

Dominik Głowacki

e-mail: 177530@student.ue.wroc.pl

Uniwersytet Ekonomiczny we Wrocławiu

## Analiza wyceny piłkarzy za pomocą modelowania ekonometrycznego

DOI: 10.15611/2024.53.6.06

JEL Classification: C12, C51, C52, Z23

© 2024 Dominik Głowacki

Praca opublikowana na licencji Creative Commons Uznanie autorstwa-Na tych samych warunkach 4.0 Międzynarodowe (CC BY-SA 4.0). Skrócona treść licencji na <https://creativecommons.org/licenses/by-sa/4.0/deed.pl>

**Cytuj jako:** Głowacki, D. (2024). Analiza wyceny piłkarzy za pomocą modelowania ekonometrycznego. W: A. Grześkowiak, P. Peternek (red.), *Zastosowanie metod ilościowych w ekonomii i finansach* (s. 82-98). Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu.

**Streszczenie:** W artykule podjęto tematykę związaną z piłką nożną, czyli najpopularniejszym sportem na świecie. Główną osią badania była analiza wycen piłkarzy. W pierwszej części przedstawiono ogólne informacje na temat wycen piłkarzy, w tym najdroższe transfery w piłce nożnej. Druga część poświęcona została opisowi metody badawczej – w tym przypadku był to model ekonometryczny i jego weryfikacja. Z kolei w trzeciej części przedstawiono analizę danych, która przeprowadzona została w programie MS Excel. Celem badania było wskazanie zmiennych najsilniej wpływających na wartość piłkarza oraz stworzenie modelu ekonometrycznego pomagającego w wycenie zawodnika. Po zweryfikowaniu współczynnika determinacji, a także po przeprowadzeniu testów normalności i homoskedastyczności wektora reszt, stwierdzono, że pomimo niespełnienia założeń, osiągnięto cel, jakim było wytypowanie zmiennych mających istotny wpływ na wycenę zawodnika. W końcowej części artykułu zaproponowano możliwe działania, które poprawiłyby model pod względem dopasowania do danych oraz spełnienia założeń.

**Słowa kluczowe:** wycena, modelowanie ekonometryczne, piłka nożna

### 1. Wstęp

Bez wątpienia piłka nożna jest jednym z najpopularniejszych sportów w Polsce i na świecie. Zwłaszcza w ostatnim czasie, wobec rosnącej popularności sztucznej inteligencji oraz dzięki rozwojowi zaawansowanych metod badawczych, pojawiło się wiele opracowań dotyczących pozyskiwania danych w sporcie, w tym w piłce nożnej. Sumpter (2022, s. 181) wskazuje, że podczas meczu piłkarskiego rejestruje się ponad 200 milionów liczb. Z kolei Biermann (2022, s. 15) stwierdza, że „piłka nożna stała się grą liczb”. Pomimo tego piłka nożna wciąż uważana jest za dyscyplinę sto-

sunkowo nieskomplikowaną, a jej amatorskie uprawianie nie niesie za sobą znacznych nakładów finansowych jak w przypadku tenisa ziemnego czy sportów zimowych. Nie oznacza to jednak, że piłka nożna nie generuje znacznych pieniędzy, a beneficjentami są zarówno federacje piłkarskie, kluby, trenerzy i zawodnicy. Kontrakty najlepszych piłkarzy na świecie są absolutnie nieporównywalne z zarobkami nawet dobrze opłacanych zawodów. W roku 2023 najlepiej zarabiającym piłkarzem w polskiej ekstraklasie był zawodnik Lecha Poznań, który pobierał niemal 400 tysięcy złotych miesięcznie (Tubacki, 2023). Nierozłączną częścią piłki nożnej są transfery zawodników, czyli kupowanie i sprzedawanie piłkarzy przez kluby piłkarskie. Dotychczas (marzec 2024) najdroższym piłkarzem w historii był Brazylijczyk Neymar da Silva Santos Júnior, który w 2017 roku został kupiony przez Paris Saint-Germain z klubu FC Barcelona za 222 miliony euro. W przypadku polskiej ekstraklasy, czyli najwyższej klasy rozgrywkowej w Polsce, kwoty transferów są zdecydowanie niższe. Najdrożej sprzedanymi piłkarzami z ekstraklasy byli Kacper Kozłowski (w 2022 roku) i Jakub Moder (w 2020 roku). Obaj zostali kupieni przez angielski klub Brighton & Hove Albion za 11 milionów euro, odpowiednio z Pogoni Szczecin i Lecha Poznań. Znacznie mniejszą kwotą transferu był najdrożej kupiony piłkarz przez klub ekstraklasy. W lipcu 2023 roku do Lecha Poznań wytransferowany został Irańczyk Ali Gholizadeh. Klub z Poznania zapłacił belgijskiemu Charleroi S.C. około 1,8 miliona euro. Łatwo zauważalna jest ogromna dysproporcja pomiędzy tym, ile wydać mogą kluby z najlepszych lig świata, zwłaszcza gdy tak jak w przypadku Paris Saint-Germain są zarządzane przez katarskich biznesmanów, a tym, ile mogą zapłacić kluby z Polski. Przedmiotem niniejszego artykułu nie będzie jednak badanie tej dysproporcji, a próba sprawdzenia, skąd biorą się wyceny piłkarzy. Jako podstawowe źródło wiedzy zarówno do danych historycznych, jak i do weryfikacji statystyk posłuży portal Transfermarkt. Analizie poddane zostaną wyceny napastników (piłkarzy, którzy są tak klasyfikowani przez Transfermarkt) z ekstraklasy na dzień 31 stycznia 2024 roku.

## 2. Metodyka badań

W niniejszym rozdziale zaprezentowane zostaną teoretyczne aspekty dotyczące miar, metod i testów statystycznych, które zostaną zastosowane w analizie. Biorąc pod uwagę specyfikę badanego zjawiska, czyli szacowanie wartości piłkarzy, postanowiono, że dobór zmiennych objaśniających nastąpi na podstawie subiektywnej oceny autora po uwzględnieniu korelacjami między poszczególnymi zmiennymi niezależnymi.

Jednym z podstawowych sposobów badania zależności pomiędzy cechami jest współczynnik korelacji liniowej Pearsona o wzorze (Ostasiewicz i in., 2006):

$$r_{xy} = \frac{C(X, Y)}{s_x s_y} \quad (1)$$

gdzie:  $s_x$  – odchylenie standardowe zmiennej  $X$ ,  $s_y$  – odchylenie standardowe zmiennej  $Y$ ,  $C(X,Y)$  kowariancja między cechami  $X$  i  $Y$ .

Na podstawie współczynników korelacji liniowej Pearsona utworzona zostanie macierz korelacji. Głównie na jej podstawie zostaną wytypowane zmienne, które będą tworzyły liniowy model ekonometryczny. Ogólna postać liniowego modelu ekonometrycznego przedstawiona jest poniżej (Dziechciarz, 2003).

$$y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_n X_m + \varepsilon \quad (2)$$

Oznaczone symbolami  $\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m$  są parametrami strukturalnymi modelu i należy je wyestymować, natomiast  $\varepsilon$  to składnik losowy. Estymacja parametrów modelu ekonometrycznego zostanie przeprowadzona za pomocą klasycznej metody najmniejszych kwadratów. Postępowanie w kolejnych krokach od budowy modelu aż po wnioskowanie przeprowadzono, opierając się na opracowaniach naukowych (Dziechciarz, 2003; Bobowski, 2004; Gładysz i Mercik, 2007).

Jednym z podstawowych kryteriów oceny dobroci dopasowania modelu do danych empirycznych jest współczynnik determinacji, który oblicza się wzorem:

$$R^2 = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2} \quad (3)$$

Współczynnik ten przyjmuje wartości z przedziału od zera do jedności. Gdy  $R^2$  osiąga wartość jeden, oznacza to idealne dopasowanie skonstruowanego modelu do danych, a co za tym idzie pełne wyjaśnienie przez model zmienności zmiennej zależnej. Dobrze skonstruowany model powinien osiągać wartość  $R^2$  bliską jedności, należy jednak pamiętać, że zbyt dobre dopasowanie może być przejawem tzw. przeuczenia (Kufel i Lewandowski, 2023). Nie jest to zjawisko korzystne ze względu na możliwe trudności z zastosowaniem modelu do wnioskowania lub prognozowania. Po wstępnym zweryfikowaniu modelu poprzez wartość współczynnika determinacji, należy przystąpić do wykonania testu istotności statystycznej wektora parametrów strukturalnych (test Fishera). Hipoteza zerowa w tym teście zakłada, że cały wektor parametrów strukturalnych modelu jest nieistotny statystycznie. Natomiast hipoteza alternatywna w tym teście mówi, że istnieje przynajmniej jeden z parametrów strukturalnych, który jest istotnie różny od zera. Statystyka testowa:

$$F = \frac{n-m-1}{m} \cdot \frac{R^2}{1-R^2} \quad (4)$$

ma rozkład F-Snedecora z  $n_1 = m$  oraz  $n_2 = m - 1$  stopniami swobody, gdzie  $n$  to liczba obserwacji, a  $m$  to liczba zmiennych objaśniających (Dziechciarz, 2003). W celu zweryfikowania istotności statystycznej poszczególnych parametrów należy zastosować test Studenta. W hipotezie zerowej zakłada się, że badany parametr jest nieistotnie różny od zera, natomiast hipoteza alternatywna jest tego zaprzecze-

niem. Dla każdego kolejnego parametru strukturalnego statystyka testowa wygląda następująco (Górecki, 2010):

$$t_i = \frac{|a_i|}{S(a_i)} \quad (5)$$

i ma rozkład  $t$ -Studenta z  $n - m - 1$  stopniami swobody, gdzie  $a_i$  to ocena parametru, natomiast  $S(a_i)$  to standardowy błąd szacunku parametrów.

Testem na zgodność z rozkładem normalnym wektora reszt, który zostanie zastosowany w niniejszym artykule, jest powszechnie znany test Shapiro-Wilka. Hipoteza zerowa zakłada normalność wektora reszt, natomiast hipoteza alternatywna jest tego zaprzeczeniem. Aby przeprowadzić test, należy uporządkować reszty w sposób niemalejący, a następnie odejmować od siebie zgodnie z regułą „ostatni minus pierwszy, przedostatni minus drugi” itd. Następnie otrzymane w ten sposób różnice trzeba przemnożyć przez odpowiednie współczynniki z tablic. Statystyka testowa ma postać (Dębowska-Mróż i in., 2018):

$$W = \frac{\left[ \sum_{i=1}^{\lfloor \frac{n}{2} \rfloor} a_{n,i} * (e_{(n-i+1)} - e_{(i)}) \right]^2}{\sum_{i=1}^n (e_i - \bar{e})^2} \quad (6)$$

Z tablic testu Shapiro-Wilka należy odczytać wartość krytyczną, a następnie zestawić ją z obliczoną wcześniej statystyką testową. Jeśli wartość statystyki testowej będzie większa niż wartość krytyczna, oznacza to, że nie ma podstaw do odrzucenia hipotezy zerowej, zatem reszty mają rozkład normalny (Dębowska-Mróż i in., 2018).

Ostatnim, niezbędnym do weryfikacji modelu, testem jest sprawdzenie homoskedastyczności składnika resztowego. Zastosowany zostanie w tym celu test Goldfelda-Quandt. Aby przeprowadzić test, należy podzielić obserwacje na dwie grupy, a następnie zbadać, czy wartości wariancji w otrzymanych podpopulacjach są zbliżone do siebie. Liczebności grup powinny być do siebie zbliżone. Podział następuje przeważnie arbitralnie, jednak należy uprzednio przeanalizować wektor reszt (Dziechciarz, 2003). Hipoteza zerowa mówi o tym, że wariancje reszt w obu podpopulacjach nie różnią się od siebie, natomiast hipoteza alternatywna jest negacją tego stwierdzenia. Statystyką testową jest iloraz obu wariancji, przy czym wartość większa znajduje się w liczniku. Wartość krytyczną testu należy wyznaczyć z rozkładu  $F$  – Snedecora z parą stopni swobody:  $n_1 - m - 1$  oraz  $n_2 - m - 1$ . Gdy wartość statystyki testowej jest większa niż wartość krytyczna, należy odrzucić hipotezę o równości wariancji reszt, co oznacza występowanie zjawiska heteroskedastyczności (Dziechciarz, 2003). W związku z tym, że w wielu artykułach naukowych uważa się właśnie normalność i homoskedastyczność składnika resztowego za kluczowe dla poprawności modelu, postanowiono skupić się jedynie na tych dwóch testach, a ich wyniki uznać za najistotniejsze w przypadku oceny jakości modelu.

### 3. Wyniki badań

Analiza dotyczyła 50 najwyżej wycenianych napastników według portalu Transfermarkt. Imiona, nazwiska i kluby, które reprezentowali piłkarze, zostały zaprezentowane w tabeli 1.

**Tabela 1.** 50 najwyżej wycenianych piłkarzy ofensywnych (napastników) PKO BP Ekstraklasy

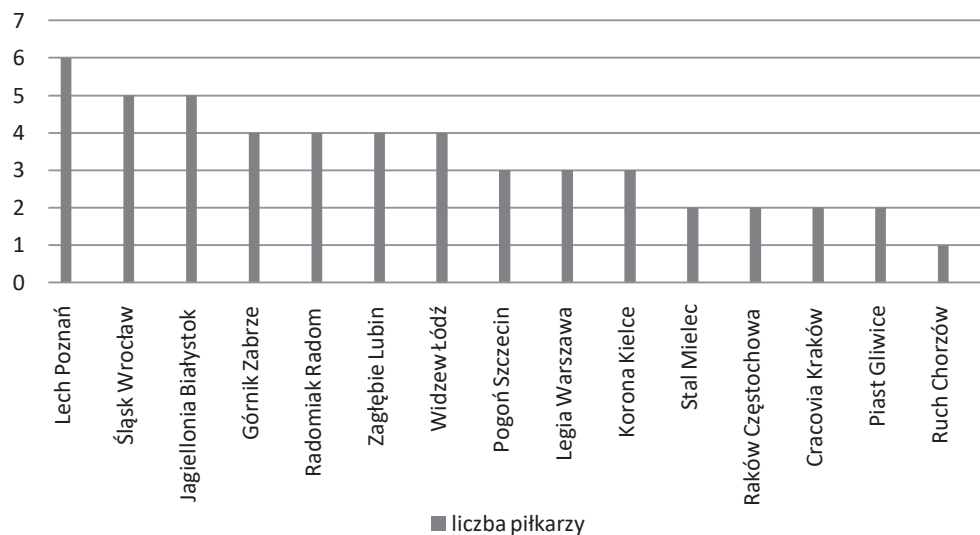
Imię	Nazwisko	Klub
1	2	3
Kristoffer	Velde	Lech Poznań
Filip	Szymczak	Lech Poznań
Erik	Exposito	Śląsk Wrocław
Patryk	Klimala	Śląsk Wrocław
Mikael	Ishak	Lech Poznań
Dino	Hotic	Lech Poznań
Ali	Gholizadeh	Lech Poznań
Efthymios	Koulouris	Pogoń Szczecin
Dominik	Marczuk	Jagiellonia Białystok
Marc	Gual	Legia Warszawa
Adrian	Kapralik	Górnik Zabrze
Pedro	Henrique	Radomiak Radom
Vahan	Bichakhchyan	Pogoń Szczecin
Ilya	Shkurin	Stal Mielec
Ante	Crnac	Raków Częstochowa
Afimico	Pululu	Jagiellonia Białystok
Kacper	Chodyna	Zagłębie Lubin
Jose	Naranjo	Jagiellonia Białystok
Piotr	Samiec-Talar	Śląsk Wrocław
Benjamin	Kallman	Cracovia Kraków
Adriel	Ba Loua	Lech Poznań
Michael	Ameyaw	Piast Gliwice
Mateusz	Wdowiak	Zagłębie Lubin
Vagner	Goncalves	Radomiak Radom
Maciej	Rosółek	Legia Warszawa
Burak	Ince	Śląsk Wrocław
Lisandro	Semedo	Radomiak Radom
Błaz	Kramer	Legia Warszawa
Dawid	Kurminowski	Zagłębie Lubin
Kristoffer	Hansen	Jagiellonia Białystok
Łukasz	Zwoliński	Raków Częstochowa
Jordi	Sanchez	Widzew Łódź
Daniel	Szczepan	Ruch Chorzów
Mateusz	Żukowski	Śląsk Wrocław
Kai	Meriluoto	Stal Mielec

Tabela 1. cd.

1	2	3
Fabian	Piasecki	Piast Gliwice
Bartłomiej	Pawłowski	Widzew Łódź
Mariusz	Fornalczyk	Korona Kielce
Jakub	Łukowski	Korona Kielce
Soichiro	Kozuki	Górnik Zabrze
Kamil	Grosicki	Pogoń Szczecin
Kaan	Caliskaner	Jagiellonia Białystok
Kacper	Śmiglewski	Cracovia Kraków
Kamil	Lukoszek	Górnik Zabrze
Ernest	Terpiłowski	Widzew Łódź
Antoni	Klimek	Widzew Łódź
Sebastian	Musiolik	Górnik Zabrze
Leonardo	Rocha	Radomiak Radom
Juan	Munoz	Zagłębie Lubin
Adrian	Dalmau	Korona Kielce

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

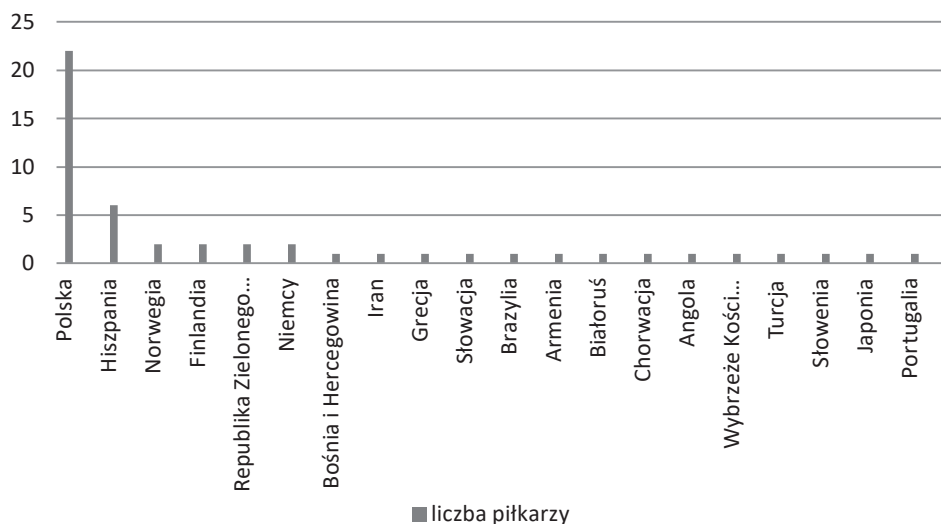
Spśród piłkarzy przedstawionych w tabeli 1, najwięcej reprezentuje Lecha Poznań (6 piłkarzy), a w dalszej kolejności Śląsk Wrocław i Jagiellonię Białystok (po 5 piłkarzy). Szczegółowe informacje co do liczby piłkarzy z poszczególnych klubów przedstawia rysunek 1.



Rys. 1. Liczba najwyżej wycenianych piłkarzy w podziale na kluby

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

Jak wspomniano wcześniej, najliczniejszą grupę stanowią piłkarze z poznańskiego klubu. Wartym odnotowania jest fakt, iż spośród 18 zespołów grających w najwyższej klasie rozgrywkowej w Polsce, wśród 50 najwyższej wycenianych napastników swojego przedstawiciela nie mają trzy kluby. Są to: Puszcza Niepołomice, ŁKS Łódź i Warta Poznań. Ponadto warto przedstawić podział według narodowości piłkarzy, który zaprezentowano na rysunku 2.



**Rys. 2.** Liczba najwyższej wycenianych piłkarzy w podziale na narodowości

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

Zdecydowanie najwięcej piłkarzy spośród wybranych do badania pochodzi z Polski (22 piłkarzy). Oprócz Polski licznie reprezentowanymi nacjami są: Hiszpania (6 piłkarzy) i Norwegia, Finlandia i Republika Zielonego Przylądka (po 2 piłkarzy).

Po ogólnym przedstawieniu podstawowych informacji o piłkarzach, zaprezentowane zostaną dane, które posłużą do modelowania ekonometrycznego. W celu zachowania czytelności danych, zostaną one przedstawione w dwóch częściach w tabelach 2 i 3. Oprócz omówionych wcześniej informacji o klubie, w którym gra dany zawodnik oraz jego narodowości, w tabelach 2 i 3 zawarto wiek zawodnika, długość kontraktu z obecnym klubem (wyrażoną w latach), informację, czy zawodnik jest aktualnym reprezentantem kraju (tylko reprezentacje seniorskie), procent rozegranych minut w ekstraklasie w stosunku do łącznej możliwej liczby minut z rundy jesiennej sezonu 2023/2024 (od 1 lipca do 31 grudnia 2023), liczbę strzelonych goli, liczbę asyst, miejsce drużyny na koniec rundy jesiennej sezonu 2023/2024 (według oficjalnej strony PKO BP Ekstraklasy) oraz wartość zawodnika wyrażoną w milionach euro.

W tabeli 2 zaprezentowano 25 najwyższej wycenianych napastników grających w polskiej ekstraklasie (stan na 31 stycznia 2024).

**Tabela 2.** Dane do analizy – piłkarze ofensywni z miejsc 1-25 w rankingu

Imię	Nazwisko	Klub	Narodowość	Wiek	Pozostały kontrakt (lata)	Czy reprezentant	% minut	Gole	Asysty	Miejsce drużyny	Wartość (mln euro)
Kristoffer	Velde	Lech Poznań	Norwegia	25	2,5	TAK	90	7	4	3	6,0
Filip	Szymczak	Lech Poznań	Polaka	21	3,5	NIE	29	1	2	3	3,0
Erik	Exposito	Śląsk Wrocław	Hiszpania	27	0,5	NIE	100	14	3	1	2,5
Patryk	Klimala	Śląsk Wrocław	Polaka	25	3,5	NIE	9	1	0	1	1,8
Miakael	Ishak	Lech Poznań	Szwecja	30	1,5	TAK	62	6	3	3	1,5
Dino	Hotic	Lech Poznań	Bośnia i Hercegowina	28	2,5	TAK	27	0	2	3	1,5
Ali	Gholizadeh	Lech Poznań	Iran	27	2,5	TAK	10	0	0	3	1,5
Eftymios	Koulouris	Pogoń Szczecin	Grecja	27	2,5	TAK	91	8	1	6	1,5
Dominik	Marczuk	Jagiellonia Białystok	Polaka	20	2,5	NIE	72	2	9	2	1,5
Marc	Gual	Legia Warszawa	Hiszpania	27	2,5	NIE	47	1	2	5	1,3
Adrian	Kapralik	Górnik Zabrze	Słowacja	21	0,5	TAK	64	2	1	7	1,2
Pedro	Henrique	Radomiak Radom	Brazylia	27	2,5	NIE	95	8	4	10	1,2
Vahan	Bichakhchyan	Pogoń Szczecin	Armenia	24	1,5	TAK	75	4	2	6	1,2
Ilya	Shkurin	Stal Mielec	Białoruś	24	0,5	NIE	84	8	5	12	1,2
Ante	Crnac	Raków Częstochowa	Chorwacja	20	4,5	NIE	43	2	2	4	1
Afimico	Pululu	Jagiellonia Białystok	Angola	24	1,5	NIE	55	6	2	2	0,9
Kacper	Chodyna	Zagłębie Lubin	Polaka	24	1,5	NIE	76	3	5	8	0,9
Jose	Naranjo	Jagiellonia Białystok	Hiszpania	29	1,5	NIE	75	6	1	2	0,85
Piotr	Samiec-Talar	Śląsk Wrocław	Polaka	22	0,5	NIE	57	4	5	1	0,8
Benjamin	Kallman	Cracovia Kraków	Finlandia	25	1,5	TAK	83	5	4	13	0,7
Adriel	Ba Loua	Lech Poznań	Wybrzeże Kości Słoniowej	27	1,5	NIE	46	4	2	3	0,7
Michael	Ameyaw	Piast Gliwice	Polaka	23	2,5	NIE	92	2	3	9	0,65
Mateusz	Wdowiak	Zagłębie Lubin	Polaka	27	2,5	NIE	56	4	1	8	0,65
Vagner	Goncalves	Radomiak Radom	Republika Zielonego Przylądka	28	2,5	TAK	0	0	0	10	0,6
Maciej	Rosolek	Legia Warszawa	Polaka	22	1,5	NIE	36	0	2	5	0,6

Źródło: opracowanie własne na podstawie Transfermarkt.pl.



W tabeli 3 przedstawiono napastników z miejsc 26-50 według wycen z portalu Transfermarkt.

Aby możliwa była analiza powyższych danych, zwłaszcza stworzenie modelu ekonometrycznego, dokonano przekodowania zmiennych opisujących narodowość piłkarzy oraz ich uczestnictwo w reprezentacjach narodowych. W pierwszym przypadku „1” oznaczono piłkarzy z Polski, natomiast „0” piłkarzy posiadających obywatelstwo innego kraju. Natomiast w drugiej zmiennej zerowej „1” oznaczono zawodników będących reprezentantami swoich krajów, natomiast „0” piłkarzy, którzy nie zagraли w seniorskich reprezentacjach. W tabelach 4 i 5 zaprezentowano dane do analizy po przekodowaniu dwóch zmiennych.

Na podstawie danych zaprezentowanych w tabelach 4 i 5 stworzono macierz korelacji (tab. 6). Na podstawie przedstawionej w tabeli 6 macierzy korelacji można stwierdzić, że zdecydowana większość wartości współczynników korelacji liniowej Pearsona jest, co do wartości bezwzględnej, niższa niż 0,4, co oznacza brak lub niską zależność liniową pomiędzy zmiennymi. Wyjątek stanowią wartości dla zmiennej opisującej procent rozegranych minut oraz liczbę goli, a także liczbę asyst. Wartości współczynnika korelacji liniowej Pearsona wynoszą w tych przypadkach odpowiednio 0,637 i 0,75. Można jednak stwierdzić, że zależność ta nie jest żadnym zaskoczeniem, gdyż logicznym wydaje się, że jeśli piłkarz grający w ofensywie przebywa dłużej na boisku, to powinien w tym czasie zanotować więcej goli i asyst.

Wobec niskich lub bardzo niskich wartości współczynników korelacji liniowej Pearsona postanowiono stworzyć model ze wszystkimi zmiennymi objaśniającymi, gdzie zmienną zależną jest zmienna wartość. Dla takiego modelu wartość statystyki Fishera wynosi 3,53, natomiast p-value 0,003. Na poziomie istotności  $\alpha = 0,1$  oznacza to, że wektor parametrów strukturalnych modelu jest istotny statystycznie. Następnie przeprowadzono test istotności poszczególnych parametrów modelu – test Studenta. Wartości statystyki  $t$  i p-value dla każdej zmiennej są zamieszczone w tabeli 7.

W tabeli 7 kursywą zaznaczono te zmienne, dla których p-value wynosi mniej niż 0,1, a co za tym idzie, są one istotne statystycznie. W przypadku doboru zmiennych do modelu zdecydowano, aby przyjąć poziom istotności  $\alpha = 0,1$ , a nie stosowany przeważnie  $\alpha = 0,05$  po to, aby do modelu włączyć zmienną *gole*, która w przypadku piłkarzy ofensywnych wydaje się być ważna. Ponowna estymacja modelu zostanie zatem przeprowadzona dla zmiennych objaśniających, które opisują: wiek, pozostały czas kontraktu, granie w reprezentacji narodowej, liczbę goli strzelonych w rundzie jesiennej PKO BP Ekstraklasy oraz miejsce drużyny w tabeli. Ponadto w modelu znajdzie się *wyraz wolny*. Przy powtórznym oszacowaniu parametrów modelu, wszystkie zmienne okazały się istotne statystycznie (na poziomie istotności  $\alpha = 0,05$ ), co przedstawia tabela 8.

Tabela 3. Dane do analizy – piłkarze ofensywni z miejsc 26-50 w rankingu

Imię	Nazwisko	Klub	Narodowość	Wiek	Pozostały kontrakt (lata)	Czy reprezentant	% minut	Gole	Asysty	Miejsce drużyny	Wartość (mln euro)
Burak	Ince	Śląsk Wrocław	Turcja	20	1,5	NIE	22	2	1	1	0,6
Lisandro	Semedo	Radomiak Radom	Republika Zielonego Przylądka	27	0,5	TAK	61	2	2	10	0,6
Blaž	Kramer	Legia Warszawa	Słowenia	27	1,5	TAK	21	2	0	5	0,6
Dawid	Kurminowski	Zagłębie Lubin	Polska	24	1,5	NIE	63	3	0	8	0,55
Kristoffer	Hansen	Jagiellonia Białystok	Norwegia	29	1,5	NIE	39	5	1	2	0,55
Łukasz	Zwoliński	Raków Częstochowa	Polska	30	2,5	NIE	42	4	0	4	0,55
Jordi	Sanchez	Widzew Łódź	Hiszpania	29	0,5	NIE	81	4	3	11	0,55
Daniel	Szczepan	Ruch Chorzów	Polska	28	1,5	NIE	84	7	1	17	0,55
Mateusz	Żukowski	Śląsk Wrocław	Polska	22	0,5	NIE	54	0	0	1	0,5
Kai	Meriluoto	Stal Mielec	Finlandia	21	0,5	TAK	17	2	0	12	0,5
Fabian	Piasecki	Piast Gliwice	Polska	28	3	NIE	29	0	0	9	0,5
Bartłomiej	Pawłowski	Widzew Łódź	Polska	31	2,5	NIE	67	4	3	11	0,5
Mariusz	Formalczak	Korona Kielce	Polska	21	2,5	NIE	14	2	0	16	0,5
Jakub	Łukowski	Korona Kielce	Polska	27	0,5	NIE	13	0	0	16	0,5
Soichiro	Kozuki	Górnik Zabrze	Japonia	23	1	NIE	10	0	0	7	0,5
Kamil	Grosicki	Pogoń Szczecin	Polska	35	2,5	TAK	95	8	8	6	0,5
Kaan	Caliskaner	Jagiellonia Białystok	Niemcy	24	0,5	NIE	24	0	0	2	0,5
Kacper	Śmiglewski	Cracovia Kraków	Polska	19	3,5	NIE	17	0	1	13	0,5
Kamil	Lukoszek	Górnik Zabrze	Polska	21	0,5	NIE	51	0	0	7	0,5
Ernest	Terpifowski	Widzew Łódź	Polska	22	0,5	NIE	38	1	0	11	0,5
Antoni	Klimek	Widzew Łódź	Polska	21	1,5	NIE	46	1	0	11	0,45
Sebastian	Musiolić	Górnik Zabrze	Polska	27	2,5	NIE	50	2	4	7	0,45
Leonardo	Rocha	Radomiak Radom	Portugalia	26	2,5	NIE	19	2	0	10	0,45
Juan	Munoz	Zagłębie Lubin	Hiszpania	28	1,5	NIE	31	2	0	8	0,45
Adrian	Dalmau	Korona Kielce	Hiszpania	29	1,5	NIE	28	3	1	16	0,45

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

Tabela 4. Dane do analizy po zakodowaniu zmiennych zerojedynkowych (zawodnicy z miejsc 1-25)

Imię	Nazwisko	Klub	Narodo- wość	Wiek	Pozostały kontrakt (lata)	Czy reprezentant	% minut	Gole	Asysty	Miejsce drużyny	Wartość (mln euro)
Kristoffer	Velde	Lech Poznań	0	25	2,5	1	90	7	4	3	6,0
Filip	Szymczak	Lech Poznań	1	21	3,5	0	29	1	2	3	3,0
Erik	Exposito	Śląsk Wrocław	0	27	0,5	0	100	14	3	1	2,5
Patryk	Klimala	Śląsk Wrocław	1	25	3,5	0	9	1	0	1	1,8
Miakael	Ishak	Lech Poznań	0	30	1,5	1	62	6	3	3	1,5
Dino	Hotic	Lech Poznań	0	28	2,5	1	27	0	2	3	1,5
Ali	Gholizadeh	Lech Poznań	0	27	2,5	1	10	0	0	3	1,5
Eftymios	Koulouris	Pogoń Szczecin	0	27	2,5	1	91	8	1	6	1,5
Dominik	Marczuk	Jagiellonia Białystok	1	20	2,5	0	72	2	9	2	1,5
Marc	Gual	Legia Warszawa	0	27	2,5	0	47	1	2	5	1,3
Adrian	Kapralik	Górnik Zabrze	0	21	0,5	1	64	2	1	7	1,2
Pedro	Henrique	Radomiak Radom	0	27	2,5	0	95	8	4	10	1,2
Vahan	Bichakhchyan	Pogoń Szczecin	0	24	1,5	1	75	4	2	6	1,2
Ilya	Shkurin	Stal Mielec	0	24	0,5	0	84	8	5	12	1,2
Ante	Crnac	Raków Częstochowa	0	20	4,5	0	43	2	2	4	1,0
Afimico	Pululu	Jagiellonia Białystok	0	24	1,5	0	55	6	2	2	0,9
Kacper	Chodyna	Zagłębie Lubin	1	24	1,5	0	76	3	5	8	0,9
Jose	Naranjo	Jagiellonia Białystok	0	29	1,5	0	75	6	1	2	0,85
Piotr	Samiec-Talar	Śląsk Wrocław	1	22	0,5	0	57	4	5	1	0,8
Benjamin	Kalliman	Cracovia Kraków	0	25	1,5	1	83	5	4	13	0,7
Adriel	Ba Loua	Lech Poznań	0	27	1,5	0	46	4	2	3	0,7
Michael	Ameyaw	Piast Gliwice	1	23	2,5	0	92	2	3	9	0,65
Mateusz	Wdowiak	Zagłębie Lubin	1	27	2,5	0	56	4	1	8	0,65
Vagner	Goncalves	Radomiak Radom	0	28	2,5	1	0	0	0	10	0,6
Maciej	Rosolek	Legia Warszawa	1	22	1,5	0	36	0	2	5	0,6

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

**Tabela 5.** Dane do analizy po zakodowaniu zmiennych zerojedynkowych (zawodnicy z miejsc 26-50)

Imię	Nazwisko	Klub	Narodowość	Wiek	Pozostały kontrakt (lata)	Czy reprezentant	% minut	Gole	Asysty	Miejsce drużyny	Wartość (w mln euro)
Burak	Ince	Śląsk Wrocław	0	20	1,5	0	22	2	1	1	0,60
Lisandro	Semedo	Radomiak Radom	0	27	0,5	1	61	2	2	10	0,60
Blaz	Kramer	Legia Warszawa	0	27	1,5	1	21	2	0	5	0,60
Dawid	Kurminowski	Zagłębie Lubin	1	24	1,5	0	63	3	0	8	0,55
Kristoffer	Hansen	Jagiellonia Białystok	0	29	1,5	0	39	5	1	2	0,55
Łukasz	Zwoliński	Raków Częstochowa	1	30	2,5	0	42	4	0	4	0,55
Jordi	Sanchez	Widzew Łódź	0	29	0,5	0	81	4	3	11	0,55
Daniel	Szczepan	Ruch Chorzów	1	28	1,5	0	84	7	1	17	0,55
Mateusz	Żukowski	Śląsk Wrocław	1	22	0,5	0	54	0	0	1	0,50
Kai	Meriluoto	Stal Mielec	0	21	0,5	1	17	2	0	12	0,50
Fabian	Piasecki	Piast Gliwice	1	28	3	0	29	0	0	9	0,50
Bartłomiej	Pawłowski	Widzew Łódź	1	31	2,5	0	67	4	3	11	0,50
Mariusz	Fornalczyk	Korona Kielce	1	21	2,5	0	14	2	0	16	0,50
Jakub	Łukowski	Korona Kielce	1	27	0,5	0	13	0	0	16	0,50
Soichiro	Kozuki	Górnik Zabrze	0	23	1	0	10	0	0	7	0,50
Kamil	Grosicki	Pogoń Szczecin	1	35	2,5	1	95	8	8	6	0,5
Kaan	Caliskaner	Jagiellonia Białystok	0	24	0,5	0	24	0	0	2	0,50
Kacper	Śmiglewski	Cracovia Kraków	1	19	3,5	0	17	0	1	13	0,50
Kamil	Lukoszek	Górnik Zabrze	1	21	0,5	0	51	0	0	7	0,50
Ernest	Terpitiowski	Widzew Łódź	1	22	0,5	0	38	1	0	11	0,50
Antoni	Klimek	Widzew Łódź	1	21	1,5	0	46	1	0	11	0,45
Sebastian	Musioliik	Górnik Zabrze	1	27	2,5	0	50	2	4	7	0,45
Leonardo	Rocha	Radomiak Radom	0	26	2,5	0	19	2	0	10	0,45
Juan	Munoz	Zagłębie Lubin	0	28	1,5	0	31	2	0	8	0,45
Adrian	Dalmáu	Korona Kielce	0	29	1,5	0	28	3	1	16	0,45

Źródło: opracowanie własne na podstawie Transfermarkt.pl.

Tabela 6. Macierz korelacji

	Narodo- wość	Wiek	Pozostały kontrakt (lata)	Czy repre- zentant	% minut	Gole	Asysty	Miejsce drużyny
Narodowość	1,00	-0,18	0,18	-0,43	-0,01	-0,26	0,09	0,18
Wiek	-0,18	1,00	0,05	0,22	0,22	0,40	0,12	0,04
Pozostały kontrakt (lata)	0,18	0,05	1,00	-0,03	-0,15	-0,11	0,09	-0,09
Czy Reprezentant	-0,43	0,22	-0,03	1,00	0,08	0,09	0,08	-0,04
% minut	-0,01	0,22	-0,15	0,08	1,00	0,75	0,64	-0,04
Gole	-0,26	0,40	-0,11	0,09	0,75	1,00	0,45	-0,08
Asysty	0,09	0,12	0,09	0,08	0,64	0,45	1,00	-0,15
Miejsce drużyny obecny	0,18	0,04	-0,09	-0,04	-0,04	-0,08	-0,15	1,00

Źródło: opracowanie własne.

Tabela 7. Wyniki testu Studenta i p-value

Zmienna	Statystyka t	p-value
Wyraz wolny	2,415	0,020
Narodowość	-0,149	0,883
Wiek	-2,101	0,042
Pozostały kontrakt (lata)	2,218	0,032
Czy reprezentant	2,050	0,047
% Minut	0,271	0,788
Gole	1,769	0,084
Asysty	0,022	0,982
Miejsce drużyny	-2,208	0,033

Źródło: opracowanie własne.

Tabela 8. Wartości statystyk testu Studenta, p-value i symbole zmiennych

Zmienna	Symbol zmiennej	Statystyka t	p-value
Wyraz wolny	-	2,742	0,009
Wiek	$X_1$	-2,238	0,030
Pozostały kontrakt (lata)	$X_2$	2,357	0,023
Czy reprezentant	$X_3$	2,478	0,017
Miejsce drużyny	$X_4$	-2,379	0,022
Gole	$X_5$	3,331	0,002

Źródło: opracowanie własne.

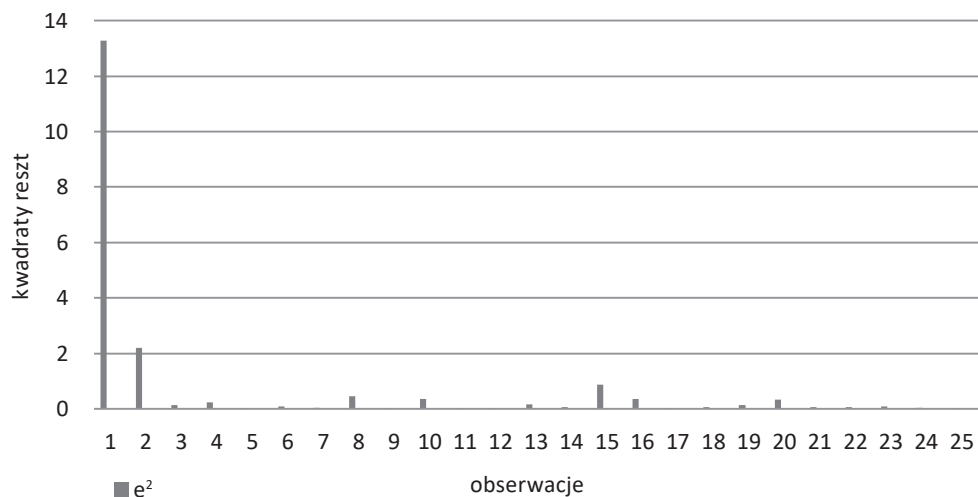
Ponadto w tabeli 8 wskazano symbole poszczególnych zmiennych. Zastosowanie symboli ma na celu uproszczenie zapisu modelu z wyestymowanymi parametrami strukturalnymi:

$$\hat{y} = 2,255 - 0,076X_1 + 0,256X_2 + 0,604X_3 - 0,056X_4 + 0,132X_5 \quad (7)$$

Współczynnik determinacji wyniósł 0,41, co świadczy o słabym (41%) dopasowaniu modelu do danych. W dalszej kolejności reszty modelu zostały poddane sprawdzeniu dwóch własności: normalności i heteroskedastyczności. Uznano, że wykonanie tych dwóch testów (Shapiro-Wilka i Goldfeldta-Quandta) pozwoli na ostateczną weryfikację modelu. Jako pierwszy przeprowadzono test normalności wektora reszt. Statystyka testowa wyniosła 0,18, co na poziomie istotności  $\alpha = 0,05$  oraz wartości krytycznej równej 0,947 oznacza konieczność odrzucenia hipotezy zerowej i przyjęcia alternatywnej. Należy zatem stwierdzić, że reszty nie mają rozkładu normalnego. W dalszej kolejności przeprowadzono test homoskedastyczności składnika resztowego. W tym celu podzielono obserwacje na dwie równoliczne grupy (po 25 obserwacji). Na rysunkach 3 i 4 zaprezentowano wykresy kwadratów reszt modelu w podziale na dwie grupy po 25 obserwacji.

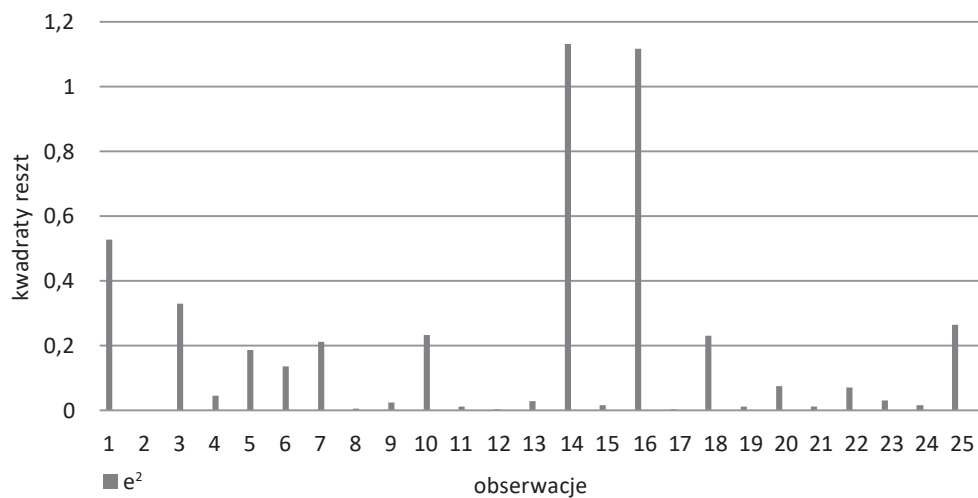
Zestawiając oba wykresy zaprezentowane poniżej, można podejrzewać występowanie zjawiska heteroskedastyczności. Jednak, aby to potwierdzić, należy zastosować odpowiedni test. Statystyka testowa wyniosła 468,76, co jest wartością zdecydowanie większą niż wartość krytyczna, która wynosi 2,01. Można zatem stwierdzić, test Goldfeldta-Quandta wskazał na poziomie istotności  $\alpha = 0,05$ , że w analizowanym modelu występuje zjawisko heteroskedastyczności.

Biorąc pod uwagę trzy poruszone wcześniej kryteria oceny modelu: współczynnik determinacji (dopasowanie modelu do danych), normalność i homoskedastyczność składnika resztowego, należy stwierdzić, że utworzony model jest niesatysfakcjonujący. Wspomniane kryteria nakazują dość ostrożnie podchodzić do wnioskowania na podstawie modelu. Jednak należy uznać, że założony cel – identyfikacja zmiennych mających istotny wpływ na wycenę zawodników – został osiągnięty. Co więcej, analizując znak poszczególnych zmiennych w modelu, można uznać, że mają one swoje logiczne uzasadnienie. Przeważnie wyżej wyceniani są piłkarze młodszy, gdzie często eksperci piłkarscy mówią o „kupowaniu potencjału zawodnika”. W przypadku dłuższego kontraktu również piłkarze są wyceniani wyżej, ponieważ zgodnie z tzw. prawem Bosmana piłkarz, którego kontrakt z danym klubem wygaś, może przenieść się do innego klubu za darmo (Klimczewski, 2020). W przypadku zawodnika będącego reprezentantem kraju również można zauważyć wyższe wyceny portalu Transfermarkt. Podobnie zawodnik, który strzela dużo goli jest wyżej wyceniany niż strzelający mniej, zwłaszcza że analiza skupiona była na piłkarzach ofensywnych. Analogiczny wniosek da się wysnuć w przypadku drużyn zajmujących niskie miejsce w tabeli – taki piłkarz będzie przeważnie niżej wyceniany niż zawodnik drużyny będącej na pozycji lidera tabeli. Można zatem przyjąć, że pomimo niespełnienia założeń, model jest wart uwagi. Zwłaszcza w aspekcie wytypowania dających się logicznie i intuicyjnie uzasadnić zmiennych, które wpływają na wartość zawodnika.



Rys. 3. Wykres kwadratów reszt dla pierwszej grupy

Źródło: opracowanie własne.



Rys. 4. Wykres kwadratów reszt dla drugiej grupy

Źródło: opracowanie własne.

## 4. Podsumowanie

Celem przeprowadzonej analizy było sprawdzenie, co w głównej mierze odpowiada za wartość piłkarza. Miał posłużyć do tego model ekonometryczny, który w przypadku bardzo dobrego dopasowania do danych oraz spełnienia warunków normalności i homoskedastyczności wektora reszt mógłby spełniać rolę pomocniczą w wycenach zawodników. Należy jednakże zauważyć, że sam dobór zmiennych był obciążony ryzykiem niepowodzenia. Do modelu zostały bowiem włączone jedynie zmienne dotyczące bieżącego sezonu. W przypadku piłkarzy można przypuszczać, że ich wartość zależy także od tego, w jakich klubach poprzednio grali, ile notowali bramek i asyst, czy odnosili poważne kontuzje oraz jaki wpływ na drużynę miała ich gra. Z drugiej strony można rozważyć teorię, iż model ekonometryczny nie jest najlepszą metodą do wyceny piłkarzy. Poważnie należałoby rozważyć zastosowanie bardziej zaawansowanych metod, np. z użyciem technik uczenia maszynowego. Innym możliwym rozwiązaniem jest zwiększenie próby i ponowna weryfikacja modelu bazującego na większej liczbie zawodników.

## Literatura

- Biermann, C. (2022). *Piłkarscy hakerzy. O rewolucji w futbolu i sztuce zbierania danych*. Wydawnictwo SQN.
- Bobowski, Z. (2004). Wybrane metody statystyki opisowej i wnioskowania statystycznego. *Nauki Społeczne. Prace Dydaktyczne*, (5).
- Dębowska-Mróz, M., Ferensztajn-Galardos, E., Krajewska, R., i Rogowski, A. (2018). Ocena możliwości zastosowania rozkładu normalnego do opisu wybranych parametrów ruchu drogowego w miastach na przykładzie Radomia. *Autobusy: Technika, Eksploatacja, Systemy Transportowe*, (19). <https://doi.org/10.24136/atest.2018.419>
- Dziechciarz, J. (red.). (2003). *Ekonometria. Metody, przykłady, zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Gładysz, B. i Mercik, J. (2007). *Modelowanie ekonometryczne. Studium przypadku*. Oficyna Wydawnicza Politechniki Wrocławskiej.
- Górecki, B. (2010). *Ekonometria. Podstawy teorii i praktyki*. Key Text.
- Klimczewski, W. (2020, 14 sierpnia). *Największa rewolucja w historii futbolu – Prawo Bosmana*. Pobrano z <https://prawosportowe.pl/a/najwieksza-rewolucja-w-historii-futbolu-prawo-bosmana>
- Kufel, J., i Lewandowski, P. (red.). (2023). *Medycyna Przyszłości: rewolucja sztucznej inteligencji w opiece zdrowotnej*. ArchaeGraph.
- Ostasiewicz, S., Rusnak, Z. i Siedlecka, U. (2006). *Statystyka: Elementy teorii i zadania*. Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu.
- Sumpter, D. (2022). *Piłkomatyka. Matematyczne piękno futbolu*. Copernicus Center Press.
- Tubacki, P. (2023, 8 listopada). *Zarobki w Ekstraklasie. Ile pieniędzy krąży w polskiej lidze?* Pobrano z <https://www.sts.pl/blog/zarobki-w-ekstraklasie/>
- TransferMarkt.pl. (2024). *PKO BP Ekstraklasa*. Pobrano z [https://www.transfermarkt.pl/pko-bp-ekstraklasa/marktwerte/wettbewerb/PL1/plus//galerie/0?pos=Sturm&detailpos=&altersklasse=all&land\\_id=0](https://www.transfermarkt.pl/pko-bp-ekstraklasa/marktwerte/wettbewerb/PL1/plus//galerie/0?pos=Sturm&detailpos=&altersklasse=all&land_id=0)



## Analysis of Football Players Valuation Using Econometric Modeling

**Abstract:** The article discusses topics related to football, which is the most popular sport in the world. The main focus of the research was the analysis of football players valuations. The first part provided general information on player valuations, including the most expensive transfers in football. The second part was devoted to describing the research method – in this case, it was an econometric model and its verification. Meanwhile, the third part presented data analysis, which was conducted using MS Excel software. The aim of the study was to identify variables that most strongly influence the value of a football player and to create an econometric model to assist in player valuation. After verifying the coefficient of determination and conducting tests for the normality and homoscedasticity of the residual vector, it was found that the model did not meet basic requirements, and further inference was deemed unwarranted.

**Keywords:** valuation, econometric modelling, football