

ANÁLISIS DEL VALOR DE MERCADO DE JUGADORES DE FUTBOL DE ELITE

Agustina Margaritis¹, Natalia Macchi¹, Román Tanducci¹, Santiago García
Sánchez¹, Javier Bussi^{1,3}, and Lucía Andreozzi^{1,2}

¹Escuela de Estadística, Facultad de ciencias Económicas y Estadística, Universidad
Nacional de Rosario.

²Consejo Nacional de Investigaciones Científicas y Técnicas.

³Instituto de Investigaciones Teóricas y Aplicadas de la Escuela de Estadística,
Facultad de Ciencias Económicas y Estadística, Universidad Nacional de Rosario.

RESUMEN

El fútbol es un deporte con gran cantidad de fanáticos alrededor del mundo, pero además es una “industria” que genera millones de dólares para sus participantes. El presente estudio se plantea como objetivo indagar qué factores influyen en el valor de mercado de un jugador. Como objetivo secundario se postula estudiar la influencia de una red social en dicho valor. La muestra seleccionada corresponde a los 200 jugadores con mayor valor de mercado en ligas profesionales de fútbol masculino al mes de junio de 2022. Los datos se obtuvieron mediante la técnica de web scraping. Para el análisis se estimaron modelos de regresión lineal múltiple y el modelo final seleccionado determinó que las variables que influyen sobre el valor de mercado del jugador son la edad, el número de partidos jugados, el número de goles, asistencias y amarillas por partido, el número de ingresos sin ser titular, el salario y la cantidad de seguidores de Instagram. El modelo final estimado presentó un buen ajuste en términos de adecuación de los residuos a los supuestos del modelo. El presente trabajo representa un aporte científico de relevancia en la determinación de los factores que influyen en el valor de mercado de los jugadores de élite y puede servir como disparador de estudios futuros que consideren esta temática.

Palabras claves: Fútbol, Regresión Lineal Múltiple, Valor de Mercado.

ABSTRACT

Football is a sport with a large number of fans around the world, but it is also an “industry” that generates millions of dollars for its participants. The objective of this study is to investigate which factors influence the market value of a player. A secondary objective is to study the influence of a social network on such value. The selected sample corresponds to the 200 players with the highest market value in professional men’s football leagues as of June 2022. The data was obtained using the web scraping technique. For the analysis, multiple linear regression models were estimated and the final selected model determined that the variables that influence the player’s market value are age, the number of games played, the number of goals, assists and yellow cards per game, the number of substitutions, the salary and the number of Instagram followers. The final estimated model presented a good fit in terms of adequacy of the residuals given the model assumptions. This work represents a relevant scientific contribution in determining the factors that influence the market value of elite players and could serve as a basis for future studies that consider this issue.

Keywords: Football, Multiple Linear Regression, Market Value.

1 Introduccion

El fútbol es un deporte con gran cantidad de fanáticos alrededor del mundo, pero además es una “industria” que genera millones de dólares para sus participantes, ya sea clubes u organizaciones como la FIFA, por medio de derechos televisivos, publicidad y merchandising. Se proyecta que los ingresos de las cinco mayores ligas (Alemania, España, Francia, Inglaterra e Italia) superarán los 15 mil millones de euros en la temporada 2021/2022. Es por esto que los clubes que mejor se posicionan a nivel deportivo pueden generar mayores ingresos. Esto da lugar a “mercados” de compraventa en los que los clubes pagan por los jugadores cifras que pueden parecer irrisorias para el común de las personas, tanto por el pase del jugador de un club a otro como por salarios para el jugador, porcentajes de ingresos publicitarios, etc.

Ahora bien, no se trata de mercados “transparentes” en los que la libre concurrencia de oferentes y demandantes determina un precio, sino que son acuerdos contractuales entre privados en los que muchas cuestiones particulares del jugador y de los clubes pueden tener influencia, y en los que tampoco puede hablarse de un único precio. En este contexto, el presente estudio se plantea como objetivo indagar qué factores hacen más valioso a un jugador, entendiendo su “valor” no como un precio efectivamente pagado por un producto, sino desde un sentido amplio, como una referencia a lo que un club debería ceder (o podría ganar) al incorporar al jugador a su plantel. Como objetivo secundario se postula estudiar la influencia de una red social en el valor del jugador. Dicha variable no ha sido tenida en cuenta tradicionalmente en estimaciones de esta clase, pero investigaciones recientes sugieren que puede tener un efecto significativo.

Según estudios tradicionales, el valor de los jugadores de fútbol lo determinan múltiples factores, es decir que no solo depende de la habilidad del jugador en cuestión —no es un criterio absoluto—, sino también de una serie de condiciones como la edad -la retribución aumenta hasta los 27 años y después vuelve a descender-, el número de partidos que hayan jugado la temporada anterior y las apariciones en competiciones internacionales, así como los goles marcados.

En la actualidad a esos factores se suma la relación

del futbolista con los medios de comunicación, un factor que hasta ahora no había sido tenido en cuenta a la hora de estudiar las retribuciones de los deportistas. La presencia de los futbolistas en las redes sociales podría influir en su valor de mercado y, por extensión, en el sueldo que se les paga, ya que una mayor popularidad entre el público se traduce en mayores ingresos para el club en concepto de marketing.

Hay algunas plataformas que se encargan de analizar todos los factores útiles para reconstruir el valor de mercado y, en consecuencia, predecir cuál podría ser la cantidad pagada por un posible fichaje, Transfermarkt es una de ellas, es una empresa alemana que tiene como objetivo cotizar a todos los futbolistas profesionales del mundo. Se estima que tiene una base de datos con más de 833.000 jugadores, 1.150.000 fichas de partidos y valores de mercado.

Su funcionamiento se basa en la premisa de que el valor de mercado o de posible venta de un futbolista va a ser muy parecido al promedio de las cotizaciones hechas por decenas de expertos que realizan los estudios estadísticos y evalúan a cada uno de ellos. *Transfermarkt* no utiliza un algoritmo, sino que se basa en el saber de la comunidad, se alega que dicha página web posee más de un 95% de acierto de los precios finales de venta de jugadores. Algunos de los factores que son tenidos en cuenta para establecer el valor de mercado del futbolista son: perspectivas de futuro, edad, desempeño en el club y la selección, nivel y estado de la liga, tanto en términos deportivos como financieros, reputación/prestigio, potencial de desarrollo y rendimiento, valor de marketing, número y reputación de los clubes interesados, nivel de experiencia, susceptibilidad a las lesiones, demanda general y tendencias del mercado y evolución general de las tasas de traspaso.

2 Material y Métodos

La investigación comienza con la fase de construcción del conjunto de datos. La muestra seleccionada corresponde a los 200 jugadores con mayor valor de mercado en ligas profesionales de fútbol masculino al mes de junio de 2022, con lo cual son considerados jugadores de élite. La mayoría de los datos fueron extraídos de la página *Transfermarkt*, sitio web alemán propiedad del grupo

editorial Axel Springer SE que provee información sobre el fútbol. Dicho portal cuenta con un equipo de profesionales y voluntarios de todas partes del mundo y es considerada una fuente de información confiable y completa, siendo utilizada en diversos artículos tales como éste. En particular, los valores de mercado que la página asigna a los jugadores son utilizados por el sitio web *FiveThirtyEight* (fundado por el estadístico Nate Silver) para realizar predicciones sobre el desempeño de los equipos de las distintas ligas de fútbol. Los datos se obtuvieron mediante la técnica de web scraping (“raspado web”), que consiste en crear un programa que escanee el código HTML de un sitio web y automáticamente guarde la información de interés en un formato cómodo para trabajar. Dicho proceso se llevó a cabo con el programa *RStudio* mediante el paquete *rvest*.

Los datos correspondientes al salario de los jugadores provienen del sitio web *Salary Sport*. Esta página fue creada por un equipo de investigadores universitarios que estaban trabajando en un artículo sobre los salarios en el fútbol y decidieron hacer su base de datos pública. Las fuentes que utiliza esta página incluyen comunicados oficiales de los equipos o jugadores, enciclopedias especializadas, artículos de publicaciones confiables, expertos en la industria, entre otros. Finalmente, el número de seguidores de Instagram de cada jugador fue recolectado directamente de su página personal. La recolección de datos se llevó a cabo en junio del 2022. Aunque algunos datos pueden haber cambiado desde esta fecha, se considera que la ventana de tiempo transcurrida es muy pequeña para que haya un cambio significativo en las relaciones observadas con este conjunto de datos en comparación a uno actualizado a julio del 2022. Para el análisis se detallan las variables estudiadas en la Tabla 1. El código en R utilizado para el web scraping como el conjunto de datos analizado se encuentra disponible en el enlace <https://github.com/SGS2000/valor-de-mercado-de-jugadores>.

Para el análisis se estimaron modelos de regresión lineal múltiple utilizando también *R Studio*. Se consideró que era interesante utilizar tanto la variable salario como la variable valor de mercado como variable respuesta, siendo el resto explicativas. En el primer caso, se trata del salario en libras más reciente que se conoce del jugador. Por otra parte, la segunda variable es una estimación

del valor actual de un jugador en el mercado, construida en forma colectiva entre los participantes de *Transfemarket*. Para su construcción se tienen en cuenta múltiples aspectos, que van desde cuestiones monetarias como el salario del jugador o el valor del último pase del jugador pagado por un club a otro, hasta cuestiones más generales como su desempeño tanto en su club como en su selección, su experiencia, susceptibilidad a lesiones, imagen pública, situación financiera del club y la liga en general, tendencias del mercado, factores externos (como por ejemplo el COVID-19), entre muchos otros.

3 Resultados

Se obtuvieron datos de los 200 jugadores mejor valuados según el sitio *Transfermarkt*. Debido a no poseer datos del salario para todos ellos, la muestra se redujo entonces a 188 jugadores. Posteriormente, la muestra disminuyó ligeramente a 184 jugadores. Todos los jugadores de la muestra pertenecen a clubes de la primera división de las ligas europeas de mayor relevancia: Inglaterra, España, Italia, Francia y Alemania. Las nacionalidades de los jugadores son más diversas, habiendo jugadores de más de 30 países de todos los continentes salvo Oceanía, tal y como se observa en la Fig. 1. El valor de mercado para los jugadores de la muestra abarca desde un mínimo de 30 hasta un máximo de 160 millones de euros. El valor mediano es de 50 millones, mientras que la media es de 52.7. Si bien se cubrió un importante rango de edades, siendo el jugador más joven de 18 años y el mayor de 37, los jugadores mayores a 30 y menores a 20 son muy poco frecuentes (Fig. 2), lo que se intuye es un problema de selección de la muestra (ver sección *Discusión*). La edad media fue de 25.47 años con un desvío estándar de 3.2 años.

Los gráficos construidos dan cuenta de la composición de la muestra de acuerdo a las diversas variables consideradas. En la Fig. 3a se observa que el 68% de los jugadores tienen como pierna hábil a la derecha y el 28% a la izquierda, mientras que el 4% restante son ambidiestros. En la Fig. 3b se observa que la mayor parte de los jugadores pertenecen a clubes de Inglaterra, también se observan jugadores de clubes de España e Italia y en menor medida de Alemania y Francia.

Table 1: Variables en el conjunto de datos.

Variable	Nombre en Dataset	Tipo de Variable	Descripción
Nombre del Jugador	Nombre	Cualitativa Identificadora	Nombre y Apellido del Jugador
Valor de Mercado	Valor	Cuantitativa Continua	Estimación (en millones de euros) del valor monetario actual que costaría transferir jugador, según transfermarkt.com.
Salario del Jugador	Último Salario	Cuantitativa Continua	Último salario conocido (en libras) del jugador, según salarysport.com.
Altura	Altura	Cuantitativa Continua	Altura en metros del jugador.
Edad	Edad	Cuantitativa Discreta	Edad actual del jugador.
Pierna Dominante	Pierna	Cualitativa	Indica si la pierna dominante del jugador es la izquierda, derecha o ambas.
Selección	Selección	Cualitativa	Selección nacional a la que pertenece el jugador.
Club	Club	Cualitativa	Club de fútbol que emplea actualmente al jugador.
País del Club	País del Club	Cualitativa	País al que pertenece dicho club.
Posición del Jugador	Posicion	Cualitativa	Tipo de posición en la que juega el individuo (defensor, mediocampista o delantero).
Representante	Representante	Cualitativa Dicotómica	Indica si el jugador es representado por sus familiares o por una empresa.
N° de Partidos Jugados	Partidos	Cuantitativa Discreta	Número de partidos de la temporada actual en los que el jugador participó.
N° de Goles	Goles	Cuantitativa Discreta	Número total de goles realizados por el jugador en la temporada actual.
N° de Goles en Contra	Goles en contra	Cuantitativa Discreta	Indica cuántos de estos goles fueron en contra.
N° de Asistencias	Asistencias	Cuantitativa Discreta	Número de pases realizados por el jugador que resultaron en gol en la temporada actual
N° de Tarjetas Amarillas	Amarillas	Cuantitativa Discreta	Número de tarjetas amarillas recibidas por el jugador en la temporada actual.
N° de Segundas Amarillas	Segundas Amarillas	Cuantitativa Discreta	Indica cuántas de estas tarjetas fueron segundas amarillas.
N° de Tarjetas Rojas	Rojas	Cuantitativa Discreta	Número de tarjetas rojas directas recibidas por el jugador en la temporada actual.
N° de Ingresos	Ingresos	Cuantitativa Discreta	Número de partidos en los cuáles el jugador ingresó al campo de juego sin ser titular.
N° de Salidas	Salidas	Cuantitativa Discretas	Número de partidos en los cuáles el jugador salió del campo de juego.
Seguidores de Instagram	Seguidores	Cuantitativa Continua	Número de seguidores del jugador en Instagram.

En la Fig. 3c se observa que de los jugadores considerados en este estudio las posiciones de juego son relativamente similares: delanteros (35%), mediocampistas (35%) y en menor medida defen-

sores (30%). Se desestimaron del análisis los arqueros, que eran 7, por tratarse de una posición de juego muy diferente a las restantes. En la Fig. 3d se observa que el 8% de los jugadores tiene como

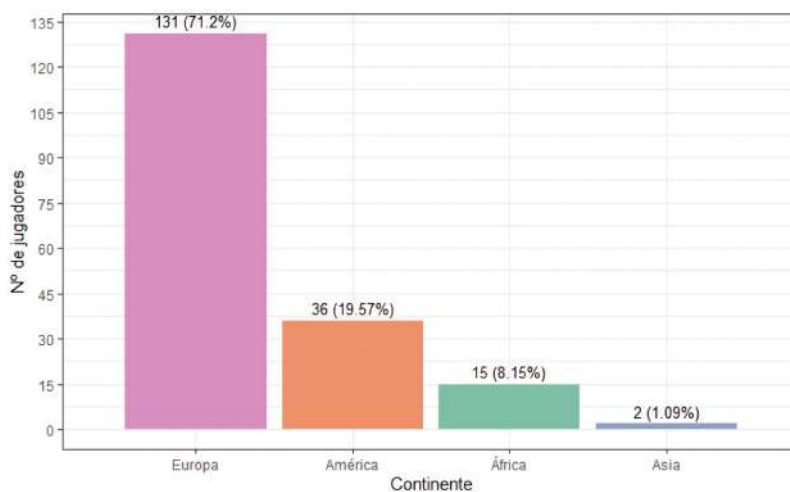


Figure 1: Continente de origen de los jugadores de la muestra.

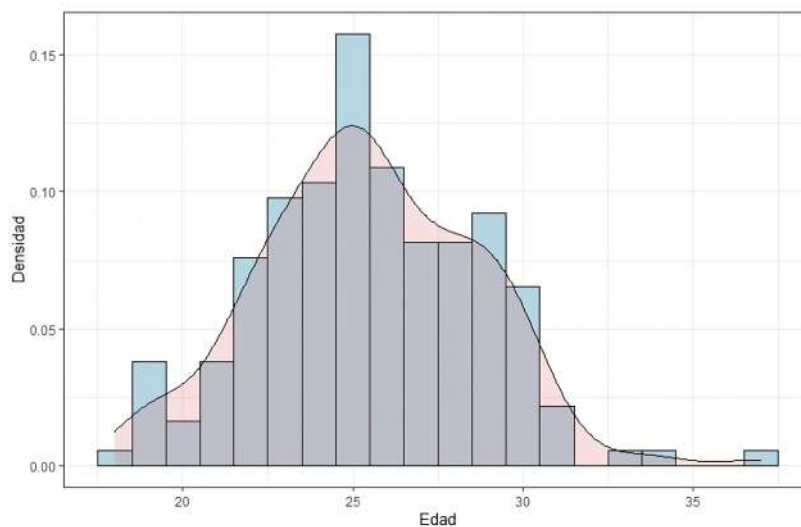


Figure 2: Continente de origen de los jugadores de la muestra.

representante a un familiar.

En la Fig. 4 y Fig. 5 se evalúan asociaciones entre las variables explicativas y el logaritmo del salario con el logaritmo del valor de mercado. En todos los casos parece haber presente una gran variabilidad, pero igualmente se aprecian patrones leves de asociación, tanto positiva como negativa con el logaritmo del valor de mercado. Además, estos patrones parecen ser lineales y no hay indicios fuertes de otras formas de asociación. En el caso del número de partidos jugados, goles por partido, asistencias por partido, $\log(\text{salario})$ y $\log(\text{seguidores})$ se aprecian asociaciones positivas. Es decir, cuando estas variables aumentan el

valor de mercado parece aumentar. Por otro lado, para la edad del jugador, el número de ingresos sin ser titular y el número de amarillas por partido se observa una relación inversa. A medida que éstas variables aumentan el valor de mercado parece disminuir.

Para evaluar asociación con variables que no son cuantitativas, se presentan diagramas de caja de la distribución del logaritmo del valor de mercado del jugador según distintas variables categóricas. Los triángulos dentro de los gráficos representan a la media, las líneas negras a la mediana y los círculos a las observaciones atípicas (*outliers*).

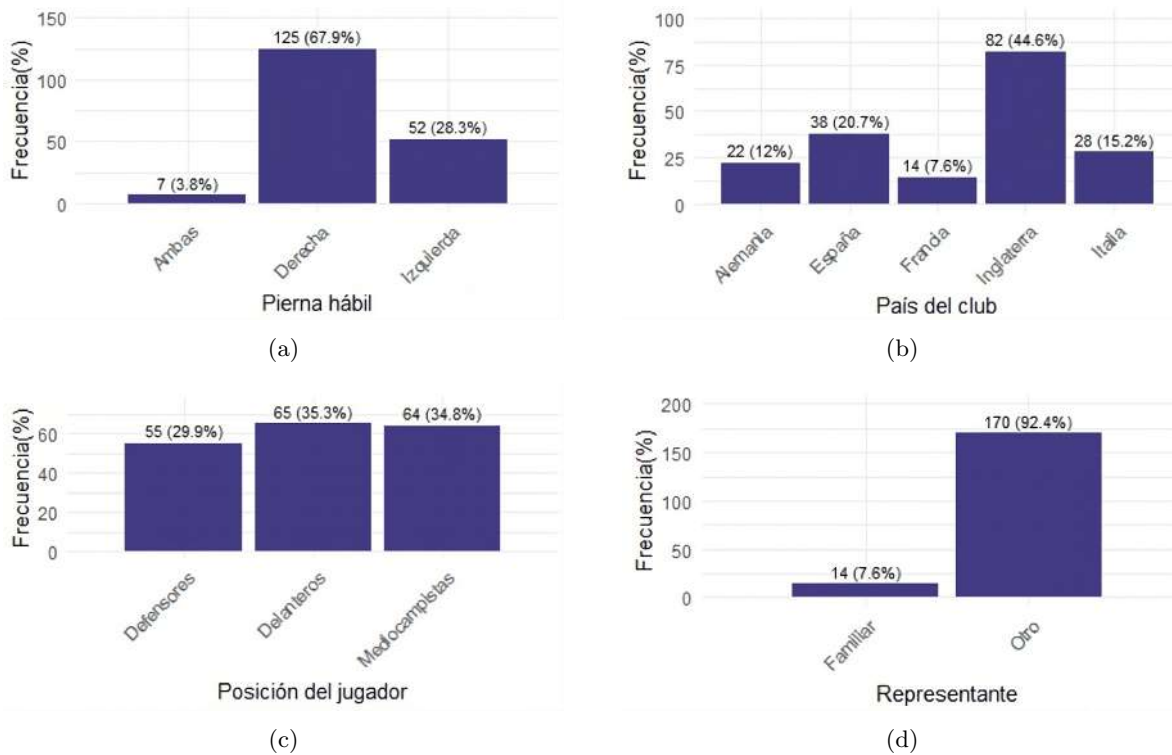


Figure 3: (a) Pierna hábil, (b) país del club, (c) posición y (d) tipo de representante de los jugadores en la muestra

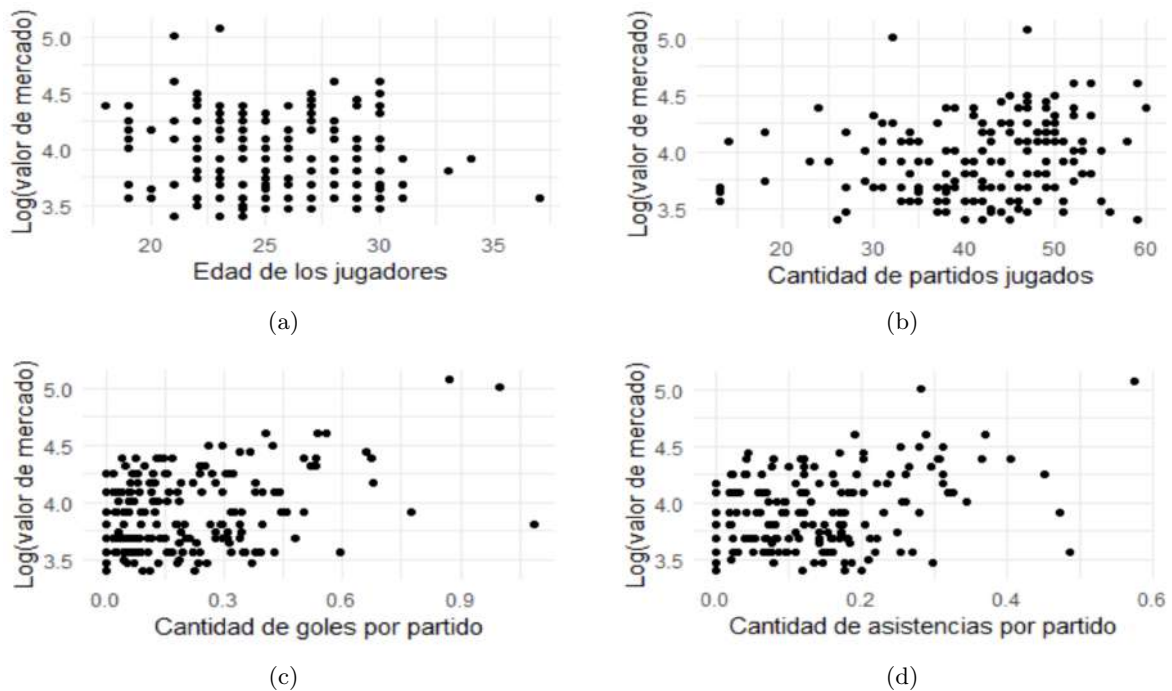


Figure 4: (a) Edad del jugador, (b) cantidad de partidos jugados, (c) cantidad de goles por partido y (d) cantidad de asistencias por partido vs. logaritmo del valor de mercado.

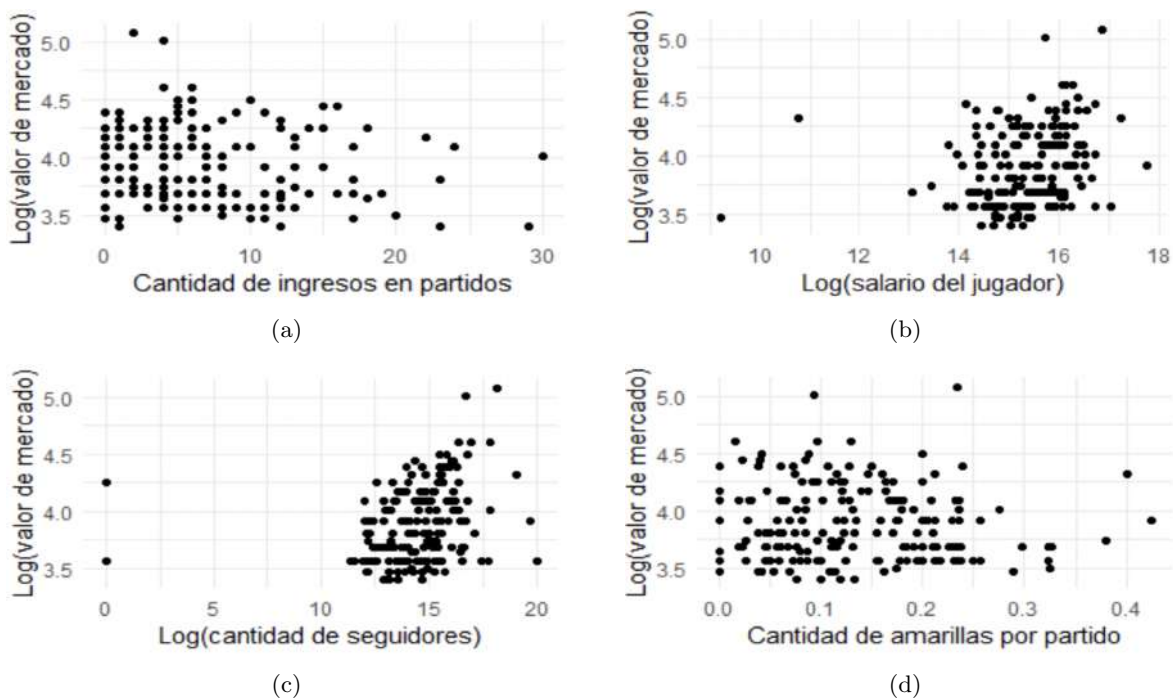


Figure 5: (a) Cantidad de ingresos al campo de juego, (b) logaritmo del salario, (c) logaritmo de la cantidad de seguidores de Instagram y (d) cantidad de tarjetas amarillas por partido vs. logaritmo del valor de mercado.

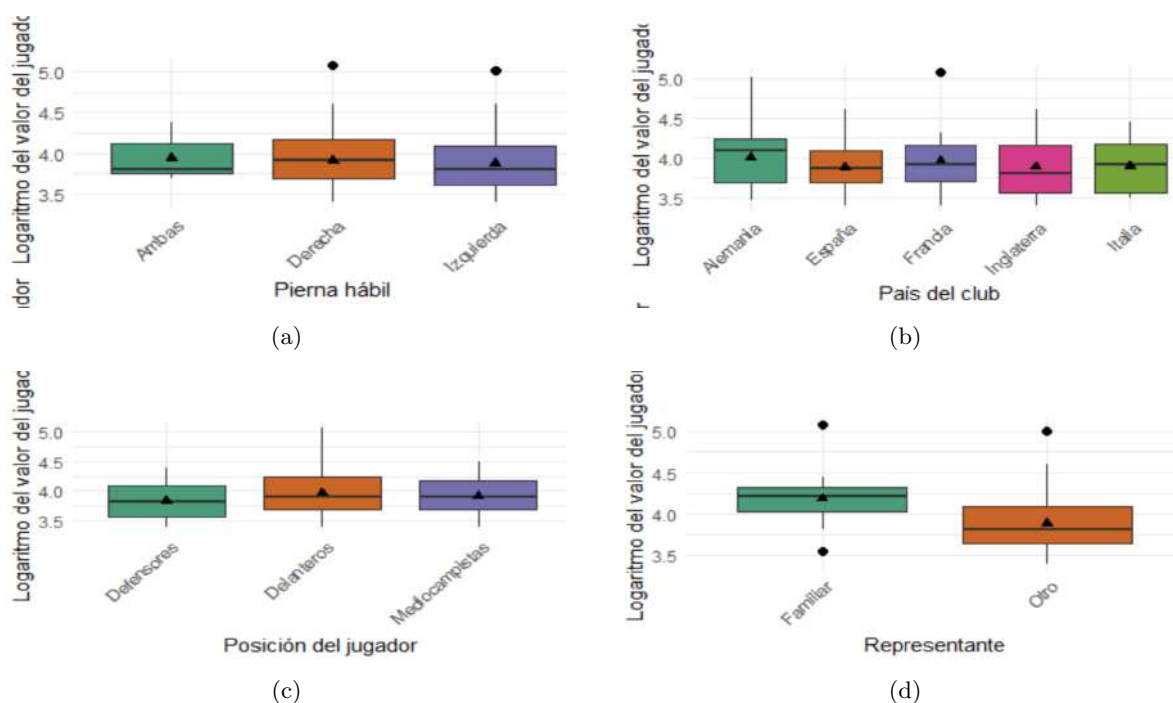


Figure 6: (a) Pierna hábil del jugador, (b) país del club, (c) posición del jugador y (d) tipo de representante vs. logaritmo del valor de mercado.

En la Fig. 6a se observa que el valor medio del logaritmo del valor de mercado es similar para los tres grupos existentes en cuanto a pierna hábil. El valor mediano de aquellos jugadores con ambas piernas hábiles es menor al de los que tienen una sola pierna hábil, y su amplitud es menor. También se puede notar la presencia de outliers. En la Fig. 6b se observa que el promedio del logaritmo del valor de mercado es similar para los 5 países. El logaritmo del valor del jugador parece tener mayor dispersión (rango intercuartílico) para los clubes provenientes de Alemania, Inglaterra e Italia. Alemania parece tener una mayor mediana. Se observa un valor atípico.

En la Fig. 6c se observa que la media del logaritmo del valor de mercado es similar para el grupo de Delanteros y Mediocampistas y levemente menor para el grupo de los defensores. Los jugadores que son defensores parecen tener un valor mediano menor al resto. En la Fig. 6d se observa que el valor medio del logaritmo del valor de mercado difiere según si el representante es un familiar o no. Además se observan outliers. El valor de los jugadores cuyo representante no es un familiar tienen una mayor dispersión.

El proceso de selección del modelo comienza con la modificación de algunas variables, se realizan transformaciones logarítmicas en la variable respuesta valor y en las explicativas salario y seguidores

- En el caso de la variable seguidores, al haber 2 jugadores que no tienen a la fecha cuenta de Instagram se les asigna el valor 1, que luego al aplicarse logaritmo pasa a ser 0.
- Las variables goles, asistencias y amarillas se hacen relativas a la cantidad de partidos donde el jugador participa. Por ejemplo, no se habla ya de cantidad de goles en la temporada sino de goles por partido en la temporada.
- Algunas variables se desestiman para el modelo, como ser Selección por tratarse de una variable categórica con muchas categorías. También Goles en contra, Segundas amarillas y Tarjetas rojas, por tener mayoritariamente valores iguales a cero.
- Además, algunos jugadores se desestiman por el hecho de presentarse variables categóricas muy desbalanceadas. Por ejemplo, hay

únicamente 7 arqueros y además no se cuenta con variables de performance futbolística para estos. Otro caso es el relativo al país al cual pertenece el club del jugador, ya que hay sólo 4 casos de Países Bajos. Además hay dos jugadores para los que no se contaba con información de salarios. Luego, se plantea un modelo con todas las variables candidatas (Tabla 2)

Table 2: Coeficientes estimados del modelo completo.

Variable	Estimación	Error est.	Valor t	p-value
(Intercepto)	2.949	0.812	3.633	<0.001
Altura	-0.061	0.385	-0.159	0.873
Edad	-0.046	0.007	-6.250	<0.001
Pierna (Derecha)	-0.064	0.107	-0.599	0.550
Pierna (Izquierda)	-0.090	0.110	-0.814	0.417
Pais del club (Francia)	-0.125	0.099	-1.263	0.208
Pais del club (España)	-0.003	0.078	-0.036	0.972
Pais del club (Inglaterra)	-0.027	0.071	-0.376	0.708
Pais del club (Italia)	-0.068	0.079	-0.861	0.391
Partidos	0.010	0.002	4.139	<0.001
Goles por (partidos)	0.561	0.166	3.388	0.001
Asist. por (partidos)	0.346	0.242	1.428	0.155
Amarillas por (partidos)	-0.332	0.251	-1.327	0.186
Ingresos	-0.013	0.004	-3.208	0.002
Salidas	-0.002	0.004	-0.520	0.604
log(salario)	0.097	0.025	3.902	<0.001
log(seguidores)	0.030	0.010	2.924	0.004
rep	0.134	0.079	1.690	0.093
Tipo posición (Delanteros)	-0.048	0.087	-0.551	0.582
Tipo posición (Mediocampo)	0.050	0.065	0.764	0.446

Posteriormente, se plantea un modelo con las variables significativas a un nivel del 10% y además se retienen Amarillas por partido y Asistencias por partido ya que son medidas importantes de la performance deportiva de los jugadores, y los conocimientos a nivel teórico indican que deberían tener influencia sobre el valor de mercado (Tabla 3).

Table 3: Coeficientes estimados del segundo modelo estimado.

Variable	Estimación	Error est.	Valor t	p-value
(Intercepto)	2.756	0.329	8.377	< 0.001
edad	-0.045	0.007	-6.354	< 0.001
Partidos	0.010	0.002	4.524	< 0.001
Goles por partido	0.421	0.118	3.578	< 0.001
asist por partido	0.368	0.212	1.736	0.084
amarillas por partido	-0.396	0.238	-1.664	0.098
rep [1]	0.119	0.076	1.557	0.121
ingresos	-0.013	0.003	-3.887	< 0.001
log salario	0.097	0.024	4.005	< 0.001
log seguidores	0.028	0.010	2.744	0.007

Table 4: Coeficientes estimados del modelo final.

Variable	Estimación	Error est.	Valor t	p-value
(Intercepto)	2.683	0.327	8.206	< 0.001
edad	-0.046	0.007	-6.493	< 0.001
partidos	0.009	0.002	4.407	< 0.001
goles por partido	0.431	0.118	3.650	< 0.001
asist por partido	0.408	0.211	1.931	0.055
amarillas por partido	-0.413	0.239	-1.726	0.086
ingresos	-0.013	0.003	-3.880	< 0.001
log salario	0.102	0.024	4.246	< 0.001
log seguidores	0.030	0.010	2.974	0.003

Ahora resulta que dichas variables importantes a nivel teórico sí resultan significativas al 10%, cosa que ya no ocurre con la variable Representante, por lo que se la retira. Luego se evalúan efectos cuadráticos (por ej en la variable Edad) e interacciones de interés (por ejemplo entre la posición y goles/asistencias/amarillas) resultando todos no significativos. Mediante un proceso de transformación de algunas variables y de selección del modelo (Tabla 4), se llegó al modelo final:

$$\begin{aligned} \log(\text{valor}) = & \beta_0 + \beta_1 \text{edad} + \beta_2 \text{partidos} + \beta_3 \text{goles} \\ & + \beta_4 \text{asistencias} + \beta_5 \text{amarillas} \\ & + \beta_6 \text{ingresos} + \beta_7 \log(\text{salario}) \\ & + \beta_8 \log(\text{seguidores}) + u, \end{aligned} \quad (1)$$

donde:

- valor es el valor de mercado del jugador.

- edad es la edad del jugador.
- partidos es el número de partidos en los que participó el jugador en la última temporada.
- goles es la cantidad de goles por partido realizados por el jugador en la última temporada.
- asistencias es el número de asistencias por partido realizadas por el jugador
- amarillas es el número de amarillas por partido obtenidas por el jugador.
- ingresos es la cantidad de veces en la que el jugador ingresó al campo de juego (número de partidos en los que participó sin ser titular)
- salario es el sueldo anual del jugador (en libras esterlinas).
- seguidores es el número de seguidores del jugador en Instagram.

Es importante destacar que se realizaron transformaciones logarítmicas en las variables Valor de Mercado, Salario y Seguidores. En la Fig. 7 se representan dichas variables, tanto en su escala original como en logaritmo. Se observa que éstas variables presentan una marcada asimetría hacia valores positivos, que puede corregirse considerando en escala logarítmica obteniendo distribuciones más simétricas. Considerando además que el uso de transformaciones logarítmicas es una práctica usual en variables económicas y tiene ventajas interpretativas, se decide utilizar estas variables transformadas. La estimación de los parámetros del modelo resultó:

$$\begin{aligned} \log(\text{valor}) = & 2.68 - 0.05 \text{edad} + 0.01 \text{partidos} \\ & + 0.43 \text{goles} + 0.41 \text{asistencias} \\ & - 0.41 \text{amarillas} - 0.01 \text{ingresos} \\ & + 0.1 \log(\text{salario}) + 0.03 \log(\text{seguidores}) \end{aligned} \quad (2)$$

Por lo tanto, las interpretaciones del modelo seleccionado se detallan a continuación. Dadas las restantes variables consideradas en el modelo, se obtiene como resultado que:

- Por cada año que pasa en la carrera del jugador su valor de mercado disminuye casi un 5%.
- Por cada partido adicional en el cual que el jugador participe en la temporada, su valor de mercado se incrementa un 1%.

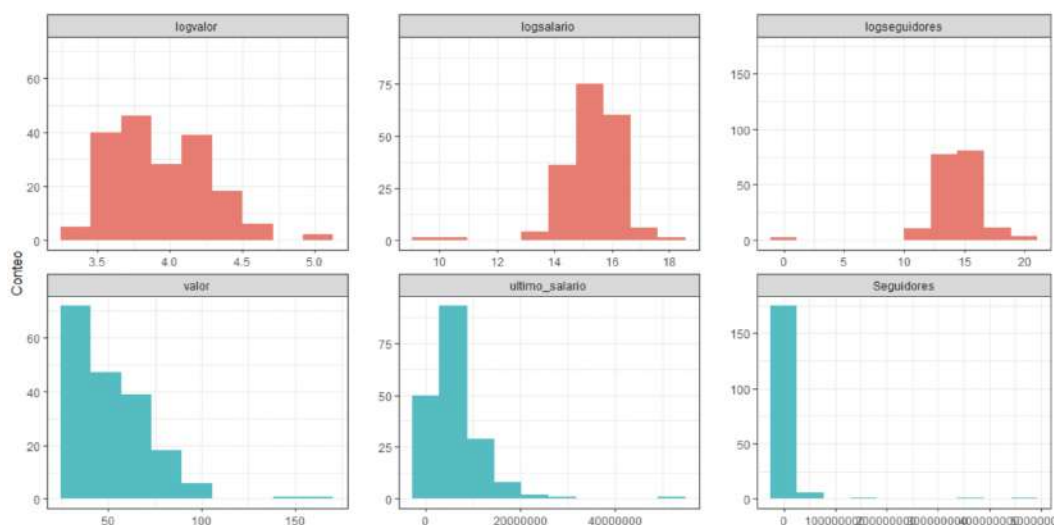


Figure 7: Distribución de las variables logaritmo del valor de mercado, logaritmo del salario, logaritmo de los seguidores de Instagram, valor de mercado, salario y seguidores de Instagram.

- En cuanto a la performance futbolística, un gol adicional cada 10 partidos jugados representa un incremento del valor de mercado del 4,3%. Un incremento en una asistencia cada 10 partidos jugados incrementa el valor de mercado en un 4,1%, y una amarilla adicional cada 10 partidos disminuye su valor de mercado en un 4,1%.
- Por cada partido adicional en la temporada en que el jugador ingrese sin ser titular, su valor de mercado disminuye un 1%. Ante un incremento en un 1% del salario del jugador, su valor de mercado se incrementa un 0,1%.
- Un incremento en un 1% en el número de seguidores de Instagram del jugador representa un incremento en un 0,03% en su valor de mercado. En este punto hay que destacar que hubo dos jugadores que no tenían a la fecha del estudio cuenta en Instagram; si se quita del análisis los mismos el incremento en el valor de mercado asciende a 0,08% ante un aumento en un 1% en el número de seguidores en la mencionada red social.

La medida R^2 ajustado del modelo estimado es de 0,39, indicativa de que el 39% de la variabilidad en el logaritmo del valor de mercado es explicada por las variables seleccionadas. El modelo estimado se adecúa a los supuestos de los modelos lineales.

El modelo final estimado presentó un buen ajuste

en términos de adecuación de los residuos a los supuestos del modelo. Tal y como se aprecia en la Fig. 8 y Fig. 9:

- No hay patrones remanentes de no linealidad en el ajuste: el gráfico de residuos vs. valores predichos presenta una nube de puntos alrededor del cero.
- No hay indicios de variancia no constante ya que no se observan patrones en la amplitud de los puntos que indiquen crecimiento o decrecimiento de la variabilidad.
- No se observan observaciones extremas que tengan un efecto sobre las estimaciones del modelo: todos los residuos se sitúan dentro de los límites, incluso los de alto leverage.
- No existe multicolinealidad entre las variables explicativas: todos los factores de incremento de variancia (VIF) para las variables explicativas son menores a 5.
- No hay indicios de distanciamiento de la normalidad en los residuos: los puntos se ordenan cercanos a la línea de referencia teórica de la distribución normal.

También se decidió plantear otro modelo utilizando como variable respuesta al logaritmo del salario. Nuevamente, se utilizó la escala logarítmica en las variables explicativas valor de mercado y seguidores. Las variables que resul-

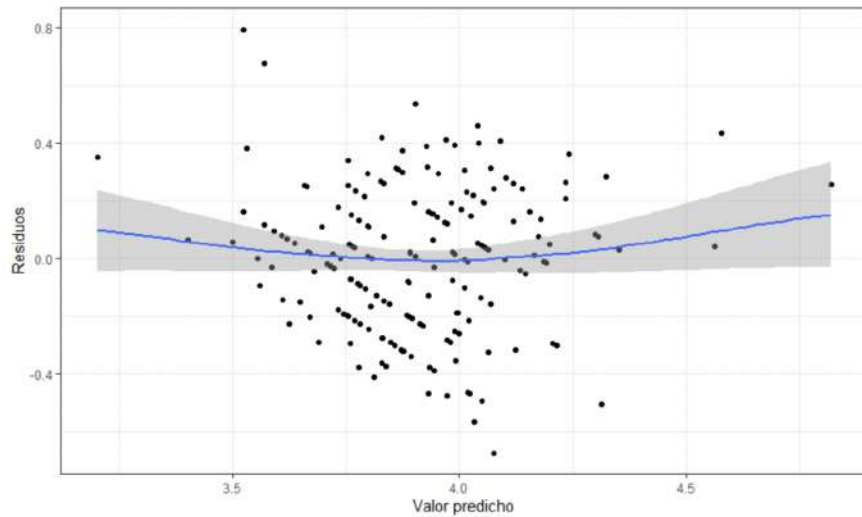


Figure 8: Valores predichos por el modelo vs. residuos.

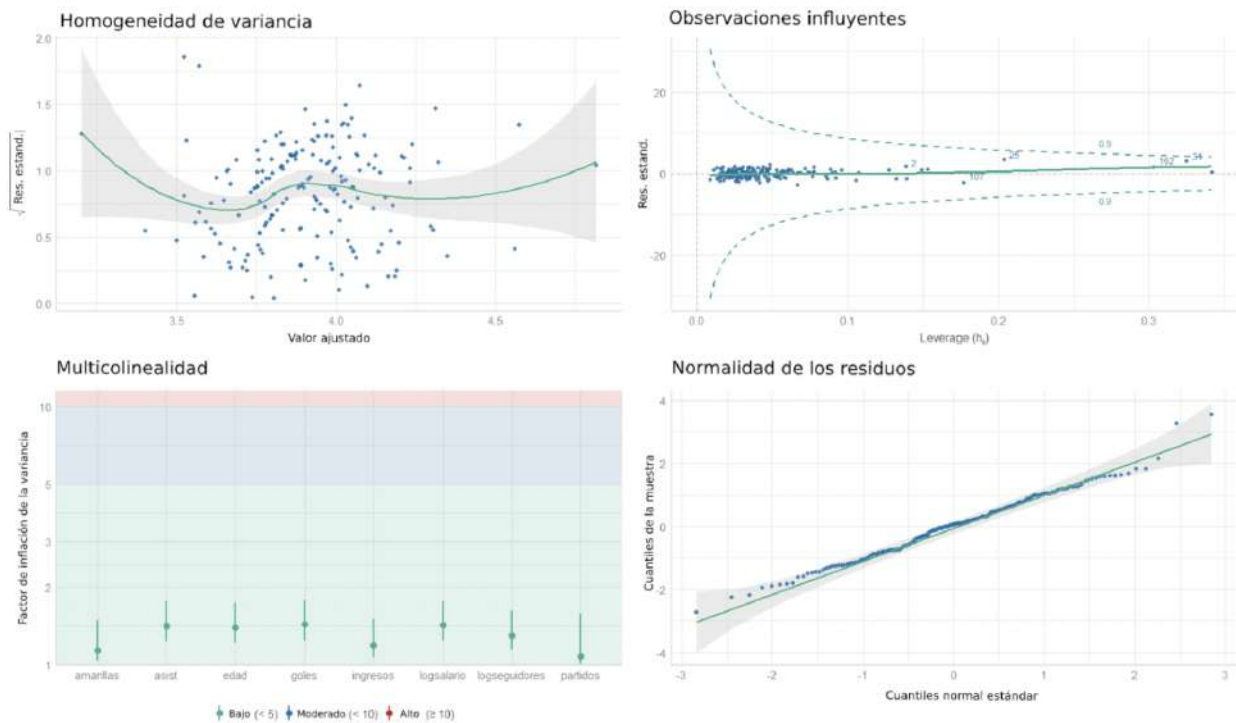


Figure 9: Comprobación de los supuestos del modelo.

taron significativas a un nivel del 10% fueron:

- Edad
- Partidos
- Logaritmo del valor de mercado

- Logaritmo de los seguidores

Al momento de realizarse el análisis de residuos, se observa que no se cumple el supuesto de homogeneidad de variancias, por lo tanto, se decide desestimar este modelo.

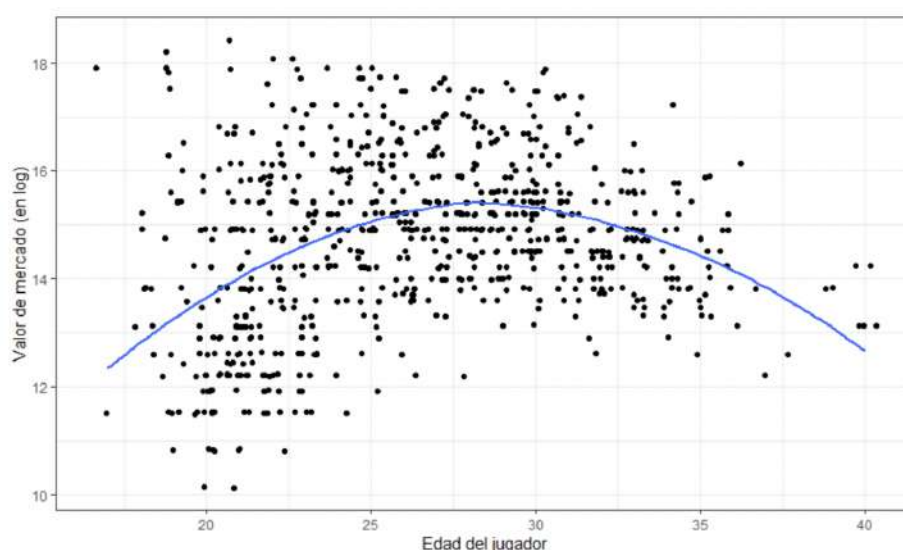


Figure 10: Comprobación de los supuestos del modelo.

4 Conclusiones

Las variables que influyen sobre el valor de mercado del jugador son la edad, el número de partidos jugados, el número de goles, asistencias y amarillas por partido, el número de ingresos sin ser titular, el salario y la cantidad de seguidores de Instagram.

Se logró postular un modelo que cumple con el objetivo de responder a la interrogante de cuáles son los factores que hacen que un jugador sea valioso, cumpliendo dicho modelo con los supuestos correspondientes. El modelo seleccionado presentó un R^2 ajustado igual a 0,39 que si bien es menor al 50%, toma un valor relativamente alto considerando que se trata de un estudio observacional. El modelo seleccionado incluye la variable cantidad de seguidores en Instagram con lo cual no se descarta la hipótesis sugerida de que las redes sociales efectivamente impactan en el valor de los jugadores. Se pudo responder a los dos objetivos establecidos inicialmente de manera satisfactoria.

5 Discusión

El acceso a la información es una gran limitante del análisis. La muestra no fue obtenida aleatoriamente sino por conveniencia, y se circunscribe a jugadores de élite, es decir los 200 jugadores mejor valuados correspondiente a las ligas más im-

portantes, lo que hace que los resultados deban tomarse con precaución: no pueden hacerse extensibles ni a la totalidad de los jugadores de las ligas estudiadas, ni a otras ligas de menor relevancia.

Los problemas que surgen de este limitante se manifiestan en 3 aspectos: en los sesgos en los que se podría incurrir, en los métodos posibles de aplicar y en variables no consideradas en el estudio.

5.1 Sesos

Por el hecho de contar sólo con jugadores de élite, ciertas relaciones conocidas no se dieron tal cual se esperaba, como por ejemplo con la edad. Es de esperar un efecto cuadrático de la edad sobre el valor de los jugadores: su valor aumenta cuando son jóvenes (< 20 años), se estabiliza cuando alcanzan su plenitud para el juego (entre 20 y 30 años) y luego disminuye cuando se acercan a la finalización de su carrera (> 30 años). Esta relación no fue la que se dio en el análisis, pero sí se aprecia en el siguiente gráfico, obtenido con datos de Transfermarkt, para todos los jugadores de la temporada 2021 en la liga de España (Fig. 10).

Además, el hecho de contar sólo con jugadores mejor valuados implica contar con datos más homogéneos con respecto al valor de mercado, lo que pudo haber tenido consecuencias en efectos no reconocidos como importantes. Un ejemplo de esto

es la posición del jugador, que de acuerdo a los conocimientos previos del tema, era de esperarse que tenga mayor relevancia.

Otro ejemplo de las consecuencias de la muestra en el análisis es lo ocurrido con la variable Pierna Hábil, que no resultó significativa. Esto contradice resultados obtenidos en investigaciones previas, tales como este artículo que afirma que los jugadores en Europa que utilizan ambas piernas tienen mayores salarios en promedio. Esto puede deberse a que en la muestra del presente trabajo, únicamente 7 jugadores (3% del total) eran ambidiestros. En un futuro trabajo, se podrían incluir más jugadores que utilicen ambas piernas para evaluar correctamente el efecto de ser ambidiestro sobre el valor de mercado y/o el salario del jugador.

5.2 Metodología de análisis

El estudio es de corte transversal, por lo que se cuenta con información de cada jugador en un momento determinado. Esto no permite incorporar la experiencia y trayectoria del jugador: se sabe que el valor de un jugador hoy, depende del valor que haya tenido el jugador previamente (salarios previos, pases pagados de un club a otro), así como de covariables dependientes en el tiempo (como desempeño en torneos internacionales, lesiones, etc). Se considera que esta autocorrelación en el valor de los jugadores y este tipo de covariables podrían haberse aprovechado mejor mediante un estudio longitudinal.

5.3 Variables no consideradas

Dentro de la muestra por conveniencia hubo variables que no pudieron ser incorporadas, tales como el último valor pagado por un club a otro por el pase del jugador, variable conocida por tener influencia en el valor de mercado. Tampoco se consideraron variables de performance futbolística que abarquen por ejemplo a los arqueros, motivo por el cual se eliminaron de la muestra y variables que incorporen los ingresos publicitarios generados por el jugador.

A pesar de estas limitaciones, el presente trabajo representa un aporte científico de relevancia en la determinación de los factores que influyen en el valor de mercado de los jugadores de élite y puede servir como disparador de estudios futuros

que consideren esta temática.

Referencias

- [1] Bryson, A., Frick, B., & Simmons, R. (2013). The returns to scarce talent: Footedness and player remuneration in European soccer. *Journal of Sports Economics*, 14(6), 606-628.
- [2] Estudio revela que redes sociales influye en sueldo de los futbolistas. (28 de mayo 2014) ESPN. https://www.espn.com.ar/noticias/nota/_/id/2097850/estudio-revela-que-redes-sociales-influye-en-sueldo-de-los-futbolistas
- [3] FiveThirtyEight. Global Club Soccer Rankings. <https://projects.fivethirtyeight.com/soccer-predictions/global-club-rankings/>
- [4] Statista. Revenue of the biggest (Big Five) European soccer leagues from 1996/97 to 2021/22. <https://www.statista.com/statistics/261218/big-five-european-soccer-leagues-revenue/>