

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych

Artur Skowroński

**222 mln euro za piłkarza- determinanty wartości
rynkowych piłkarzy na rynku transferowym**

Praca zaliczeniowa
na przedmiot: Ekonometria

Warszawa, styczeń 2021

Streszczenie

W pracy zostały zbadane determinanty wartości rynkowych 150 piłkarzy z pięciu najlepszych lig świata w sezonie 2020/2021 na dzień 28.12.2020. Analiza została przeprowadzona przy użyciu Klasycznego Modelu Regresji Liniowej, a dane pochodziły ze stron transfermarkt.de, instagram.com oraz ea.com. Moje badanie wykazało, że istotnymi zmiennymi, które mają wpływ na cenę piłkarza to wiek do kwadratu, liczba obserwujących w aplikacji Instagram, rating karty piłkarza z gry komputerowej Fifa21 oraz liczba strzelonych goli.

Słowa kluczowe: model ekonometryczny, piłka nożna, wartość rynkowa, social media

Spis treści

WSTĘP	4
1. PRZEGLĄD LITERATURY	5
2. HIPOTEZY BADAWCZE	8
3. OPIS DANYCH	9
4. DIAGNOSTYKA MODELU	16
5. WERYFIKACJA HIPOTEZ	22
ZAKOŃCZENIE	25
BIBLIOGRAFIA	27

WSTĘP

W historii piłki nożnej możemy wyróżnić dwa przełomowe transfery, które zdecydowanie miały wpływ na późniejsze szacowanie wartości rynkowych piłkarzy. W 2009 roku Real Madryt wykupił portugalskiego napastnika Cristiano Ronaldo za około 94 miliony euro. Na początku, ta kwota mogła wydawać się przesadzona, jednakże po tym historycznym wydarzeniu, takie sumy powoli stawały się normą, a transfery nawet powyżej 100 mln euro były odnotowywane praktycznie co kolejne okno transferowe.

Prawdziwym przełomem był natomiast rok 2017, w którym francuski klub Paris Saint Germain postanowił kupić Neymara za astronomiczną kwotę 222 mln euro. Transfer Neymara wywołał wiele pytań, głównie natury etycznej. Wzbudził także wiele emocji, w tym niedowierzanie i złość¹. Jednak według najnowszego badania przeprowadzonego przez firmę KPMG² wybuch pandemii spowodował spadki na rynku transferowym, aż o 26,5%. Pomimo tego, ceny za piłkarzy nadal przekraczają barierę 100 miliona euro, nie należy spodziewać się, że rekord transferowy za brazylijskiego piłkarza zostanie szybko pobity.

W związku z powyższym, celem pracy jest uzyskanie odpowiedzi na pytanie, co kształtuje tak wysokie ceny piłkarzy na rynku transferowym. Badanie powinno być przydatne dla osób interesujących się sportem, a przede wszystkim piłką nożną oraz dla ludzi którzy chcą zrozumieć, czym kierują się kluby wydając tak ogromne sumy pieniędzy na najlepszych zawodników.

Pierwszy rozdział pracy jest poświęcony szczegółowemu omówieniu literatury, w której już analizowano podobny problem. Następnie na podstawie artykułów naukowych oraz własnej intuicji przedstawiono hipotezę główną oraz dwie hipotezy poboczne. W trzecim rozdziale opisano wszystkie dane, które znalazły się w modelu oraz przeprowadzono weryfikacje poprawności zmiennych. W kolejnym rozdziale zdiagnozowano poprawność modelu przeprowadzając testy zgodne z założeniem KMRL. Ostatni rozdział zawiera interpretację modelu oraz wnioski końcowe.

¹ Yashiv E..2017 The value of top footballers, bubbles, and pitfalls of the free market. <https://voxeu.org/article/neymar-bubble>

² Wiśniewski T, Bernatek A. 2020. Raport KPMG-Player value not immune to pandemic. <https://home.kpmg/pl/pl/home/insights/2020/05/raport-kpmg-pt-player-value-not-immune-to-pandemic.html>

1. PRZEGLĄD LITERATURY

Jednymi z pierwszych osób, które naukowo chciały wyjaśnić co kształtuje tak wysokie ceny za piłkarzy byli Carmichael and Thomas³. Niestety, baza danych do takich badań była w tamtym okresie niedostępna bądź mocno ograniczona. W związku z tym musieli oni opierać swoje analizy na podstawie różnych publikacji prasowych⁴. Badali oni rynek transferowy ligi angielskiej (Premier League), a w swoim modelu posłużyli się takimi zmiennymi jak: wiek piłkarza, liczba zdobytych goli, ogólna charakterystyka piłkarza (waga, wzrost), ilość rozegranych spotkań oraz pozycja na boisku. Według nich, głównymi zmiennymi, które miały wpływ na kształtowanie się rynkowych cen była liczba zdobytych goli oraz liczba ligowych występów (p. 1475). Natomiast jeśli chodzi o wiek, badający doszli do konkluzji, iż w momencie przekroczenia 25 lat, cena rynkowa danego piłkarza zaczyna spadać. W tym miejscu warto zaznaczyć, że aktualnie piłkarze kończą swoje kariery zdecydowanie później niż piłkarze w latach 90. Dlatego też istnieje prawdopodobieństwo, że w dzisiejszych czasach średni wiek dla którego, dany piłkarz osiąga najwyższą wartość rynkową w swojej karierze mógł wzrosnąć.

W swoim następnym badaniu Carmichael, Forrest oraz Simmons⁵ chcieli przewidzieć, które zmienne mogą mieć wpływ na kształtowanie się rynkowych cen. W przeciwieństwie do poprzedniego badania, autorzy nie uwzględnili piłkarzy, którzy nigdy nie zmienili klubu. Po uwzględnieniu zmiennych takich jak wiek, wiek², pozycja zawodnika, międzynarodowe występy oraz cechy drużyny, autorzy doszli do wniosku, iż kluczową zmienną, dzięki której dałoby się przewidzieć jak kształtuje się rynek transferowy są cechy kontraktu. Niestety, w przeciwieństwie do dzisiejszych czasów, nie można było w modelu uwzględnić takich zmiennych jak długość trwania kontraktu oraz data wygaśnięcia kontraktu ponieważ dane te były ściśle chronione.

Kolejnymi osobami, które postanowiły zbadać zależność kształtowania się cen byli Gerrard, Dobson, Howe (2000)⁶. Swój model oparli na wcześniej przeprowadzonym badaniu (Dobson, Gerrard, 1999)⁷, w którym wykazali, iż średnia stopa inflacji stanowi około 11% ceny wartości

³ Carmichael, F. & Thomas, D. (1993). Bargaining in the transfer market: Theory and evidence. *Applied Economics*, 25 (12), 1467-1476.

⁴ Ruijg, J., & van Ophem, H. (2015). Determinants of football transfers. *Applied Economics Letters*, 22(1), 12-19.

⁵ Carmichael, F., Forrest, D. & Simmons, R. (1999). The labour market in association football: Who gets transferred and for how much? *Bulletin of Economic Research*, 51 (2), 125-150.

⁶ Dobson, S., Gerrard, B., & Howe, S. (2000). The determination of transfer fees in English nonleague football. *Applied Economics*, 32 (9), 1145-1152.

⁷ Dobson, S., & Gerrard, B. (1999). The determination of player transfer fees in English professional soccer. *Journal of Sport Management*, 13 (4), 259-279.

piłkarza. W celu kontynuacji badania, chcieli oni sprawdzić, co może kształtować cenę piłkarza. Swoje badania oparli głównie na angielskim amatorskim rynku piłkarskim analizując kariery 114 zawodników w okresie od 1988 do 1997. Model zawierał takie zmienne jak wiek, ilość spotkań w wyjściowej jedenastce, liczba goli, liczba występów na arenie międzynarodowej oraz „goal ratio”. Badacze dowiedli, iż schematy oraz algorytmy tworzenia się cen rynkowych w ligach amatorskich są podobne do tych, które można zaobserwować w ligach profesjonalnych.

Trochę innym tokiem rozumowania w swojej pracy postanowili posłużyć się Lucifora and Simmons (2003), którzy badali ligę włoską (Serie A). Skupili się oni na tym, co wyróżnia danego piłkarza mianem „superstar effect”. W przeciwieństwie do swoich poprzedników, autorzy uznali, że zmienną zdobyte gole w sezonie, należy zamienić na ilość goli na mecz. Dodatkowo, sądzili także, że to co znacząco wpływa na wartość danego piłkarza to nie tylko jego charakterystyki i umiejętności, ale także popularność. Udało im się potwierdzić, że najbardziej wartościowi są Ci piłkarze, którzy zdobywają najwięcej goli na mecz oraz wytwarzają najwięcej okazji dla swoich kolegów z drużyny. Oni generują dla klubów największe zyski i przyciągają publikę, przez co ich cena znacząco wzrasta. Dzięki temu badaniu, można również potwierdzić powszechną opinię, że to napastnicy są zawodnikami o najwyższej wartości rynkowej.

W latach 2004-2007 badacze Barrio and Pujol⁸ podjęli próbę udowodnienia, iż popularność ma zdecydowany wpływ na cenę piłkarza na rynku. Swoje badania oparli na sezonie 2001/2002 ligi hiszpańskiej, a dane uzyskali z popularnej gazety sportowej „Marca”. Oprócz tradycyjnych zmiennych takich jak liczba goli oraz ilość rozegranych meczów, po raz pierwszy uwzględnili oni zmienną opisującą ilość wyszukiwań danego piłkarza w przeglądarce Google. Przeprowadzona analiza danych nie pozostawiała złudzeń, że internetowa popularność danego piłkarza ma znaczący wpływ na jego cenę rynkową oraz znacznie przyczynia się do osiągnięcia korzyści majątkowych przez klub, w którym aktualnie występuje.

W późniejszych latach do podobnych wniosków doszli badacze Lehmann, Schulz⁹ oraz Franck, Nuesch¹⁰ (2012), których badania opierały się na lidze niemieckiej (Bundesliga). Dane,

⁸ Garcia-del-Barrio, P., & Pujol, F. (2004). Pay and performance in the Spanish soccer league: Who gets the expected monopsony rents? Working Paper 05/04.

⁹ Lehmann, E., & Schulze, G. (2008). What does it take to be a star? - The role of performance and the media for German soccer players. *Applied Economics Quarterly*, 54 (1), 59-70.

¹⁰ Franck, E., & Nuesch S. (2012). Talent and/or popularity: What does it take to be a superstar? *Economic Inquiry*, 50 (1), 202–216.

które wykorzystali do wykonania swoich badań pochodziły z gazety Kicker oraz strony internetowej transfermarkt.de obecnie uważanej za największą bazę posiadającą piłkarskie dane statystyczne. Konkluzje, które wynikły z ich badań potwierdzają, że indywidualne umiejętności oraz popularność w social media, są pozytywnie skorelowane z wartością danego piłkarza na rynku.

Ważną zmienną, która badacze brali pod uwagę była również narodowość piłkarza lub jego rasa. Lennart Ante¹¹ w swoim badaniu udowodnił, że narodowość piłkarza ma pozytywny lub negatywny wpływ na transfery w zależności od ligi. Wynika to z faktu, że w niektórych ligach jak np. angielska Premier League, piłkarz spoza UE musi uzyskać pozwolenie na pracę. W większości wypadków, wiąże się to z długimi i złożonymi procesami prawnymi, które mogą być uciążliwe dla klubów. Ze względu na powyższe restrykcje, piłkarze mający podwójne obywatelstwo (przeważnie francusko-afrykańskie), bardzo często decydują się na reprezentowanie barw kraju europejskiego.

Wśród badaczy znajdowali się również przeciwnicy poprzednich badań. W 2007 roku Bernd Frick¹², który jako jeden z pierwszych osób, poddał analizie 5 największych lig na świecie. Stwierdził on, że większość badanych statystyk jest „raczej pośrednia” (p.426) przez co wartość piłkarzy miała ciężki do określenia wpływ na drużynę, za wyjątkiem takich zmiennych jak liczba asyst lub liczba odbiorów. Kluczowymi danymi, które w swoim modelu wyodrębnił Frick, była unikatowa zmienna o nazwie „footedness” obrazująca czy dany piłkarz jest lewo lub prawonożny oraz zmienna obrazująca ilość czasu pozostałą do zakończenia kontraktu. W swoim kolejnym badaniu Bryson, Frick, Simmons¹³ dowiedli, że umiejętność gry lewą nogą jest zdecydowanie rzadsza, co pozytywnie wpływa na rynkową wartość piłkarza.

¹¹ Ant, L. (2019) Determinants of Transfer Fees: Evidence from the Five Major European Football Leagues. University of Hamburg.

¹² Frick, B. (2007). The football players' labor market: Empirical evidence from the major European leagues. *Scottish Journal of Political Economy*, 54 (3), 422-446.

¹³ 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. HIPOTEZY BADAWCZE

W odniesieniu do wyżej przedstawionej literatury oraz bazując na własnych obserwacjach, zdecydowałem się wyróżnić 3 różne hipotezy bazujące na różnych aspektach piłkarskiego życia.

- Hipoteza główna: Podstawowe charakterystyki piłkarza określone poprzez wiek, wpływają dodatnio na wartość rynkową piłkarza.

W artykule napisanym przez Lennarta Ante, autor podjął próbę określenia wpływu podstawowych zmiennych piłkarskich na cenę zawodnika w pięciu najlepszych ligach na świecie. Na podstawie badania można wywnioskować, iż kluczowy wpływ na cenę miała zmienna określająca wiek². W podsumowaniu zaznaczono, że model może odzwierciedlać rzeczywistość jednakże musiałby zostać przeprowadzony bardziej szczegółowo. Uważam, że może to być całkiem logiczne stwierdzenie ponieważ nowoczesna piłka nożna bazuje przede wszystkim na świetnym przygotowaniu motorycznym, a wraz z przyrostem wieku, elementy takie jak dynamika, zwinność lub mobilność powoli zanikają.

- Hipoteza poboczna pierwsza: Na wartość rynkową piłkarza wpływają pozaboiskowe sprawy, a przede wszystkim popularność w social media.

Szerokie wykorzystanie Internetu, a w szczególności aktywność w social media, ma kluczowe znaczenie w dzisiejszych czasach. Aktualnie wiele osób, a w szczególności młodzież, swój jakikolwiek czas wolny przeznaczają na przeglądanie informacji w Internecie. W związku z tym zakładam, że popularność w Internecie może być zmienną, która ma wpływ na cenę piłkarza. Co więcej wydaje mi się, że najpopularniejsze piłkarskie nazwiska na globie nie dość, że przyciągają kibiców na mecze swoimi umiejętnościami, osiągają również status celebryty będąc twarzami wielkich międzynarodowych firm, co może wpływać na ich rynkową wartość. Potwierdzenie tej hipotezy znalazło swoje odzwierciedlenie w artykule Lukasa Barbuscaka, który wykazał iż ilość wyszukiwań danego piłkarza w wyszukiwarce Google.com dodatnio wpływa na cenę rynkową piłkarza.

- Hipoteza poboczna druga: Cena piłkarza uzależniona jest od jego statystyk boiskowych, określonych poprzez liczbę zdobytych goli.

Kluby nie mogą odnosić sukcesów oraz generować zysków nie mając utalentowanych zawodników. To oni stanowią o sile klubu i to dzięki ich umiejętnościom można liczyć na sukces. W związku z tym wydają mi się, że podstawowe statystyki piłkarza takie jak liczba strzelonych goli w dotychczasowym sezonie może mieć zdecydowane przełożenie na jego wartość rynkową. W artykułach wypisanych powyżej, to właśnie ta zmienna przedstawiona w różnych postaciach przeważnie okazywała się być zmienną kluczową.

3. OPIS DANYCH

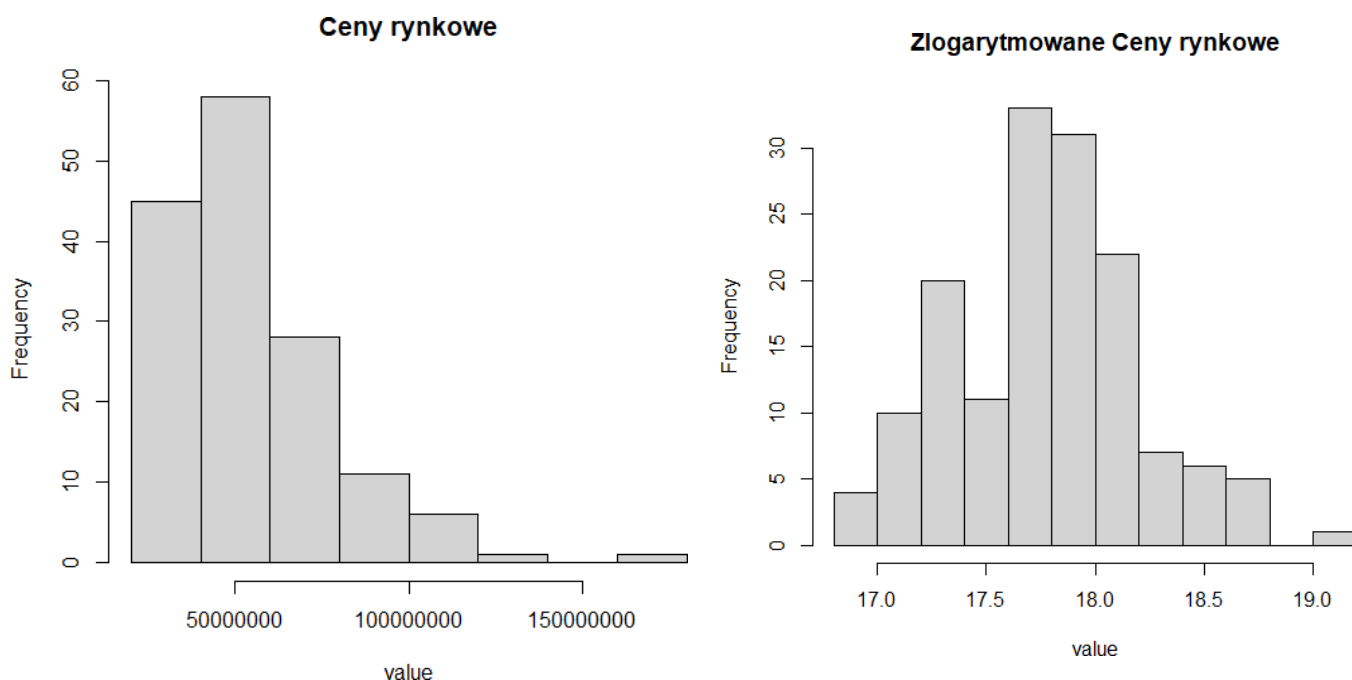
Większość danych pochodzi z polskiej wersji strony internetowej transfermarkt.de, która jest największą ogólnodostępną bazą piłkarską zawierającą podstawowe statystyki piłkarzy. Dane dotyczące liczby obserwatorów w słynnej aplikacji Instagram, zostały zaczerpnięte z kont każdego z badanych piłkarzy dostępnych na stronie instagram.com. Dzięki stronie <https://www.ea.com/pl-pl/games/fifa/fifa-21/ratings> uzyskałem dostęp do indeksów kart zawodników z popularnej gry FIFA21. Model został oparty na 30 aktualnie najdroższych zawodnikach (stan z 28.12.2020) z 5 największych lig na świecie (Premier League, Bundesliga, Primera Division, Ligue 1, Serie A). W ogólnym rozrachunku daje to łącznie 150 obserwacji.

Zmienna zależna:

- Wartość rynkowa (value, $\log(\text{value})$) – zmienna ciągła określająca aktualną wartość rynkową piłkarza w euro (Hendriks 2017, Barbuscak 2018).¹⁴

¹⁴ W nawiasach podano przykładowe artykuły, w których również użyto tych zmiennych.

Rysunek 1-2. Histogramy dla cen rynkowych piłkarzy oraz dla zlogarytmowanych cen.



Źródło: Opracowanie własne na podstawie danych z transfermarkt.pl

Średnia cena za piłkarza wynosi 57 136 667 €. Bazując na histogramie dla wartości rynkowych zawodników jestem skłonny stwierdzić, iż rozkład jest zbliżony do lognormalnego oraz ma prawostronną asymetrię. W związku z tym w modelu użyto logarytmu, co również znajduje swoje odzwierciedlenie w literaturze.

Zmienne niezależne:

- Wiek (age, age²) – zmienna ilościowa ciągła określająca aktualny wiek piłkarza (Carmichael, F., Forrest, D. & Simmons, R. 1999).

W piłce nożnej wiek na pewno stanowi kluczową zmienną, która determinuje na jakim etapie kariery znajduje się obecnie piłkarz. Średnia wieku dla badanych piłkarzy wynosi blisko 25 lat.

- Index (rating_fifa21) – zmienna ilościowa ciągła przyjmująca wartości od 74 do 93, która ma odzwierciedlać umiejętności piłkarza (Majewski 2014).

Zdolności piłkarskie są uznawane za wartości niemierzalne i ciężko je opisać odpowiednim równaniem lub liczbą. Niemniej jednak ten czynnik wydaje się być bardzo

istotny w celu oszacowania poziomu umiejętności piłkarza, które mogą mieć bezpośredni wpływ na jego rynkową wartość. Średnia wartość indeksu wynosi 83,27.

- Pozycja – zmienna dyskretna przyjmująca 4 poziomy, rozkodowana na 3 zmienne zero-jedynkowe: striker-gdy, zawodnik jest napastnikiem, midfielder – gdy, zawodnik jest pomocnikiem, defender – gdy, zawodnik jest obrońcą (Carmichael, F., Forrest, D. & Simmons, R. 1999).

W piłce nożnej można wyróżnić bardzo wiele różnych pozycji. Zawodnik może mieć charakter bycia defensywnym zawodnikiem, ofensywnym, wysuniętym, skrzydłowym itp. W modelu przyjęto klasyczny podział, w którym 41% najdroższych graczy to napastnicy, 33% to pomocnicy, 21% obrońcy i tylko 6% stanowili bramkarze.

- Wzrost (height) – zmienna ilościowa quasi-ciągła określająca wzrost piłkarza zawierająca się w przedziale od 163cm do 199cm (Carmichael, F. & Thomas, D. 1993).

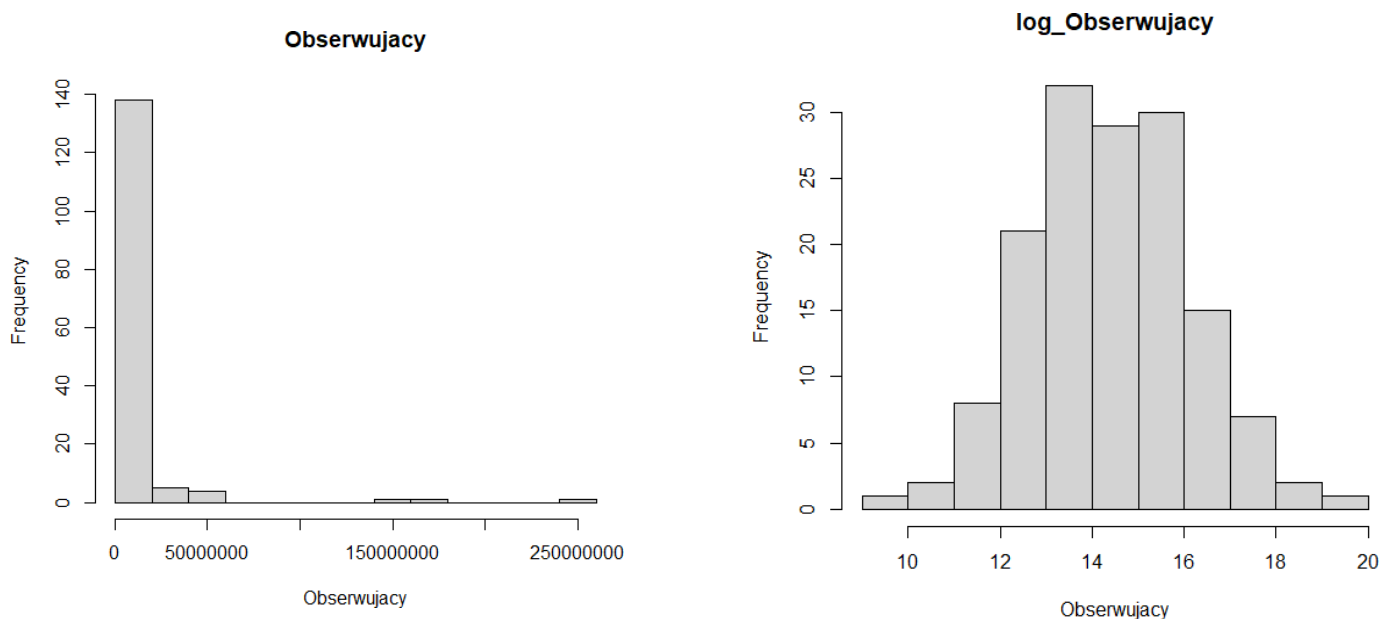
Ta zmienna na pewno jest istotna w sportach takich jak koszykówka lub siatkówka, jednakże postanowiono sprawdzić czy w piłce nożnej również może mieć to przełożenie na wartość rynkową zawodników. Średni wzrost piłkarza wynosi 182,3 cm, a najwyżsi zawodnicy są w grupie bramkarzy.

- Pozostały kontrakt (contract_left) – zmienna ilościowa ciągłą określająca pozostałą liczbę lat do zakończenia kontraktu (Dobson, S., Gerrard, B., & Howe, S 2000).

Bardzo często kluby podpisują z piłkarzami umowy aż na 5 lat, a w zapisach kontraktowych zapisują wysokie klauzule wykupu w celu zniechęcenia innych klubów do podkupywania zawodników. Aktualnie, średnia długość kontraktu wynosi około 3.5 lat.

- Obserwujący (instagram_followers, $\log_{1p}(\text{instagram_followers})$) – zmienna ilościowa quasi-ciągła określająca liczbę obserwatorów danego piłkarza w aplikacji Instagram (Ante, L. 2019).

Rysunek 3-4. Histogramy dla zmiennej obserwujący oraz log(obserwujący).



Źródło: Opracowanie własne na podstawie danych z instagram.com

Średnia liczba obserwujących piłkarza w aplikacji Instagram wynosi 8 489 358. Histogram dla liczby obserwujących jest zbliżony do lognormalnego, więc ostatecznie zlogarytmowano tą wartość, dzięki czemu zbliżono się do rozkładu normalnego.

W przeprowadzonym badaniu 2 piłkarzy nie posiadało konta na tym portalu, w związku z czym przypisano im wartość 0, a zastosowany logarytm dla wektora obserwujących miał wzór $\log(1+x)$.

- Reprezentant – zmienna dyskretna przyjmująca trzy poziomy, rozkodowana na dwie zmienne zero-jedynkowe: europe – gdy zawodnik pochodzi z Europy, south_america- gdy zawodnik pochodzi z Ameryki Południowej (Ante, L. 2019).

W analizie, aż 73% piłkarzy pochodziło z Europy, 19% z Ameryki Południowej. Pozostałe 9% to reprezentanci Ameryki Północnej, Azji lub Afryki. Niestety w żadnej z 5 najsilniejszych lig europejskich nie znalazł się piłkarz z Australii i Ocenii.

- „Footedness” – zmienna dyskretna przyjmująca trzy poziomy, rozkodowana na dwie zmienne zero-jedynkowe: twolegs-gdy zawodnik jest obunożny, leftleg-gdy zawodnik jest lewonożny (Frick, B. 2007).

Zdecydowana większość piłkarzy jest tylko prawonożnych i stanowi ona 67% całej próbki. Natomiast 27% piłkarzy jest tylko lewnonożnych, zaś piłkarze obunożni stanowią 6% badanych.

- Liczba goli (goals) – zmienna ilościowa quasi-ciągła określająca liczbę ligowych goli zdobytych w sezonie 2020/2021 do dnia 28.12.2020 (Dobson, S., Gerrard, B., & Howe, S 2000).

Niewątpliwie, gole to najbardziej charakterystyczny symbol piłki nożnej, a klubowym włodarzom zdecydowanie powinno zależeć na pozyskiwaniu bramkostrzelnych zawodników. W momencie przeprowadzania analizy, badani zawodnicy mieli na koncie strzelone średnio 2.4 gola od początku sezonu.

- Napastnik z Europy (europe_striker) – zmienna binarna przyjmująca wartość 0 jeśli piłkarz nie jest z Europy i nie gra na pozycji napastnika oraz 1 w przeciwnym przypadku.

Bardzo często kluby walczące o utrzymanie w lidze kupują piłkarzy na ostatnią chwilę, którzy mają zapewnić drużynom kluczowe zwycięstwa. Od zawodników spoza Europy wymaga się uprzedniego uzyskania pozwolenia na pracę w danym kraju europejskim. Zagrożone spadkiem kluby chcąc działać szybko, zawężają swoje poszukiwania zawodników tylko do skali europejskiej. Pozwala im to mieć gracza gotowego do gry już po zapłaceniu kwoty odstępnego. Napastnicy, którzy jednocześnie pochodzą z Europy stanowią 27% wszystkich badanych.

Współliniowość

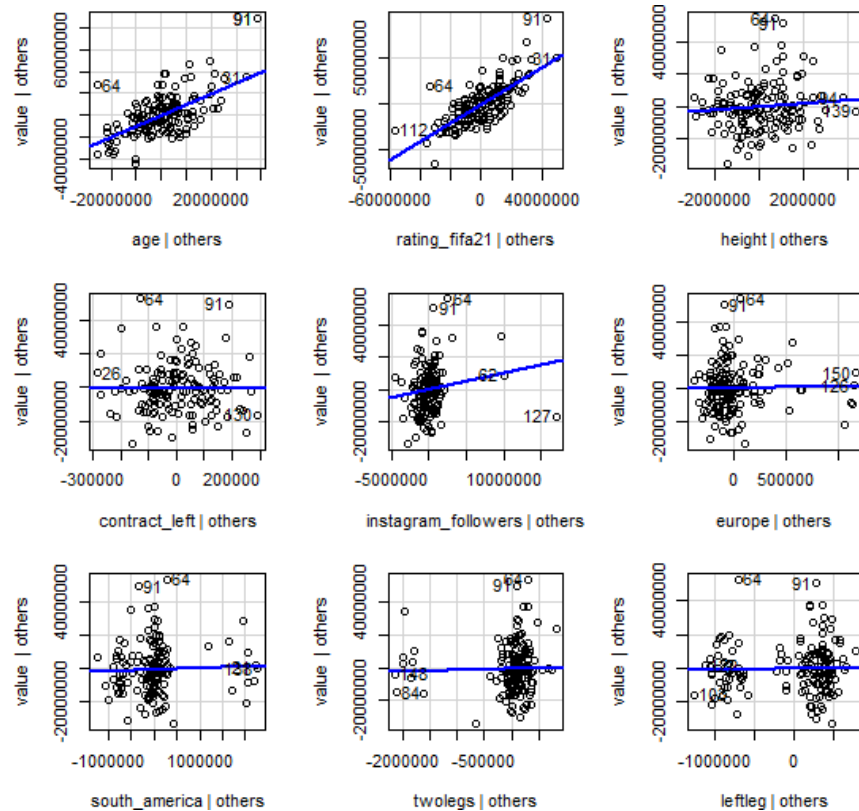
Podstawowym zagadnieniem w celu sprawdzenia czy moje zmienne są ze sobą silnie skorelowane jest współczynnik VIF, dzięki któremu jestem w stanie stwierdzić czy w moim modelu nie występuje niedokładna współliniowość. Taka sytuacja miałaby miejsce, gdyby współczynnik ten był większy od 10. W modelu podstawowym żadna ze zmiennych nie przekraczała wartości VIF równej 5,48 tak więc śmiało mogłem wykluczyć problem współliniowości.

Obserwacje nietypowe i błędne

W celu zidentyfikowania zmiennych nietypowych lub błędnych użyłem 3 miar: standaryzowane reszty (outliers), dźwigni oraz odległości Cooka. Używając pierwszej statystyki należy pamiętać, że obserwacje należy uznać za nietypową, gdy wartość

bezwzględna standaryzowanej reszty jest większa od 2. W swoim modelu zaobserwowałem 9 takich obserwacji, w związku z czym postanowiłem przyjrzeć się im bliżej w celu sprawdzenia ich poprawności. Przeszedłem do zbadania nietypowości korzystając ze statystyki dźwigni.

Rysunek 5. Wykresy dźwigni dla niektórych wyodrębnionych zmiennych.

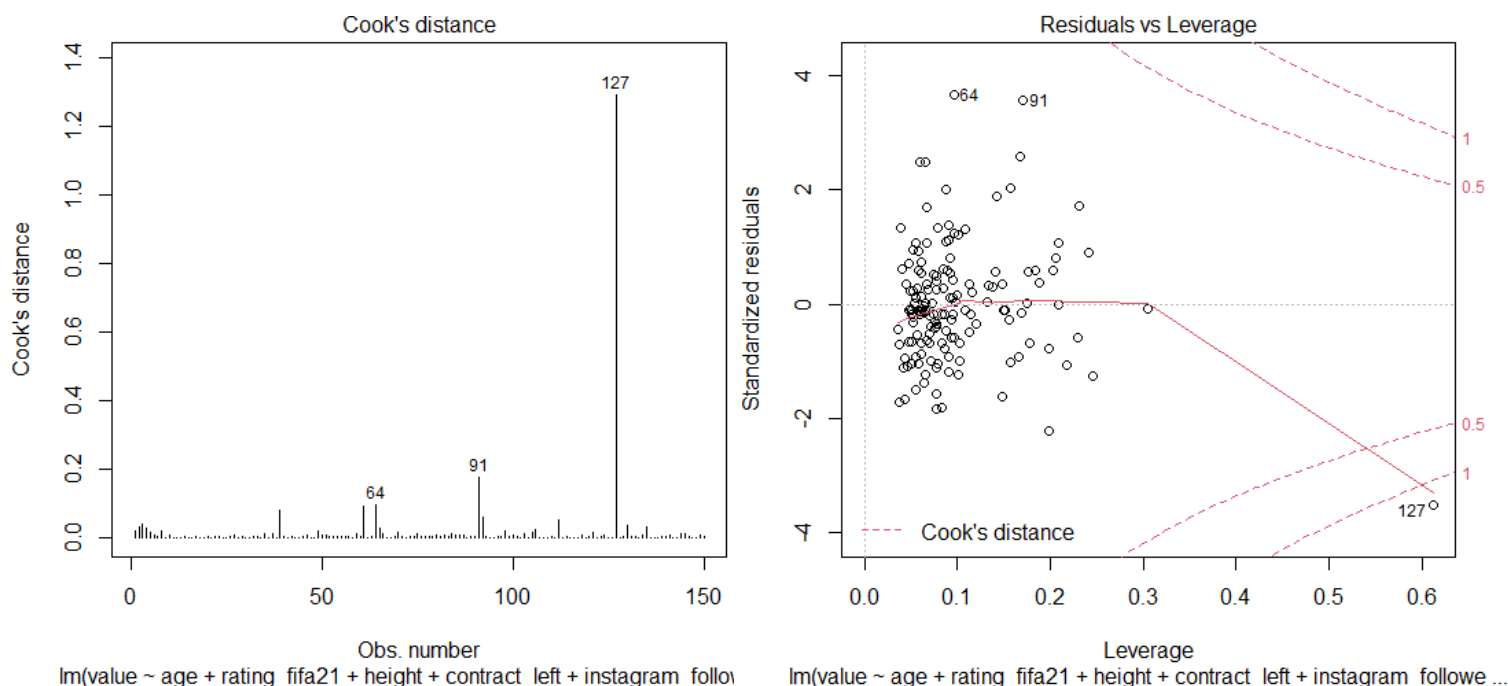


Źródło: Opracowanie własne na podstawie danych z serwisu transfermarkt.pl, ea.com, instagram.com

Pamiętając o tym, że obserwacja nietypowa charakteryzuje się wartością wyższą od $2K/N$ za pomocą statystyki dźwigni otrzymałem kilka nietypowych obserwacji, jednakże najbardziej nietypową okazała się być obserwacja numer 64 oraz obserwacja numer 91, które pojawiły się na wszystkich wykresach.

Ostatnią statystyką, która pomogła mi ustalić czy w moim modelu występuje problem z błędnymi obserwacjami jest odległość Cooka, w której wg teorii w zależności od podręcznika obserwacja nietypowa jest wtedy, gdy $CD > 1$ lub $CD > 4/N$.

Rysunek 6-7. Odległość Cooka oraz stosunek reszt z odległością Cooka.



Źródło: Opracowanie własne na podstawie danych z serwisu transfermarkt.pl, ea.com, instagram.com

Tym razem program wyróżnił jako wątpliwe 3 obserwacje, a przede wszystkim ponownie obserwację numer 64 oraz 91. Stwierdzam, że wyróżnione obserwacje na pewno nie są błędne i nie wynikają z niezamierzonego błędu, a są po prostu nietypowe. Odchylenia mogą wynikać z tego, że Antoine Griezmann (obs. 64) strzelił w tym sezonie mało goli, natomiast jego wartość rynkowa jest nadal wysoka. Natomiast Kylian Mbappe (obs.91) ma tylko 22 lata, a jest aktualnie najdrożej wycenianym zawodnikiem na rynku transferowym. Warto jednak zwrócić uwagę na obserwację nr. 127, która znacznie przekracza granicą wartość 1 na wykresie porównującym stosunek reszt z odległością. Tym zawodnikiem okazuje się być Cristiano Ronaldo, który pomimo bycia najstarszym w stawce jest nadal jednym z najbardziej bramkostrzelnych zawodników. Co więcej jego zasięgi w social media znacznie przekraczają możliwości innych piłkarzy. Pomimo swojego wieku, nadal jest on bardzo cennym piłkarzem na rynku transferowym. Jest to obserwacja jak najbardziej nietypowa, dlatego nie usunięto jej z modelu.

W modelu nie stwierdzono nieprawidłowości danych, a obserwacje które budziły wątpliwości są jak najbardziej autentyczne i nie wynikają z błędu. W związku z tym przystąpiono do procesu uzyskania najlepszego możliwego modelu.

4. DIAGNOSTYKA MODELU

W badaniu wykorzystano Metodę Najmniejszych Kwadratów OLS, dzięki której mogłem wyestymować nieznane parametry w różnych wyodrębnionych wersjach modelu regresji liniowej. W celu poprawy modelu dokonywałem kolejnych iteracji, które przedstawiam poniżej:

Model bazowy:

$$\begin{aligned} value_i = & \beta_0 + \beta_1 age_i + \beta_2 rating_fifa21_i + \beta_3 height_i + \beta_4 contract_left_i + \\ & \beta_5 instagram_followers_i + \mu_1 europe_i + \mu_2 south_america_i + \gamma_1 twolegs_i + \gamma_2 leftleg_i + \phi_1 striker_i \\ & + \phi_2 midfielder_i + \phi_3 defender_i + \beta_6 goals_i + \beta_7 europe_striker_i + \epsilon_i \end{aligned}$$

Pierwsza iteracja:

$$\begin{aligned} value_i = & \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 rating_fifa21_i + \beta_4 height_i + \beta_5 contract_left_i + \\ & \beta_6 instagram_followers_i + \mu_1 europe_i + \mu_2 south_america_i + \gamma_1 twolegs_i + \gamma_2 leftleg_i + \phi_1 striker_i \\ & + \phi_2 midfielder_i + \phi_3 defender_i + \beta_7 goals_i + \beta_8 europe_striker_i + \epsilon_i \end{aligned}$$

Druga iteracja:

$$\begin{aligned} \log(value_i) = & \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 rating_fifa21_i + \beta_4 height_i + \beta_5 contract_left_i + \\ & \beta_6 \log(instagram_followers_i) + \mu_1 europe_i + \mu_2 south_america_i + \gamma_1 twolegs_i + \gamma_2 leftleg_i + \phi_1 striker_i \\ & + \phi_2 midfielder_i + \phi_3 defender_i + \beta_7 goals_i + \beta_8 europe_striker_i + \epsilon_i \end{aligned}$$

Trzecia iteracja:

$$\begin{aligned} \log(value_i) = & \beta_0 + \beta_1 age_i + \beta_2 age_i^2 + \beta_3 rating_fifa21_i + \beta_4 height_i + \beta_5 contract_left_i + \\ & \beta_6 \log(instagram_followers_i) + \mu_1 europe_i + \mu_2 south_america_i + \gamma_1 twolegs_i + \gamma_2 leftleg_i + \\ & \phi_1 striker_i + \phi_2 midfielder_i + \phi_3 defender_i + \beta_7 goals_i + \beta_8 europe_striker_i + \epsilon_i \end{aligned}$$

Model ostateczny:

$$\log(value_i) = \beta_0 + \beta_1 age_i^2 + \beta_2 rating_fifa21_i + \beta_3 \log(instagram_followers_i) + \beta_4 goals_i + \epsilon_i$$

Kroki podjęte w celu poprawy modelu

Po przeanalizowaniu poprawności zmiennych, przeprowadzono odpowiednie testy, które mają na celu zbadać spełnienie założeń Klasycznego Modelu Regresji Liniowej (KMRL). Dzięki temu jestem w stanie stwierdzić, czy mój model jest nieobciążony oraz efektywny tzn. czy parametry obrazują nam rzeczywistość oraz czy zmienne w moim modelu są istotne. Gdybym nie spełnił większości założeń KMRL moje wyjaśnienie modelu mogłoby okazać się nieprawidłowe oraz podważałoby możliwość poprawnej interpretacji ekonomicznej wartości modelu.

W celu sprawdzenia prawidłowości formy funkcyjnej przeprowadzono test RESET, którego hipoteza zerowa zakłada liniową postać modelu, a alternatywna przeciwnie. Dla poziomu istotności $\alpha = 0,05$ zarówno testy przy wykorzystaniu potęg zmiennych niezależnych (regressor) oraz potęg wartości dopasowanych (fitted) wyszły odpowiednio: wartość testu $F=20,327$ oraz $F=2,2644$, a ich $p\text{-value}=0,0000$ oraz $p\text{value}=0,0014$, przez co jestem zmuszony do odrzucenia hipotezy zerowej.

Następnym regułą KMRL jest założenie o nielosowości zmiennych. W związku z tym przeprowadziłem test na homoskedastyczność Breusch'a-Peagan'a, z którego wartość BP wyszła 32,53, a $p\text{-value}=0,0033$, tak więc także i w tym teście muszę odrzucić hipotezę zerową mówiącą stałości wariancji. W związku z niespełnieniem dwóch podstawowych założeń KMRL już w tym momencie mogę uznać, że model należy poprawić ponieważ moje estymatory mogą być nieefektywne.

Chcąc sprawdzić, czy spełniam pozostałe reguły KMRL sprawdziłem za pomocą testu Jarque-Berra założenie o rozkładzie normalnym błędu losowego. Hipoteza zerowa testu zakłada normalność zaburzenia losowego. Wyniki testu przy poziomie istotności $\alpha = 0,05$ wyszły następujące: $X\text{-squared}=26,435$ przy $p\text{value} = 0,0000$ Nie spełniając tego założenia narażam się na nieprawidłowe wnioskowanie statystyczne co wiąże się z niemożliwością znalezienia najlepszego estymatora b, dlatego w tym przypadku również jestem zmuszony do odrzucenia hipotezy zerowej.

Ostatnie założenie KMRL mówi o braku autokorelacji reszt. Na tym etapie mogę stwierdzić, że mając dane przekrojowe problem autokorelacji mnie nie dotyczy, w przeciwieństwie do przypadku posiadania danych panelowych. Nie zmienia to faktu, iż postanowiłem sprawdzić jak to wygląda w moim sytuacji. Chcąc sprawdzić czy w moim modelu występuje ten problem przeprowadziłem test Durbina-Watsona, który bada

kowariancję reszt między punktem i , a $i-1$. Również i w tym teście przy poziomie istotności $\alpha = 0,05$ p-value wyszło mniejsze. Zostałem zatem zmuszony do ponownego odrzucenia hipotezy zerowej, która zakłada kowariancję między błędami losowymi równą 0 na rzecz hipotezy przeciwnej.

Iteracja pierwsza

W celu poprawy modelu, bazując na literaturze¹⁵ oraz teorii dotyczącej przede wszystkim braku spełnienia warunku liniowości dokonałem pierwszej iteracji, w której podniosłem do kwadratu moją zmienną niezależną wiek (age), a następnie przeprowadziłem ponownie testy sprawdzające zgodność z założeniami KMRL. Jako pierwszy przeprowadziłem test RESET przy wykorzystaniu potęg wartości dopasowanych $F(2,132)=22,22$ oraz $p\text{-value}=0,000$ na poziomie istotności $\alpha = 0,05$ musiałem odrzucić hipotezę zerową mówiącą o prawidłowej formie funkcyjnej. Dla pewności przeprowadziłem jeszcze test RESET wykorzystując potęgi zmienne niezależne otrzymałem wartość $F(30, 104)=2.0095$ przy $p\text{-value}=0,0019$ w związku z tym dla dwóch sposobów, nie udało mi się uzyskać prawidłowej formy funkcyjnej.

Kolejny test miał na celu sprawdzenie homoskedastyczności mojego modelu czyli założenia o stałości wariancji błędu losowego dla wszystkich obserwacji. W tym celu przeprowadziłem test Breusch'a-Peagan'a, z którego statystyka testowa wyszła $\chi^2(15)=29,83$, przy $p\text{-value} = 0,0125$ na poziomie istotności $\alpha = 0,05$. W związku z tym musiałem odrzucić hipotezę zerową.

Następnym testem jaki wykonałem był test Jarque-Berra, który jest podstawowym testem sprawdzającym normalność reszt. Pomimo wprowadzenia potęgi do zmiennej niezależnej wiek (age), nie udało mi się spełnić założenia osiągając $X\text{-squared}=27,726$ oraz $p\text{-value}=0,0000$ w związku z tym po raz kolejny musiałem odrzucić hipotezę zerową.

Ostatnim założeniem, które postanowiłem sprawdzić jest reguła o braku autokorelacji reszt czyli moment, w którym kowariancja między błędami losowymi $= 0$. W moim przypadku po zastosowaniu testu Durбина-Watsona przy poziomie istotności $\alpha = 0,05$ p-value wyszło mi 0.000 tak więc po raz kolejny miałem w swoim modelu autokorelację reszt.

¹⁵ Ante, L. (2019). Determinants of Transfers Fees: Evidence from the Five Major European Football Leagues. University of Hamburg, p.6.

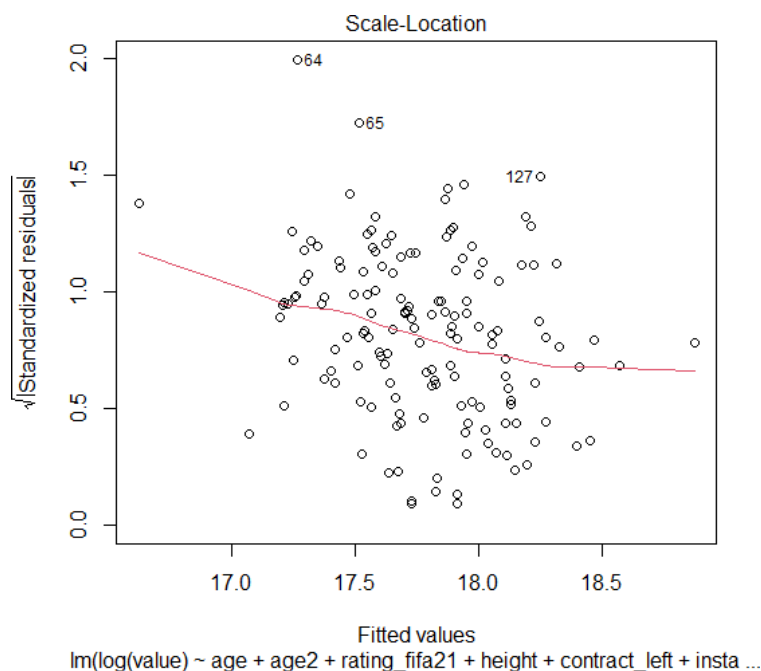
Iteracja druga

W ogólnym rozrachunku dodając zmienną wiek^2 niestety nie poprawiłem swojego modelu. Tym razem postanowiłem zająć się problemem heteroskedastyczności. Według literatury, teorii oraz własnej intuicji postanowiłem zlogarytmować zmienną objaśnianą wartość rynkową (value).

Przeprowadziłem test RESET w celu sprawdzenia poprawności mojej formy funkcyjnej przy wykorzystaniu potęg wartości dopasowanych $F(2, 133)=21,615$ oraz $p\text{-value}=0,076$ na poziomie istotności $\alpha = 0,05$ nie mogłem odrzucić hipotezę zerowej mówiącą o prawidłowej formie funkcyjnej. W celu zweryfikowania pierwszego zadowalającego wyniku przeprowadziłem ponownie test RESET, ale dla potęg zmiennych zależnych. Statystyka testowa, która otrzymałem wynosi $F(30, 104) = 1,0688$ a $p\text{-value} = 0,3892$, tak więc dla obu testów mogę stwierdzić, że udało mi się spełnić warunek liniowości.

Następnie postanowiłem sprawdzić homogeniczność wariancji. Oprócz testów, które już wyżej wymieniłem dzięki którym można zbadać ten warunek, postanowiłem tym razem również narysować wykres *scale-location*.

Rysunek 8. Wykres zależności między resztami a wartościami dopasowanymi.



Źródło: Opracowanie własne na podstawie danych z transfermarkt.pl, instagram.com oraz ea.com

Na podstawie wykresu badającym związek pomiędzy resztami, a wartościami dopasowanymi jestem w stanie stwierdzić, że wykres może sprawiać wrażenie sferycznego. Dla potwierdzenia mojego przypuszczenia przeprowadziłem test Breusch'a-Peagan'a $\chi^2(15)=22,59$, przy $p\text{-value} = 0,0932$ na poziomie istotności $\alpha = 0,05$ nie mogę odrzucić hipotezy zerowej, więc moje przypuszczenia okazały się być prawdziwe.

Iteracja trzecia

Po przeprowadzeniu odpowiednich testów w celu weryfikacji założeń KMRL, co prawda udało mi się spełnić warunki o prawidłowości formy funkcyjnej oraz homoskedastyczności wariancji, jednakże problem z założeniami o normalności rozkładu reszt oraz braku autokorelacji reszt pozostawał niezmieniony. Cechą testu Jarque-Bera jest wysoka wrażliwość na niewielkie odchylenia od rozkładu normalnego. Zgodnie z teorią, mógłbym rozważyć czy moja próbka jest dostatecznie duża, żeby nie uwzględniać założenia o normalności rozkładu reszt w moim modelu mając spełnione dwa pierwsze warunki. Jednakże, po przeanalizowaniu literatury¹⁶ postanowiłem zlogarytmować zmienną obserwowaną (instagram_followers) funkcją $\log(1+x)$, ponieważ w dwóch obserwacjach, piłkarze nie mieli założonych kont w słynnej aplikacji, a wartość $\log(0)$ nie istnieje.

Zarówno po przeprowadzeniu testów na prawidłowość formy funkcyjnej jak i stałości wariancji składnika losowego, nie miałem podstaw do odrzucenia hipotez zerowych, tak więc podobnie jak w iteracji drugiej udało mi się spełnić dwa podstawowe założenia.

Następnie przeszedłem do analizy normalności rozkładu reszt. Po przeprowadzeniu testu Jarque-Bera udało mi się uzyskać zadowalające wartości $X\text{-squared}=5,5106$ oraz $p\text{-value} = 0,06$, dzięki czemu po raz pierwszy nie miałem podstaw do odrzucenia hipotezy zerowej zakładającej normalność składnika losowego.

Czysto formalnie sprawdziłem założenie o braku autokorelacji reszt. Przeprowadzając test Durбина-Watsona statystyka testowa wyszła mi 1,40, przy $p\text{-value} = 0,0001$ przez co musiałem po raz kolejny odrzucić hipotezę zerową.

¹⁶ Ante, L. (2019). Determinants of Transfers Fees: Evidence from the Five Major European Football Leagues. University of Hamburg, p.7.

Iteracja czwarta

Bazując na formie funkcyjnej modelu z trzeciej iteracji przeprowadziłem procedurę „General to specific” chcąc sprawdzić istotność wszystkich zmiennych użytych w regresji. Na początku sprawdziłem łączną istotność wszystkich zmiennych nieistotnych. Hipoteza zerowa zakłada łączną nieistotność zmiennych nieistotnych. Niestety w moim modelu $p\text{-value} = 0,0134$ co przy poziomie istotności $\alpha = 0,05$ nakazuje mi odrzucić hipotezę zerową. Przystąpiłem więc do usuwania kolejnych zmiennych nieistotnych. Na każdym kroku sprawdzałem łączną nieistotność za pomocą testu anova. Do porównań, używałem modelu z iteracji czwartej z tym samym modelem, jednakże na każdym kroku zmniejszałem go o zmienne nieistotne, których $p\text{-value}$ było największe.

Mój końcowy wyestymowany model poddałem po raz ostatni badaniu spełnienia założeń KMRL. Jako pierwsza została sprawdzona poprawność formy funkcyjnej. Test RESET dla wartości dopasowanych wyniósł 1,625 przy $p\text{-value}$ równym 0,2005. Natomiast ten sam ten test przeprowadzony dla potęg zmiennych niezależnych dał odmienne wyniki, w którym $p\text{-value}$ wyniosło 0,0020. W związku z tym posiadając dwa różne wyniki nie jestem w stanie stwierdzić czy w ogólnym rozrachunku mój model jest liniowy. Powinienem więc odrzucić prawidłowość formy funkcyjnej, co może skutkować obciążeniem estymatorów, przez co uzyskane parametry nie będą przybliżać rzeczywistych wartości.

W przypadku analizy drugiego założenia tj. homoskedastyczności reszt statystyka testowa jest równa $BP=6,881$, co przy $p\text{-value}=0,1423$ nie daje podstaw do odrzucenia hipotezy zerowej. Przeprowadziłem również test White’a, ponieważ w moim modelu znalazły się nieliniowe zmienne. Jest to bardziej restrykcyjna postać modelu Breusch’a-Peagan’a, jednakże hipoteza zerowa brzmi dokładnie tak samo tj. występowanie w modelu homoskedastyczności. $P\text{-value}$ wynosi 0,7916 w związku z tym teoretycznie zarówno statystyki t i F oraz ich $p\text{-value}$ ’a nie powinny być przeszacowane. Kolejnym ważnym warunkiem KMRL jest normalność rozkładu reszt. W moim modelu udało się spełnić te założenia, ponieważ $X\text{-Squared}$ wyniosło 3,1783 przy $p\text{-value} = 0,2041$, tak więc nie mogłem odrzucić hipotezy zerowej zakładającej normalność składnika losowego. Ostatnim testem jaki wykonałem, był test Durbina-Watsona badający czwarte założenie o braku autokorelacji reszt, jednakże i tym razem musiałem odrzucić hipotezę zerową na korzyść hipotezy alternatywnej. Przyczynami, z powodu których występuje ten problem mogą być pominięte istotne zmienne w modelu lub charakter przeprowadzonych przekształceń.

5. WERYFIKACJA HIPOTEZ

Tabela 1. Porównanie modeli 1-5.

	<i>Dependent variable:</i>				
	value		log(value)		
	(1)	(2)	(3)	(4)	(5)
age	-4,889,400.000*** (539,940.100)	2,097,673.000 (5,724,949.000)	-0.006 (0.092)	-0.035 (0.077)	
age2		-138,931.900 (113,330.100)	-0.001 (0.002)	-0.001 (0.002)	-0.002*** (0.0002)
rating_fifa21	5,938,483.000*** (453,694.400)	5,749,041.000*** (478,492.400)	0.101*** (0.008)	0.094*** (0.008)	0.093*** (0.007)
height	-248,818.900 (222,973.400)	-260,507.400 (222,763.400)	-0.003 (0.004)	-0.001 (0.004)	
contract_left	-110,066.500 (1,187,605.000)	-119,675.500 (1,185,426.000)	0.019 (0.019)	0.017 (0.018)	
instagram_followers	0.097* (0.056)	0.140** (0.066)	0.000 (0.000)		
log1p(instagram_followers)				0.036*** (0.010)	0.038*** (0.009)
europe	-1,290,512.000 (5,843,728.000)	-1,860,646.000 (5,851,386.000)	0.005 (0.094)	0.023 (0.090)	
south_america	-2,992,104.000 (5,718,884.000)	-3,757,588.000 (5,742,314.000)	0.009 (0.092)	0.004 (0.088)	
twolegs	-2,302,654.000 (5,483,918.000)	-2,287,914.000 (5,473,747.000)	0.022 (0.088)	0.022 (0.084)	
leftleg	-1,312,952.000 (2,998,996.000)	-1,149,520.000 (2,996,393.000)	-0.016 (0.048)	-0.016 (0.046)	
striker	-1,137,225.000 (5,959,318.000)	-1,027,360.000 (5,948,926.000)	-0.029 (0.096)	-0.010 (0.093)	
midfielder	-3,843,170.000 (3,847,839.000)	-3,585,305.000 (3,846,449.000)	-0.065 (0.062)	-0.049 (0.060)	
defender	-6,097,008.000 (7,357,330.000)	-4,672,143.000 (7,435,077.000)	-0.105 (0.120)	-0.122 (0.114)	
goals	1,498,429.000***	1,610,388.000***	0.017* (0.017)	0.015* (0.015)	0.014** (0.014)

	(535,065.100)	(541,823.900)	(0.009)	(0.008)	(0.006)
europe_striker	-2,979,413.000	-2,957,923.000	-0.043	-0.066	
	(5,869,592.000)	(5,858,717.000)	(0.094)	(0.091)	
Constant	- 269,273,123.000*	- 337,927,781.000*	10.963**	10.876**	10.414**
	**	**	*	*	*
	(54,279,716.000)	(77,921,252.000)	(1.254)	(1.168)	(0.457)
Observations	150	150	150	150	150
R ²	0.661	0.665	0.681	0.703	0.695
Adjusted R ²	0.626	0.627	0.645	0.669	0.687
Residual Std. Error	15,349,202.000	15,320,695.000	0.247	0.238	0.232
	(df = 135)	(df = 134)	(df = 134)	(df = 134)	(df = 145)
F Statistic	18.811*** (df = 14; 135)	17.723*** (df = 15; 134)	19.085** (df = 15; 134)	21.107** (df = 15; 134)	82.583** (df = 4; 145)
Note:	* p < 0.05, ** p < 0.01, *** p < 0.001				

Współczynnik R², który uzyskałem w modelu końcowym wyjaśnia 69,5% zmienności całkowitej, a model jest oczywiście statystycznie istotny. W ostatecznym modelu wzrósł również wskaźnik dopasowanego R², dzięki czemu pozwala to stwierdzić uzyskanie najlepszego możliwego modelu, w stosunku do wszystkich przeprowadzonych iteracji.. Zmienna zależna w modelu końcowym pozostała zlogarytmowana, tak więc wszystkie zmienne z pominięciem obserwujący (log1p(instagram_followers)) powinienem interpretować zgodnie z zagadnieniem semielastyczności, a przy otrzymanych oszacowaniach parametrów należy stosować wzór :

$$(e^{\beta} - 1) * 100\%.$$

Udało się mi się potwierdzić moją hipotezę główną, iż podstawowe charakterystyki piłkarza wpływają na jego wartość rynkową, chociaż zmienna określająca wzrost uległa zmianie postaci. Wzrost zmiennej age² przy średniej wieku badanego piłkarza równej 25 lat, powoduje spadek jego wartości rynkowej o około -9,5%. Mając na uwadze, że jest to wartość paraboliczna, istotność tej zmiennej potwierdza wnioski wysunięte w różnej literaturze (Ante, L. 2019). Paraboliczna zależność ceny piłkarza, może wyjaśniać fakt, iż do pewnego wieku piłkarz tak naprawdę stale się rozwija, a w szczególności dotyczy to młodych zawodników, którzy dopiero zaczynają stanowić o sile zespołu. Ma to swoje odzwierciedlenie także w rzeczywistości, ponieważ kluby coraz chętniej stawiają na młodych zawodników tuż po ukończonych

akademiach piłkarskich wiedząc, że to oni mogą w przyszłości stanowić o sile klubu, wpływać znacznie na sukcesy drużyny oraz generować klubowe zyski. Zdecydowanie, kluby nie boją się wydawać dużych pieniędzy na pozyskanie młodego zawodnika. Świadczy o tym m.in. przykład pozyskania dwa lata temu, zaledwie 19-letniego piłkarza przez klub Atletico Madryt za kwotę 127,20 mln €. Jednakże, w pewnym momencie kariery każdy piłkarz dochodzi do punktu, w którym wiek zaczyna ograniczać jego szybkość oraz dynamikę. W związku z tym szansa na to, że starzejącemu się piłkarzowi uda się wrócić do szczytu formy sprzed kilku lat, jest niewielka. Staje się on mniej atrakcyjny dla klubów, a co za tym idzie, jego cena rynkowa zaczyna spadać.

Ciekawym wnioskiem, może być ocena umiejętności piłkarza używając indexu z popularnej gry FIFA21. Wraz ze wzrostem ratingu o 1, wartość rynkowa piłkarza rośnie o najwięcej spośród wszystkich zmiennych tzn. o 9,74%. Warto mieć na uwadze, że bardzo ciężko ocenić indywidualne umiejętności piłkarza na danej pozycji, ponieważ to co wyróżnia danego piłkarza nie musi być zawsze tożsame. Dla przykładu, to co będzie cechować statystycznie dobrego bramkarza to przykładowo liczba obronionych strzałów lub straconych goli. Ciężko jest natomiast określić jego umiejętność pracy nóg lub predykcję toru lotu piłki. Statycznego dobrego odbiorcę może cechować liczba udanych wślizgów lub wygranych pojedynków jeden na jeden, natomiast napastnika liczba strzelonych goli lub asyst. Jednakże, niemożliwe jest oszacowanie liczbowe tego jak dany zawodnik w polu przyswaja taktykę trenera lub jaka jest jego skala wytrzymałości. Dzięki sztucznie stworzonemu rankingowi przez twórców gry komputerowej, pomimo że w rzeczywistości nie jest to zmienna uwzględniana w szacowaniu wartości rynkowej piłkarza, może być bardzo pomocna. W literaturze bardzo często do oszacowania umiejętności piłkarza stosowano index stworzony przez analityków sportowych strony squawka.com (Hendriks, 2017). Niemniej jednak należy pamiętać, że takie rankingi mogą bazować nie tylko na predykcji, ale i własnych upodobaniach autora. Możliwość odzwierciedlenia umiejętności danego piłkarza wartościami liczbowymi jest na pewno bardzo interesująca, jednakże do tych indeksów należy podchodzić z lekkim dystansem ze względu na niemierzalność niektórych umiejętności piłkarskich.

Wzrost liczby obserwujących na Instagramie o 1% powoduje wzrost wartości rynkowej piłkarza tylko o 0.038%, co należy uznać za nieoczekiwany wynik. Postać zmiennej określającej popularność została zmieniona, jednakże uznano, iż hipoteza o wpływie popularności na cenę piłkarza została potwierdzona. Potwierdza to również wnioski opisane w literaturze (Garcia del Barrio & Pujol, 2007). W rzeczywistość patrząc na takie nazwiska jak Leo Messi czy Cristiano Ronaldo, którzy mają najwięcej obserwujących na całym świecie,

pomimo ich wysokiego wieku nadal są wartościowymi piłkarzami na rynku transferowym. Oczywiście na te wartości wpływają także ich wysokie umiejętności, jednakże dla każdego klubu stanowiłoby oni wartość dodaną przyciągając nowych sponsorów. Co więcej to oni są głównymi powodami, dla których kibice chcą oglądać na boisku Barcelonę lub Juventus, a co za tym idzie to oni generują dla klubów najwyższe zyski poprzez sprzedaż biletów na mecze, praw telewizyjnych lub gadżetów klubowych. Faktem jest, że bycie bardzo droгим piłkarzem wiąże się ze statusem bycia wręcz celebryty, ponieważ są to osoby do których fani chcą się upodabniać.

Ostatnią hipotezą poboczną, którą udało mi się udowodnić jest wpływ podstawowych statystyk boiskowych na cenę piłkarza. Strzelenie gola w ciągu sezonu wiąże się ze wzrostem wartości rynkowej o 1,4%, co ma swoje odwzorowanie w rzeczywistości. Ten wynik oczywiście również znalazł swoje odzwierciedlenie w literaturze (Carmichael and Thomas 1993), jednakże istnieją też badania, które są w opozycji do tego twierdzenia (Ruijg, J., & van Ophem, H. 2015). Dodatkowo, analizując dane, w każdej z pięciu topowych lig miano najbardziej wartościowego piłkarza należy do napastników, którzy są „bramkostrzelni”. Niewątpliwie gole są kwintesencją piłki nożnej, a zarazem to one decydują która drużyna w ostatecznym rozrachunku wygrywa mecz. Tak jak w przypadku popularności również i gole są główną z rzeczy, która przyciąga ludzi na stadiony i generuje zyski dla klubów.

ZAKOŃCZENIE

W świetle założeń KMRL, otrzymując mój model ostateczny $\log(value_i) = \beta_0 + \beta_1 age_i^2 + \beta_2 rating_fifa21_i + \beta_3 \log(instagram_followers_i) + \beta_4 goals_i + \epsilon_i$ nie mogę jednoznacznie stwierdzić, czy otrzymałem prawidłową formę funkcyjną ze względu na zróżnicowanie wyników po wykonaniu testu RESET. Wartym zaznaczenia jest fakt, że przed przeprowadzeniem procesu General to specific wyniki w obu przypadkach nie odrzucały hipotezy zerowej. Ze względu na problem z określeniem poprawności pierwszego założenia KMRL konsekwencjami może być nieuznanie zasadności wniosków pochodzących z modelu lub uniemożliwienie udowodnienia własności estymatora MNK.

Udało się natomiast uzyskać homoskedastyczność w modelu bez tworzenia macierzy odpornej White’a oraz nie rozważano o tym czy próbka jest wystarczająco duża, ponieważ w końcowej wersji modelu uzyskano normalność reszt. Dzięki spełnieniu tych dwóch założeń, teoretycznie zarówno statystyki T jak i p-value zmiennych, powinny być poprawne. Model jest

oparty na danych przekrojowych w związku z czym posiadanie autokorelacji reszt nie jest istotnym problemem.

Stwierdzam, że udało mi się spełnić postawione hipotezy, a zmiennymi które okazały się być istotne są wiek², zlogarytmowany wektor określający liczbę obserwujących na Instagramie, rating fifa21 oraz liczba goli. Należy zaznaczyć, że w związku z niespełnieniem wszystkich założeń w modelu, wyniki powinny być interpretowane z przymrużeniem oka.

Podjmując się przyszłych prac można podjąć próbę zwiększenia próbki zarówno o liczbę badanych piłkarzy jak i lig. Ligi takie jak portugalska lub holenderska przyciągają do siebie coraz więcej kluczowych piłkarzy, a klubowi wychowankowie coraz bardziej zaczynają być drogimi, a zarazem atrakcyjnymi zawodnikami na rynku transferowym. Warto również uwzględnić zmienne określające prestiż klubu lub liczbę rozegranych meczów w wyjściowej jedenastce, które w tym modelu nie zostały zbadane.

BIBLIOGRAFIA

- Ant, L. 2019. Determinants of Transfer Fees: Evidence from the Five Major European Football Leagues. University of Hamburg.
- 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.
- Carmichael, F. & Thomas, D. 1993. Bargaining in the transfer market: Theory and evidence. *Applied Economics*, 25 (12), 1467-1476.
- Carmichael, F., Forrest, D. & Simmons, R. 1999. The labour market in association football: Who gets transferred and for how much? *Bulletin of Economic Research*, 51 (2), 125-150.
- Dobson, S., & Gerrard, B. 1999. The determination of player transfer fees in English professional soccer. *Journal of Sport Management*, 13 (4), 259-279.
- Dobson, S., Gerrard, B., & Howe, S. 2000. The determination of transfer fees in English nonleague football. *Applied Economics*, 32 (9), 1145-1152.
- Franck, E., & Nuesch S. 2012. Talent and/or popularity: What does it take to be a superstar? *Economic Inquiry*, 50 (1), 202–216.
- Frick, B. 2007. The football players' labour market: Empirical evidence from the major European leagues. *Scottish Journal of Political Economy*, 54 (3), 422-446.
- Garcia-del-Barrio, P., & Pujol, F. 2004. Pay and performance in the Spanish soccer league: Who gets the expected monopsony rents? Working Paper 05/04.
- Lehmann, E., & Schulze, G. 2008. What does it take to be a star? - The role of performance and the media for German soccer players. *Applied Economics Quarterly*, 54 (1), 59-70.
- Majewski, S. 2014. Szacowanie wartości rynkowe piłkarskich kart zawodniczych przy wykorzystaniu modeli ekonometrycznych. *Zeszyty Naukowe Uniwersytetu Szczecińskiego* nr 803. 663-673
- Ruijg, J., & van Ophem, H. 2015. Determinants of football transfers. *Applied Economics Letters*, 22(1), 12-19.
- Wiśniewski T, Bernatek A. 2020. „Raport KPMG-Player value not immune to pandemic”.<https://home.kpmg/pl/pl/home/insights/2020/05/raport-kpmg-pt-player-value-not-immune-to-pandemic.html>

- Yashiv E. 2017. The value of top footballers, bubbles, and pitfalls of the free market.<https://voxeu.org/article/neymar-bubble>
- Instagram, <https://www.instagram.com/>, dostęp 28 grudnia 2020.
- EA Games, <https://www.ea.com/pl-pl/games/fifa/fifa-21/ratings>, dostęp 28 grudnia 2020.
- Transfermarkt, <https://www.transfermarkt.pl/>, dostęp 28 grudnia 2020.