	0,24	0.28	0.22	0.26	0.19
Ligue 1	$\hat{\lambda}$	1,21	1,17	1,26	1,27	1,23	1,25	1.26	1.31	1.36	1.28
	$d_{\hat{\lambda}}$	0,28	0,28	0,27	0,27	0,27	0,27	0.27	0.26	0.26	0.27
	$\hat{E}$	0,26	0,34	0,28	0,28	0,28	0,23	0.28	0.25	0.25	0.29

Taula 1: Resultats mitjana de gols i probab.empats. Elaboració pròpia

Si analitzem la taula que acabem de veure, podem observar que els resultats entre la probabilitat teòrica i la probabilitat real són força similars. De fet, si mirem les desviacions d'un i altre amb una precisió de $\pm 0,05$ el que obtenim és que només la diferència es supera o és igual en 5 ocasions les quals són:

- Premier League
  - Temporada 2013/14 diferència de 0,05
  - Temporada 2018/19 diferència de 0,06
- La Liga
  - Temporada 2010/11 diferència de 0,05
- Ligue 1
  - Temporada 2010/11 diferència de 0,06
- Serie A
  - Temporada 2014/15 diferència de 0,06

Cal destacar, que en els càlculs de la probabilitat teòrica, només estem tenint en compte la mitjana de gols marcats en un partit. I com bé sabem, en un partit de futbol intervenen molt més factors que podrien explicar les diferències d'aquestes probabilitats teòriques amb les probabilitats reals d'empat. Alguns exemples podrien ser l'estat d'ànim dels jugadors, si s'ha expulsat a algun jugador durant el partit, quants aficionats tant de l'equip local com del visitant han anat a l'estadi i d'aquests, molts més factors on la majoria d'ells són difícils de quantificar.

4.2 Càlculs Mesures Balanç Competitiu

A continuació, calcularem el nostre índex basat en els gols com també les altres mesures definides a la primera secció i veurem el seu comportament al llarg de les deu temporades per a les 5 grans lligues.

Per a calcular el nostre **IBG**, no es detallen tots els càlculs, perquè amb les dades de la pàgina web *DataHub* [6] i amb l'ajut de l'*Excel* els càlculs són molt senzills.

Només cal destacar per tal d'evitar errors en els càlculs que el número d'equips en les 5 lligues és de 20 equips, excepte la Bundeslliga, on el número d'equips és de 18. Per tant, la N serà 306 per a la Bundeslliga i 380 per a les altres lligues.

Dit això, a continuació es mostra a la taula l'IBG de les 5 grans lligues en els 10 anys que estem estudiant.

Més endavant calcularem la prova de raó de versemblança, sota el supòsit teòric de la secció 2, i veurem si en alguna temporada de les 5 grans lligues, obtenim que hi ha hagut balanç competitiu.

IBG 5 Lligues	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
La Liga	1,10	1,21	1,30	1,17	1,30	1,34	1,29	1,20	1,25	1,03
Bundeslliga	1,14	1,14	1,19	1,13	1,12	1,18	1,12	1,17	1,08	1,23
Premier League	1,30	1,03	1,18	1,11	1,24	1,10	1,09	1,17	1,21	1,12
Ligue 1	1,11	0,97	0,97	1,05	0,99	1,14	1,20	1,23	1,15	1,11
Serie A	0,93	1,05	1,12	1,09	1,08	1,01	1,05	1,13	1,18	1,05

Taula 2: Resultats dels IBG. Elaboració pròpia

Seguidament, calcularem els valors de les mesures (NC6, NHICB i ND).

1. **NC6**

   Els càlculs d'aquest índex també són molt senzills un cop es tenen totes les dades. Per aquest motiu no es mostra amb detall com s'obtenen. No obstant, s'ha de tenir en compte que només es sumen els punts dels 6 primers equips de cada lliga.

   Els resultats són els següents:

NC6 5 Lligues	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
La Liga	0,42	0,38	0,31	0,42	0,45	0,54	0,45	0,48	0,40	0,27
Bundeslliga	0,28	0,31	0,37	0,34	0,44	0,31	0,36	0,31	0,27	0,38
Premier League	0,42	0,27	0,41	0,43	0,52	0,38	0,30	0,54	0,49	0,52
Ligue 1	0,33	0,16	0,31	0,27	0,35	0,35	0,26	0,42	0,39	0,33
Serie A	0,32	0,32	0,27	0,37	0,42	0,28	0,41	0,50	0,51	0,39

Taula 3: Taula del NC6 5 Lligues. Elaboració pròpia

2. **IHNBC**

   Si mirem la definició d'aquesta mesura tenim que:

$$IHNBC = \left(m(n) - \frac{1}{n}\right)^{-1} \left[\frac{IHBC}{n} - \frac{1}{n}\right]. \quad (4.1)$$

Per a calcular l'IHBC, amb les dades que tenim de totes les lligues, els càlculs són senzills. Ara bé, per a poder calcular el valor de $m(n)$, utilitzarem els valors que apareixen en [13], els quals aquest valor de $m(n)$ l'hem referenciat amb més detall a les definicions de la primera secció per tal de veure d'on sortien. Aquests valors són

Number of Teams (N)	Scoring Pattern P = {2, 1, 0}		Scoring Pattern P = {3, 1, 0}	
	Cascade ($q = N - 1$)	Maximum of HHI	Truncated Cascade (q^*)	Maximum of HHI
2	1	1.0000000	1	1.0000000
3	2	0.5555556	1	0.5937500
4	3	0.3888889	1	0.4133333
5	4	0.3000000	2	0.3251029
6	5	0.2444444	2	0.2662722
7	6	0.2063492	3	0.2243767
8	7	0.1785714	3	0.1953981
9	8	0.1574074	3	0.1723899
10	9	0.1407407	4	0.1541667
11	10	0.1272727	4	0.1397569
12	11	0.1161616	4	0.1275433
13	12	0.1068376	5	0.1174003
14	13	0.0989011	5	0.1087967
15	14	0.0920635	5	0.1012346
16	15	0.0861111	6	0.0947846
17	16	0.0808824	6	0.0890706
18	17	0.0762527	7	0.0839378
19	18	0.0721248	7	0.0794709
20	19	0.0684211	7	0.0754015

Figura 1: Càlculs de m(n) segons el número d'equips.

Com ve podem observar a la Figura 1, a nosaltres ens interessen els valors de la columna amb sistema de puntuació {0,1,3}, ja que a totes les 5 lligues que estem estudiant tenen aquest sistema de puntuació.
Ara, com que la Bundeslliga té 18 equips, i les altres 4 grans lligues en tenen 20, tenim que els valors de m(n) que nosaltres farem servir seran 0,0839378 i 0,0754015 respectivament.
Per tant, els resultats de l'IHNBC són els que es mostren a la següent taula

IHNBC 5 Lligues	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
La Liga	0,24	0,19	0,19	0,21	0,23	0,29	0,22	0,29	0,22	0,14
Bundeslliga	0,15	0,13	0,20	0,22	0,25	0,16	0,20	0,16	0,17	0,24
Premier League	0,25	0,12	0,21	0,22	0,25	0,18	0,17	0,26	0,25	0,29
Ligue 1	0,16	0,10	0,15	0,13	0,18	0,15	0,16	0,21	0,21	0,18
Serie A	0,15	0,16	0,16	0,20	0,26	0,19	0,20	0,28	0,28	0,24

Taula 4: Resultats IHNBC de les 5 grans Lligues. Elaboració pròpia

3. **MND**
   Si tornem a la definició de la Mesura Normalitzada de la dinàmica,

$$MND = \frac{2}{n^2} \sum_{i=1}^{n} |C_{i,1} - C_{i,0}|. \tag{4.2}$$

   observem que els càlculs tornen a ser molt senzills, però amb la particularitat que haurem de fer servir les dades dels punts obtinguts de tots els equips per a la temporada 2008/09, ja que necessitem l'any anterior al del 2009/10.
   A continuació es mostra la taula amb els resultats.

MND 5 Lligues	2009/10	2010/11	2011/12	2012/13	2013/14	2014/15	2015/16	2016/17	2017/18	2018/19
La Liga	0,35	0,38	0,42	0,31	0,34	0,19	0,30	0,25	0,39	0,26
Bundeslliga	0,53	0,51	0,50	0,29	0,42	0,25	0,44	0,53	0,36	0,34
Premier League	0,28	0,25	0,27	0,26	0,30	0,19	0,27	0,31	0,34	0,23
Ligue 1	0,43	0,40	0,38	0,41	0,39	0,32	0,26	0,35	0,27	0,48
Serie A	0,34	0,31	0,28	0,26	0,38	0,36	0,29	0,28	0,12	0,24

Taula 5: Taula dades MND 5 Lligues. Elaboració pròpia

4.2.1 Coeficients Correlació

Si observem els resultats obtinguts a les quatre taules que acabem de veure, el que veiem és que els resultats de NC6 i l'IHNBC tendeixen a comportar-se de la mateixa manera, és a dir, que segueixen una tendència força similar durant les 10 temporades de les 5 grans lligues.
Per a veure amb més detall aquest fet, el que farem serà calcular la correlació que hi ha entre totes les mesures que hem calculat.
Aquests càlculs els farem utilitzant la funció COEF.DE.CORREL de l'*Excel*, la qual el que ens calcula és el coeficient de correlació de Pearson.
Aquest coeficient és una mesura de dependència lineal entre dues variables aleatòries (en el nostre cas aquestes variables seran tots els valors calculats de dues mesures del balanç competitiu). Donades dues variables aleatòries (X, Y) i sigui $\{(x_i, y_i)\}$ $i = 1, ..., n$ els parells de n dades aleshores la fórmula del coeficient de Pearson és la següent:

$$r_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \bar{x})^2 \sum_{i=1}^{n} (y_i - \bar{y})^2}} \tag{4.3}$$

On $\bar{x}$ és la mitjana mostral de les x_i, i el mateix amb les variables y_i.
Aquest coeficient de Pearson (r_{xy}) pot prendre valors entre -1 i $+1$ i la interpretació d'aquest coeficient varia en funció del valor que prengui. Si el valor s'apropa a 1 les variables es diu que estan directament relacionades, mentre que quan tendeix cap a -1 es diu que les variables es relacionen

inversament. I quan el valor val 0 això ens vol dir que no existeix una relació lineal. Cal destacar que aquest coeficient no ens diu res sobre la relació no lineal entre les variables, és a dir que aquest coeficient podria valdre 0, i que les variables tinguessin una relació no lineal, però aquesta relació la determinen altres coeficients que aquí no veurem.
Dit això a continuació es mostra amb una taula la relació entre les mesures del balanç competitiu calculades anteriorment usant la funció de l'*Excel* esmentada anteriorment.

	NC6	IHNBC	MND	IBG
NC6	1,00	0,91	-0,39	0,52
IHNBC	0,91	1,00	-0,39	0,52
MND	-0,39	-0,39	1,00	-0,08
IBG	0,52	0,52	-0,08	1,00

Com podem veure en aquesta taula, les mesures NC6 i l'IHNBC tenen una correlació de 0,91 fet que ens indica que segueixen una relació lineal. Quan augmenta una de les dues mesures, l'altre mesura també augmenta amb una proporció similar.
En canvi, si ens fixem en l'IBG la correlació és de 0,52 tant amb el NC6 com amb l'IHNBC fet que ens diu que la correlació ja no és tan evident.
Per últim si ens fixem en la mesura MND, el que veiem és que sorprenentment té una correlació negativa amb les altres 3 mesures. I aquest fet es veurà amb més detall quan es mostri l'evolució de les mesures de les 10 temporades a les 5 grans lligues en un gràfic.

4.2.2 Càcul prova de raó de versemblança

Pel Teorema 3.3, sabem que $(2N-1)IBG$ és asimptòticament equivalent a una variable aleatòria khi-quadrat amb $2N-1$ graus de llibertat.
Aleshores com que acabem de calcular tots els IBG de les 5 grans lligues durant els 10 anys que estem estudiant, el que farem serà veure si en aquests anys acceptem o no la hipòtesi nul·la.
Si s'accepta la hipòtesi nul·la tindrem que la lliga estava perfectament equilibrada, però si rebutgem la hipòtesi nul·la aleshores ens trobarem en una situació on la lliga no estava perfectament equilibrada.
En tots els càlculs farem servir un nivell de significació $\alpha = 0,05$. Per tant, rebutjarem la hipòtesi nul·la si el valor de $(2N-1)IBG$ és més gran que el valor crític $\chi^2_{2N-1;\alpha}$, i l'acceptarem en cas contrari.
Per a fer els càlculs de $\chi^2_{2N-1;\alpha}$, al tenir un nombre de graus de llibertat molt elevat (611 a la Bundeslliga i 729 a les altres 4 grans lligues), el que hem fet ha sigut usar la plataforma excel la qual ens calcula aquest valor crític que busquem. La fórmula per a calcular-lo a l'Excel és la següent:

$$
\begin{aligned}
&= PRUEBA.CHI.INV(0,05;611) = 669,6 \\
&= PRUEBA.CHI.INV(0,05;759) = 824,2
\end{aligned} \tag{4.4}
$$

on 611 i 759 són els corresponents graus de llibertat i 0,05 és el nivell de significació.
Per tant, rebutjarem la hipòtesi nul·la si $(2N-1)IBG > 669,6$ en el cas de la Bundeslliga i si $(2N-1)IBG > 824,2$ en el cas de les altres 4 lligues.
A continuació es mostren els càlculs de tots els anys i totes les lligues.

Els primers resultats que analitzarem són els de la Lliga Espanyola.

Prova khi quadrada La Liga				
Anys	$(2N-1)IBG$	$\chi^2_{2N-1;\alpha}$	Diferència	Decisió
2009/10	836,67	824,20	12,47	rebutgem H_0
2010/11	917,53	824,20	93,33	rebutgem H_0
2011/12	987,78	824,20	163,58	rebutgem H_0
2012/13	890,66	824,20	66,45	rebutgem H_0
2013/14	986,91	824,20	162,71	rebutgem H_0
2014/15	1015,94	824,20	191,74	rebutgem H_0
2015/16	982,04	824,20	157,84	rebutgem H_0
2016/17	912,33	824,20	88,13	rebutgem H_0
2017/18	952,59	824,20	128,38	rebutgem H_0
2018/19	780,54	824,20	-43,66	**acceptem** H_0

La Lliga espanyola s'ha caracteritzat durant tots aquests anys que estem estudiant per ser una competició on el Barça i el Madrid sempre han tingut un domini sobre la resta d'equips de la Lliga. És per aquest motiu que a priori, diríem que la Liga no està perfectament equilibrada. De fet, com bé podem veure a la taula, l'única temporada on s'accepta la hipòtesi nul·la (supòsit teòric de lliga equilibrada) és la temporada 2018/19.
Si analitzem amb més detall que va succeir aquella temporada, el que podem veure és que la quarta posició (última per aconseguir una plaça per a jugar a la "Champions League"), es va decidir a l'última jornada, i fins a 3 equips podien aconseguir aquesta quarta posició (València, Getafe i Sevilla). Seguidament, el que observem és que de la 7a posició (Espanyol) fins a l'11a posició (Deportivo Alavés), tots els 5 equips estaven separats només per 3 punts, és a dir, que la 7a posició (última plaça per a la 2a lliga europea) també va estar molt disputada fins a l'última jornada. I per últim també es pot veure que les posicions de mitja taula també van estar molt igualades sense diferències gaire grosses, fins i tot el Villareal (14) i el Levante (15) van empatar a punts i es van quedar més o menys als mateixos punts de

descendir de categoria com classificar-se per a competicions europees, fet que indica que va ser una lliga molt competida, i sota el supòsit teòric que s'està treballant, va ser una temporada perfectament equilibrada.

Pel que fa a la Bundeslliga s'han obtingut els següents resultats.

Prova khi quadrada Bundeslliga				
Anys	$(2N-1)IBG$	$\chi^2_{2N-1;\alpha}$	Diferència	Decisió
2009/10	695,92	669,61	26,31	rebutgem H_0
2010/11	695,70	669,61	26,08	rebutgem H_0
2011/12	728,45	669,61	58,83	rebutgem H_0
2012/13	689,80	669,61	20,19	rebutgem H_0
2013/14	686,61	669,61	17,00	rebutgem H_0
2014/15	719,88	669,61	50,27	rebutgem H_0
2015/16	684,64	669,61	15,02	rebutgem H_0
2016/17	718,84	669,61	49,22	rebutgem H_0
2017/18	658,59	669,61	-11,02	**acceptem** H_0
2018/19	750,95	669,61	81,33	rebutgem H_0

La Bundeslliga és la lliga alemana, on durant aquests anys estudiats hi ha hagut un clar domini del Bayer de Munich, i del Borussia Dormunt. I pel que fa a les 4 primeres posicions també han acompanyat a aquests dos grans equips el Schalke 04 el Bayern Leverkussen i el Wolsburg en la majoria de les temporades amb una clarividència bastant notable. Aquest és un dels principals motius pels quals la lliga Alemana no hagi sortit en gairebé cap temporada que sigui equilibrada.
De fet, l'única temporada on s'ha acceptat la hipòtesi nul·la ha sigut la 2017/18.
Una explicació d'aquest fet és que en aquell any de la 3a posició (Hoffenheim) fins a la 7a posició (Stuttgart) la diferència només va ser de 4 punts. Això va implicar que fins a l'última jornada, totes les places per a les competicions europees estaven en joc. Però si mirem més en concret, veiem que el 3r, el 4t i el 5è van obtenir els mateixos punts (55), i el que es va quedar fora de la Champions League (màxima competició europea) va ser el Bayern Leverkusen degut a la diferència de gols. A mitjana taula la lliga també va estar bastant disputada, on des de la posició 10 fins a la 15 la diferència de punts obtinguts només va ser de 7 punts, la qual indica que va haver-hi una gran igualtat entre els equips de la lliga alemana.
En resum, pel tot el que s'acaba d'observar tant a la taula de la khi quadrada, com a les posicions finals de la competició, diem que la Bundeslliga a la temporada 2017/18 va estar perfectament equilibrada.

A continuació es mostra la taula de resultats de la Prova khi quadrada de la Premier League.

Prova khi quadrada Premier League				
Anys	$(2N-1)IBG$	$\chi^2_{2N-1;\alpha}$	Diferència	Decisió
2009/10	985,51	824,20	161,31	rebutgem H_0
2010/11	781,17	824,20	-43,03	**acceptem** H_0
2011/12	893,18	824,20	68,97	rebutgem H_0
2012/13	839,82	824,20	15,62	rebutgem H_0
2013/14	938,79	824,20	114,59	rebutgem H_0
2014/15	834,05	824,20	9,85	rebutgem H_0
2015/16	826,94	824,20	2,74	rebutgem H_0
2016/17	892,89	824,20	68,69	rebutgem H_0
2017/18	922,42	824,20	98,22	rebutgem H_0
2018/19	852,13	824,20	27,92	rebutgem H_0

La Premier League és una lliga que sempre s'ha caracteritzat per ser molt competitiva, però durant l'última dècada aquesta competitivitat s'ha vist reduïda per la gran dominància de 6 equips, els quals se'ls denomina 'Big Six', i aquests equips són el Manchester United, el Manchester City, l'Arsenal, el Chelsea, el Liverpool i per últim el Tottenham Hotspur, els quals han provocat que hi hagi certa diferència entre aquests equips i la resta de competidors de la lliga.
Aquest, doncs, és un dels motius principals pels quals podem veure als resultats que només a la temporada 2010/11 s'accepta la hipòtesi nul·la, fet que ens indica que la lliga no ha estat gaire equilibrada durant aquests anys que estem estudiant.
Si mirem amb més detall que va passar a la temporada 2010/11 el que veiem és que el Chelsea i el Manchester City van obtenir els mateixos punts (71), i des de la 8a posició (Fulham) fins a la 19a posició (Blackpool), la diferència de punts només va ser de 10 punts. Més concretament, 12 equips de la lliga van estar en aquest rang de 10 punts, fet que fa que aquesta lliga fos molt competitiva, i en conseqüència que en la prova de raó de versemblança s'accepti la hipòtesi nul·la.

Seguidament analitzarem els resultats de la lliga francesa de futbol.

Prova khi quadrada Ligue 1				
Anys	$(2N-1)IBG$	$\chi^2_{2N-1;\alpha}$	Diferència	Decisió
2009/10	842,43	824,20	18,22	rebutgem H_0
2010/11	740,94	824,20	-83,26	**acceptem** H_0
2011/12	740,22	824,20	-83,98	**acceptem** H_0
2012/13	799,97	824,20	-24,23	**acceptem** H_0
2013/14	755,16	824,20	-69,04	**acceptem** H_0
2014/15	867,54	824,20	43,34	rebutgem H_0
2015/16	910,23	824,20	86,02	rebutgem H_0
2016/17	935,55	824,20	111,35	rebutgem H_0
2017/18	876,84	824,20	52,64	rebutgem H_0
2018/19	840,84	824,20	16,64	rebutgem H_0

Observant els resultats, veiem que la Ligue 1 durant la temporada 2010/11 fins a la temporada 2013/14 s'accepta la hipòtesi nul·la, i per tant es diu que la lliga estava equilibrada.
Si analitzem aquestes 4 temporades amb més detall, tenim que a la temporada 2010/11 els equips classificats de la 7a posició fins a la 18 ena posició estaven separats només per 7 punts, i les competicions europees també es van decidir per menys de 4 punts.
Pel que fa a la temporada 2011/12 també es veu que des de la 12a posició fins a l'última (20), només els van separar 9 punts, és a dir que fins a les últimes 5 jornades tots aquests equips encara tenien opcions de quedar en les tres últimes posicions les quals són els que baixen de categoria.
A la 2012/13 tenim que les posicions europees van estar disputades fins al final, i també les diferències a mitja taula van ser mínimes.
I per últim a la 2013/14 també es pot veure que des de la 8a posició fins a la 18a posició només els van separar 9 punts, fet que indica que la lliga va estar molt competida.
Pel que fa a les altres temporades que no s'accepta la hipòtesi nul·la de la khi quadrada, una de les explicacions és que a finals de l'any 2011 el grup Qatar Investment Authority (QIA) va comprar el 70% del club PSG, i va invertir una gran quantitat de milions per a tal que el seu club pogués guanyar la màxima competició europea. Per aquest motiu el projecte va tardar uns anys a consolidar-se i no va ser aproximadament fins a l'any 2015 que el club va començar a tenir una superioritat molt evident davant els seus altres competidors de la Ligue 1. Per posar un exemple evident, a la temporada 2015/16 el PSG va obtenir 96 punts, mentres que el seu primer perseguidor va obtenir un total de 65 punts, és a dir, una diferència de 31 punts fet que indica que la lliga no era gaire competida.

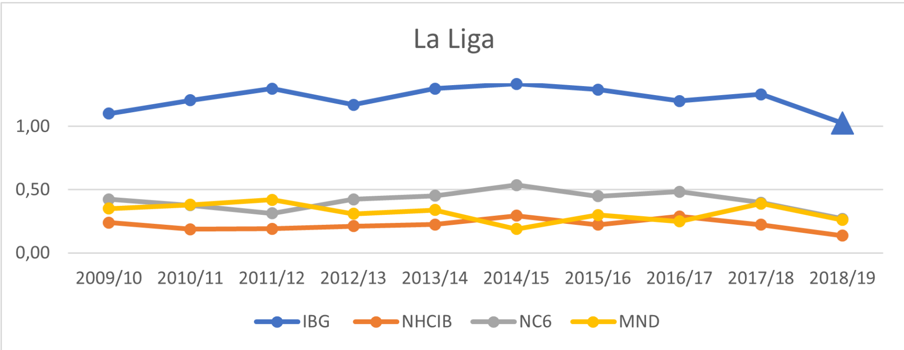
Un dels altres motius també és la inversió multimilionària que va dur a terme l'empresari rus Dmitri Rybolóvlev per tal d'aconseguir que el Mònaco arribés a les primeres posicions de la lliga. I ho va aconseguir fins al punt que la temporada 2016/17 el Mònaco va quedar primer de la lliga amb 95 punts seguit del PSG amb 87 punts, i el tercer ja va quedar només amb 78 punts, fet que indica que no hi havia gran competència aquella temporada.

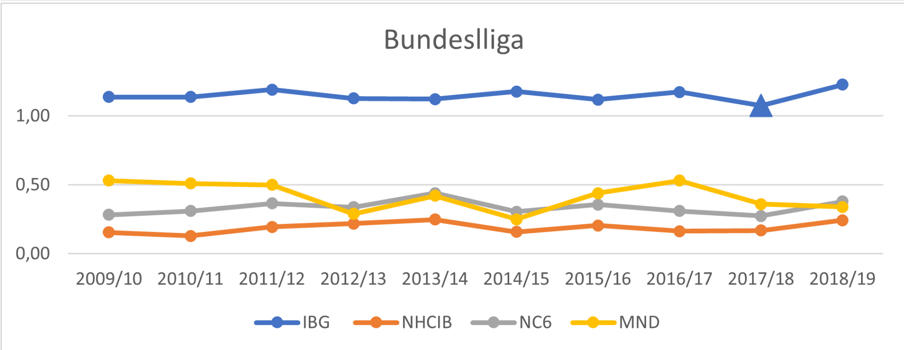
I per últim, analitzarem els resultats de la khi quadrada de la lliga italiana.

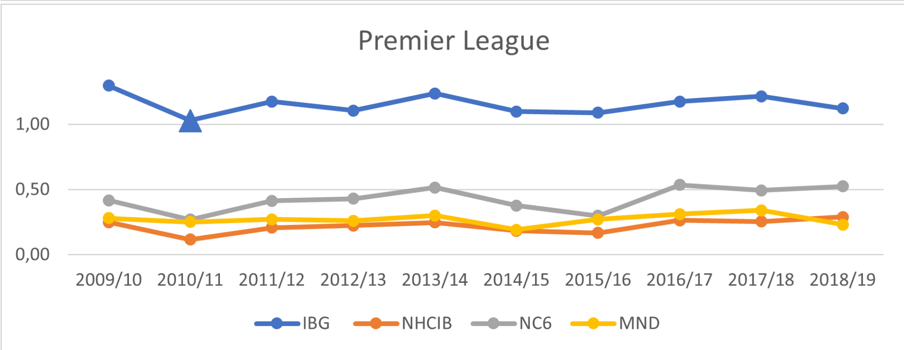
Prova khi quadrada Serie A				
Anys	$(2N-1)IBG$	$\chi^2_{2N-1;\alpha}$	Diferència	Decisió
2009/10	707,19	824,20	-117,02	**acceptem** H_0
2010/11	795,47	824,20	-28,73	**acceptem** H_0
2011/12	851,80	824,20	27,59	rebutgem H_0
2012/13	826,32	824,20	2,11	rebutgem H_0
2013/14	822,77	824,20	-1,43	**acceptem** H_0
2014/15	767,16	824,20	-57,05	**acceptem** H_0
2015/16	800,97	824,20	-23,23	**acceptem** H_0
2016/17	861,69	824,20	37,49	rebutgem H_0
2017/18	895,28	824,20	71,08	rebutgem H_0
2018/19	795,63	824,20	-28,58	**acceptem** H_0

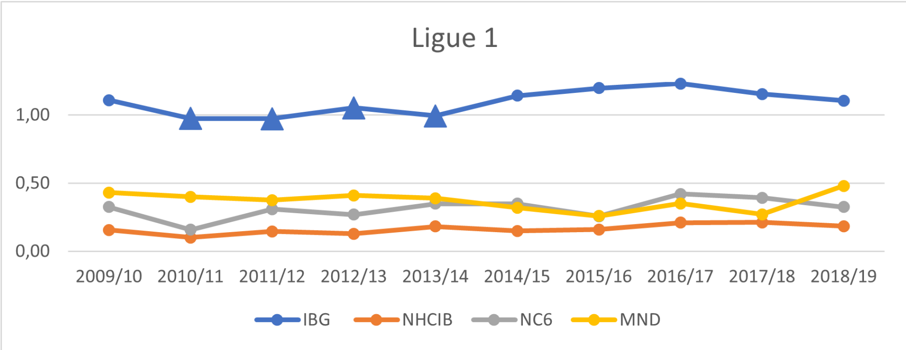
El primer que podem observar és que la Serie A és la lliga en què s'ha acceptat més vegades la hipòtesi nul·la, i per tant és la lliga on hi ha hagut més equilibri competitiu durant aquestes deu temporades que estem estudiant.
No analitzarem totes les temporades una per una on hi ha hagut balanç competitiu, però el que si que podem extreure són unes conclusions generals, les quals fan referència a l'alta igualtat que hi ha entre els equips de mitja taula a la Serie A, i també en certs casos en la igualtat per a les posicions de competició europea com per exemple la temporada 2018/19 on del tercer al sisè només els van separar 6 punts.
En gairebé totes les temporades on s'ha acceptat la hipòtesi nul·la, les diferències de punts entre uns 9 equips (gairebé la meitat), no és superior als 10 punts, fet que indica que hi ha una alta competitivitat, perquè en pocs punts tant pots acabar a la part mig/alta de la classificació, com ben a prop de les posicions de descens.

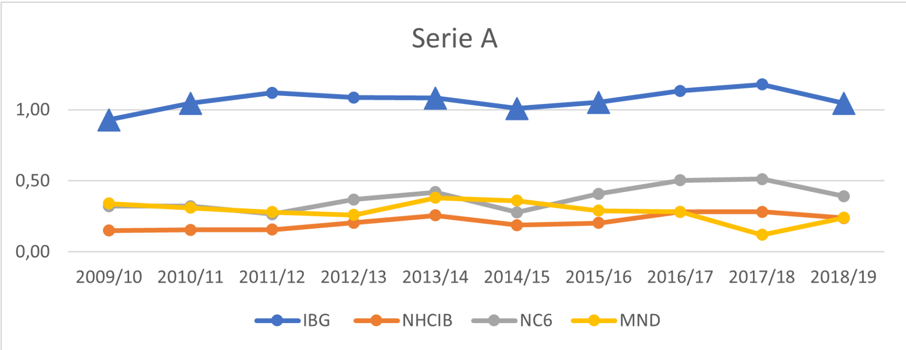
Per tal d'il·lustrar tots els resultats que hem estat comentant fins ara, el que veurem seran 5 gràfics amb tots els valors de les mesures del balanç competitiu calculades durant les 10 temporades i separades per a cada una de les 5 grans lligues.
En aquests gràfics el que podem veure és que les temporades on s'accepta la hipòtesi nul·la (i, per tant, indica que la lliga va estar equilibrada) estan identificades amb el símbol d'un triangle (▲), mentre que la resta de temporades estan identificades amb un cercle (•)








Ligue 1
1,00
0,50
0,00
2009/10
2010/11
2011/12
2012/13
2013/14
2014/15
2015/16
2016/17
2017/18
2018/19
IBG
NHCIB
NC6
MND


Serie A
1,00
0,50
0,00
2009/10
2010/11
2011/12
2012/13
2013/14
2014/15
2015/16
2016/17
2017/18
2018/19
IBG
NHCIB
NC6
MND

5 Estudi variacions dels ingressos

5.1 Introducció

L'econometria consisteix a estudiar i analitzar les característiques d'una variable utilitzant informació d'altres variables que puguin explicar el comportament de la primera. Dit això, en el nostre cas el que analitzarem seran com varien els ingressos d'una lliga de futbol, en funció d'altres variables que considerarem a continuació.

5.2 Variables de l'estudi

Quan parlem de futbol, les variables que infereixen en els ingressos d'una lliga de futbol són moltes; ara bé, en aquest estudi en mirarem tres d'elles les quals considero que són bastant rellevants.
En primer de tot, com que estem estudiant 5 lligues de països diferents és adient observar les dades del producte interior brut (PIB) de cada país per tal de veure si la riquesa d'un país influeix en l'augment dels ingressos d'una lliga o viceversa.
Seguidament el que estudiarem també serà la influència dels espectadors als estadis de futbol, la qual també pot ser un factor clau per tal d'explicar les variacions en els ingressos de la lliga.
I per últim, el que analitzarem serà el nostre objectiu principal d'aquest estudi. Es tracta de totes les mesures de balanç competitiu que hem vist fins ara, i veurem si els ingressos depenen significativament o no d'aquestes quatre mesures.

5.3 Dades de panell

Qualsevol estudi economètric està influenciat per la naturalesa de les dades que s'estigui treballant. És per aquest motiu que a continuació es presenten 2 tipus de dades per definir posteriorment les dades de panell que serà amb el que farem la nostre anàlisis.

- **Sèries Temporals**
  Aquestes dades són un conjunt d'observacions sobre els valors d'una variable en diferents moments temporals. Aquests moments temporals solen ser en dies, setmanes, mesos o anys. En el nostre estudi en centrarem en els anys, ja que observarem les dades de les deu temporades (2009/10 - 2018/19) que estem utilitzant durant tot el treball.

- **Dades Transversals**
  Aquestes dades proporcionen informació de diferents agents en un mateix instant del temps. Si ho apliquem al nostre estudi, aquests agents

seran les 5 grans lligues de futbol, les quals tenen una naturalesa similar entre elles.

Usant la informació anterior, a continuació es presenta la definició de les dades de panell.

Definició 5.1 (Dades de Panell)**.** *Les dades de panell són una combinació entre les dades de sèrie temporals i les transversals. És a dir, són les dades on un conjunt d'agents són observats en diversos moments del temps.*

Observació 5.2. A l'utilitzar dades de panell el que estem fent és augmentar la nostra mostra, és a dir, que gràcies a combinar dades de sèrie temporal amb transversals recopilem més informació, i això ens permetrà ser més precisos quan haguem de treure conclusions del model.

Observació 5.3. Tenint en compte tot el que s'ha esmentat a continuació, el que estudiarem serà l'evolució del PIB, de l'assistència als partits i de les mesures del balanç competitiu durant les deu temporades de les 5 grans lligues per tal d'explicar quina relació hi ha amb els ingressos de les lligues.

5.3.1 Obtenció dades

Per tal de dur aquest estudi el que farem serà recollir totes les dades d'aquestes variables que acabem de parlar.
Quan parlem d'ingressos, les dades que he fet servir són els ingressos comercials de les 5 grans lligues i les he obtingut d'una pàgina web anomenada *Statista* [7].
Per obtenir el PIB acurat des del 2010 fins al 2019, el que he fet ha sigut agafar les dades de la pàgina web *The World Bank* [14]. Les dades estan en trilions d'euros però jo les he utilitzat amb milions d'euros per tal d'evitar números massa llargs.
El número d'assistents en un partit estan registrats a moltes pàgines web relacionades amb el futbol, però jo he fet servir les dades de *FootyStats* [11], la qual el que em proporciona són les dades de la mitjana d'espectadors durant la temporada en les 5 grans lligues que estic estudiant.

5.4 Model proposat

Si ens fixem en els ingressos dels últims anys en les 5 grans lligues (mirar la Figura 2), el que veiem és una tendència lineal en les 5 lligues, és per aquest motiu que tractarem amb un model lineal de dades de panell.

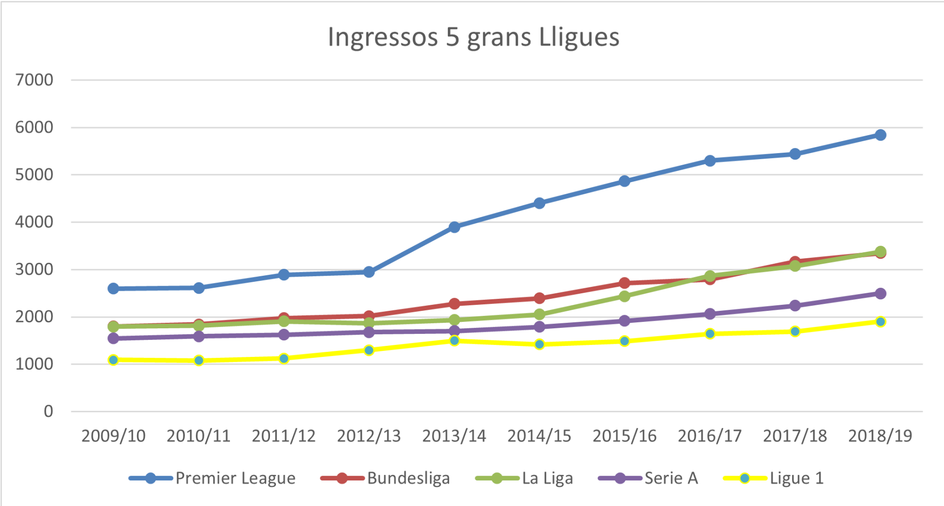


Figura 2: Ingressos 5 grans lligues. Elaboració pròpia

Primer de tot veurem ràpidament com és el model de regressió lineal clàssic per tal d'entendre el model de dades de panell, ja que és una extensió d'aquest model clàssic.
El model de regressió lineal relaciona la variable dependent "Y"amb n variables dependents X_i, el qual ens serveix per a veure aproximadament quina relació hi ha entre aquestes variables.
La fórmula d'aquest model és la següent:

$$Y = \beta_0 + \sum_{j=1}^{n} \gamma_j X_j + \epsilon \tag{5.1}$$

on els β són paràmetres (mesuren la influència que tenen les variables X_i sobre Y) i el ϵ és una variable aleatòria la qual recull tots els factors no controlables i observables.
Un cop esmentat com és el model de regressió lineal simple el que introduirem serà el model de dades de panell, el qual és de la següent forma:

$$Y_{it} = \beta_i + \gamma_1 X_{1it} + \gamma_2 X_{2it} + \cdots + \gamma_n X_{nit} + \epsilon_{it} \tag{5.2}$$

El que podem observar, doncs és que per a cada variable tenim dos subíndexs: el t que fa referència al temps i, per tant, a la sèrie temporal, i la i que ens indica de la dimensió transversal (el número d'agents que intervenen).

Observació 5.4. En el nostre estudi, com estem analitzant la variació dels ingressos en una lliga de futbol en funció de les mesures de balanç competitiu, el PIB i l'assistència als camps de futbol, el que hem de tenir en compte és l'efecte d'aquelles característiques en cada lliga que existeixen, però que no estem tenint en compte en el nostre model com ara podrien ser la cohesió dels equips dintre d'una lliga, l'estat d'ànim dels jugadors en termes de drets dintre d'una lliga, i d'aquests molts més exemples, els quals els recollim en la nostra variable β_i.
Aquesta variable β_i ens està dient que són aquelles característiques dels individus que generalment no es poden observar i que són fixes en el temps. Més endavant veurem que en funció de com es consideri β_i farem una anàlisi diferent del nostre model.

Pel que fa al nostre cas en particular, a la variable dependent dels ingressos la denotarem per ING, i a les variables explicatives les denotarem per PIB (producte interior brut de cada país), ESP (mitjana d'espectadors anuals) i MBC (mesura balanç competitiu).
Per tant, el model de dades de panell a estudiar és el següent:

$$\begin{gathered}\ln ING_{lt} = \beta_i + \gamma_1 t + \gamma_2 PIB_{lt} + \gamma_3 ESP_{lt} + \alpha MBC_{lt} + \epsilon_{lt}, \\ 1 \leq i \leq 5, 1 \leq t \leq T\end{gathered} \tag{5.3}$$

on i denota la i-èssima lliga, i la T fa referència al període de 10 anys que estem estudiant (2009/10 fins al 2018/19).
El nostre objectiu és veure si els ingressos depenen significativament o no de les mesures de balanç competitiu que hem estudiat, és a dir que ens fixarem en el paràmetre α.

5.4.1 Efectes fixos o aleatoris

Els dos models que analitzarem seran els d'efectes fixos i els d'efectes aleatoris, els quals són els dos models més utilitzats en la teoria de dades de panell.
La principal diferència entre ells està a veure si els efectes inobservables β_i estan correlacionats o no amb les variables explicatives observables (PIB, ESP i MBC).

- El model d'**efectes fixos (EF)** suposa que hi ha una correlació entre els efectes inobservables dels individus (β_i) amb les variables explicatives del model, és a dir que $corr(\beta_i, X) \neq 0$ on la X fa referència al PIB, MBC i ESP del nostre model.
- Mentre que el model d'**efectes aleatoris (EA)** el que suposa és just el contrari. No hi ha correlació entre els efectes β_i amb les variables explicatives. Per tant, tenim que $corr(\beta_i, X) = 0$.

5.4.2 Test de Hausman

Per tal de veure quin dels dos models serà més adient en el nostre estudi, el que realitzarem serà el test de Hausman.
Aquest test el que compara són les estimacions dels coeficients en funció de si estem en un model d'efectes fixos o d'efectes aleatoris.
De manera més visual, el test de Hausman és el següent:

$$H_0 : corr(\beta_i, X) = 0$$
$$H_1 : corr(\beta_i, X) \neq 0$$

Si li diem t_1 a l'estimador calculat amb efectes aleatoris, i t_2, l'estimador amb efectes fixos, aleshores l'estadístic utilitzat pel test de Hasman és:

$$H = (t_1 - t_2)'((Var(t_1) - Var(t_2))^{-1}(t_1 - t_2), H \sim \chi_n^2 \tag{5.4}$$

On $Var(t_1)$ és la matriu de variàncies i covariàncies corresponents a un model d'efectes aleatoris, mentre que $Var(t_2)$ correspon a la matriu de variàncies i covariàncies corresponents a un model d'efectes fixes.
No veurem amb més detall totes les característiques perquè aquest test el podem calcular de manera eficient amb l'aplicació *Rstudio*, la qual té una funció explícita per aquest text.
Cal tenir en compte que en aquesta funció els resultats s'interpreten de la següent manera.

	S'accepta H_0	**S'accepta H_1**
t_1 (Estimador EA)	Consistent i eficient	Inconsistent
t_2 (Estimador EF)	Consistent i Ineficient	Consistent

Per tant, com bé veiem a la taula, si s'accepta la hipòtesi nul·la és preferible el model d'efectes aleatoris, mentre que si es rebutja, el model més adient és el d'efectes fixos.

5.4.3 Càlculs

Tenint en compte tot el que acabem de veure en aquesta secció, a continuació calcularem el nostre model de dades de panell definit a l'equació (5.3), i anirem canviant el valor de MBC per les quatre mesures que estem estudiant i analitzarem quina de les mesures és més consistent.
Primer de tot el que calcularem és el test de Hausman per tal de decidir si treballarem amb un model d'efectes fixos o amb un model d'efectes aleatoris. Aquests càlculs els faré tal com he comentat a la secció 5.4.2, amb l'aplicació *RStudio*.
Per tal de treballar amb dades de panell al *RStudio*, usarem el paquet 'plm' el qual és una versió extensa del que *Yves Croissant* i *Giovanni Millo* [24]

van publicar l'any 2008.
Primer de tot, el que s'ha de fer per tal de poder fer servir el paquet plm, és estructurar les dades d'una manera concreta perquè el programa les llegeixi correctament.
L'estructura de les nostres dades serà de la següent forma:

LLIGA	INGRESSOS	ANYS	PIB	ESP	MBC
La Liga	x	2010	x	x	x
La Liga	x	2011	x	x	x
La Liga	x	2012	x	x	x
...	...	...	...	...	...
Bundeslliga	x	2010	x	x	x
Bundeslliga	x	2011	x	x	x
Bundeslliga	x	2012	x	x	x
...	...	...	...	...	...

Observació 5.5. Cal tenir en compte que a la taula falten les dades de les altres 3 grans lligues i també de totes les temporades en què estem estudiant, però per brevetat, només es mostren les que es veuen a la taula.

5.4.4 Codi *RStudio*

Un cop tenim les dades ben estructurades al programa del *RStudio*, el que farem serà cridar al paquet 'plm', per tal així de poder fer els càlculs del nostre model de dades de panell. A continuació es mostra quin és el codi que he fet servir al Rstudio per a dur a terme els càlculs.

```
library("plm")
Dadesordenades <- pdata.frame(DadesPanell, index = c("Lliga", "Anys"))
EA <- plm(Ingressos ~ Anys+ PIB + ESP + MBC, data = Dadesordenades, model = "random")
EF<- plm(Ingressos ~ Anys+ PIB + ESP + MBC, data = Dadesordenades, model = "within")
phtest(EA,EF)
```

Figura 3: Codi *Rstudio*

Les dades de la taula descrita anteriorment en el meu cas estan en un *Excel* anomenat *DadesPanell*, el qual he importat al *RStudio*. Un cop s'han importat les dades al Rstudio, el que fem és cridar la llibreria 'plm'.

Seguidament, el que fem és usar la funció *pdata.frame* la qual adapta les dades que li hem proporcionat per tal que ho reconegui com a dades de panell.

Aquesta funció *pdata.frame* necessita que s'especifiquin quines són les dades a estudiar (*DadesPanell*), i també necessita saber quin és l'índex per a identificar els individus, i quin és l'índex per a identificar el període de les nostres dades. En el nostre cas com es pot veure a la Figura 3, la Lliga fa referència a l'índex dels individus que estem treballant (5 grans lligues), i l'índex temporal el qual està com a Anys són les 10 temporades en les quals estem fent l'estudi.

Si ens seguim fixant en el codi de la Figura 3, el que fem a continuació és cridar la funció *plm*, la qual ens calcula el model de regressió per a les nostres dades de panell.
En aquesta funció *plm* el que necessitem és que se li especifiqui com és el nostre model. Per tant, hem de reescriure el que teníem a l'equació (5.3) de tal forma que el programa sàpiga fer les estimacions del model.
També veiem que en aquesta funció, se li ha de dir d'on treu les dades, i en aquest cas ja són les dades que el propi *RStudio* ha transformat per a poder fer els càlculs.
Per últim, s'ha de dir amb quin model volem que es faci l'aproximació del nostre estudi. En el nostre cas ens interessen els d'efectes aleatoris ("random") i el d'efectes fixos ("within").

Seguidament, només ens falta cridar la funció *phtest*, la qual ens farà decidir si escollim un model d'efectes fixos o un d'efectes aleatoris.

5.4.5 Resultats

Un cop tenim totes les funcions i les dades per tal de calcular el nostre model, a continuació es presenten els resultats del test de Hausman.

Mesura Balanç	P-valor	Model escollit
IBG	0,98	EA
IHNBC	0,44	EA
MND	2.2×10^{-16}	EF
NC6	0,008	EF

Taula 6: Taula dades resultats Test Hausman. Elaboració pròpia

Observació 5.6. La hipòtesi nul·la del test de Hausman s'accepta quan el p-valor és més gran que 0,05 (nivell de significació), que com hem comentat amb anterioritat ens indica que el model d'efectes aleatoris és consistent i eficient.
És per aquest motiu que per al nostre IBG i a l'IHNBC al ser $p{-}valor > 0,05$ acceptem la hipòtesi nul·la del test, i afirmem que el model amb efectes aleatoris és més adient per a les nostres dades observades.

Seguidament el que és mostra a continuació és el resultat del model amb Efectes Aleatoris usant el nostre IBG, ja que és la mesura que ens interessa més estudiar.

Variables	Estimació
PIB	2.53×10^{-7}
ESP	1.55×10^{-5}
IBG	-0.30256*
R^2	0.94
R^2 ajustat	0.918

Taula 7: Resulats de coeficients a estimar i valor de R^2. Elaboració pròpia

Observació 5.7. El primer que veiem és un * al valor de l'estimació de la variable IBG. Aquest * ens diu que en el nostre model estudiat, aquesta variable té una relació estadísticament significativa amb la variable dependent que en el nostre cas és la d'ingressos.

Observació 5.8. Aquest valor del R^2 és un coeficient de determinació el qual determina la qualitat del model per a replicar resultat, és a dir, que sí que es poden fer unes bones prediccions en funció d'aquest model. Pren valors entre 0 i 1, i com més proper a 1 voldrà dir que el model estarà més ben ajustat.
El problema d'aquest R^2, és que a l'afegir variables explicatives no significatives en el model aquest valor de R^2 augmenta, és per aquest motiu que és millor ajustar aquest coeficient per tal d'obtenir un resultat amb més credibilitat.

Al coeficient de determinació ajustat li direm $\bar{R}^2$ i es calcula de la següent manera.

$$\bar{R}^2 = 1 - \frac{N-1}{N-k-1}(1-R^2), \tag{5.5}$$

on N és la dimensió de la mostra, i k el número de variables explicatives.

Un cop explicat breument en què consisteix el coeficient R^2, fent els càlculs del model al *RStudio* amb les altres mesures de Balanç Competitiu el que he obtingut és que els valors de $\bar{R}^2$ són inferiors en les 3 altres mesures de Balanç (IHNBC, MND, NC6).

Si ens continuem fixant en els resultats de la Taula 7, el que veiem és que tant el PIB com el ESP (mijtana d'espectadors), tenen una relació positiva

amb els ingressos, és a dir, que si augmenten aquestes dues variables, els ingressos de la lliga també augmentaran.
En canvi, en el nostre índex basat en els gols, el coeficient estimat és negatiu, i per tant si augmenta el nostre Índex basat en els gols els ingressos de la lliga disminuiran.

Observació 5.9. Aquest fet és molt rellevant, ja que sabem que a mesura que augmenta el nostre índex basat en els gols, les competicions es tornen menys equilibrades, i per tant la competició esdevé menys atractiva, ja que sabem que l'interès per a qualsevol esdeveniment com va estudiar Rottenberg(1956) augmenta si el grau d'incertesa dels resultats és molt elevat. Per tant, tenim que si disminueix el nostre índex basat en els gols, els ingressos de la lliga augmentaran. En conclusió, si la lliga de futbol està equilibrada els ingressos augmentaran, i si, per contra, el nostre índex augmenta, els ingressos de la lliga disminuiran.

Recapitulant tot el que acabem de comentar, sota el supòsit teòric del nostre treball, l'índex basat en els gols ofereix una millor aproximació a l'hora d'explicar els ingressos de la lliga de futbol respecte a les altres mesures de balanç competitiu.

6 Extrapolació a l'Hoquei patins

A mode de completar el meu treball, en aquesta secció, el que veurem és si el nostre índex basat en els gols també és adient per aplicar-lo en un altre esport completament diferent del futbol com és l'hoquei patins.
L'elecció d'aquest esport en concret, ha sigut per alguns motius, el primer és perquè a l'hoquei es marquen molts més gols que al futbol, és per això que m'agradaria veure si l'índex s'adapta a aquests canvis. I també perquè des de petit he sigut aficionat al Reus Deportiu d'hoquei, i m'ha semblat oportú relacionar les dues grans aficions que tinc com són el futbol i l'hoquei.

6.1 Càlculs Probabilitats d'Empat

A continuació, realitzaré els mateixos càlculs que a la secció 4.1, per tal de veure si el model proposat segons el mateix marc teòric és adient, o pel contrari, al ser un altre esport totalment diferent, el nostre índex no és adient.

Observació 6.1. L'hoquei és un esport on el sistema de puntuació en la lliga funciona exactament igual que en el futbol, és a dir, que per victòria aconsegueixes 3 punts, per empat 1 i per derrota 0 punts. És per aquest motiu que estem en el mateix context teòric que en el futbol, però amb la particularitat que en l'hoquei el número d'equips és inferior.

Totes les dades han estat extretes de la pàgina web oficial de la *Reial Federació Espanyola de Patinatge* [16]. I el que analitzarem serà l'efecte d'aquest índex a la primera divisió masculina d'hoquei patins, durant 7 temporades les quals van de la 2015/16 fins a la temporada 2021/22 .

Primer de tot calcularem la mitjana de gols per partit, i posteriorment calcularem la probabilitat teòrica d'empat i la compararem amb la probabilitat real observada d'empat.
Dit això, els càlculs es fan idènticament igual que en la secció 4.1 i a manera de no repetir exactament el mateix procediment per a dur a terme els càlculs, a continuació es mostren els càlculs.

Lliga	Var.	2021-22	2020-21	2019-20	2018-19	2017-18	2016-17	2015-16
	$\hat{\lambda}$	3,53	3,38	3,02	3,08	3,08	3,11	3,44
OK Lliga	$d_{\hat{\lambda}}$	0,15	0,16	0,17	0,16	0,16	0,16	0,16
	$\hat{E}$	0,13	0,17	0,16	0,19	0,15	0,20	0,16

Taula 8: Resultats mitjana de gols i prob.empats. Elaboració pròpia

Observació 6.2. El primer que observem és que la mitjana de gols per partit ha augmentat considerablement respecte al futbol. Hem passat de tenir aproximadament una mitjana d'entre $1,2 \leq \hat{\lambda} \leq 1,6$ en el futbol, a ara tenir una mitjana d'entre $3 \leq \hat{\lambda} \leq 3,6$ gols per partit en l'hoquei.

Aquest augment dels gols és causat per diversos motius, però un dels quals és que en l'hoquei patins hi ha més penals, jugadors expulsats durant 2 minuts, faltes directes i d'aquests exemples molts més que fan que en l'hoquei el número de gols per partit augmenti més del doble que el futbol.

Observació 6.3. Si ens fixem ara en la probabilitat teòrica (calculada amb el Wolfram Alpha com en la secció 4.1) i amb la probabilitat real observada, el primer que tenim és que la probabilitat d'empat es redueix aproximadament 10 punts percentuals respecte al futbol, i aquest fet està relacionat clarament amb el número de gols per partit. A mesura que hi ha més gols en un partit, la probabilitat que acabi en empat es redueix. Els exemples on s'evidencia aquest fet són en el rugbi, o l'handbol, on es marquen molts gols, i en conseqüència la lliga té molts menys empats.

Per altra banda, si mirem la desviació de la nostra probabilitat teòrica amb la probabilitat real d'empat, en cap cas supera els 0,05 punts percentuals de diferència. Més concretament, l'única temporada en què la diferència és de 0,04 punts percentuals és la 2016/17.

Per tant, podem afirmar que aquesta probabilitat teòrica és bastant acurada pel que fa a aproximar la probabilitat real d'empat que hi ha en una temporada en funció dels gols.

6.2 Càlcul Índex basat en els gols

Per tal d'analitzar si l'índex basat en els gols calculat per a les 5 grans lligues també és eficient per a l'hoquei, procedirem a fer els càlculs d'aquest. Només recordar que la fórmula d'aquest índex la qual és la següent.

$$IBG = \frac{1}{(2N-1)\hat{\lambda}} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \sum_{k=0}^{1} (X_{ijk} - \hat{\lambda})^2. \tag{6.1}$$

La pàgina web oficial de la Federació Espanyola de Patinatge ofereix tots els resultats dels partits en una temporada, és per aquest que aleshores els càlculs amb un programa com l'*Excel* són senzills, per tant, aquí no es mostraran amb detall.

Observació 6.4. Per a dur a terme els càlculs d'aquest índex correctament s'han de tenir en compte els següents aspectes:

- **Temporada 2021/22**
  El número d'equips a la lliga va ser de 14 per tant la $N = 182$.
- **Temporada 2020/21**
  El número d'equips a la lliga va ser de 16 equips, i el número de partits va ser de 239. Hauria d'haver sigut de 240, però va haver-hi un partit a la tercera jornada on no es va jugar per resolució disciplinària i per això no l'he tingut en compte per a fer els càlculs.
- **Temporada 2019/20**
  El número d'equips aquesta temporada va ser de 14 equips. Tanmateix, el número de partits va ser de 175 degut a que a l'última jornada no es va jugar cap partit. Per tant, en aquesta temporada $N = 175$.
- **Temporades 2015/16 - 2018/19**
  En aquestes 4 temporades el número d'equips va ser de 16 (n=16) i en aquestes quatre temporades si que es van disputar tots els partits, per tant la $N = 240$.

Un cop havent fet aquestes precisions a continuació es mostren els resultats de l'índex basat en els gols per a les 7 temporades.

IBG	2021-22	2020-21	2019-20	2018-19	2017-18	2016-17	2015-16
OK LLIGA	1,452	1,631	1,618	1,340	1,467	1,116	1,064

Taula 9: Resultats IBG OK Lliga. Elaboració pròpia

Com podem observar, exceptuant la temporada 2015/16 i la 2016/17, el nostre índex basat en els gols pren valors més elevats que en el cas de les

5 grans lligues de futbol. Més endavant comentarem aquests resultats amb més detall.

6.2.1 Càlcul prova raó de versemblança

Com que hem vist en l'hoquei, estem en el mateix context teòric que en el cas de les 5 grans lligues de futbol, és per aquest motiu que també podem calcular la prova de raó de versemblança descrita en el Teorema 3.3.

Només recordar que en aquest teorema 3.3, el que vam demostrar és que $(2N-1)IBG$ és asimptòticament equivalent a una variable aleatòria khi quadrada amb $2N-1$ graus de llibertat.

En la taula de continuació, es mostren els càlculs de la khi quadrada i la decisió corresponent en funció dels resultats.

Temporades	$(2N-1)IBG$	$\chi^2_{2N-1;\alpha}$	Diferència	Decisió
2021-22	525,48	408,43	117,05	rebutgem H_0
2020-21	776,15	528,92	247,23	rebutgem H_0
2019-20	562,91	393,56	169,34	rebutgem H_0
2018-19	640,71	531,02	109,69	rebutgem H_0
2017-18	701,19	531,02	170,17	rebutgem H_0
2016-17	533,37	531,02	2,35	rebutgem H_0
2015-16	508,37	531,02	-22,65	**acceptem** H_0

Taula 10: Taula de resulats Prova khi quadrada OK Lliga. Elaboració pròpia

Observació 6.5. Per a fer els càlculs, s'ha utilitzat un nivell de significació de 0,05 ($\alpha = 0,05$), i per obtenir els valors de la $\chi^2_{2N-1;\alpha}$ s'ha fet servir la funció d'*Excel* descrita en l'equació (4.4).

6.3 Anàlisis dels Resultats

El que podem observar en la taula 10, és que només en la temporada 2015/16 s'ha acceptat la hipòtesi nul·la i, per tant, és l'única temporada en què teòricament estem en el supòsit de lliga perfectament equilibrada.

Si analitzem amb més detall aquestes 7 temporades, el que veiem és que en totes les temporades el Barça ha dominat amb una clarividència molt evident. Tant és així, que en totes les temporades que estem estudiant excepte en les dues que tenen l'índex basat en els gols inferior (2015/16 i 2016/17), la diferència de gols del Barça és igual o superior a 100. Això vol dir que la diferència entre els gols marcats i rebuts ha estat de 100 gols, fet que indica que la lliga pel que fa a la primera posició no ha estat gens equilibrada durant aquests anys.

El que sí que podem veure és que pel que fa a la resta d'equips, a la temporada on s'accepta la hipòtesis nul·la del test, els equips han obtingut una puntuació més igualada al llarg de la temporada respecte a les altres temporades en què es rebutja la hipòtesi nul·la.
Aquest fet, però no acaba d'explicar per què la lliga a la temporada 2015/16 està equilibrada en comparació a les altres de forma clara.

Com bé hem vist a la Taula 9, els valors de l'IBG han estat bastant superiors en comparació amb els del futbol, ja que s'ha superat el valor de 1,63 en la temporada 2020/21, mentre que en el futbol el valor màxim d'aquest índex ha sigut de 1,34 (temporada 2014/15 de la Lliga Espanyola), on en aquesta temporada sí que hi ha evidència que la lliga estava desequilibrada.

Aquesta diferència en els valors de l'índex s'explica bàsicament perquè en un partit d'hoquei es marquen molts gols, i això fa, que en molts partits, com per exemple els que juga el Barça, aquest número de gols s'allunya de la mitjana de gols de la temporada ($\hat{\lambda}$), i en conseqüència el nostre índex basat en els gols augmenta.

Dit això, podria passar, però que la lliga estigués equilibrada sota el nostre marc teòric encara que en certs partits els gols marcats fossin molt elevats respecta la mitjana de gols per partit de la temporada, fet que el nostre índex això no ho té en compte.

És per aquest motiu que l'índex basat en els gols proposat per a l'hoquei no dona una consistència en els resultats obtinguts per tal d'explicar si una lliga està equilibrada o no.

7 Conclusions

En aquest treball s'ha abordat l'eficiència de l'índex basat en els gols (IBG) per provar si les lligues de futbol estaven perfectament equilibrades o no.

Soc una persona a qui l'apassionen els esports, en especial el futbol i l'hoquei. A través d'aquest treball he pogut combinar les matemàtiques amb els esports i he aprofundit l'aprenentatge en aquests dos conceptes.

Gràcies a l'article de Soudeep Deb. [1], vaig decidir que podria combinar la passió dels esports amb les matemàtiques i dur a terme un treball estadístic interessant.
El treball s'ha realitzat sota el supòsit de què les variables "gols"són independents i segueixen una distribució de Poisson. No és l'única suposició que existeix, però he volgut fer un estudi sobre aquest supòsit, ja que tractava conceptes que havia vist a l'assignatura d'Estadística i trobava molt interessant dur a terme tant l'estudi teòric com pràctic d'aquest cas.

Els resultats del treball han estat força interessants. S'ha provat i vist que l'índex basat en els gols és una mesura teòrica la qual pot explicar si una lliga de futbol està equilibrada o no, i també hem vist que té una relació significativa amb els ingressos de la lliga de futbol. A mesura que augmenta l'índex, la lliga està més desequilibrada i els ingressos disminueixen, mentre que si l'índex disminueix, això resulta en un augment dels ingressos perquè la lliga és més competida i atrau a més espectadors.

Per últim, quan s'ha calculat el mateix índex per a l'hoquei, els resultats obtinguts no han sigut tan eficients. S'ha vist que el supòsit teòric serveix per a calcular amb força certesa la probabilitat d'empat que hi ha en una temporada, però si ens fixem en els resultats de quan la lliga està equilibrada o no, hem vist que els resultats de l'índex eren molt elevats. Aquest fet és degut a que a l'hoquei es marquen molts més gols que al futbol, i per aquest motiu hi ha hagut més desviació, però aquest fet no vol dir que la lliga no estigui equilibrada. S'haurien de tenir en compte més mesures per tal de reduir la influència dels gols, però en aquest treball no s'han pogut abordar.

Referències

[1] Soudeep Deb. (2022). "A goal based index to analyze the competitive balance of a football league". *J.Quant.Anal.Sports* 18(3):171-186.

[2] Michael Bromberg. (2023). "Herfindahl-Hirschman Index (HHI) Definition, Formula". *Investopedia*

[3] P. Dorian Owen. (2007). "Measuring Competitive Balance in Professional Team Sports Using the Herfindahl-Hirschman Index". *https://www.jstor.org/stable/41799360.*

[4] Simon Rottenberg. (1956). "The Baseball Players' Labor Market". *https://www.jstor.org/stable/1825886.*

[5] Will Kenton. (2020). "Concentration Ratio Definition, How to Calculate With Formula". *Investopedia.*

[6] DataHub. (2019). "Data set 5 European football leagues". *Datahub Spanish la liga*

[7] Statista: "Revenue of the Big Five soccer leagues in Europe from 2012/13 to 2021/22, with a forecast to 2023/24, by league". *https://www.statista.com/statistics/261218/big-five-european-soccer-leagues-revenue/*

[8] LiveFutbol. "Standings top 5 european football leagues". *https://www.livefutbol.com/calendario/esp-primera-division-2015-2016-spieltag/38/.*

[9] Irwin, J. O. "The Frequency Distribution of the Difference between Two Independent Variates Following the Same Poisson Distribution." Journal of the Royal Statistical Society, vol. 100, no. 3, 1937, pp. 415–16. JSTOR, *https://doi.org/10.2307/2980526.*

[10] J.G.Skellam. (1946). "The Frequency Distribution of the Difference Between Two Poisson Variates Belonging to Different Populations".*Journal of the Royal Statistical Society.*

[11] FootyStats. "Football stats, Tables and results". *https://footystats.org/.*

[12] Fotmob."Criterios de desempate en la clasificación de LaLiga en España". *https://www.fotmob.com/news/qeucx2l253e71r3544ngh2pdl-criterios-de-desempate-en-la-clasificaci%C3%B3n-de-laliga-en-espa%C3%B1a.*

[13] Antonio Avila-Cano. (2021)."Identifying the Maximum Concentration of Results in Bilateral Sports Competitions".

[14] The-World-Bank. (2024). "World Bank Open Data". *https://data.worldbank.org/.*

[15] Universidad Waterloo. Bessel Functions of the First and Second Kind. *http://mhtlab.uwaterloo.ca/courses/me755/web_chap4.pdf*

[16] Real Federación Española de Patinaje. *https://www.hockeypatines.fep.es/*

[17] J James and Manasis. (2011). "Measurement of Competitive Balance in Professional Team Sports Using the Normalized Concentration Ratio". Volume 31. Economics Bulletin.

[18] Neale, W.C., (1964). The Peculiar Economics of Professional Sports: A Contribution to the Theory of the Firm in Sporting Competition and in Market Competition. The Quarterly journal of economics, vol. 78, no. 1, pp. 1-14.

[19] Quirk,J.P. y Fort, R.D., (2018). Pay dirt: the business of professional team sports. 1. Princeton: Princeton University Press.

[20] Hall,M. y Tideman,N. (1967). Measures of Concentration. Journal of the American Statistical Association, vol. 62, no. 317, pp. 162-168.

[21] KOOPMAN, S.J. y LIT, R., (2015). A dynamic bivariate Poisson model for analysing and forecasting match results in the English Premier League. Journal of the Royal Statistical Society. Series A, Statistics in society, vol. 178, no. 1, pp. 167-186.

[22] PELECHRINIS, K. y WINSTON, W., 2(021). A Skellam regression model for quantifying positional value in soccer. Journal of quantitative analysis in sports, vol. 17, no. 3, pp. 187-201.

[23] Haan, M. A., Koning, R. H., & van Witteloostuijn, A. (2007). Competitive balance in national European soccer competitions. In J. Albert, & R. H. Koning (Eds.), Statistical Thinking in Sports (pp. 57 - 68).

[24] CROISSANT, Y. y MILLO, G., 2008. Panel Data Econometrics in R: The plm Package. Journal of statistical software, vol. 27, no. 2, pp. 1-43.