

DETERMINANTS DELS SALARIS A LA BUNDESLIGA

TREBALL FINAL DE GRAU

MAIG 2024

Alumne: Massa Rahola, Pol
Grau en Economia
Tutora: Berga Colom, Maria Dels Dolors

ÍNDIX

ÍNDIX.....	1
RESUM.....	2
ABSTRACT.....	2
1. INTRODUCCIÓ.....	3
2. LITERATURA PRÈVIA.....	4
3. DADES.....	6
4. ESPECIFICACIÓ DEL MODEL ECONOMÈTRIC.....	9
- MODEL POOLED.....	9
Especificació del model.....	9
Estimació del model.....	10
- MODEL D'EFECTES FIXOS.....	10
Especificació del model.....	10
Estimació del model.....	10
- MODEL D'EFECTES ALEATORIS.....	11
Especificació del model.....	11
Estimació del model.....	11
Validació del model.....	12
- COMPARACIÓ ENTRE MODELS.....	13
5. RESULTATS.....	15
6. CONCLUSIONS.....	19
ANNEXOS:.....	20
BIBLIOGRAFIA.....	31
WEBGRAFIA.....	32

RESUM

Els salaris en el món del futbol professional han estat objecte de debat durant les últimes dècades. En aquest treball, s'analitzen els determinants dels salaris dels jugadors professionals que competeixen en la Bundesliga. L'objectiu és observar si les estadístiques esportives i alguna variable econòmica poden influir en els salaris. Utilitzant dades des de la temporada 2013-2014 fins a la 2022-2023, hem trobat que les variables de rendiment, com ara els partits jugats, les assistències, les targetes vermelles, la competició europea i la classificació en la lliga afecten en els salaris. Però també variables com l'edat i la posició en el camp també poden ser importants. Addicionalment, les variables econòmiques com el PIB per càpita i els ingressos per drets televisius també tenen un impacte en els salaris dels futbolistes de la Bundesliga.

Paraules clau: Bundesliga, salaris, dades de panell, rendiment.

ABSTRACT

The salaries in the world of professional football have been a subject of debate for the past few decades. In this study, the determinants of salaries of professional players competing in the Bundesliga are analyzed. The objective is to observe if sports statistics and some economic variables can influence salaries. Using data from the 2013-2014 season to the 2022-2023 season, we have found that performance variables, such as games played, assists, red cards, European competition, and league ranking, affect salaries. However, variables such as age and position on the field are also important. Additionally, economic variables such as GDP per capita and income from television rights also impact the salaries of Bundesliga footballers.

Key words: Bundesliga, salaries, panel data, performance.

1. INTRODUCCIÓ

El futbol professional és conegut per oferir alguns dels salaris més alts en el món de l'esport, sobretot en l'àmbit europeu, on és l'esport més important. Els jugadors professionals de futbol poden guanyar enormes sumes de diners a través de contractes amb equips, patrocinis i altres oportunitats comercials. Els salaris varien enormement segons la lliga, el club, el país i el prestigi del jugador. Els jugadors més reconeguts i talentosos poden arribar a guanyar milions de dòlars o euros anuals, mentre que els jugadors de nivell inferior o en lligues menys conegudes poden guanyar menys, però encara poden tenir ingressos substancials comparats amb altres professions, això fa que qualsevol nen somiï en ser futbolista professional i representar el seu equip preferit. Tanmateix, els salaris del futbol professional són un tema de gran interès i debat, ja que sovint reflecteixen les estructures econòmiques i les desigualtats socials en el món de l'esport.

El debat sobre els salaris dels futbolistes professionals és intens i dividit. Mentre alguns veuen les altes remuneracions com a justificades pel valor que aporten, l'entreteniment que generen i les grans audiències que atrauen, altres consideren els salaris excessius i fora de proporció. Els estudis han suggerit una correlació entre el salari i el rendiment del jugador, però també han destacat la influència de factors com l'eficiència del club i la durada del contracte en la determinació dels salaris.

Un altre tema de controvèrsia al món del futbol són els ingressos per televisió. Per una banda, els acords multimilionaris han impulsat el creixement econòmic dels clubs, permetent inversions en jugadors de primer nivell i infraestructures de qualitat. No obstant això, aquesta prosperitat econòmica ha generat desigualtats significatives entre els clubs, les lligues més riques reben acords lucratius, mentre que les menys privilegiades troben dificultats per competir econòmicament, erosionant la igualtat d'oportunitats com veiem en l'article publicat per Carreras i Garci.

En aquest treball, analitzem els determinants dels salaris dels jugadors de la Bundesliga entre els anys 2013 i 2023. L'objectiu és comprendre com les variables esportives i de rendiment, com ara el nombre de gols, assistències i minuts jugats, influeixen en els salaris dels jugadors. Però també, volem observar la influència de variables econòmiques, com el PIB per càpita de les regions alemanyes, per entendre com la situació econòmica pot afectar la capacitat dels clubs per pagar salaris competitius, i l'impacte dels ingressos per drets televisius mencionats abans, per mesurar la seva importància.

2. LITERATURA PRÈVIA

En aquest apartat mencionarem algunes investigacions i recerca prèvia sobre la relació entre els salaris dels jugadors en el futbol i diverses variables, vinculades amb el seu rendiment i el context de la lliga.

Primer, ens centrarem en l'estudi del 2017 de Yaldo i Shamir, aquest ens dona una visió bastant general sobre la relació entre els salaris i diferents variables determinants sobre el rendiment, on s'utilitza mètodes computacionals per estimar els salaris basant-se en el rendiment esportiu dels jugadors i proposa un mètode objectiu i quantitatiu per determinar els salaris dels futbolistes basant-se en les seves habilitats. Els resultats experimentals amb dades de 6.082 jugadors mostren que hi ha una forta relació entre les habilitats dels jugadors i els seus salaris, troben que variables de rendiment com els gols, assistències i passes, variables de comportament com l'agressivitat i el lideratge, i les variables físiques com la velocitat i la força, tenen una influència en els salaris dels jugadors.

Explorem també diferents treballs més centrats en un territori concret i que van més enllà del rendiment dels jugadors, buscant altres variables que poden ser rellevants a l'hora d'estudiar els salaris dels futbolistes professionals. Per exemple, l'any 2012, Lee i Harris, es van centrar en la Major League Soccer (MLS) dels Estats Units, analitzant com el rendiment individual dels jugadors es relaciona amb el seu salari. Els resultats van mostrar que hi ha una correlació positiva entre les dues variables, és a dir, els jugadors que tenen un millor rendiment individual guanyen més diners. D'altra banda, en l'àmbit sud-americà, trobem la recerca de De Menezes i de De Santana de l'any 2022, on analitzen la primera divisió del futbol brasiler, entre l'any 2003 i 2019, examinant com les desigualtats regionals influeixen en els salaris dels jugadors.

En el context europeu, l'estudi de Caruso, Di Domizio i Rossignoli de l'any 2017, examina la lliga italiana i la relació entre el salari dels jugadors i el rendiment de l'equip, revelant una correlació positiva entre la quantitat total de diners invertits en els sous dels jugadors i l'èxit dels equips a la lliga italiana. Mentrestant, l'article de Ribeiro i Lima de l'any 2012, se centra en la lliga portuguesa, investigant la relació entre l'eficiència dels equips i els salaris dels jugadors, mesurada com la relació entre recursos invertits i resultats obtinguts. Observen que alguns clubs de la Primera Lliga Portuguesa podrien estar gastant més diners en sous dels jugadors del que és necessari per aconseguir els seus objectius.

Finalment, comentem dos treballs addicionals sobre la lliga alemanya de futbol que ofereixen perspectives interessants, com que el nostre estudi es basa en dades de la Bundesliga és important veure els estudis previs en el territori alemany. Frick l'any 2011 estudia la relació entre el rendiment individual dels jugadors i la durada dels seus contractes amb els seus salaris individuals, obté una correlació positiva entre les variables de rendiment i la remuneració tal com hem vist en la majoria dels treballs citats, també comenta que el rendiment dels jugadors augmenta en el seu últim any de contracte. Mentre que Wilkesmann

l'any 2017 investiga com els diners invertits en salaris es relacionen amb el rendiment esportiu dels equips. Els seus resultats mostren com els equips amb majors pressupostos assoleixen una millor classificació a la lliga, i com aquest factor genera una desigualtat econòmica entre els equips, deguda en part pels ingressos de les competicions europees.

En termes generals, aquests estudis suggereixen una correlació positiva entre el salari d'un futbolista i el seu rendiment individual. No obstant això, també assenyalen que altres factors com el rendiment de l'equip, l'eficiència del club, les desigualtats regionals i la durada del contracte poden influir de manera significativa en els salaris dels jugadors. Aquests resultats posen de manifest la complexitat dels determinants salarials en el món del futbol i la importància de considerar múltiples variables en analitzar aquesta relació.

A partir del que hem pogut observar a la literatura, la motivació d'aquest treball és comprovar quines variables sobre el rendiment dels jugadors són importants per determinar el salari que rep un jugador de la Bundesliga. Com hem mencionat anteriorment, també volem veure si algunes variables fora del merament esportiu tenen una influència sobre els salaris, siguin variables territorials o econòmiques.

3. DADES

En el nostre estudi, hem utilitzat dades de la Bundesliga des de la temporada 2013-2014 fins a la temporada 2022-2023. Aquest període s'ha determinat segons la disponibilitat de dades sobre els salaris dels jugadors de la lliga alemanya. Això significa que les altres dades recopilades estan limitades als anys corresponents a aquesta franja temporal. Aquesta selecció específica de dades ens permet centrar-nos en els canvis i tendències observats durant aquests anys a la Bundesliga.

Les variables que són estadístiques, és a dir, que fan referència a aspectes del rendiment esportiu del jugador, estan basades en els seus números en la lliga domèstica. Això significa que únicament tenim en compte les seves estadístiques en partits de lliga, excluint altres competicions com la copa nacional o les competicions europees. Hem pres aquesta decisió principalment per assegurar la consistència i precisió de les dades. Si incloguéssim partits de la copa nacional, on poden enfrontar-se a equips de divisions inferiors, o competicions europees amb rivals de major nivell, es podria distorsionar la comparabilitat i, per tant, la precisió de les nostres variables. D'aquesta manera, garantim que el número de partits potencials i l'exigència competitiva siguin constants per a tots els jugadors analitzats. Aquesta metodologia ha sigut utilitzada en alguns treballs de la literatura prèvia i ens serà útil pel nostre estudi per oferir una igualtat de condicions.

Per construir la nostra base de dades, hem recopilat dades de diverses fonts citades en la bibliografia. Les estadístiques de rendiment, a les quals ens referim en el paràgraf anterior, juntament amb les dades particulars dels jugadors com ara nom, equip o nacionalitat, han estat extretes directament de la llibreria del R-Studio anomenada "Worldfootballr". Aquesta eina ha estat instrumental per a la recopilació d'aquest tipus de dades, ja que agrupa informació de diverses pàgines web en un únic lloc, facilitant així la creació de la nostra base de dades. En el nostre cas, hem utilitzat dades extretes de la web FBRef, que proporciona estadístiques detallades dels jugadors per temporada. Aquestes estadístiques es refereixen a accions quantificables durant un partit, com ara gols o assistències. Això vol dir que cada vegada que un futbolista marca un gol, aquesta informació s'incorpora a les estadístiques i podem determinar el nombre total de gols que ha marcat durant la temporada analitzada. A més, de la mateixa plataforma, obtenim informació detallada sobre el nom, l'edat, l'equip, la posició en el camp, la nacionalitat i la temporada dels jugadors. La variable principal del nostre estudi, que és el salari dels jugadors, també s'extreu d'aquesta web, proporcionant-nos dades en tres monedes diferents i en format setmanal i anual.

Les altres variables de la base de dades, han sigut afegides manualment i extretes de diferents pàgines webs. Les variables sobre el rendiment de l'equip com la posició de la lliga i la participació en competicions europees s'han recopilat de la pàgina oficial de la Bundesliga. Les dades econòmiques com el PIB per càpita, el PIB regional i el nom de la regió s'han obtingut de l'Eurostat, ja que, ens donava el PIB i PIB per càpita de cada regió d'Alemanya durant els anys analitzats en el treball. Per últim, la variable que fa referència als ingressos

anuals per televisió s'ha extret de la pàgina "Fernsehgelde.de", una web alemanya on es troba documentat els ingressos que rep cada equip de la Bundesliga i de la segona divisió des de la temporada 2012/2013. Per acabar la base de dades, hem corregit les dades econòmiques que tenim per l'IPC de cada regió, l'IPC l'hem aconseguit de la pàgina oficial d'estadística del país alemany, que també podem trobar a la bibliografia.

Amb la nostra base de dades completada, seleccionem el salari anual en euros, ajustat per l'IPC, com a variable dependent principal. Per a l'anàlisi, considerem les temporades com a sèrie temporal i els noms dels jugadors com a secció creuada. Això indica que hem fet seguiment de 1560 jugadors durant el període d'estudi.

En referència a les variables independents, inicialment en teníem un total de 27. No obstant això, en analitzar la seva correlació, vam identificar-ne diverses que mostraven correlacions fortes entre elles. Les variables que són creades a partir d'altres i les variables que feien referència a l'equip, la nacionalitat i la regió ens indicaven problemes potencials de multicolinealitat i autocorrelació, per la qual cosa van ser eliminades del nostre conjunt de variables. Això ens deixa amb un total de 12 variables independents per a la nostra anàlisi de regressió posterior, les quals trobem definides a la Taula 1. La resta de variables les podem trobar definides a la Taula 4 de l'Annex 1.1.

Taula 1: Descripció de les variables. Font: Elaboració pròpia

Variable	Definició
<i>AnnualWageEUR_ipc</i>	Salari anual del jugador en € corregit per l'IPC de la seva regió.
<i>Player</i>	Nom i primer cognom del jugador.
<i>Season</i>	Any en què es duu a terme la temporada de lliga.
<i>Pos</i>	Variable qualitativa sobre la posició que ocupa en el camp. Hi ha quatre grups: Porter (GK), Defensa (DF), Migcampista (MF) i Davanter (FW).
<i>Age</i>	Edat del jugador en cada temporada.
<i>Matches_played</i>	Número de partits en què participa el jugador.
<i>Gls</i>	Número de gols totals en una temporada.
<i>Ast</i>	Número de assistències totals en una temporada.
<i>PK</i>	Número de gols de penal totals en una temporada.
<i>CrdY</i>	Número de targetes grogues totals en una temporada.
<i>CrdR</i>	Número de targetes vermelles totals en una temporada.
<i>pib_capita_ipc</i>	PIB per càpita de la regió on es troba situat l'equip del jugador, corregit per l'IPC de la regió. (en euros)
<i>eu_comp</i>	Variable qualitativa sobre la competició europea en què participen els jugadors. Hi ha quatre grups: Champions League (UCL), Europa League (UEL), Conference League (CFL) i no participants (NO).
<i>league_position</i>	Posició final en la classificació de la lliga nacional.
<i>tv_income_ipc</i>	Ingrés que rep l'equip del jugador per drets televisius cada temporada, corregit per l'IPC de la regió.

4. ESPECIFICACIÓ DEL MODEL ECONOMÈTRIC

El nostre estudi adopta l'anàlisi de dades de panell, una metodologia que fusiona conceptes de les sèries temporals (relacionades amb les temporades) i dels models de secció creuada (associats als jugadors). Aquesta tècnica ens permet explorar la dinàmica de les variables tant en el temps com entre diferents jugadors, oferint una perspectiva més completa i detallada dels fenòmens estudiats. Permet ajustar-se a les característiques úniques de cada jugador i entendre com aquestes varien al llarg del temps.

Dins d'aquest marc, podem aplicar diversos models per analitzar les dades de panell, concretament, veurem tres models en específic: el model pooled, el model d'efectes aleatoris i el model d'efectes fixos. Al final de l'apartat, comparem aquests models amb l'objectiu de seleccionar i justificar el model que millor s'ajusta a les nostres dades i objectius d'investigació.

- MODEL POOLED

El model pooled, també conegut com a model de dades agrupades, es combinen totes les observacions de les diferents entitats (en aquest cas, jugadors de futbol) en un únic conjunt de dades. Això significa que es consideren tots els individus com si pertanguessin a una única mostra gran. En el model pooled, no es tenen en compte les diferències individuals entre els futbolistes professionals al llarg del temps, ja que es tracta totes les observacions com a independents entre elles. Això pot ser útil per a estudis inicials on l'objectiu principal és avaluar la relació entre les variables sense tenir en compte les diferències individuals o temporals. No obstant això, aquest model pot no ser adequat si hi ha heterogeneïtat entre les entitats o si les variables independents no són veritablement exògenes.

Per especificar el nostre model pooled, tenim en compte que la variable dels salaris és la variable dependent, les temporades és la sèrie temporal, els jugadors com a secció creuada i la resta de variables com a independents. A conseqüència del fet que les variables econòmiques com el salari dels jugadors, el PIB per càpita i els ingressos per televisió, numèricament, estan molt desproporcionades de la resta de variables numèriques, els hi aplicarem una transformació logarítmica. L'especificació del model pooled serà el següent:

Especificació del model

$$\log(\text{AnnualWageEUR_ipc})_{jt} = \beta_0 + \beta_1 \text{Age}_{jt} + \beta_2 \text{Pos}_{jt} + \beta_3 \text{Matches_played}_{jt} + \beta_4 \text{Gls}_{jt} + \beta_5 \text{Ast}_{jt} + \beta_6 \text{Crdr}_{jt} + \beta_7 \log(\text{pib_capita_ipc})_{jt} + \beta_8 \text{eu_comp}_{jt} + \beta_9 \text{league_position}_{jt} + \beta_{10} \log(\text{tv_income_ipc})_{jt} + u_{jt}$$

On j denota els jugadors i t la temporada.

Estimació del model

Basant-nos en l'anàlisi del nostre model pooled, que es detalla als annexos, observem que el model explica al voltant del 51,68% de la variabilitat en el logaritme del salari anual ajustat per l'IPC (vegeu el Multiple R-squared). Tenint en compte el nombre de predictors, l'Adjusted R-squared revela que el model cobreix aproximadament el 51,53% de la variabilitat d'aquesta mateixa variable, això és un percentatge bastant alt en termes estadístics. L'anàlisi de l'estadístic F i el p-valor, confirma la significativitat global del model, suggerint que almenys una de les variables predictores és rellevant per predir la nostra variable dependent. En aquest cas, totes les variables són significatives, tal com podem veure en l'estimació del model.

En resum, aquest model proporciona una visió detallada de com diferents factors com l'edat, la posició del jugador, el rendiment en el camp, el PIB per càpita, la participació en competicions europees, la posició a la lliga i els ingressos televisius influeixen en el salari dels jugadors de futbol.

- MODEL D'EFECTES FIXOS

El model d'efectes fixos és un mètode d'anàlisi de dades de panell que controla les variacions constants específiques de cada entitat (en aquest cas, jugadors de futbol) en l'equació de regressió. Aquest model elimina aquestes variacions inobservables per analitzar les relacions entre les variables independents i la dependent sense la seva interferència. L'objectiu és proporcionar estimacions més precises de l'impacte de les variables independents sobre la variable dependent en dades de panell.

En aquest model hem utilitzat les mateixes variables que el model pooled, però hem trobat que moltes variables ja no eren significatives. Concretament, l'especificació del model d'efectes fixos ha sigut la següent:

Especificació del model

$$\log(\text{AnnualWageEUR_ipc})_{jt} = \beta_0 + \beta_1 \text{Age}_{jt} + \beta_2 \text{Matches_played}_{jt} + \beta_3 \text{CrdR}_{jt} + \beta_4 \log(\text{pib_capita_ipc})_{jt} + \beta_5 \text{league_position}_{jt} + \beta_6 \log(\text{tv_income_ipc})_{jt} + u_j + \epsilon_{jt}$$

On j denota els jugadors, t la temporada, u_j són els efectes individuals inobservables constants i ϵ_{jt} són els termes d'error residuals.

Estimació del model

Basant-nos en l'anàlisi del nostre model d'efectes fixos, que també es detalla als annexos, observem que el valor de R quadrat és de 0,31442, cosa que significa que aproximadament el 31,44% de la variabilitat en el logaritme del salari anual ajustat per l'IPC es pot explicar per les variables independents en aquest model. Això indica una capacitat moderada del model per explicar la variabilitat de la variable dependent. L'Adjusted R-squared és -0,050269, que

és inusualment baix i pot suggerir que alguna de les variables podria no ser rellevant o que el model podria estar sobreajustant-se. L'estadístic F és de 224.881 amb un p-valor molt baix ($< 2.22 \cdot 10^{-16}$), el que indica que el model en conjunt és significatiu i millor que un model sense cap variable.

En resum, el model té una capacitat moderada per explicar la variabilitat en el salari dels jugadors de futbol basat en les variables seleccionades. No obstant això, seria aconsellable revisar l'Adj. R-squared i investigar més profundament la presència d'heteroscedasticitat o la necessitat d'incorporar altres variables que podrien millorar la qualitat del model.

- MODEL D'EFECTES ALEATORIS

El model d'efectes aleatoris és un mètode en anàlisi de dades de panell que considera que els coeficients de les variables independents poden variar entre les unitats d'anàlisi. A diferència del model d'efectes fixos, aquest model incorpora un component aleatori per capturar la variabilitat no observada entre les unitats. Això proporciona una visió més flexible de les relacions entre les variables i és útil per fer inferències en poblacions més àmplies.

En el cas d'aquest model, també utilitzarem les mateixes variables que el model pooled, però hem trobat que la variable Gls, que mesura els gols d'un jugador deixa de ser significativa, per tant, l'especificació del model és la següent:

Especificació del model

$$\log(\text{AnnualWageEUR_ipc})_{jt} = \beta 1_j + \beta 2 \text{ Age}_{jt} + \beta 3 \text{ Pos}_{jt} + \beta 4 \text{ Matches_played}_{jt} + \beta 5 \text{ Ast}_{jt} + \beta 6 \text{ CrdR}_{jt} + \beta 7 \log(\text{pib_capita_ipc})_{jt} + \beta 8 \text{ eu_comp}_{jt} + \beta 9 \text{ league_position}_{jt} + \beta 10 \log(\text{tv_income_ipc})_{jt} + u_{jt}$$

$$\beta 1_j = \beta 1 + n_j$$

Estimació del model

Un cop tenim el model d'efectes aleatoris especificat, analitzem la seva estimació per poder veure'n la qualitat. L'estimació d'aquest model amb r-studio també la podem trobar als annexos.

El model d'efectes aleatoris presentat mostra diverses característiques que indiquen la seva qualitat i adequació. El R-quadrat d'aquest model és de 0,86549, que és bastant alt. Aquest valor indica que aproximadament el 86,55% de la variabilitat en el logaritme del salari anual ajustat per l'IPC es pot explicar pels predictors inclosos en el model. El R quadrat ajustat també és elevat (0,8651), corroborant la robustesa del model. La majoria dels predictors del model són estadísticament significatius amb valors de $p < 0,05$, la qual cosa suggereix que aquests predictors tenen un efecte significatiu sobre el salari dels jugadors. L'estadístic F del model és de 2810 amb un p-value de $< 2,22 \cdot 10^{-16}$, aquest resultat, proposa que el model en

conjunt és altament significatiu i que almenys un dels predictors és útil per predir el logaritme del salari anual. La distribució dels residuals mostra que la majoria d'aquests són propers a zero, indicant que el model no presenta un patró clar d'heteroscedasticitat o no-linealitat, la qual cosa és un bon signe per a la seva fiabilitat.

En resum, el model d'efectes aleatoris ofereix una explicació forta i coherent de la relació entre els diferents predictors i el salari dels jugadors.

Validació del model

En aquest apartat, mirem si el model compleix una sèrie de supòsits per ser validat. Primer elaborem el diagnòstic dels residuals, si observem els dos gràfics dels annexos, el gràfic de residuals contra valors ajustats i el gràfic de variància dels residus contra els valors predits, podem observar que el model pot presentar problemes d'heteroscedasticitat, però no sembla molt preocupant segons veiem en els gràfics. En el gràfic de residuals contra valors ajustats, els punts estan repartits al voltant del zero i no hi ha cap patró important, que ens generi un problema d'heteroscedasticitat important. El gràfic de variància dels residus contra els valors predits, podem considerar que no hi ha una distribució constant de la variància dels residuals al llarg de l'eix de valors predits i veiem una possible tendència on la variància dels residuals augmenta o disminueix a mesura que augmenten els valors predits, això és característic de l'heteroscedasticitat.

Llavors mesurem la multicol·linealitat, es refereix a una situació en la qual dues o més variables independents en un model estadístic estan altament correlacionades, de manera que proporcionen informació similar sobre la variable dependent. Aquesta correlació pot fer que sigui difícil determinar l'efecte individual de cada variable sobre la variable dependent. Un dels mètodes més comuns és calcular el factor d'inflació de la variància (Variance Inflation Factor, VIF). Un VIF alt suggeriria alta multicol·linealitat. Els valors VIF que hem obtingut, mostrats als annexos, indiquen la presència de multicol·linealitat baixa en el nostre model. Com a regla general, si un valor VIF és significativament més gran que 1, ja pot ser considerat com a indicatiu de multicol·linealitat, però valors superiors a 5, són els valors realment vistos com a indicatius d'alta multicol·linealitat. Per tant, podem dir que el model compleix aquest supòsit, ja que, veiem en els resultats que la majoria dels valors VIF estan per sota del llindar comunament utilitzat de 5.

Finalment, analitzem la normalitat dels residus. Com es pot observar en el gràfic de normalitat inclòs en els annexos, confirmem que es compleix aquest supòsit. Els residus segueixen la bisectriu excepte als extrems, on hi ha una major dispersió. A més, amb una mostra de $n = 4508$, que supera els 30 casos recomanats pel Teorema del Límit Central (TLC), podem aproximar-nos a una distribució normal. Per tant, es verifica el supòsit de normalitat.

En resum, després d'analitzar detalladament els diferents supòsits que un model estadístic ha de complir per ser considerat vàlid, el nostre model ha demostrat una conformitat adequada

amb cadascun d'ells. Tot i les lleugeres anomalies analitzades en els gràfics de diagnòstic dels residuals, aquestes no semblen ser greus i no afecten de manera substancial la integritat del model. Així, amb la confirmació de l'absència de multicol·linealitat significativa i l'aproximació a una distribució normal dels residus, podem tenir confiança en les conclusions i interpretacions derivades del model. Aquesta validesa del model ens proporciona una base robusta per continuar amb les anàlisis i les interpretacions relacionades amb les variables d'estudi i les seves relacions amb la variable dependent.

- COMPARACIÓ ENTRE MODELS

En aquest apartat, observarem la decisió final sobre quin model hem d'escollir per continuar amb el nostre estudi. Per escollir l'adequat, compararem els tres models que hem vist fins ara, a la següent taula, trobem les estimacions d'aquests models. Per tant, per comparar aquests models, el model pooled, el model d'efectes fixos i el model d'efectes aleatoris utilitzarem diverses mètriques i criteris estadístics que avaluen la qualitat i adequació de cada model.

Taula 2: Estimacions dels models. Font: R-studio

Variables	Model Pooled		Model d'efectes fixes		Model d'efectes aleatoris	
CONSTANT	-6.865195	4.76e-10 ***			-3,3638689	0.0012052 **
AGE	0.125496	< 2e-16 ***	0.0850390	< 2.2e-16 ***	0.1233746	< 2.2e-16 ***
POS (FW)	0.021846	0.58704			0.0498770	0.2465008
POS (GK)	-0.388518	2.48e-14 ***			-0,5749515	2.487e-13 ***
POS (MF)	0.045003	0.17744			0.0359697	0.3380276
MATCHES_PLAYED	0.035678	< 2e-16 ***	0.0151246	< 2.2e-16 ***	0.0242449	< 2.2e-16 ***
GLS	0.020364	0.00017 ***				
AST	0.024524	0.00207 **			0.0155798	0.0147947 *
CRDR	0.106996	0.00682 **	0.0605324	0.0326256 *	0.0792723	0.0067976 **
LOG (PIB_CAPITA_IPC)	0.639927	1.89e-15 ***	0.3210996	0.0009123 ***	0.3905906	3.272e-06 ***
EU_COMP (NO)	0.700623	1.07e-07 ***			0.3360484	0.0006129 ***
EU_COMP (UCL)	1,295578	< 2e-16 ***			0.5924128	5.999e-09 ***
EU_COMP (UEL)	0.885217	6.17e-11 ***			0.3927236	0.0001002 ***
LEAGUE_POSITION	-0,030761	< 2e-16 ***	-0,006313	0.0175058 *	-0,0187798	7.521e-13 ***
LOG (TV_INCOME)	0.528913	< 2e-16 ***	0.6480950	< 2.2e-16 ***	0.5107378	< 2.2e-16 ***
Nº Observacions	5.008		4.508		4.508	
R² Ajustat	0.5153		-0.050269		0.8651	
P-valor	< 2.2e-16		< 2.22e-16		< 2.22e-16	

El model pooled té un R quadrat ajustat més alt (0,5153) en comparació amb el model d'Efectes Fixos (-0,050269) i el model d'efectes aleatoris té un R quadrat més alt (0,8651), això ens indica que el model d'efectes aleatoris explica una proporció més gran de la variabilitat en el logaritme del salari anual ajustada per la quantitat de predictors en el model i el nombre d'observacions. En els tres models econòmics, el p-valor de l'estadístic F és molt baix (< 0,05), indicant que tots dos models són globalment significatius i també en ambdós models, les variables independents són significatives. Però els coeficients dels predictors són més significatius en el model d'efectes aleatoris, amb valors p més baixos, indicant una major confiança en la relació entre els predictors i la variable de resposta. Per acabar, si mirem els gràfics dels residuals dels annexos, veiem que els residus del model d'efectes aleatoris i del

model d'efectes fixos estan distribuïts de manera més equilibrada al voltant de zero en comparació amb el model pooled, que presenta una dispersió més gran.

Per tant, basant-nos en aquesta anàlisi comparativa, el model d'efectes aleatoris sembla ser el més adequat i el més robust per a aquest conjunt de dades. Aquest model proporciona una millor explicació de la variabilitat observada en el salari dels jugadors de futbol i presenta una major significació estadística dels seus coeficients.

5. RESULTATS

El model seleccionat per analitzar els salaris dels futbolistes de la Bundesliga és el model d'efectes aleatoris. Aquesta elecció ens ha permès identificar les nou variables independents que tenen un impacte significatiu sobre la nostra variable dependent. A continuació, explorarem com aquestes variables poden influir en la nostra variable objectiu, a partir dels coeficients de l'estimació del model d'efectes aleatoris que hem vist anteriorment.

En primer lloc, trobem la variable *Age*, que com ja sabem aquesta fa referència a l'edat del futbolista. Observem que un augment d'un any en l'edat està associat amb un augment del 12,34% en la variable del salari, mantenint constants les altres variables. Per tant, es valora positivament els futbolistes amb més edat que els joves, en termes salarials. Segurament, això es pot explicar per l'experiència o la maduresa dels jugadors, és a dir, els jugadors més antics sovint tenen més experiència i maduresa tàctica, que són valorades pels clubs i aquest factor pot traduir-se en un salari més alt. També cal tenir en compte el prestigi o la reputació, és a dir, els jugadors amb més anys de carrera poden haver establert una reputació o un prestigi en el món del futbol, la qual cosa pot augmentar el seu valor en el mercat.

En segon lloc, trobem la variable *Pos*, que fa referència a la posició que ocupa el jugador al terreny de joc. El model agafa la posició de defensa com a referència, així que només podem interpretar la resta de posicions respecte d'aquesta. L'única posició significativa és la de porter (*GK*), observem que el coeficient és de -0,575 amb un valor p molt baix de $2,487 \times 10^{-13}$. Aquest coeficient negatiu indica que, de mitjana, els porters guanyen significativament menys que la posició de referència, és a dir, els porters cobren un 5,75% menys que els defenses. Això pot ser degut a diverses raons com ara la menor demanda per a aquesta posició o la menor exposició mediàtica que reben en comparació amb altres jugadors de camp. Pel que fa als davanters i migcampistes, semblen tenir salaris lleugerament més alts, però aquestes diferències no són estadísticament significatives, la qual cosa vol dir que poden haver altres factors no considerats en el model que influeixin en els salaris d'aquestes posicions.

En tercer lloc, trobem la variable dels partits jugats (*Matches played*), observem que el coeficient associat és de 0,0242 amb un valor p extremadament baix de $< 2,2 \times 10^{-16}$, per tant, això indica que hi ha una relació positiva significativa entre el número de partits jugats per un

jugador i el seu salari, és a dir, participar en un partit més implica un augment del 4,42% del salari. Doncs, els jugadors que han participat en més partits tendeixen a tenir salaris més alts, aquesta correlació pot ser atribuïda a diversos factors. Per exemple, els jugadors que juguen més partits solen ser considerats com a jugadors clau o titulars en els seus equips, i per tant poden tenir millors contractes i salaris més elevats. Addicionalment, la consistència i la capacitat de mantenir-se actiu en un nombre més gran de partits poden ser vistes com a indicadors de la qualitat i el rendiment del jugador, el que pot traduir-se en un salari més alt.

En quart lloc, trobem les assistències que genera un jugador al llarg de la temporada. En aquest cas, el coeficient associat a les assistències (*Ast*) en el model és de 0,0156 amb un valor p de 0,0148. Aquest coeficient positiu i significatiu suggereix que hi ha una relació positiva entre el número d'assistències d'un jugador i el seu salari, dient que una assistència més, genera un augment d'1,56% en la variable del salari. Els jugadors que realitzen més assistències tenen tendència a tenir salaris més alts. Les assistències són un indicador de la capacitat d'un jugador per proporcionar oportunitats d'anotació als seus companys d'equip. Un jugador que pot fer assistències eficaces és valorat no només per la seva capacitat individual, sinó també per la seva contribució a l'èxit col·lectiu de l'equip. Això pot reflectir-se en el seu contracte i en el seu salari. En molts casos, les assistències poden ser tan importants com els mateixos gols, ja que demostren la visió de joc, habilitat i comprensió tàctica d'un jugador. Per tant, les assistències no només són un indicador del talent individual d'un jugador, sinó que també poden influir en la seva valoració econòmica dins del mercat de futbol.

En cinquè lloc, trobem la variable *CrdR*, el número de targetes vermelles. En aquest cas, trobem un resultat bastant inesperat, però el coeficient associat a les targetes vermelles en el model és de 0,0793, això ens indica que hi ha una relació positiva entre el número de targetes vermelles rebudes per un jugador i el seu salari. És a dir, els jugadors que reben més targetes vermelles tenen tendència a tenir salaris més alts, una targeta vermella més implica un augment del 7,93% de la variable del salari. Aquesta relació pot semblar contraintuïtiva a primera vista, ja que habitualment les targetes, especialment les vermelles, són vistes com a indicadors de comportament antiesportiu o d'errors defensius greus. No obstant això, aquest resultat pot interpretar-se de diverses maneres, pot ser que alguns equips poden valorar la passió, l'agressivitat i la determinació d'un jugador, encara que això comporti un risc d'incidències que condueixin a targetes o pot ser que els jugadors defensius que juguen en posicions crítiques (com ara defensors centrals o migcampistes defensius), s'enfronten a situacions de perill més sovint i poden rebre més targetes en general que la resta. Tot i aquesta relació positiva entre targetes vermelles i salari, és essencial recordar que aquesta correlació no implica causalitat, és a dir, els jugadors no reben salaris més alts simplement per rebre targetes, sinó que altres factors relacionats amb la seva habilitat, experiència i contribució a l'equip també juguen un paper crucial en la seva valoració econòmica.

En sisè lloc, trobem la variable *pib_capita_ipc*, fa referència al PIB per càpita de la regió on està situada l'equip del jugador. El coeficient associat al logaritme del PIB per càpita en el model és de 0,3906 amb un valor p molt significatiu, aquest coeficient positiu i altament significatiu indica que hi ha una relació positiva entre el PIB per càpita de la regió on es troba

l'equip i el salari dels jugadors. Per tant, un augment de l'1% del PIB per càpita de la regió implica un augment del 0,39% dels salaris dels futbolistes de la Bundesliga. Podem interpretar que un PIB per càpita més alt pot ser indicatiu d'una economia més forta i estable, la qual cosa pot permetre als clubs de futbol invertir més en salaris elevats. També es pot argumentar que un PIB per càpita més alt, permet als aficionats poder tenir un major poder adquisitiu, que pot augmentar la recaptació del club a través de la venda d'entrades, marxandatge i drets de televisió i aquest augment pot permetre als clubs pagar salaris més alts als seus jugadors. Tot i aquesta relació positiva, és essencial entendre que la causalitat no és directa, els salaris dels futbolistes no augmenten simplement a mesura que el PIB per càpita creix.

En setè lloc, trobem la influència de les participacions en competicions europees en els salaris dels futbolistes professionals. El model agafa la competició Conference League (CFL) com a referència, així que només podem interpretar la resta de competicions respecte d'aquesta. En aquest cas, els coeficients de les competicions són positius i significatius, per tant, ens indiquen una relació positiva entre la participació dels clubs en diferents competicions europees i els salaris dels seus jugadors. El coeficient més alt associat a la participació en competicions europees, és la UEFA Champions League (UCL), ens indica que els jugadors que participen en aquesta competició europea tenen salaris un 59,24% més elevats que els jugadors que participen en la CFL. Tot i que el coeficient associat a la UEL és inferior al de la UCL, encara és significatiu i positiu, això indica que també hi ha una prima salarial per a la participació en aquesta competició, els jugadors cobren un 39,27% més que els jugadors que participen en la CFL. Els clubs que no participen en cap competició europea, també un impacte positiu en els salaris dels seus jugadors en comparació amb la Conference League, tenen uns salaris un 33,6% més alts que els jugadors que participen en la Conference. Això, pot ser degut a l'esforç econòmic que fan aquests equips en els salaris dels seus jugadors, per tal de poder classificar a la UCL o a la UEL, on obtenen beneficis més alts que en la CFL. Per tant, això suggereix que la participació dels clubs en competicions europees no només augmenta el seu prestigi i visibilitat, sinó que també té un impacte directe en la seva capacitat per pagar salaris més competitius als seus jugadors. A més, aquesta prima salarial és més significativa per a la Champions League, la qual cosa reflecteix el valor més gros i el prestigi associat a aquesta competició en comparació amb la UEFA Europa League i la falta de participació en competicions europees.

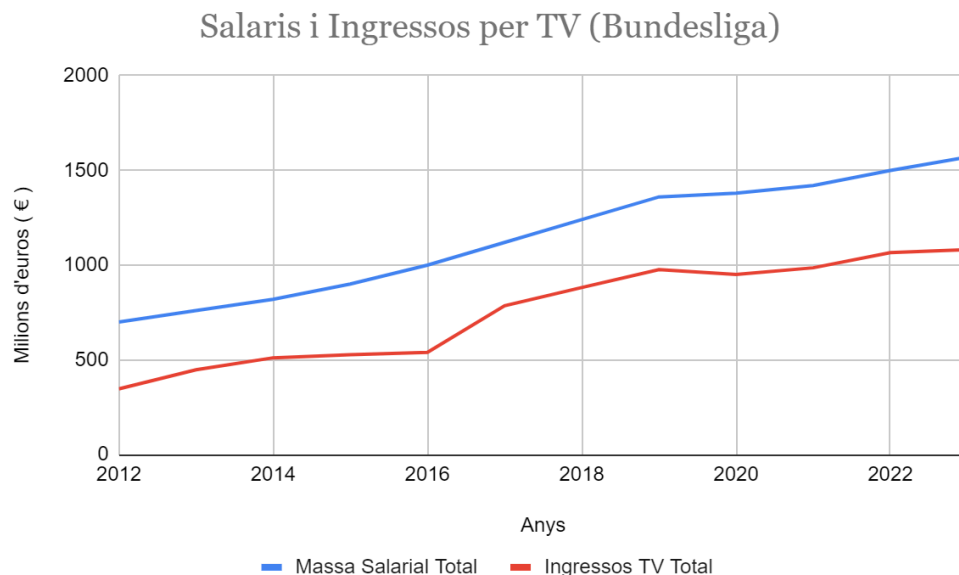
En vuitè lloc, trobem la posició en la classificació de la lliga. En aquest cas, el coeficient negatiu en la variable *league_position* suggereix que a mesura que un club s'enfonsa en la classificació de la lliga (és a dir, baixa posicions), el salari mitjà dels seus jugadors tendeix a disminuir. És a dir baixar una posició a la lliga està associada amb una disminució del salari dels jugadors en un 1,88%, segons les dades del model. Aquesta reducció podria ser una conseqüència de la disminució de la rendibilitat del club o de la necessitat d'ajustar el pressupost a la nova situació esportiva.

En últim lloc, trobem una de les variables més influents del nostre model, es tracta dels ingressos per drets televisius. En aquest cas, el coeficient associat a *log(tv_income_ipc)* és de

0,5107, que ens indica que per cada augment de l'1% en els ingressos per televisió (ajustat per l'IPC de cada regió), el salari dels jugadors augmenta en un 0,51%, un augment significatiu i suggereix que els ingressos televisius són un factor crucial que influeix en els salaris dels futbolistes de la Bundesliga. Aquesta relació pot ser explicada pel fet que els ingressos televisius són una de les principals fonts de finançament per als clubs de futbol, és a dir, si els clubs reben més diners de les vendes de drets de televisió, tenen més recursos disponibles per invertir en la contractació de jugadors i, per tant, poden oferir salaris més alts. A més, aquesta troballa reflecteix l'increment constant del valor dels drets de televisió en el món del futbol, on els transmissors estan disposats a pagar quantitats elevades per retransmetre partits de les grans lligues, com la Bundesliga. Això fa que els clubs puguin generar més ingressos, i per tant distribuir una part d'aquests ingressos en forma de salaris més alts als seus jugadors.

Taula 3: Evolució dels salaris i dels ingressos per TV a la Bundesliga.

Font: Elaboració pròpia. Dades: FBref i Fernsehgelde



En el gràfic anterior (taula 3), podem observar l'evolució de les despeses totals en salaris de jugadors de la lliga i dels ingressos totals per drets televisius de la lliga. Veiem clarament una tendència positiva de les variables, podem visualitzar com les dues línies van augmentant en el període estudiat en el treball (2013-2023), un augment en els ingressos per televisió és respost amb un augment de la despesa en salaris.

El model d'efectes aleatoris revela que l'experiència dels futbolistes, mesurada per l'edat i el rendiment en el camp, com les assistències i els partits jugats, és clau per a determinar els seus salaris a la Bundesliga. Els factors econòmics com el PIB per càpita de la regió i l'ingrés per drets televisius també influeixen significativament la remuneració dels futbolistes. A més, la participació en competicions europees i la posició en la lliga tenen un impacte directe en els salaris dels jugadors. En resum, diversos factors, des de l'experiència fins als ingressos del club, determinen la valoració econòmica dels futbolistes a la Bundesliga.

6. CONCLUSIONS

En aquest estudi, hem identificat 9 factors determinants en la fixació dels salaris dels futbolistes de la Bundesliga durant les temporades 2013-2014 a 2022-2023. Aquests factors inclouen l'edat del jugador, la seva posició en el camp, el número de partits disputats, les assistències realitzades en una temporada, les targetes vermelles rebudes, el PIB per càpita ajustat per l'IPC, la participació en competicions europees, la seva classificació a la lliga i els ingressos obtinguts per drets televisius. Més enllà dels aspectes purament esportius, en aquest estudi, també hem considerat variables econòmiques. Hem constatat que el PIB per càpita i els ingressos televisius poden influir significativament, de manera positiva, en l'escala salarial dels jugadors.

Amb el nostre model, hem vist que aquestes variables identificades tenen un impacte significatiu en la determinació dels salaris dels jugadors analitzats. Sorprenentment, aquestes variables conjuntament expliquen una alta proporció de la variabilitat dels salaris dels jugadors, concretament el 86,51%. Aquesta capacitat explicativa del nostre model subratlla la importància d'aquests factors en la negociació i determinació dels salaris dins del món del futbol professional alemany. També hem de reconèixer que encara hi ha àmbits on es podria millorar de cara a estudis futurs. Seria beneficiós ampliar la mostra d'estudi per incloure més jugadors i temporades, així com considerar altres variables potencialment rellevants que no han estat incloses en aquesta anàlisi, com per exemple, la inclusió de més variables de rendiment utilitzades en estudis anteriors. Aquestes millores podrien aportar una comprensió més completa i refinada de la dinàmica salarial dins de la Bundesliga.

ANNEXOS:

ANNEX 1: Variables

Annex 1.1 - Descripció de les variables

Taula 4: Descripció de totes les variables disponibles. Font: Elaboració pròpia.

Variable	Definició
<i>AnnualWageEUR_ipc</i>	Salari anual del jugador en € corregit per l'IPC de la seva regió.
<i>Player</i>	Nom i primer cognom del jugador.
<i>Season</i>	Any en què es duu a terme la temporada de lliga.
<i>Team</i>	Nom de l'equip del jugador
<i>Nation</i>	Nacionalitat principal del jugador.
<i>Pos</i>	Variable qualitativa sobre la posició que ocupa en el camp. Hi ha quatre grups: Porter (GK), Defensa (DF), Migcampista (MF) i Davanter (FW).
<i>Age</i>	Edat del jugador en cada temporada.
<i>Matches_played</i>	Número de partits en què participa el jugador.
<i>Starts_playing</i>	Partits començant en l'once titular.
<i>Min_playing</i>	Minuts totals disputats en la temporada.
<i>Mins_per_90</i>	Minuts totals jugats en tota la temporada dividits per 90 minuts.
<i>Gls</i>	Número de gols totals en una temporada.
<i>Ast</i>	Número de assistències totals en una temporada.
<i>G_minus_PK</i>	Número de gols totals en una temporada menys els gols marcats de penal.
<i>PK</i>	Número de gols de penal totals en una temporada.
<i>PKatt</i>	Número total de penals tirats.
<i>CrdY</i>	Número de targetes grogues totals en una temporada.
<i>CrdR</i>	Número de targetes vermelles totals en una temporada.

<i>Gls_per_minutes</i>	Número de gols marcats per cada 90 minuts en una temporada.
<i>Ast_per_minutes</i>	Número d'assistències generades per cada 90 minuts en una temporada.
<i>G_minus_PK_mins</i>	Número de gols menys els gols marcats de penal per cada 90 minuts en una temporada.
<i>pib_capita</i>	PIB per càpita de la regió on es troba situat l'equip del jugador. (en euros)
<i>pib_capita_ipc</i>	PIB per càpita de la regió on es troba situat l'equip del jugador, corregit per l'IPC de la regió. (en euros)
<i>regional_pib</i>	PIB total de la regió on es troba situat l'equip del jugador. (en euros)
<i>regional_pib_ipc</i>	PIB total de la regió on es troba situat l'equip del jugador, corregit per l'IPC de la regió. (en euros)
<i>eu_comp</i>	Variable qualitativa sobre la competició europea en què participen els jugadors. Hi ha quatre grups: Champions League (UCL), Europa League (UEL), Conference League (CFL) i no participants (NO).
<i>league_position</i>	Posició final en la classificació de la lliga nacional.
<i>tv_income</i>	Ingrés que rep l'equip del jugador per drets televisius cada temporada.
<i>tv_income_ipc</i>	Ingrés que rep l'equip del jugador per drets televisius cada temporada, corregit per l'IPC de la regió.
<i>region</i>	Regió on es troba l'equip del jugador, seguint la divisió territorial donada pel sistema NUTS2.

Annex 1.2 - Factor d'inflació de la variància (FIV) de les variables

Taula 5: Factor d'inflació de la variància (FIV) de totes les variables disponibles. Font: R-studio.

	GVIF	Df	GVIF ^{1/(2*Df)}
factor(Team)	1.185381e+06	27	1.295624
factor(Nation)	1.179294e+01	92	1.013501
factor(Pos)	2.786311e+00	3	1.186237
Age	1.249662e+00	1	1.117883
Matches_played	1.322338e+01	1	3.636397
Starts_Playing_Time	1.206464e+02	1	10.983914
Min_Playing_Time	1.539347e+05	1	392.345130
Mins_Per_90	1.538581e+05	1	392.247513
Gls	4.735454e+02	1	21.761099
Ast	1.873087e+00	1	1.368608
G_minus_PK	3.881447e+02	1	19.701389
PKatt	1.465552e+01	1	3.828253
CrDY	1.969872e+00	1	1.403521
CrDR	1.124927e+00	1	1.060626
Gls_Per_Minutes	5.717032e+02	1	23.910315
Ast_Per_Minutes	1.092173e+00	1	1.045071
G_minus_PK_Per_Minutes	5.600630e+02	1	23.665650
regional_pib	8.240481e+02	1	28.706237
regional_pib_ipc	2.085338e+03	1	45.665506
pib_capita	2.380402e+02	1	15.428551
pib_capita_ipc	3.616611e+02	1	19.017388
factor(eu_comp)	4.917783e+00	3	1.304052
league_position	2.959289e+00	1	1.720258
tv_income	8.541268e+02	1	29.225448
tv_income_ipc	7.481943e+02	1	27.353140

Observant els resultats del FIV, veiem les variables que ens mostren els problemes de multicol·linealitat i d'autocorrelació. Per exemple, les variables que són creades a partir d'altres, com els gols per cada 90 minuts (Gls_Per_Minutes), mostraven una forta correlació que ja podríem esperar i, per tant, totes aquestes variables semblants van ser eliminades de qualsevol model. També vam observar una alta correlació entre els variables partits jugats, minuts jugats, partits de titular i minuts totals jugats dividits per 90, decidint conservar només una d'elles. Addicionalment, les variables relacionades amb el PIB i la seva versió per càpita mostraven una alta correlació entre elles i ens quedem només amb el PIB per càpita. Les variables que feien referència a l'equip, la nacionalitat i la regió ens indicaven problemes potencials de multicol·linealitat i autocorrelació, per la qual cosa van ser eliminades del nostre conjunt de variables.

ANNEX 2: Especificació del model Bundesliga

Annex 2.1 - Factor d'inflació de la variància (FIV) dades escollides.

Taula 6: Factor d'inflació de la variància (FIV) de les dades escollides. Font: R-studio.

	GVIF	Df	GVIF^(1/(2*Df))
factor(Pos)	1.561765	3	1.077133
Age	1.092661	1	1.045304
Matches_played	2.100400	1	1.449276
Goals	2.647119	1	1.626997
Ast	1.673503	1	1.293639
PK	1.589323	1	1.260683
CrdY	1.752326	1	1.323754
CrdR	1.078599	1	1.038556
log(pib_capita_ipc)	1.024349	1	1.012101
factor(eu_comp)	1.798321	3	1.102752
league_position	1.709134	1	1.307338
log(tv_income_ipc)	1.305756	1	1.142697

Observant els valors del FIV trobem que cap de les variables escollides té valors elevats, per tant, podem confirmar que el model pooled no té problemes de multicol·linealitat.

Annex 2.2 - Estimació del model pooled amb totes les variables.

Taula 7: Estimació model pooled amb variables no significatives. Font: R-studio.

```
Call:
lm(formula = log(AnnualWageEUR_ipc) ~ factor(Pos) + Age + Matches_played +
  Gls + Ast + PK + CrdY + CrdR + log(pib_capita_ipc) + factor(eu_comp) +
  league_position + log(tv_income_ipc), data = dades_panell)

Residuals:
    Min       1Q   Median       3Q      Max
-4.2061 -0.5287  0.0674  0.6078  2.9081

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)      -6.838146   1.100485  -6.214 5.64e-10 ***
factor(Pos)FW      0.012918   0.040866   0.316 0.751938
factor(Pos)GK     -0.401435   0.052338  -7.670 2.09e-14 ***
factor(Pos)MF      0.044469   0.033386   1.332 0.182934
Age                0.125795   0.003306  38.049 < 2e-16 ***
Matches_played    0.036537   0.001822  20.049 < 2e-16 ***
Gls                0.022457   0.006441   3.487 0.000494 ***
Ast                0.024270   0.007963   3.048 0.002318 **
PK                -0.016501   0.025417  -0.649 0.516248
CrdY              -0.007542   0.007233  -1.043 0.297109
CrdR               0.116611   0.040410   2.886 0.003924 **
log(pib_capita_ipc) 0.638733   0.080290   7.955 2.24e-15 ***
factor(eu_comp)NO  0.700637   0.131614   5.323 1.07e-07 ***
factor(eu_comp)UCL 1.293702   0.133849   9.665 < 2e-16 ***
factor(eu_comp)UEL 0.884996   0.135045   6.553 6.26e-11 ***
league_position   -0.030496   0.003367  -9.057 < 2e-16 ***
log(tv_income_ipc) 0.527745   0.041953  12.580 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8983 on 4491 degrees of freedom
Multiple R-squared:  0.517,    Adjusted R-squared:  0.5153
F-statistic: 300.4 on 16 and 4491 DF,  p-value: < 2.2e-16
```

Observem que hi ha variables no significatives, la variable PK i la variable CrdY no són significatives pel nostre model, així que elaborarem un model sense aquestes.

Annex 2.2 - Estimació del model pooled definitiu

Taula 8: Estimació del model pooled. Font: R-studio.

```
Call:
lm(formula = log(AnnualWageEUR_ipc) ~ factor(Pos) + Age + Matches_played +
    Gls + Ast + CrdR + log(pib_capita_ipc) + factor(eu_comp) +
    league_position + log(tv_income_ipc), data = dades_panell)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.2044	-0.5293	0.0679	0.6067	2.9117

Coefficients:

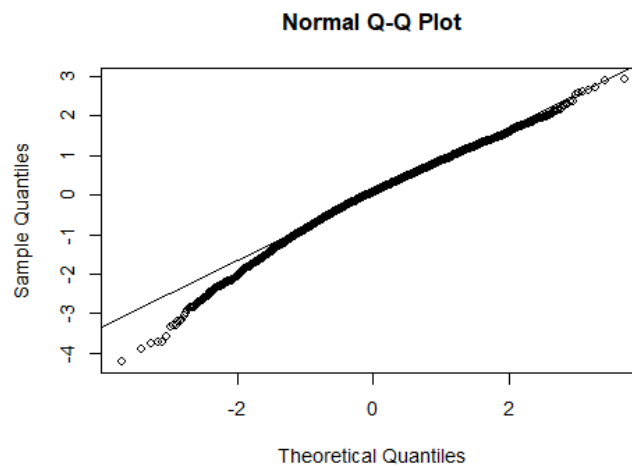
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-6.865195	1.100076	-6.241	4.76e-10 ***
factor(Pos)FW	0.021846	0.040220	0.543	0.58704
factor(Pos)GK	-0.388518	0.050800	-7.648	2.48e-14 ***
factor(Pos)MF	0.045003	0.033363	1.349	0.17744
Age	0.125496	0.003294	38.096	< 2e-16 ***
Matches_played	0.035678	0.001559	22.890	< 2e-16 ***
Gls	0.020364	0.005411	3.764	0.00017 ***
Ast	0.024524	0.007959	3.081	0.00207 **
CrdR	0.106996	0.039532	2.707	0.00682 **
log(pib_capita_ipc)	0.639927	0.080223	7.977	1.89e-15 ***
factor(eu_comp)NO	0.700623	0.131608	5.324	1.07e-07 ***
factor(eu_comp)UCL	1.295578	0.133833	9.681	< 2e-16 ***
factor(eu_comp)UEL	0.885217	0.135034	6.556	6.17e-11 ***
league_position	-0.030761	0.003360	-9.156	< 2e-16 ***
log(tv_income_ipc)	0.528913	0.041935	12.613	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.8983 on 4493 degrees of freedom
Multiple R-squared: 0.5168, Adjusted R-squared: 0.5153
F-statistic: 343.3 on 14 and 4493 DF, p-value: < 2.2e-16

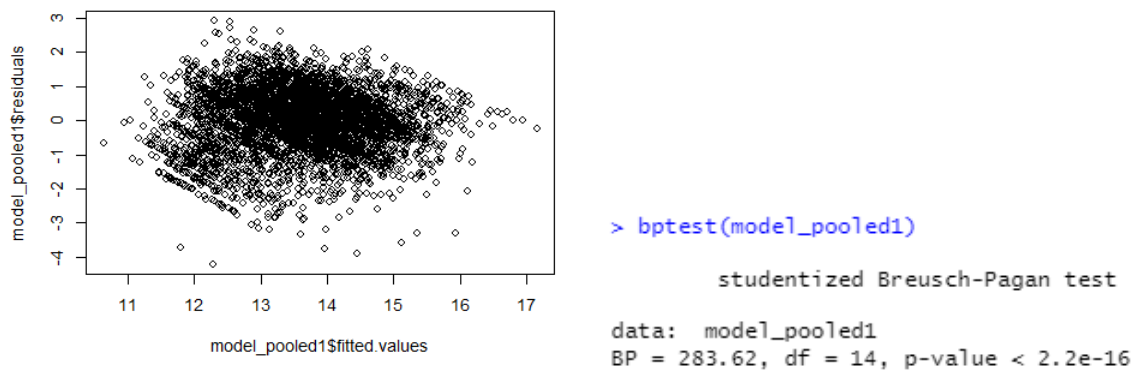
Annex 2.3 - Validació model pooled

Taula 9: Gràfic Normalitat. Font: R-studio



Es pot observar en el gràfic de normalitat que es compleix aquest supòsit. Els residus segueixen la bisectriu excepte als extrems, on hi ha una major dispersió. A més, amb una mostra de $n = 5008$, que supera els 30 casos recomanats pel Teorema del límit central (TLC), podem aproximar-nos a una distribució normal. Per tant, podem verificar el supòsit.

Taula 10: Gràfic dels residuals i Contrast Breusch-Pagan. Font: R-studio



El diagnòstic dels residuals, observem en el gràfic de residuals contra valors ajustats que el model pot presentar problemes d'heteroscedasticitat, però no sembla molt preocupant segons veiem en el gràfic. Si mirem el contrast de Breusch-Pagan, veiem que és significatiu, per tant, podem dir que hi ha evidència que la variància dels residus no és constant a través de les observacions, i això podria afectar la validesa de les inferències fetes a partir del model.

Taula 11: Contrast de Durbin Watson. Font: R-studio.

```

> durbinWatsonTest(model_pooled)
lag Autocorrelation D-W Statistic p-value
1 0.3813467 1.237199 0
  
```

Els resultats mostren un coeficient de correlació autoregressiu (lag 1) de 0,3813467 i una estadística Durbin-Watson (D-W) de 1,237199, amb un valor p associat de 0. Aquest resultat suggereix la presència de correlació serial en els residus del model.

Taula 12: Factor d'inflació de variància (FIV) del model pooled. Font: R-studio.

	GVIF	Df	GVIF^(1/(2*Df))
factor(Pos)	1.420708	3	1.060272
Age	1.084911	1	1.041591
Matches_played	1.536737	1	1.239652
GLs	1.868479	1	1.366923
Ast	1.671936	1	1.293034
Crdr	1.032337	1	1.016040
log(pib_capita_ipc)	1.022748	1	1.011310
factor(eu_comp)	1.794436	3	1.102355
league_position	1.701549	1	1.304434
log(tv_income_ipc)	1.304779	1	1.142269

Ara mesurem la multicolinealitat a partir del VIF, un VIF alt suggereix una alta multicolinealitat. Els valors VIF que hem obtinguts ens indiquen la presència de multicolinealitat baixa en el model pooled. Veiem en els resultats, que la majoria dels valors estan per sota del llindar comunament utilitzat com a referència de multicolinealitat de 5.

Annex 2.4 - Estimació del model d'efectes fixos

Taula 13: Estimació del model d'efectes fixos. Font: R-studio.

```
Oneway (individual) effect Within Model

Call:
plm(formula = log(AnnualWageEUR_ipc) ~ Age + Matches_played +
     CrdR + log(pib_capita_ipc) + league_position + log(tv_income_ipc),
     data = dades_panell, model = "within")

Unbalanced Panel: n = 1560, T = 1-10, N = 4508

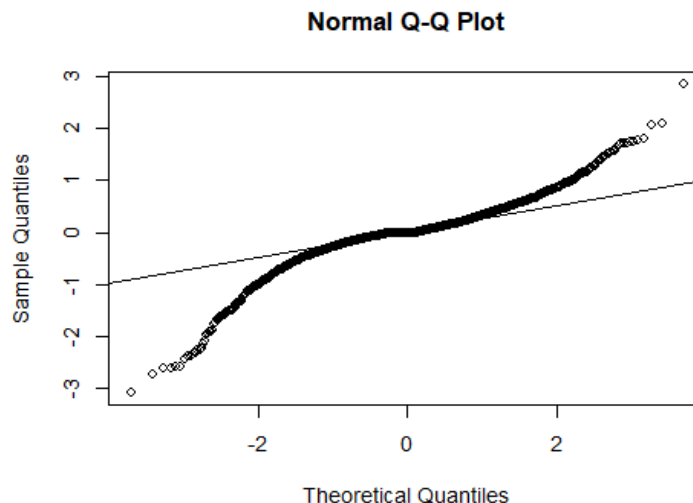
Residuals:
    Min.  1st Qu.  Median    3rd Qu.    Max.
-3.06794 -0.15071  0.00000  0.18007  2.82889

Coefficients:
                Estimate Std. Error t-value Pr(>|t|)
Age             0.0850390  0.0063735  13.3425 < 2.2e-16 ***
Matches_played  0.0151246  0.0012065  12.5354 < 2.2e-16 ***
CrdR            0.0605324  0.0283170   2.1377 0.0326256 *
log(pib_capita_ipc) 0.3210996  0.0967267   3.3197 0.0009123 ***
league_position -0.0063129  0.0026555  -2.3773 0.0175058 *
log(tv_income_ipc)  0.6480950  0.0518616  12.4966 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Total Sum of Squares:  1194.6
Residual Sum of Squares: 819.02
R-Squared: 0.31442
Adj. R-Squared: -0.050269
F-statistic: 224.881 on 6 and 2942 DF, p-value: < 2.22e-16
```

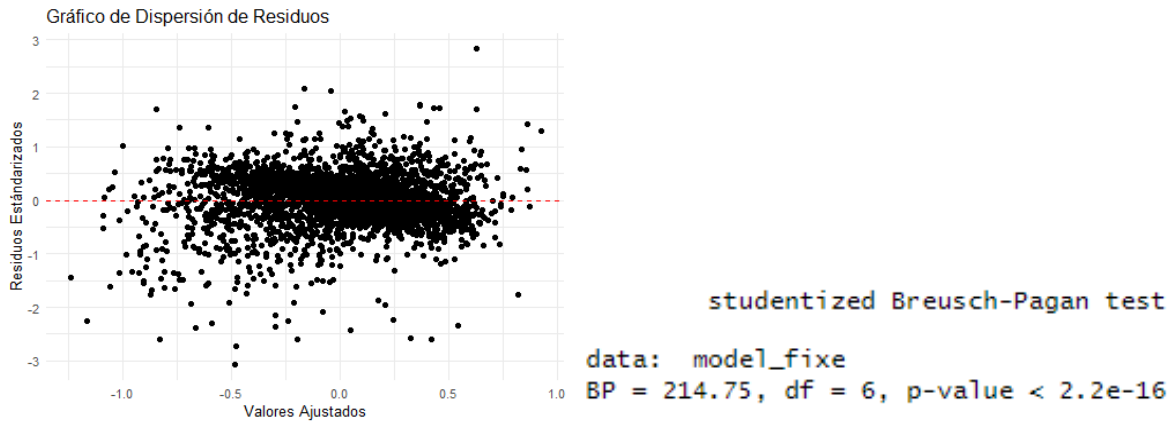
Annex 2.5 - Validació model d'efectes fixos

Taula 14: Gràfic de normalitat. Font: R-studio.



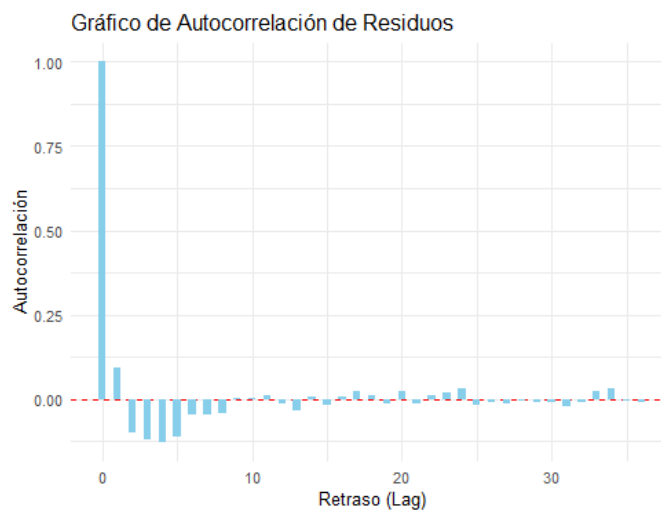
En el gràfic de normalitat es pot veure que no es compleix aquest supòsit. Els residus no segueixen la bisectriu, el gràfic presenta molta curvatura en la línia de punts. Però, amb una mostra de $n = 4508$ supera els 30 casos recomanats pel Teorema del límit central (TLC) i podem aproximar-nos a una distribució normal.

Taula 15: Gràfic dels residuals i Contrast Breusch-Pagan. Font: R-studio



El diagnòstic dels residuals, observem en el gràfic de residuals contra valors ajustats que el model pot presentar problemes d'heteroscedasticitat, però no sembla molt preocupant segons veiem en el gràfic. Però si mirem el contrast de Breusch-Pagan, veiem que és significatiu, per tant, podem dir que hi ha evidència que la variància dels residus no és constant a través de les observacions, i això podria afectar la validesa de les inferències fetes a partir del model.

Taula 16: Gràfic d'autocorrelació. Font: R-studio



En el gràfic d'autocorrelació dels residus, podem observar valors elevats en les primeres barres, però la resta d'elles no tenen uns valors molt elevats que suposin greus problemes d'autocorrelació en els residus.

Taula 17: Factor d'inflació de variància (FIV) del model d'efectes fixes. Font: R-studio.

```
> vif(model_fixe2)
      Age      Matches_played      CrdR log(pib_capita_ipc)      league_position
log(tv_income_ipc)      1.044166      1.064069      1.019788      1.007499      1.186895
1.182643
```

Si mirem la multicolinealitat a partir del VIF, trobem que els valors que hem obtingut, ens indiquen la presència de multicolinealitat baixa en el nostre model d'efectes fixes. Veiem en els resultats, que la majoria dels valors estan per sota del llindar comunament utilitzat com a referència de multicolinealitat de 5.

Annex 2.6 - Estimació del model d'efectes aleatoris

Taula 18: Estimació del model d'efectes aleatoris. Font: R-studio.

```
Unbalanced Panel: n = 1560, T = 1-10, N = 4508

Effects:
      var std.dev share
idiosyncratic 0.2780 0.5272 0.378
individual    0.4576 0.6764 0.622
theta:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3853 0.5173 0.6369 0.6191 0.7174 0.7607

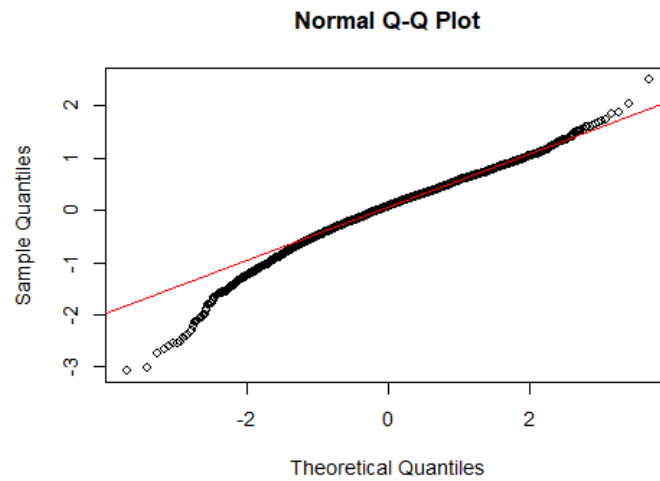
Residuals:
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-3.0695 -0.2955 0.0744 0.0274 0.3941 2.4956

Coefficients:
              Estimate Std. Error z-value Pr(>|z|)
(Intercept)  -3.3638689  1.0389863 -3.2376 0.0012052 **
Age           0.1233746  0.0039130 31.5292 < 2.2e-16 ***
factor(Pos)FW 0.0498770  0.0430386  1.1589 0.2465008
factor(Pos)GK -0.5749515  0.0785497 -7.3196 2.487e-13 ***
factor(Pos)MF 0.0359697  0.0375439  0.9581 0.3380276
Matches_played 0.0242449  0.0012874 18.8326 < 2.2e-16 ***
Ast          0.0155798  0.0063920  2.4374 0.0147947 *
CrdR         0.0792723  0.0292885  2.7066 0.0067976 **
log(pib_capita_ipc) 0.3905906  0.0839441  4.6530 3.272e-06 ***
factor(eu_comp)NO 0.3360484  0.0980925  3.4258 0.0006129 ***
factor(eu_comp)UCL 0.5924128  0.1018456  5.8168 5.999e-09 ***
factor(eu_comp)UEL 0.3927236  0.1009540  3.8901 0.0001002 ***
league_position -0.0187798  0.0026194 -7.1696 7.521e-13 ***
log(tv_income_ipc) 0.5107378  0.0387950 13.1650 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

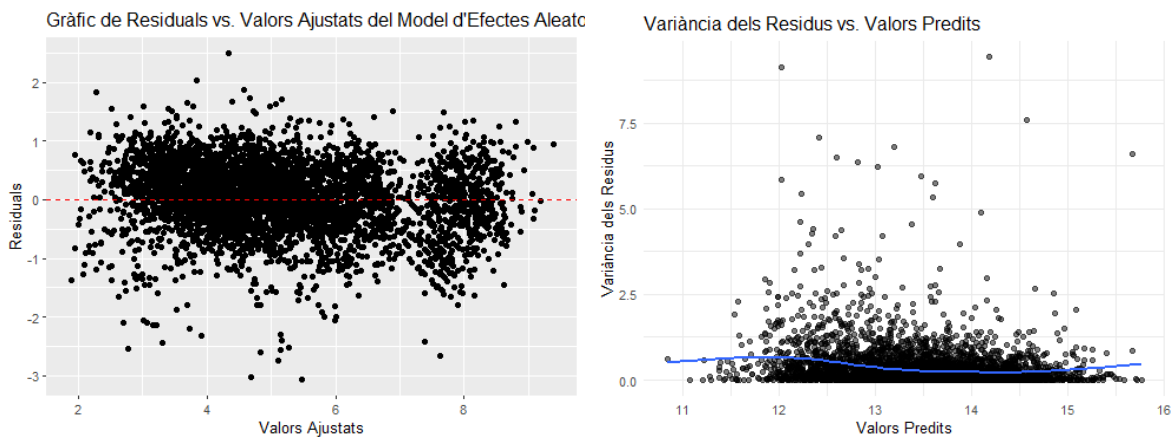
Total Sum of Squares:    10620
Residual Sum of Squares: 1470
R-Squared:               0.86549
Adj. R-Squared:         0.8651
Chisq: 2810 on 13 DF, p-value: < 2.22e-16
```

Annex 2.7 - Validació model d'efectes aleatoris

Taula 19: Gràfic de normalitat. Font: R-studio.



Taula 20: Gràfics dels residualls. Font: R-studio.



Taula 21: Factor d'inflació de variància (FIV) del model d'efectes aleatoris. Font: R-studio.

	GVIF	Df	GVIF ^{1/(2*Df)}
Age	1.235190	1	1.111391
factor(Pos)	1.052034	3	1.008490
Matches_played	1.326112	1	1.151569
Ast	1.295803	1	1.138333
Crdr	1.020031	1	1.009966
log(pib_capita_ipc)	1.046585	1	1.023028
factor(eu_comp)	1.309990	3	1.046031
league_position	1.212132	1	1.100968
log(tv_income_ipc)	1.371412	1	1.171073

BIBLIOGRAFIA

Bernd Frick (2011). Performance, Salaries, and Contract Length: Empirical Evidence from German Soccer. *International Journal of Sport Finance*, 6: 87-118.

https://www.researchgate.net/profile/Bernd-Frick-2/publication/227450873_Performance_Salaries_and_Contract_Length_Empirical_Evidence_from_German_Soccer/links/5698937908ae1c4279055f5b/Performance-Salaries-and-Contract-Length-Empirical-Evidence-from-German-Soccer.pdf

Carreras-Simó, Miquel & Garci, Jaume. (2018). TV rights, financial inequality, and competitive balance in European football: Evidence from the English Premier League and the Spanish LaLiga. *International Journal of Sport Finance*, 13: 201-224.

https://www.researchgate.net/publication/328389034_TV_rights_financial_inequality_and_competitive_balance_in_European_football_Evidence_from_the_English_Premier_League_and_the_Spanish_LaLiga

Caruso, R., Di Domizio, M., & Rossignoli, D. (2017). Aggregate wages of players and performance in Italian Serie A. *Economia Politica*, 34 (3): 515-531.

<https://doi.org/10.1007/s40888-017-0062-6>

De Menezes, C. R. C., & De Santana, J. R. (2022). Análise recente da produção científica internacional sobre gestão do conhecimento: Sciedirect – 2015 a 2019. *Gestão & Regionalidade*, 38(114). <https://doi.org/10.13037/gr.vol38n114.6713>

Lee, S. K., & Harris, J. (2012). Managing excellence in USA Major League Soccer: an analysis of the relationship between player performance and salary. *Managing Leisure*, 17(2-3): 106-123. <https://doi.org/10.1080/13606719.2012.674389>

Ribeiro, A. S., & Lima, F. (2012). Portuguese football league efficiency and players' wages. *Applied Economics Letters*, 19(6): 599-602. <https://doi.org/10.1080/13504851.2011.591719>

Wilkesmann, U. (2017). *Social inequality in German football. Does money score goals?* <http://129.217.131.68:8080/handle/2003/35782>

Yaldo, L. & Shamir, L. (2017). Computational Estimation of Football Player Wages. *International Journal of Computer Science in Sport*, 16 (1).

<https://sciendo.com/abstract/journals/ijcss/16/1/article-p18.xml>

WEBGRAFIA

Bundesliga | *el sitio web oficial*. (s. f.). bundesliga.com/es - el Sitio Web Oficial de la Bundesliga. Última visualització: 25 de Febrer de 2024.

<https://www.bundesliga.com/es/bundesliga>

Bundesliga Stats | *FBref.com*. (s. f.). FBref.com. Última visualització: 20 de Febrer de 2024.

<https://fbref.com/en/comps/20/2022-2023/2022-2023-Bundesliga-Stats>

DESTATIS Homepage. (s. f.). Federal Statistical Office. Última visualització: 28 de Febrer de 2024. https://www.destatis.de/EN/Home/_node.html

Eurostat. (s. f.). Eurostat. Última visualització: 28 de Febrer de 2024.

<https://ec.europa.eu/eurostat>

Fernsehgelder.de. (s. f.). Última visualització: 13 de Març de 2024.

<https://www.fernsehgelder.de/>

README.(s.f.). *Worldfootballr*. Última visualització: 25 de Febrer de 2024.

<https://cran.r-project.org/web/packages/worldfootballR/readme/README.html>



DETERMINANTS DELS SALARIS A LA BUNDESLIGA

Treball Final de Grau Economia



Aquest estudi analitza els determinants dels salaris dels jugadors de la Bundesliga des de la temporada 2013-2014 fins a la 2022-2023. Es troba que el rendiment del jugador, incloent factors com partits jugats i assistències, juntament amb elements com la classificació de la lliga, afecten els seus salaris. També s'observa l'impacte de variables econòmiques com el PIB per càpita i els ingressos per drets televisius en els salaris d'aquests jugadors.

01. Literatura Prèvia

Aquesta secció resum la investigació prèvia sobre la relació entre els salaris dels jugadors de futbol i diverses variables relacionades amb el seu rendiment i context de la lliga. Destaca estudis com el de Yaldo i Shamir del 2017, que utilitza mètodes computacionals per relacionar els salaris amb el rendiment dels jugadors. També s'analitzen altres investigacions centrades en territoris específics com la Major League Soccer i la lliga brasilera, així com en altres lligues europees com la italiana i la portuguesa. S'observa una correlació positiva entre el rendiment individual dels jugadors i els seus salaris, però també es consideren altres factors com l'eficiència dels equips, les desigualtats regionals i la durada dels contractes. Aquesta revisió de literatura serveix de motivació per a l'estudi de les variables determinants dels salaris dels jugadors de la Bundesliga, incloent tant factors esportius com territorials i econòmics.



02. Dades Utilitzades

Variable	Definició
AnnualWageEUR_ipc	Salari anual del jugador en € corregit per l'IPC de la seva regió.
Player	Nom i primer cognom del jugador.
Season	Any en què es duu a terme la temporada de lliga.
Pos	Variable qualitativa sobre la posició que ocupa en el camp. Hi ha quatre grups: Porter (GK), Defensa (DF), Migcampista (MF) i Davanter (FW).
Age	Edat del jugador en cada temporada.
Matches_played	Número de partits en què participa el jugador.
Glz	Número de gols totals en una temporada.
Ast	Número de assistències totals en una temporada.
PK	Número de gols de penal totals en una temporada.
CrdrT	Número de targetes grogues totals en una temporada.
CrdrR	Número de targetes vermelles totals en una temporada.
pib_capita_ipc	PIB per càpita de la regió on es troba situat l'equip del jugador, corregit per l'IPC de la regió. (en euros)
eu_comp	Variable qualitativa sobre la competició europea en què participen els jugadors. Hi ha quatre grups: Champions League (UCL), Europa League (UEL), Conference League (CFL) i no participants (NO).
league_position	Posició final en la classificació de la lliga nacional.
tv_income_ipc	Ingrés que rep l'equip del jugador per drets televisius cada temporada, corregit per l'IPC de la regió.

Worldfoorbllr
FBref
Destatis
Bundesliga
Eurostat
Fernsehgelder

03. Comparar Models

Variables	Model Pooled	Model d'efectes fixes	Model d'efectes aleatoris
CONSTANT	-6.865195 4.76e-10 ***		-3.3638689 0.0012052 **
AGE	0.125496 < 2e-16 ***	0.0850390 < 2.2e-16 ***	0.1233746 < 2.2e-16 ***
POS (FW)	0.021846 0.58704		0.0498770 0.2465008
POS (GK)	-0.388518 2.48e-14 ***		-0.5749515 2.487e-13 ***
POS (MF)	0.045003 0.17744		0.0359697 0.3380276
MATCHES_PLAYED	0.035678 < 2e-16 ***	0.0151246 < 2.2e-16 ***	0.0242449 < 2.2e-16 ***
GLS	0.020364 0.00017 ***		
AST	0.024524 0.00207 **		0.0155798 0.0147947 *
CRDR	0.106996 0.00682 **	0.0605324 0.0326256 *	0.0792723 0.0067976 **
LOG (PIB_CAPITA_IPC)	0.639927 1.89e-15 ***	0.3210996 0.0009123 ***	0.3905906 3.272e-06 ***
EU_COMP (NO)	0.700623 1.07e-07 ***		0.3360484 0.0006129 ***
EU_COMP (UCL)	1.295578 < 2e-16 ***		0.5924128 5.999e-09 ***
EU_COMP (UEL)	0.885217 6.17e-11 ***		0.3927236 0.0001002 ***
LEAGUE_POSITION	-0.030761 < 2e-16 ***	-0.0063129 0.0175058 *	-0.0187798 7.521e-13 ***
LOG (TV_INCOME)	0.528913 < 2e-16 ***	0.6480950 < 2.2e-16 ***	0.5107378 < 2.2e-16 ***
Nº Observacions	5.008	4.508	4.508
R² Ajustat	0.5153	-0.050269	0.8651
P-valor	< 2.2e-16	< 2.22e-16	< 2.22e-16

Comparem tres models econòmics per analitzar els salaris dels jugadors de futbol. El model d'efectes aleatoris destaca amb un R-quadrat ajustat més alt (0,8651), indicant una millor explicació de la variabilitat en els salaris. El p-valor de l'estadístic F és baix en tots els models, indicant significació global. Però els coeficients dels predictors són més significatius en el model d'efectes aleatoris. Per tant, el model d'efectes aleatoris sembla ser el més adequat i robust per a aquest conjunt de dades.

04. Model D'Efectes Aleatoris

$$\log(\text{AnnualWageEUR_ipc})_{jt} = \beta_1 j + \beta_2 \text{Age}_{jt} + \beta_3 \text{Pos}_{jt} + \beta_4 \text{Matches_played}_{jt} + \beta_5 \text{Ast}_{jt} + \beta_6 \text{Crdr}_{jt} + \beta_7 \log(\text{pib_capita_ipc})_{jt} + \beta_8 \text{eu_comp}_{jt} + \beta_9 \text{league_position}_{jt} + \beta_{10} \log(\text{tv_income_ipc})_{jt} + u_{jt}$$

$$\beta_{1j} = \beta_1 + \eta_j$$

El model explica aproximadament el 86,55% de la variabilitat en el salari dels jugadors.

Els predictors són estadísticament significatius amb valors de $p < 0,05$

El model ofereix una explicació forta i coherent de la relació entre els predictors i el salari dels jugadors.



05. Resultats

Edat 12,34%

Un partit jugat més implica un augment del 4,42%

Un augment de l'1% del PIB per càpita de la regió implica un augment del 0,39%

Participants UCL cobren un 59,24% més que els participants CFL.

Una targeta vermella més implica un augment de 7,93%

Augment de l'1% en els ingressos per televisió implica un augment del 0,51%

Una posició menys a la lliga implica una disminució del 1,88%

Els porters cobren un 5,75% menys que els defenses.

Una assistència més genera un augment del 1,56%



06. Conclusions

El nostre estudi sobre els salaris dels futbolistes de la Bundesliga ha identificat diversos factors clau, incloent l'edat, la posició al camp, el rendiment esportiu, els ingressos per televisió del club i el PIB per càpita. Hem observat que aquests factors expliquen aproximadament el 86,51% de la variabilitat dels sous dels jugadors. Tot i això, reconeixem que encara hi ha àrees per millorar, com ampliar la mostra d'estudi i considerar altres variables rellevants. Aquesta investigació proporciona una visió significativa de la dinàmica salarial a la Bundesliga, però suggereix que s'ha de seguir investigant per a una comprensió més profunda.

