

Gutachterin: Dr. Astrid Cullmann

Fakultät VII – Wirtschaft und Management

Institut für Volkswirtschaftslehre und Wirtschaftsrecht

Fachgebiet Ökonometrie und Wirtschaftsstatistik

**Seminararbeit: Forschungsseminar Ökonometrie und
Wirtschaftsstatistik WS 20/21**

**Forschungsthema: Marktwertschätzung
von Fußballspielern anhand von Daten
des Videospiels FIFA20**



Von: Jakob Quellhorst (Matrikelnummer: 394522)

Jakub Manczak (Matrikelnummer: 388286)

Ruben Pfundner (Matrikelnummer: 459504)

Berlin 15.03.2021

Eidesstaatliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und eigenhändig sowie ohne unerlaubte fremde Hilfe und ausschließlich unter Verwendung der aufgeführten Quellen und Hilfsmittel angefertigt habe.

Berlin, den 14.03.2021, Markus Jelab Unterschrift
Berlin, den 14.3.21 Ruben Pfandrem Unterschrift
Berlin, den 15.03.2021 J. Quellhorst Unterschrift

Inhaltsverzeichnis

| | |
|---|-----|
| Eidesstaatliche Erklärung | II |
| Inhaltsverzeichnis | III |
| Abbildungsverzeichnis | V |
| 1 Einleitung | 1 |
| 1.1 Marktwerte im Fußball | 1 |
| 1.1.1 Marktwert | 1 |
| 1.1.2 Transfermarkt.de | 1 |
| 1.1.3 Große Unterschiede zwischen kleinen und großen Vereinen | 1 |
| 1.2 Datenquelle | 2 |
| 1.3 Nicht berücksichtigte Einflussgrößen in unserem Modell | 3 |
| 1.4 Wichtige Literatur | 4 |
| 2 Datensatz und Variablen | 5 |
| 2.1 Datensatz | 5 |
| 2.1.1 Datensatzbereinigung | 5 |
| 2.1.2 Erhebungszeitraum | 5 |
| 2.2 Variablen | 5 |
| 2.2.1 Erklärende Variablen | 5 |
| 2.2.2 Allgemeinattribute | 5 |
| 2.2.3 Fähigkeitsattribute | 6 |
| 2.2.4 Overall-rating | 6 |
| 2.2.5 Ober- und Unterkategorien | 6 |
| 2.2.6 Weitere Fähigkeitsvariablen | 7 |
| 2.3 Deskriptive Analyse - nicht berücksichtigte Variablen | 7 |
| 2.3.1 Fähigkeitsattribute | 8 |
| 2.3.2 Allgemeinattribute | 10 |
| 2.3.3 Nationalität | 10 |
| 2.3.4 Alter | 12 |
| 2.4. Beobachtungseinheit Stürmer und Verteidiger | 13 |
| 3 Ökonometrisches Modell | 15 |
| 3.1 Kleinstes-Quadrat-Methode | 15 |

| | |
|---|----|
| 3.2 Annahmen für multiple lineare Regression..... | 15 |
| Exkurs Log-Level-Transformation..... | 17 |
| 3.3 Interpretation von Log-Level Ergebnissen..... | 20 |
| 3.4 Re-Transformation der geschätzten logarithmierten Werte | 21 |
| 3.5 Güte des Schätzers | 22 |
| 4 Ergebnisse | 23 |
| 4.1 Zwischenstand | 23 |
| 4.2 Regressionen | 23 |
| 4.3 Z-Standardisierung | 25 |
| 5 Fazit | 27 |
| 5.1 Zusammenfassung | 27 |
| 5.2 Reale Spielertransfers in der Saison 2019/2020..... | 27 |
| 5.2.1 Vermarktungswert..... | 29 |
| 5.2.2 Liquidität des kaufenden Vereins..... | 29 |
| 5.2.3 Vertragsrestlaufzeit und Verletzungshistorie..... | 30 |
| 5.2.4 Angebot und Nachfrage | 31 |
| 5.3 Abschluss | 31 |
| Literatur- und Quellenverzeichnis..... | V |
| Bücher | V |
| Elektronische Quellen (Internet) | V |

Abbildungsverzeichnis

| | |
|---|----|
| Abbildung 1: Ausschnitt aus dem FIFA17 Spielerprofil von Christiano Ronaldo, fifauteam.com, Zugriff v. 15.02.2021 | 3 |
| Abbildung 2: Ausgaben für internationale Spielertransfers, statista.com, Zugriff v. 15.02.2021 | 4 |
| Abbildung 3: Zusammensetzung Ober- und Unterkategorien, /fifauteam.com, Zugriff v. 27.11.2020..... | 6 |
| Abbildung 4: Zusammensetzung der Variable „pace“ aus den beiden Variablen „acceleration“ und „sprint speed“, eigene Darstellung. | 7 |
| Abbildung 5: Zusammenhang zwischen „shooting“ und Marktwert eines Spielers. Rote Gerade entspricht der linearen Regressionsfunktion. Eigene Darstellung. | 8 |
| Abbildung 6: Zusammenhang zwischen „Beidfüßigkeit“ und Marktwert. Eigene Darstellung. | 9 |
| Abbildung 7: Zusammenhang zwischen „Trickfähigkeit“ und Marktwert. Eigene Darstellung. | 9 |
| Abbildung 8: Einfluss der Variable „Overall“ auf den Marktwert, Betrachtung von Spielern einer Nationalität (Deutschland, Frankreich und Spanien). Eigene Darstellung. | 10 |
| Abbildung 9: Einfluss der Variable „Overall“ auf den Marktwert, Betrachtung von Spielern einer Nationalität (England und Brasilien). Eigene Darstellung. | 11 |
| Abbildung 10: Zusammenhang zwischen „Overall“ und Marktwert, Betrachtung von Spielern einer Nationalität. Eigene Darstellung. | 12 |
| Abbildung 11: Zusammenhang zwischen Alter und Marktwert. Eigene Darstellung. | 13 |
| Abbildung 12: Betrachtung von ausschließlich Stürmern und Verteidigern | 13 |
| Abbildung 13: Zusammenhang zwischen „Overall“ und Marktwert ist offenbar nicht linear. Eigene Darstellung. | 16 |
| Abbildung 14: Zusammenhang zwischen „Overall“ und Marktwert hat die Form einer Exponentialfunktion. Eigene Darstellung. | 16 |
| Abbildung 15: Beispiel für Anpassung der Variable Marktwert durch Logarithmierung. Eigene Darstellung. | 17 |
| Abbildung 16: Zusammenhang „Overall“ und logarithmierter Marktwert. Eigene Darstellung. | 17 |

| | |
|--|----|
| Abbildung 17: Verteilung der Residuen für den Zusammenhang zwischen „Overall“ und „Log(value)“. Eigene Darstellung..... | 18 |
| Abbildung 18: Plot mit geschätzten logarithmierten Marktwerten und Residuen. Eigene Darstellung..... | 19 |
| Abbildung 19: vif-Koeffizienten. Eigene Darstellung..... | 20 |
| Abbildung 20: Regression I: Einfluss der 6 Fähigkeitsvariablen auf den logarithmierten Marktwert (links Stürmer, rechts Verteidiger). Eigene Darstellung. | 23 |
| Abbildung 21: Regression II: Einfluss von Fähigkeitsvariablen und Reputation, Potential und Alter auf den logarithmierten Marktwert. Eigene Darstellung. | 24 |
| Abbildung 22: Gegenüberstellung von unseren geschätzten Marktwerten und den beobachteten Marktwerten. Eigene Darstellung. | 25 |
| Abbildung 23 z-standardisierter Output der Regression II | 26 |
| Abbildung 24: Vergleich unserer geschätzten Marktwerte (Regression II) mit real gezahlten Transfersummen (für 47 Spieler). Eigene Darstellung. | 28 |
| Abbildung 25: basierend auf Abbildung 24 – Transfers in die englische Premier League hervorgehoben. Eigene Darstellung. | 30 |

1 Einleitung

In unserer Forschungsarbeit möchten wir Marktwerte von Fußballspielern anhand der von dem Videospiel FIFA vergebenen Fähigkeits- und Allgemeinvariablen schätzen. Unser Ziel ist es, nachvollziehbar darzustellen, unter welchen Einflussfaktoren sich Marktwerte zusammensetzen und diese Faktoren in ein plausibles Modell zu vereinigen. Darüber hinaus gehen wir auf weitere mögliche Einflussgrößen ein, die nicht in unserem Modell Platz finden. Um zu überprüfen wie valide unsere Marktwertschätzungen sind, vergleichen wir unsere geschätzten Marktwerte mit realen Verkaufspreisen aus der Sommertransferperiode 2019.

1.1 Marktwerte im Fußball

1.1.1 Marktwert

Der Marktwert eines Gutes wird allgemein definiert als dessen Preis zum Zeitpunkt einer Transaktion. Bezogen auf das Fußballgeschäft, konkretisieren sich Marktwerte damit in den Transfersummen, die Vereine für den Wechsel eines Spielers zu ihrem Klub bezahlen.¹

1.1.2 Transfermarkt.de

Auch ohne einen Transfer, schätzen Experten wie die renommierte Netzplattform „transfermarkt.de“ die potenziellen Markt- bzw. Transferwerte einzelner Spieler.² In dessen Forum mit circa 400.000 Mitgliedern werden Marktwerte für einzelne Fußballer diskutiert und vorgeschlagen. User können hier dauerhaft Vorschläge für Marktwerte von Fußballern unterbreiten, müssen diese aber unter Berücksichtigung verschiedenster Einflussfaktoren begründen, bspw. auf aktuellen Leistungsdaten, dem Alter, der Vertragslaufzeit und der Verletzungshistorie des jeweiligen Spielers.³ Die hier entstandenen Marktwerte werden durch das Transfermarkt.de-Team geprüft und freigegeben.⁴

1.1.3 Große Unterschiede zwischen kleinen und großen Vereinen

Fakt ist, die zurzeit aufgerufenen Transfersummen für Fußballspieler nehmen astronomische Ausmaße an und die Unterschiede zwischen Top-Spielern und

¹ Siehe Gerhards et al., 2014, S. 234.

² Siehe Gerhards et al., 2014, S. 240.

³ Siehe Ackermann, P. und Follert, F., 2018, S. 4.

⁴ Siehe Fehr, R., 2018, watson.ch/sport, Zugriff v. 12.02.2021.

durchschnittlichen Spieler sind hierbei enorm. Ein einfaches Beispiel für jene Absurditäten liefert die Bundesliga: Während der Marktwert des gesamten 28-köpfigen Kaders des Erstligisten 1.FC Union Berlin mit 76,2 Mio. Euro angegeben wird, befindet sich der Marktwert des Bayern München Top-Stars Joshua Kimmich alleine schon bei 90 Mio. Euro. Damit ist Kimmich (12 Tor & Assists in dieser Saison) das 1,18fache des gesamten Profikaders von Union und gleichzeitig das 15fache des Topspielers Max Kruse (Marktwert: 6 Mio. Euro; 13 Scorerpunkte) wert.⁵ So kommen wir zu unserer **ersten Hypothese**: Der Marktwert nimmt mit steigender Spielfähigkeit exponentiell zu.

Natürlich sind Scorerpunkte (Tore & Vorlagen) nicht für jeden Spieler ein Maßstab für Güte – ein Verteidiger z.B. wird womöglich eher an seinen gewonnenen Zweikämpfen gemessen. Daraus folgt unsere **zweite Hypothese**: Die Anforderungen an einen Spieler hängen von seiner Position ab.

1.2 Datenquelle

Intuitiv sollten die Marktwerte der Spieler abhängig von ihren fußballerischen Fähigkeiten sein. Genau eine solche Bewertung der Spielerfähigkeiten nimmt das jährlich erscheinende Videospiel „FIFA“ des US-amerikanischen Herstellers „Electronic Arts, Inc.“ (EA) für alle Profi-Fußballspieler der internationalen Topligen vor. Für die Bewertung sind 25 interne EA-Produzenten verantwortlich, welche Unterstützung von über 6000 externen Datenprüfern und Talentscouts erhalten. Diese ermitteln und erfassen reale Spieldaten, wie z.B. geschossene Tore, Vorlagen oder gelaufene Kilometer und übermitteln diese an EA. Auf dieser Basis entsteht für jeden Spieler ein fein ausgearbeitetes Datenprofil.⁶

Dieses Datenprofil umfasst 50 Merkmale eines Spielers, wie z.B. Informationen zu der Schnelligkeit (pace) oder der Schussfähigkeit (shooting), aber auch Bewertungen hinsichtlich des Talents (potential) oder der internationalen Reputation (international reputation) eines Spielers. Gleichzeitig führt das Videospiel Marktwerte für jeden Spieler auf, welche mit den bereits genannten Marktwerten auf transfermarkt.de etc. übereinstimmen.

⁵ Siehe Transfermarkt, 2021, transfermarkt.de, Zugriff v. 01.03.2021.

⁶ Siehe Murphy, R., 2019, goal.com, Zugriff v. 05.03.2021.

Mit Hilfe der beschriebenen FIFA-Daten haben wir versucht Marktwerte zu schätzen.



Abbildung 1: Ausschnitt aus dem FIFA17 Spielerprofil von Cristiano Ronaldo, fifateam.com, Zugriff v. 15.02.2021

Wir gehen jedoch davon aus, dass nicht alle Variablen in unserem Datensatz tatsächlich Einfluss auf den Marktwert nehmen (**Hypothese III**)

1.3 Nicht berücksichtigte Einflussgrößen in unserem Modell

Uns ist bewusst, dass in unserem Datensatz nicht alle Einflussfaktoren für eine optimale Marktwertbestimmung gegeben sind. Wir können keine marketingtechnischen Aspekte wie potenzielle Trikotverkäufe oder den Mehrerlös von Tickets in unsere Analyse einbringen. So seien nach seinem Wechsel zu Juventus Turin im Sommer 2018 allein in der ersten 24 Stunden bis zu 520.000 Cristiano Ronaldo Trikots verkauft worden.⁷ Auch klassische Marktmechanismen um Angebot und Nachfrage oder (fehlende) Liquidität können wir nur schwer in unser Modell einbeziehen: So hatte der, durch den katarischen Geschäftsmann Nasser Al-Khelaifi finanzierte, 220-Millionen-Euro-Wechsel Neymars (damaliger Marktwert bei 100 Mio. Euro)⁸ zu Paris Saint German im Sommer 2017 offenbar Auswirkungen auf den gesamten Markt: Die weltweiten jährlichen Transferausgaben stiegen im Vergleich zum Vorjahr um knapp 1,57 Mrd. \$, das entspricht einer Steigerung um 33,26 %.⁹ Zwar stiegen die Transfersummen in den vergangenen Jahren auch unabhängig solch fremdfinanzierter Wechsel – der Anstieg im Jahr 2017 hebt sich trotzdem enorm von den Vorjahren ab. Das zusätzliche Geld, welches durch den externen Geldgeber quasi „künstlich“ dem Markt zugeführt wurde,

⁷ Siehe Manager Magazin, 2018, manager-magazin.de, Zugriff v. 15.02.2021.

⁸ Siehe Transfermarkt, 2017, transfermarkt.de, Zugriff v. 01.03.2021.

⁹ Siehe Zeppenfeld, B., 2021, statista.de, Zugriff v. 15.02.2021.

verursachte eine Art Kettenreaktion: Barcelona, das die 220 Mio. Euro für Neymar erhalten hatte, zahlte in der Folge ebenfalls überhöhte Preise – der ganze Markt wurde mutmaßlich aufgeschwemmt.¹⁰

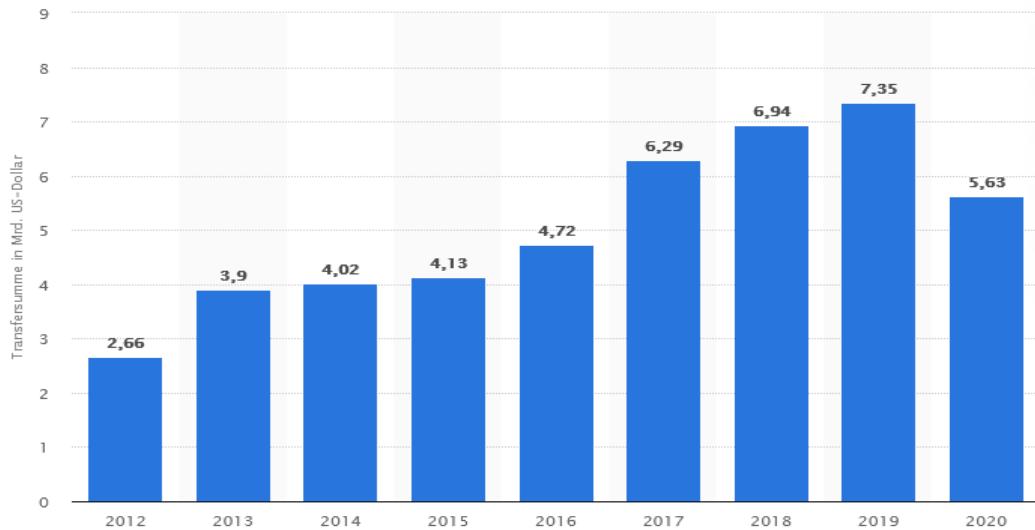


Abbildung 2: Ausgaben für internationale Spielertransfers, statista.com, Zugriff v. 15.02.2021

Die o.g. Einflussfaktoren, wie die Spielervermarktung oder die Liquidität eines Vereins, werden in dem von FIFA bereitgestellten Datensatz nicht berücksichtigt. Uns ist deshalb klar, dass der von uns geschätzte Marktwert nie einer perfekten Voraussage für den realen Verkaufspreis eines Spielers entsprechen kann, da wir nicht alle mutmaßlich relevanten Variablen in unser Modell einfließen lassen können. Unsere **vierte Hypothese** lautet deshalb: Der Marktwert eines Spielers lässt sich nicht ausschließlich anhand seiner sichtbaren/messbaren Merkmale schätzen.

1.4 Wichtige Literatur

Für unsere Schätzung verwenden wir ein multiples lineares Regressionsmodell. Die theoretische Grundlage hierfür liefert uns insbesondere das Lehrbuch „Ökonometrie – Eine Einführung“ von Ludwig von Auer.¹¹ Ebenso wichtig und hilfreich ist uns das Werk „Introductory Econometrics“ von Jeffrey Wooldridge.¹²

¹⁰ Siehe Rummel, K.H., 2017, steemit.com/sports, Zugriff v. 10.03.2021.

¹¹ Siehe Auer, L., 2016.

¹² Siehe Wooldridge, J., 2012.

2 Datensatz und Variablen

2.1 Datensatz

Der Datensatz wurde uns in Form einer Excel-Tabelle auf der Plattform „kaggle.com“ bereitgestellt. Die Daten basieren ursprünglich auf der Datenbank des Web-Portals „sofifa.com“ – eine unabhängige Plattform, auf der alle FIFA-Spieldaten gesammelt und veröffentlicht werden. Der Datensatz umfasst 18278 Spieler aus 30 internationalen Ligen.

2.1.1 Datensatzbereinigung

Insgesamt mussten wir 250 Spieler aus unserem Datensatz streichen, da diese entweder doppelt aufraten oder keine Angabe zum Marktwert des Spielers getroffen wurden.

2.1.2 Erhebungszeitraum

Unser Datensatz repräsentiert die Spielerbewertungen, die EA, zur Veröffentlichung ihres Videospiels FIFA 2020 im September 2019, vergeben hat. In die Bewertung gehen somit insbesondere die Spielerleistungen aus der Vorsaison 2018/2019 ein. Dies ist besonders für unsere abschließende Betrachtung der real gezahlten Transfersummen in Kapitel 5.2 von Bedeutung. Anders als womöglich intuitiv angenommen, betrachten wir hierbei nämlich die Spielertransfers aus dem Sommer 2019 bzw. Winter 2019/20.

2.2 Variablen

2.2.1 Erklärende Variablen

Insgesamt befinden sich in unserem Datensatz über 70 verschiedene Variablen, die verschiedene Merkmale der Fußballspieler beschreiben. Zunächst können wir die Variablen in zwei große Kategorien unterteilen, in Variablen, die die fußballerischen Fähigkeiten eines Spielers beschreiben und in solche Variablen, die allgemeine Eigenschaften eines Spielers beschreiben.

2.2.2 Allgemeinattribute

Allgemeine Eigenschaften sind z.B. das Potential, die Nationalität oder die internationale Reputation eines Spielers. Das Potential („delta_potential“) eines Spielers gibt an wie groß der (geschätzte) Entwicklungsspielraum hinsichtlich seiner Fähigkeitsvariablen ist. Die internationale Reputation („international_reputation“)

stellt einen Gradmesser für die Berühmtheit eines Spielers dar. An dieser Variable kann womöglich noch am ehesten der Marketingwert eines Spielers abgelesen werden.

2.2.3 Fähigkeitsattribute

Fähigkeitsvariablen sind beispielsweise die Laufgeschwindigkeit, die Schussstärke oder Passfähigkeit eines Spielers. Hier vergibt FIFA jedem Spieler eine Bewertung zwischen 0 und 100 Punkten. Insgesamt gibt es 42 Fähigkeitsattribute, hiervon sind 6 den Ober- und 34 den Unterkategorien zuzuordnen (siehe Kapitel 2.2.5) – zusätzlich gibt es noch 2 weitere (Spezial-)Variablen (siehe Kapitel 2.2.6).

2.2.4 Overall-rating

Für jeden Spieler wird, zusammengesetzt aus seinen einzelnen Fähigkeitsvariablen, eine Gesamtbewertung (Overall [1-100]) berechnet. Diese Berechnung folgt einer Vorschrift, die FIFA bereits veröffentlicht hat. Somit wird der Overall einheitlich berechnet und ist damit vergleichbar. Er dient uns in der Folge häufig als erster Indikator, wie gut ein Spieler ist.

2.2.5 Ober- und Unterkategorien

Die 6 übergeordneten Variablen sind Schnelligkeit (pace), Dribbelstärke (dribbling), Schussstärke (shooting), Verteidigen (defending), Passen (passing) und Physische Stärke (physical), wobei diese sich wiederum aus mehreren Unterkategorien

| | |
|-----------------------|---------------------|
| PACE | Dribbling |
| ACCELERATION 45 | AGILITY 10 |
| SPRINT SPEED 55 | BALANCE 05 |
| SHOOTING | BALL CONTROL 25 |
| FINISHING 45 | COMPOSURE 00 |
| LONG SHOTS 20 | Dribbling 50 |
| PENALTIES 05 | REACTIONS 00 |
| POSITIONING 05 | |
| SHOT POWER 20 | |
| VOLLEYS 65 | |
| PASSING | DEFENDING |
| CROSSING 20 | HEADING ACCURACY 10 |
| CURVE 05 | INTERCEPTIONS 20 |
| FREE KICK ACCURACY 05 | MARKING 30 |
| LONG PASSING 15 | SLIDING TACKLE 10 |
| SHORT PASSING 35 | STANDING TACKLE 30 |
| VISION 25 | |
| PHYSICAL | |
| | AGGRESSION 20 |
| | JUMPING 05 |
| | STAMINA 25 |

Abbildung 3: Zusammensetzung Ober- und Unterkategorien, fifauteam.com, Zugriff v. 27.11.2020.

zusammensetzen. Auch diese Zusammensetzung folgt einer allgemeinen Berechnungsvorschrift, die von FIFA veröffentlicht wurde.

So setzt sich die Oberkategorie „pace“ zu 45% aus der Unterkategorie „acceleration“ und zu 55% aus „sprint speed“ zusammen. Wir haben diesen Zusammenhang nochmal in R überprüft und konnten die von FIFA veröffentlichten Werte bestätigen.

```

call:
lm(formula = pace ~ movement_acceleration + movement_sprint_speed,
  data = data_ST)

Residuals:
    Min      1Q  Median      3Q     Max 
-0.5173 -0.2324 -0.0160  0.2440  0.4971 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept) 0.092691  0.044272  2.094   0.0364 *  
movement_acceleration 0.448594  0.001266 354.299 <2e-16 *** 
movement_sprint_speed 0.550423  0.001301 422.993 <2e-16 *** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 

Residual standard error: 0.2887 on 1780 degrees of freedom
Multiple R-squared:  0.9992, Adjusted R-squared:  0.9992
F-statistic: 1.179e+06 on 2 and 1780 DF,  p-value: < 2.2e-16

```

Abbildung 4: Zusammensetzung der Variable „pace“ aus den beiden Variablen „acceleration“ und „sprint speed“, eigene Darstellung.

Somit ist der Einsatz der Unterkategorien als erklärende Variablen nicht sinnvoll, da diese bloß eine Linearkombination der Oberkategorien darstellen. In diesem Fall spricht man von einer perfekten Multikollinearität, die es dringend zu verhindern gilt.¹³

2.2.6 Weitere Fähigkeitsvariablen

Neben den bereits beschriebenen 6 Oberkategorien, existieren noch zwei weitere Fähigkeitsvariablen, die die „Beidfüßigkeit“ oder die „Trickfähigkeit“ eines Spielers beschreiben. Hier kann jeder Spieler ein Rating zwischen 1 und 5 Punkten erhalten. Diese gehen allerdings nicht in das bereits beschrieben Overall-Rating ein.

2.3 Deskriptive Analyse - nicht berücksichtigte Variablen

Nicht alle Variablen in unserem Datensatz haben tatsächlich einen Einfluss auf den Marktwert eines Spielers. Um zu verhindern, dass jene irrelevanten Einflussgrößen in unser Modell implementiert werden, haben wir zunächst mehrere einfache lineare

¹³ Siehe Auer, L., 2016, S. 160 ff.

Regressionen durchgeführt. Hierbei ist die Unterscheidung von Fähigkeitsvariablen und Allgemeinvariablen wichtig.

2.3.1 Fähigkeitsattribute

Bei der Betrachtung der Fähigkeitsvariablen, können wir bereits mit Hilfe eines Plots und der eingezeichneten Regressionsgerade erkennen, ob eine Variable einen Einfluss auf den Marktwert eines Spielers nimmt. Die bereits zuvor beschriebenen Oberkategorien (hier am Beispiel Schussstärke in Abb. X) haben einen sichtbaren Einfluss. Der Marktwert eines Spielers steigt merklich mit zunehmender Schussstärke.

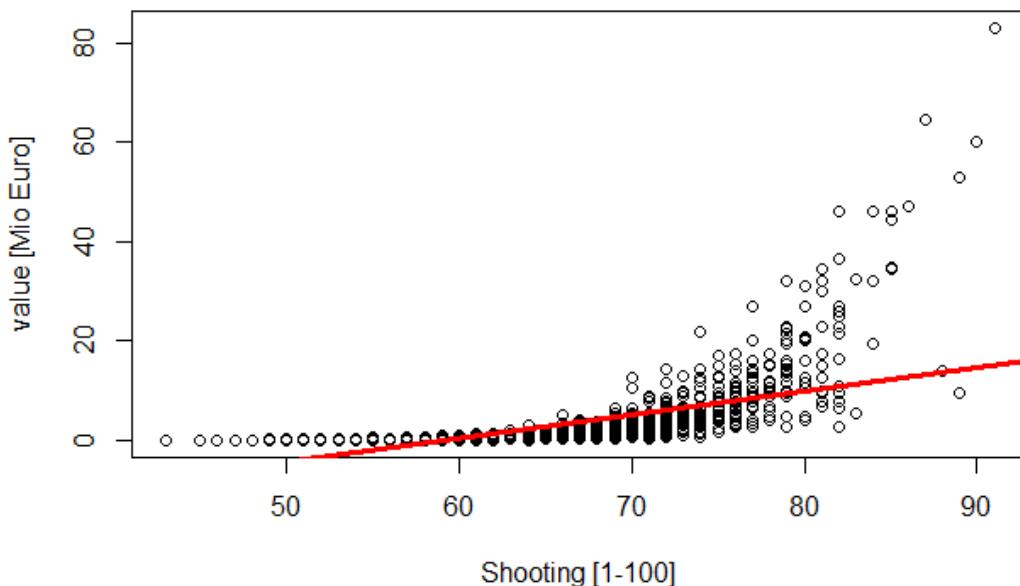


Abbildung 5: Zusammenhang zwischen „shooting“ und Marktwert eines Spielers. Rote Gerade entspricht der linearen Regressionsfunktion. Eigene Darstellung.

Einen ähnlichen Zusammenhang konnten wir für alle weiteren 5 Oberkategorien feststellen. Die Steigerung aller 6 Oberkategorien hat einen signifikanten positiven Einfluss auf den Marktwert.

Ein positiver Zusammenhang gilt offenbar nicht für die in Kapitel 2.2.6 beschriebenen Variablen „Beidfüßigkeit“ und „Trickfähigkeit“.

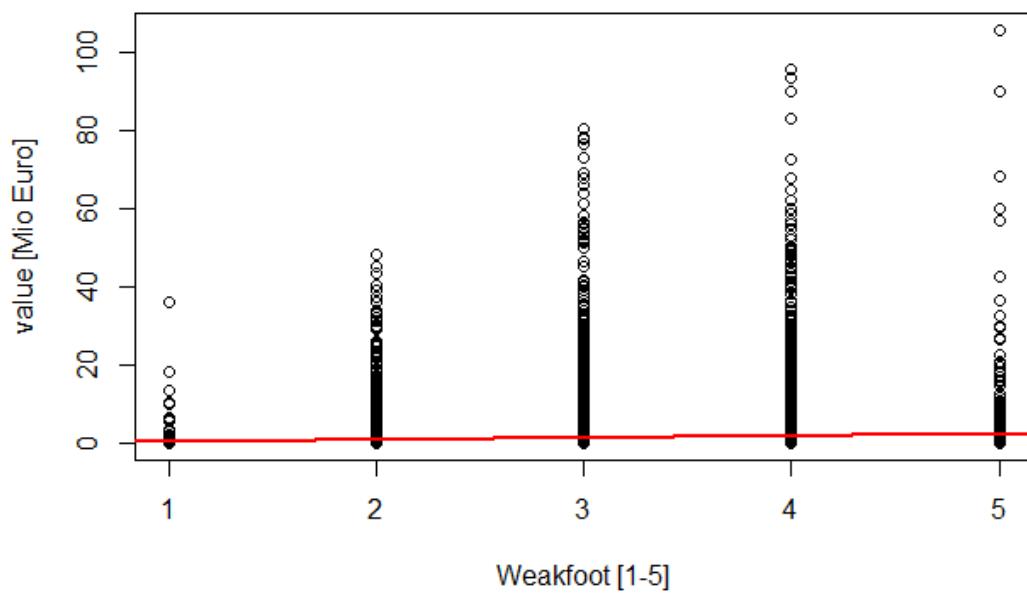


Abbildung 6: Zusammenhang zwischen „Beidfüßigkeit“ und Marktwert. Eigene Darstellung.

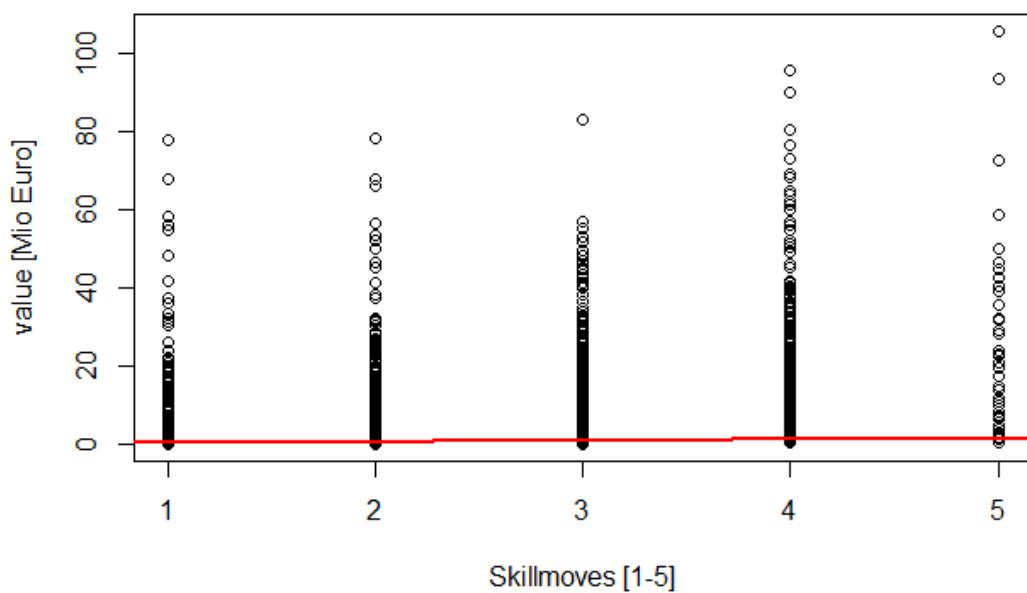


Abbildung 7: Zusammenhang zwischen „Trickfähigkeit“ und Marktwert. Eigene Darstellung.

Somit haben wir die Variablen „Beidfüßigkeit“ und „Trickfähigkeit“ in unserer weiteren Untersuchung nicht mehr berücksichtigt.

2.3.2 Allgemeinattribute

Merkmale wie die Nationalität oder die Ligazugehörigkeit sind nicht metrisch skaliert. Somit bedarf es einer anderen Herangehensweise als im vorherigen Abschnitt. Wir betrachten die Spieler getrennt voneinander in Gruppen (z.B. alle spanischen Spieler) - untersuchen innerhalb dieser Gruppe den linearen Zusammenhang zwischen dem Overall-Rating und dem Marktwert eines Spielers - und vergleichen dann die Regressionsgeraden der verschiedenen Gruppen.

2.3.3 Nationalität

In unserer Untersuchung haben wir Spieler aus 7 verschiedenen Nationen hinsichtlich ihres Overall-Ratings und ihres Marktwerts untersucht.

$$\log(\text{value}) = \beta_0 + \beta_1 * \text{overall} \mid \forall \text{nationality} = \text{germany}$$

$$\log(\text{value}) = \beta_2 + \beta_3 * \text{overall} \mid \forall \text{nationality} = \text{spain}$$

$$\log(\text{value}) = \beta_4 + \beta_5 * \text{overall} \mid \forall \text{nationality} = \text{france}$$

simultan für die restlichen 4 Nationen

$$s.d. \text{ wenn gilt: } \widehat{\beta}_0 \approx \widehat{\beta}_2 \approx \widehat{\beta}_4 \approx \dots \approx \widehat{\beta}_{12}$$

$$\text{und} \quad : \widehat{\beta}_1 \approx \widehat{\beta}_3 \approx \widehat{\beta}_5 \approx \dots \approx \widehat{\beta}_{13}$$

folgt: der Marktwert eines Spielers steigt unabhängig von seiner Nationalität geschätzt mit $\widehat{\beta}_0 + \widehat{\beta}_1 * \text{overall}$

| Coefficients: | | | | | |
|------------------|----------|------------|---------|------------|--|
| | Estimate | Std. Error | t value | Pr(> t) | |
| (Intercept) | 1.225716 | 0.103718 | 11.82 | <2e-16 *** | |
| germany\$overall | 0.188426 | 0.001565 | 120.41 | <2e-16 *** | |

| Coefficients: | | | | | |
|-----------------|----------|------------|---------|--------------|--|
| | Estimate | Std. Error | t value | Pr(> t) | |
| (Intercept) | 0.934525 | 0.123035 | 7.596 | 7.12e-14 *** | |
| france\$overall | 0.193662 | 0.001815 | 106.714 | < 2e-16 *** | |

| Coefficients: | | | | | |
|----------------|----------|------------|---------|--------------|--|
| | Estimate | Std. Error | t value | Pr(> t) | |
| (Intercept) | 0.748354 | 0.153011 | 4.891 | 1.16e-06 *** | |
| spain\$overall | 0.195753 | 0.002178 | 89.860 | < 2e-16 *** | |

Abbildung 8: Einfluss der Variable „Overall“ auf den Marktwert, Betrachtung von Spielern einer Nationalität (Deutschland, Frankreich und Spanien). Eigene Darstellung.

Wir konnten feststellen, dass insbesondere die Steigungskoeffizienten bei allen Nationen nahezu identisch sind. Der Y-Achsenschnittpunkt (intercept) variiert ein wenig (0.7 – 1.3), was möglicherweise durch die unterschiedlichen Größen der Beobachtungsgruppen verursacht wird.

Kleinere echte Ausreißer stellen Brasilien und England dar.

| Coefficients: | | | | | |
|------------------|-----------|------------|---------|----------|-----|
| | Estimate | Std. Error | t value | Pr(> t) | |
| (Intercept) | 1.704133 | 0.094525 | 18.03 | <2e-16 | *** |
| england\$overall | 0.180897 | 0.001485 | 121.81 | <2e-16 | *** |
| Coefficients: | | | | | |
| | Estimate | Std. Error | t value | Pr(> t) | |
| (Intercept) | -0.816272 | 0.241724 | -3.377 | 0.000768 | *** |
| brazil\$overall | 0.215021 | 0.003387 | 63.476 | < 2e-16 | *** |

Abbildung 9: Einfluss der Variable „Overall“ auf den Marktwert, Betrachtung von Spielern einer Nationalität (England und Brasilien). Eigene Darstellung.

Ein Grund, weshalb englische Spieler im Vergleich einen höheren Startmarktwert besitzen, könnte die sogenannte „HGP“ -Home-Grown-Player-Regel sein, die für alle englischen Premier-League-Clubs gilt. Laut dieser Regel müssen Vereine mindestens 7 Spieler in ihrem 25-köpfigen Kader haben, die bereits seit ihrer Jugend im Verein spielen.¹⁴ Dies betrifft besonders englische Spieler, die so „gezwungenermaßen“ in ersten Mannschaften von englischen Profikadern auftauchen – ihnen ein vermeintlich höherer Marktwert zugeteilt wird – obwohl diese nicht wegen ihrer spielerischen Fähigkeit im Kader sind. Diese Annahme ist allerdings rein spekulativ und kann nicht nachgewiesen werden.

Der Steigungskoeffizient für brasilianische Spieler könnte so hoch sein, da diese zum Großteil in den vermeintlich schwachen Ligen in Südamerika spielen und somit eine schlechte Spielerbewertung haben, ein kleiner Teil aber dafür sehr erfolgreich und somit hoch bewertet in den europäischen Topligen spielt.

Trotz der beschriebenen Ausreißer, gehen wir davon aus, dass der Marktwert eines Fußballspielers grundsätzlich nicht von seiner Nationalität abhängt. In einem zusammengefassten Graph für alle 7 Beobachtungen lässt sich dies nochmals verdeutlichen:

¹⁴ Siehe Premier League Squad List, 2020, premierleague.com, Zugriff vom 10.03.2021.

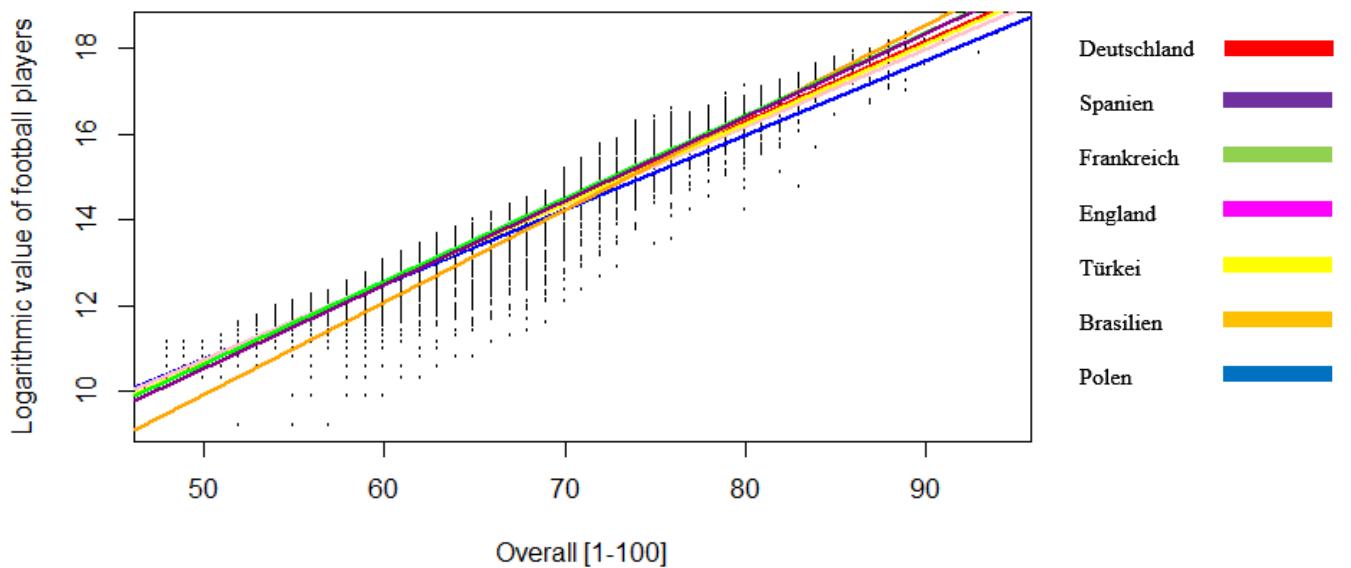


Abbildung 10: Zusammenhang zwischen „Overall“ und Marktwert, Betrachtung von Spielern einer Nationalität. Eigene Darstellung.

Somit nehmen wir die Variable Nationalität nicht in unser Modell auf.

2.3.4 Alter

Es ist logisch, dass das Alter eines Spielers direkt oder indirekt Einfluss auf die Leistungsfähigkeit eines Spielers hat und somit auch auf den Marktwert. Abhängig vom Alter ist auch, ob ein Spieler noch als talentiert, also mit Entwicklungspotenzial betrachtet wird oder bereits am oder nach seinem Leistungszenit steht.

Daraus resultiert, dass der Zusammenhang zwischen Alter und Marktwert, anders als bei allen anderen bereits beschriebenen metrischen Variablen, nicht über die gesamte Betrachtungseinheit positiv ist, sondern eher die Form einer negativ-quadratischen Funktion hat.

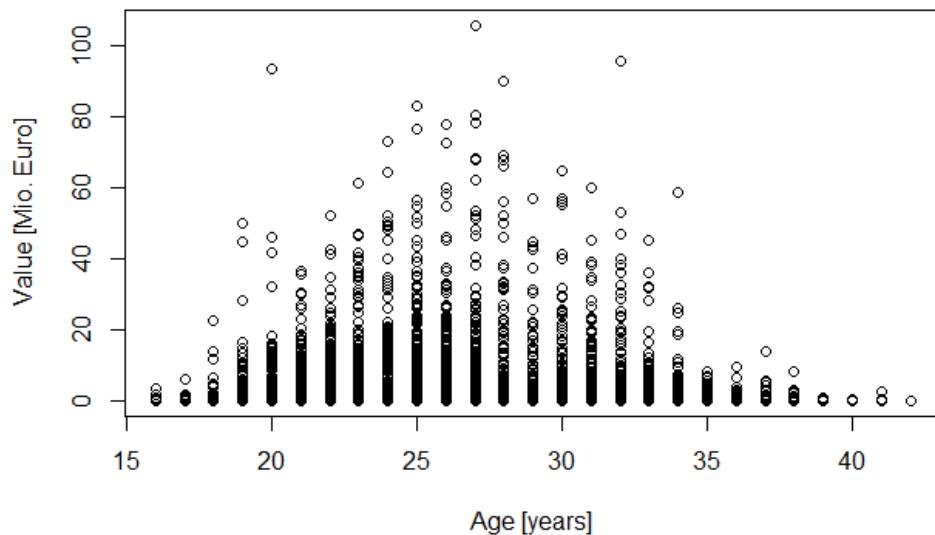


Abbildung 11: Zusammenhang zwischen Alter und Marktwert. Eigene Darstellung.

Aus diesem Grund lassen wir die Variable „Alter“ zweifach in unser Modell einfließen- einmal „age“ und „ age^2 “.

2.4. Beobachtungseinheit Stürmer und Verteidiger

Wir nehmen an, dass die Anforderungen an einen Fußballspieler stark mit seiner Position variieren. So ist der Kernanspruch an einen Stürmer, möglichst viele Tore zu erzielen, weshalb er vor allem gut im Schießen und Dribbeln sein sollte. Ein Verteidiger hingegen könnte besonders dann wertvoll sein, wenn er physisch stark ist. Eine Betrachtung aller Fußballspieler in einer einzigen Regression macht deshalb unserer Meinung keinen Sinn und würde zu einem deutlich unpräziseren Ergebnis führen. Ferner haben wir uns dazu entschlossen, jeweils eine Regression für alle Stürmer und zudem für alle Verteidiger durchzuführen.

```
Fifa20 <- read_excel("Fifa20_0_Mio.xlsx", na="NA")
data_ST <- subset(Fifa20, player_positions == "CB")
Fifa20 <- read_excel("Fifa20_0_Mio.xlsx", na="NA")
data_ST <- subset(Fifa20, player_positions == "ST")
```

Abbildung 12: Betrachtung von ausschließlich Stürmern und Verteidigern

In unserer Regression für Stürmer befinden sich somit nur noch 1783 Spieler, in der für Verteidiger 2289 Spieler. Durch die Betrachtung von zwei Spielerpositionen in Kapitel 4.2 zeigen wir, dass der grundsätzliche Zusammenhang zwischen Spielerfähigkeit und Marktwert für alle Spieler positiv ist, es jedoch abhängig von seiner Position zu einer anderen Priorisierung unter den einzelnen Fähigkeitsvariablen kommt. Eine jeweilige Regression für alle Spielerpositionen überschreitet den Umfang dieser Arbeit.

3 Ökonometrisches Modell

3.1 Kleinste-Quadrat-Methode

Unsere Schätzungen mithilfe der „lm“-Funktion in R basieren auf der Kleinste-Quadrat-Methode. Diese minimiert die Summe der quadratischen Differenzen zwischen den beobachteten und den prognostizierten Werten, den sogenannten Residuen.¹⁵ Diese Methode ist in der Ökonometrie hinlänglich bekannt und wird z.B. in Ludwig von Auers „Einführung in die Ökonometrie“ detailliert beschrieben, weshalb wir auf eine längere Herleitung verzichten. Ferner halten wir es für deutlich wichtiger, die Annahmen und daraus resultierenden Voraussetzungen für die Durchführung einer multiplen linearen Regression zu überprüfen.

3.2 Annahmen für multiple lineare Regression

Annahme A1: In unserer Regressionsgleichung fehlen keine relevanten exogenen Variablen und keine in unserer Gleichung benutzen exogenen Variablen ist irrelevant.¹⁶

Hierzu wurde bereits im Kapitel 1.3 erläutert, dass einige Einflussfaktoren, bspw. Vermarktungsaspekte nicht in unseren Datensatz und somit auch nicht in unser Modell eingehen können. Ferner werden wir die Auswirkungen des Fehlens dieser Variablen in Kapitel 5.2 erläutern.

Hinzu haben wir in Kapitel 2.3 erläutert, welche Variablen in unserem Datensatz vermutlich keine Relevanz für die Bestimmung des Marktwertes haben.

Annahme A2: Der wahre Zusammenhang zwischen den erklärenden Variablen (Schussstärke, Dribbeln etc.) und der endogenen Variable (Marktwert) ist linear. Diese Annahme ist definitiv verletzt.

¹⁵ Siehe Cullmann, A., Ökonometrie Vorlesungsfolien 10, WS 2019, S. 3 ff.

¹⁶ Siehe Auer, L., 2016, S. 160 ff.

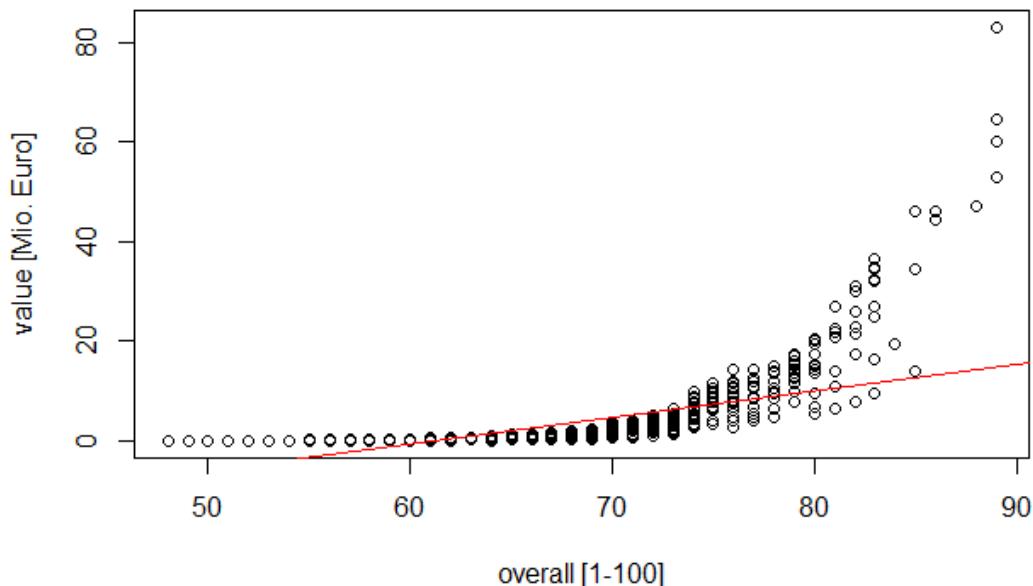


Abbildung 13: Zusammenhang zwischen „Overall“ und Marktwert ist offenbar nicht linear. Eigene Darstellung.

Es liegt eindeutig kein linearer Zusammenhang zwischen den gebündelten Fähigkeitenvariablen (overall) und dem Marktwert (value) vor.

Es liegt nahe, dass der Marktwert eines Spielers mit zunehmender Spielerfähigkeit exponentiell steigt:

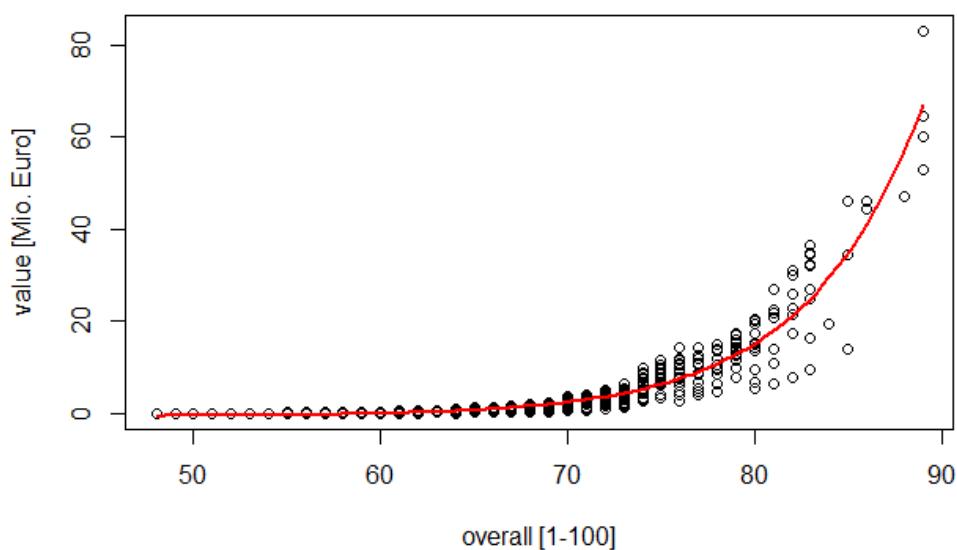


Abbildung 14: Zusammenhang zwischen „Overall“ und Marktwert hat die Form einer Exponentialfunktion. Eigene Darstellung.

Aus diesem Grund haben wir uns dazu entschieden, eine Log-Level-Transformation durchzuführen. Diese Maßnahme wird auch im Lehrbuch empfohlen.¹⁷

Exkurs Log-Level-Transformation

Da es sich in unserem Fall offensichtlich nicht um einen linearen Zusammenhang handelt, passen wir die endogene Variable Marktwert an und logarithmieren sie. Demnach wird aus der Variable „value“ nun „log(value)“:

| long_name | value_mio | log(value) |
|-----------------------------|-----------|-------------|
| Eden Hazard | 90,00 | 4,49980967 |
| Kevin De Bruyne | 90,00 | 4,49980967 |
| Harry Kane | 83,00 | 4,418840608 |
| Mohamed Salah Ghaly | 80,50 | 4,388257184 |
| Virgil van Dijk | 78,00 | 4,356708827 |
| Jan Oblak | 77,50 | 4,350277936 |
| Paulo Bruno Exequiel Dybala | 76,50 | 4,337290741 |
| Raheem Sterling | 73,00 | 4,290459441 |
| Paul Pogba | 72,50 | 4,283586562 |
| Antoine Griezmann | 69,00 | 4,234106505 |
| Christian Dannemann Eriksen | 68,00 | 4,219507705 |

Abbildung 15: Beispiel für Anpassung der Variable Marktwert durch Logarithmierung. Eigene Darstellung.

Nach unserer Log-Level-Transformation sieht der Plot folgendermaßen aus:

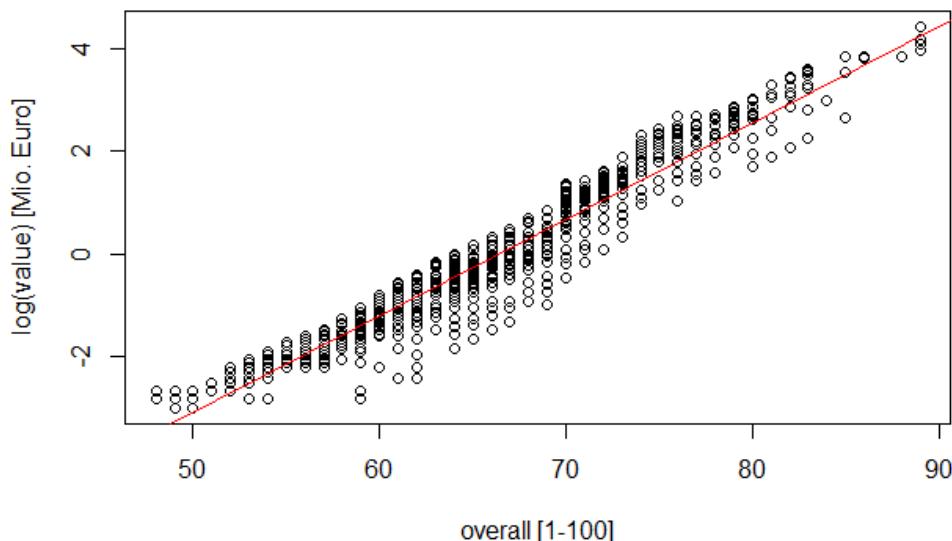


Abbildung 16: Zusammenhang „Overall“ und logarithmierter Marktwert. Eigene Darstellung.

¹⁷ Siehe Auer, L., 2016, S. 338.

Wie man sieht, konnten wir mit der Logarithmierung des Marktwertes einen linearen Zusammenhang erzeugen. Welche Auswirkungen diese Transformation auf die Interpretation unserer Ergebnisse hat, beschreiben wir in Kapitel 3.3 und 3.4.

Fakt ist, wir konnten mithilfe dieses „Tricks“ auch die zweite Annahme für die Durchführung einer multiplen linearen Regression erfüllen.

Wichtig! Ab jetzt prüfen wir alle Voraussetzungen für unser neues Log-Level-Modell.

Annahme A3: Die $n+1$ Parameter $\beta_0; \beta_1 \dots \beta_n$ sind für alle T Beobachtungen $x_{0T}; x_{1T} \dots x_{nT}$ konstant.

Eine Verletzung dieser Annahme trifft oftmals bei Zeitreihen auf.¹⁸ Unser Datensatz erfasst ausschließlich die Spielerdaten, die zum Release des Spiels FIFA20 am 26.September 2019 veröffentlicht wurden. Wir nehmen die Annahme somit als gegeben an.

Annahme B1: Der Erwartungswert der Störterme U_i liegt für alle Beobachtungen bei 0.¹⁹ Wir kennen zwar die realen Störterme nicht, können aber mithilfe der Residuen eine erste Einschätzung vornehmen. Wir schauen uns das Histogramm zur Verteilung der Residuen an.

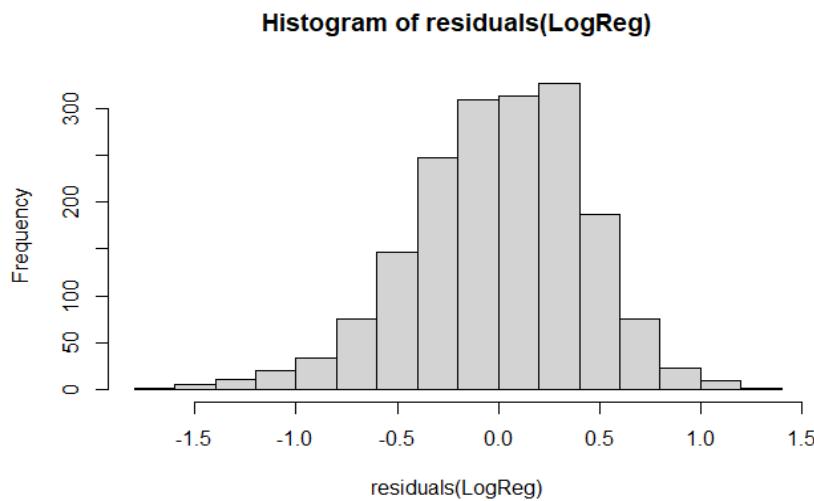


Abbildung 17: Verteilung der Residuen für den Zusammenhang zwischen „Overall“ und „Log(value)“. Eigene Darstellung.

¹⁸ Siehe Auer, L., 2016, S. 361.

¹⁹ Siehe Auer, L., 2016, S. 160 ff.

Der Erwartungswert der Residuen liegt hier bei $E(U) = -5 * 10^{-17} \approx 0$. Auch graphisch zeichnet sich eine Normalverteilung um den Wert 0 ab.

Annahme B2: Homoskedastizität: Die Störgröße U_i hat für alle Beobachtungen eine konstante Varianz: $\text{var}(U_i) = o^2$.²⁰

Auch hier kennen wir die realen Störterme nicht, können aber mithilfe der Residuen eine erste Einschätzung vornehmen. Wir schauen uns die Residuen und die geschätzten Werte gemeinsam in einem Plot an:

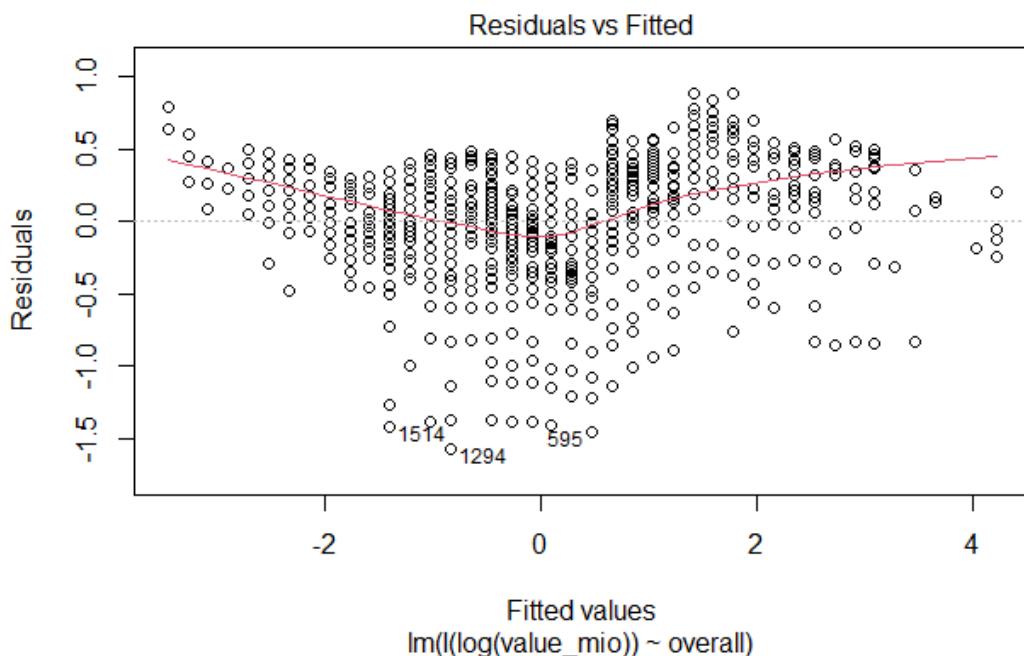


Abbildung 18: Plot mit geschätzten logarithmierten Marktwerten und Residuen. Eigene Darstellung.

Bedauerlicherweise liegt die rote Linie nicht auf ganzer Länge auf der gestrichelten Linie. Die Varianz der Residuen ist nicht konstant, ferner ist tendiert sie für geringe und sehr große Marktwerte ins Positive. Somit sprechen wir hier von leichter Heteroskedastizität.²¹

Annahme B3: Die Störgröße ist normalverteilt $u_i \sim N(E(u_i); \text{var}(u_i))$.²²

Bereits bei Annahme B1 beantwortet.

²⁰ Siehe Auer, L., 2016, S. 160 ff.

²¹ Siehe Statistik am PC, 2020, youtube.com, Zugriff v. 01.03.2021.

²² Siehe Auer, L., 2016, S. 160 ff.

Annahme C1: Die exogenen Variablen des Modells sind keine Zufallsvariablen, sondern können im Experiment kontrolliert werden.²³

Davon ist auszugehen, da FIFA die Spielerdaten kontinuierlich erfasst und auswertet, wie bereits in Kapitel 1.2 beschrieben.

Annahme C2: Freiheit von perfekter Multikollinearität: Es existieren keine Parameterwerte $\beta_0, \beta_1, \beta_2 \dots \beta_n$, s.d.: $\beta_0 + \beta_1 * x_1 + \dots + \beta_n x_k = 0$

Hierfür lassen wir uns in R den „vif“ (variance inflation factor) ausgeben:

```
> vif(LogReg)
  shooting      pace    passing   dribbling  defending     physic
  4.738779  1.614625  3.933255  5.403819  1.922484  1.847988
```

Abbildung 19: vif-Koeffizienten. Eigene Darstellung.

Die Werte für den vif sollten für alle exogenen Variablen unter 10 liegen, damit Multikollinearität gegeben ist.²⁴ Dies ist hier der Fall.

Wir haben nun alle Annahmen vollständig überprüft. Leider weisen die Residuen keine konstante Varianz auf (siehe Annahme B2), wir gehen aber von einer verkraftbaren Abweichung aus und fahren deshalb mit der Durchführung der multiplen linearen Regression fort.

3.3 Interpretation von Log-Level Ergebnissen

Da in unseren späteren Regressionen in Kapitel 4.2 um Variable Marktvalor logarithmiert geschätzt wird, muss man die errechneten Steigungsschätzparameter als Elastizitäten der abhängigen Variable (“value”) in Bezug auf die erklärenden Variablen lesen.²⁵

$$\log(value) = \beta_0 + \beta_1 * overall + U$$

$$value = e^{\beta_0 + \beta_1 * overall + U}$$

$$\frac{\partial value}{\partial overall} = (e^{\beta_0 + \beta_1 * overall + U}) * \beta_1 = value * \beta_1$$

$$\frac{\partial value}{value} = \beta_1 * \partial overall \rightarrow \% * \Delta \widehat{value} \approx (100 * \widehat{\beta}_1) * \Delta overall$$

²³ Siehe Auer, L., 2016, S. 160 ff.

²⁴ Siehe Wooldridge, J., 2012, S.98.

²⁵ Siehe Wooldridge, J., 2012, S. 191 ff.

Diese Interpretation der prozentualen Veränderung ist allerdings für unsere Zielsetzung, den Marktwert eines Spielers zu schätzen, nicht ausreichend. Es ließe sich in diesem Fall nur eine Marktwertveränderung in Relation zu einem anderen Spieler bestimmen: Spieler X hat einen Bewertungspunkt mehr im Overall als Spieler Y, so sollte der Marktwert von Spieler X im Mittel geschätzt um $100 * \beta_1 * \%$ höher liegen, als der Marktwert von Spieler Y. Diese Schätzung wird für große Marktwerte ungenauer.²⁶

Um mit unserem Log-Level-Modell eine Schätzung des nicht logarithmierten Marktwertes vorzunehmen, bedarf es einer Re-Transformation.²⁷

3.4 Re-Transformation der geschätzten logarithmierten Werte

Intuitiv ergeben sich die Level-Schätzungen für den Marktwert mit der Aufhebung des Logarithmus:

$$\widehat{\text{value}} = e^{\widehat{\beta}_0 + \widehat{\beta}_1 * \text{overall}} = e^{\log(\text{value})}$$

Das ist jedoch nicht richtig, denn dies führt zu einer systematischen Unterschätzung des Marktwertes.²⁸ Wir können dieses Problem beheben, indem wir den konstanten Parameter a_0 einführen. Dieser stützt nicht auf der Annahme, dass die Fehlerterme U normalverteilt sind (Methode I), sondern beschreibt den unabhängigen Erwartungswert aller Fehlerterme (Methode II).²⁹

Methode I: $\widehat{\text{value}} = E\left(\frac{\sigma^2}{2}\right) * e^{\log(\text{value})} \mid U \sim N(E(U); \text{var}(U))$

Methode II: $a_0 = E(U)$

Es folgt: $E(\text{value}|\text{overall}) = a_0 * e^{\beta_0 + \beta_1 * \text{overall}}$

Wir schätzen: $\widehat{a_0} = n^{-1} * \sum_{i=1}^n E(\widehat{U})$

Es folgt: $\widehat{\text{value}} = \widehat{a_0} * e^{\log(\text{value})}$

$\widehat{a_0}$ ist ein konsistenter Schätzer von $a_0 = E(U)$, allerdings nicht erwartungstreu, da wir die unbekannten Störgrößen U mit den Residuen \widehat{U} ersetzt haben.³⁰ Dies ist ein

²⁶ Siehe Auer, L. 2016, S. 338 ff.

²⁷ Siehe Wooldridge, J., 2012, S. 212.

²⁸ Siehe Wooldridge, J., 2012, S. 212.

²⁹ Siehe Wooldridge, J., 2012, S. 212.

³⁰ Siehe Wooldridge, J., 2012, S. 213.

Grund, weshalb die Schätzungen mit Methode II für zunehmend große Marktwerte schlechter wird.³¹

3.5 Güte des Schätzers

Nun gilt es also, aus den verbliebenen Variablen, das bestmögliche Regressionsmodell zu bilden und somit den bestmöglichen Schätzer zur Erklärung des Marktwertes zu finden. Eine Methode, eine Aussage über die Güte eines Schätzers zu treffen, ist das bekannte Bestimmtheitsmaß R^2 . Es zeigt, wie viel Prozent der Variation der endogenen Variablen durch die exogenen Variablen erklärt werden kann.³²

$$R^2 \equiv \frac{SQE}{SQT} = \frac{\sum (\hat{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2}$$

Die Re-Transformation der $\log(\text{value})$ -Variable (siehe Kapitel 3.4) und damit die Einführung des Parameters $\widehat{a_0}$, zwingt uns zu einer Anpassung des R^2 . Ferner betrachten wir nun die Korrelation zwischen unseren beobachteten Marktwerten und unseren geschätzten Marktwerten.³³

$$R^2_{adjusted} = (Cor(\text{value}; \widehat{\text{value}} = \widehat{a_0} * e^{\log(\text{value})}))^2$$

³¹ Siehe Wooldridge, J., 2012, S. 214.

³² Siehe Auer, L., 2016, S. 200.

³³ Siehe Wooldridge, J., 2012, S. 214 ff.

4 Ergebnisse

4.1 Zwischenstand

Wie wir bereits in Kapitel 3.2 ausgeführt haben, bedarf es einer Log-Level-Transformation, also einer Umwandlung der Variable „value“ in „log(value)“. Ferner haben wir in Kapitel 3.3 und 3.4 notwendige Hinweise zur Interpretation gegeben. Wir konnten bereits einige Variablen in unserem Datensatz als irrelevant für die Schätzung des Marktwertes einstufen (siehe Kapitel 2.3).

4.2 Regressionen

Grundgerüst für unsere Regression bilden die 6 Fähigkeitsvariablen. Aus ihnen bilden wir unser erstes multiples Regressionsmodell:

$$I \quad \log(\text{value}) = \beta_0 + \beta_1 * \text{shooting} + \beta_2 * \text{pace} + \beta_3 * \text{passing} + \beta_4 * \text{defending} + \beta_5 * \text{physic} + \beta_6 * \text{dribbling}$$

R errechnet uns folgende Koeffizienten für Stürmer und Verteidiger:

| Coefficients: | | | | | | | | | |
|---------------|-----------|------------|---------|--------------|-------------|---------------|-----------|--------|--------------|
| | Estimate | Std. Error | t value | Pr(> t) | | Coefficients: | | | |
| (Intercept) | 0.851646 | 0.111663 | 7.627 | 3.89e-14 *** | (Intercept) | 0.3990033 | 0.1147365 | 3.478 | 0.000515 *** |
| shooting | 0.127135 | 0.002858 | 44.479 | < 2e-16 *** | shooting | -0.0060802 | 0.0014173 | -4.290 | 1.86e-05 *** |
| passing | -0.007305 | 0.002268 | -3.220 | 0.0013 ** | passing | 0.0008309 | 0.0020055 | 0.414 | 0.678702 |
| pace | 0.015270 | 0.001217 | 12.549 | < 2e-16 *** | pace | 0.0196334 | 0.0009146 | 21.466 | < 2e-16 *** |
| defending | -0.003305 | 0.001838 | -1.799 | 0.0723 . | defending | 0.1536101 | 0.0019376 | 79.278 | < 2e-16 *** |
| physic | 0.016120 | 0.001377 | 11.703 | < 2e-16 *** | physic | 0.0244889 | 0.0017427 | 14.052 | < 2e-16 *** |
| dribbling | 0.048081 | 0.003205 | 15.004 | < 2e-16 *** | dribbling | 0.0102011 | 0.0018618 | 5.479 | 4.74e-08 *** |

Abbildung 20: Regression I: Einfluss der 6 Fähigkeitsvariablen auf den logarithmierten Marktwert (links Stürmer, rechts Verteidiger). Eigene Darstellung.

Die Ergebnisse bestätigen unsere Hypothese II aus Kapitel 1.1.3, dass die Einflussstärke einer Fähigkeitsvariable abhängig von der jeweiligen Position ist. Anhand unserer Schätzparameter erkennen wir, dass für Stürmer insbesondere die Schussstärke, für Innenverteidiger das Verteidigen und die physische Stärke, einen Einfluss auf den Marktwert nehmen. Konkret steigert ein zusätzlicher Punkt im Shooting-Rating, den Marktwert eines Stürmers im Mittel geschätzt um 12,71 %. Ferner lässt das geringe Signifikanzniveau $\alpha = 7,23\% > 5\%$ und der marginale Wert des Steigungskoeffizienten $\beta_4 = -0,0033$ vermuten, dass die Variable Verteidigen kaum Einfluss und damit irrelevant für die Schätzung des Marktwertes eines Stürmers ist. Wir streichen sie somit aus der folgenden Regression X. Gleiches gilt für die Variable Passen für Innenverteidiger ($\alpha = 67,87\%$ und $\beta_2 = 0,0008$).

Wir haben gezeigt, dass es einen großen Unterschied macht, auf welcher Position ein Spieler spielt und werden ab jetzt ausschließlich Stürmer in unserer Analyse betrachten.

In der Folge ergänzen wir unsere ursprüngliche Regression um die Variablen Potential (delta_potential), Internationale Bekanntheit (international_reputation) und das Alter (age und age²). (siehe Kapitel 2.2.2 und Kapitel 2.3.4)

Unsere zweite Regressionsfunktion für die logarithmierten Marktwerte der Stürmer nimmt also folgende Form an:

$$\text{II } \log(\text{value}) = \beta_0 + \beta_1 * \text{shooting} + \beta_2 * \text{pace} + \beta_3 * \text{passing} + \beta_4 * \text{dribbling} + \beta_5 * \text{physic} + \beta_6 * \text{reputation} + \beta_7 * \text{potential} + \beta_8 * \text{age} + \beta_9 * \text{age}^2$$

| Coefficients: | | | | | | |
|---|------------|------------|---------|----------|-----|--|
| | Estimate | Std. Error | t value | Pr(> t) | | |
| (Intercept) | -3.0471515 | 0.4392673 | -6.937 | 5.60e-12 | *** | |
| shooting | 0.1307435 | 0.0020926 | 62.480 | < 2e-16 | *** | |
| pace | 0.0068784 | 0.0008991 | 7.650 | 3.27e-14 | *** | |
| passing | 0.0054395 | 0.0015853 | 3.431 | 0.000614 | *** | |
| dribbling | 0.0462607 | 0.0022957 | 20.151 | < 2e-16 | *** | |
| physic | 0.0257944 | 0.0010072 | 25.611 | < 2e-16 | *** | |
| international_reputation | 0.1414113 | 0.0196113 | 7.211 | 8.21e-13 | *** | |
| delta_potential | 0.0357015 | 0.0043617 | 8.185 | 5.13e-16 | *** | |
| age | 0.2494411 | 0.0294189 | 8.479 | < 2e-16 | *** | |
| I(age^2) | -0.0055993 | 0.0004939 | -11.337 | < 2e-16 | *** | |
| --- | | | | | | |
| Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 | | | | | | |
| Residual standard error: 0.3038 on 1773 degrees of freedom | | | | | | |
| Multiple R-squared: 0.9529, Adjusted R-squared: 0.9526 | | | | | | |
| F-statistic: 3983 on 9 and 1773 DF, p-value: < 2.2e-16 | | | | | | |

Abbildung 21: Regression II: Einfluss von Fähigkeitsvariablen und Reputation, Potential und Alter auf den logarithmierten Marktwert. Eigene Darstellung.

Alle geschätzten Steigungsparameter erreichen ein Signifikanzniveau von $< 5\% = \alpha$.

Wir errechnen wie in Kapitel 3.4 beschrieben das $\widehat{a}_0 = 1,047341$

Es folgt: $\widehat{\text{value}} = 1,047341 * e^{-3.047+0.131*\text{shooting}+...+0.249*\text{age}-0.006*\text{age}^2}$

Mit einem angepassten $R^2_{adjusted} = 0,9187213$ (Siehe Kapitel 3.5).

Dieses Ergebnis ist für uns sehr zufriedenstellend.

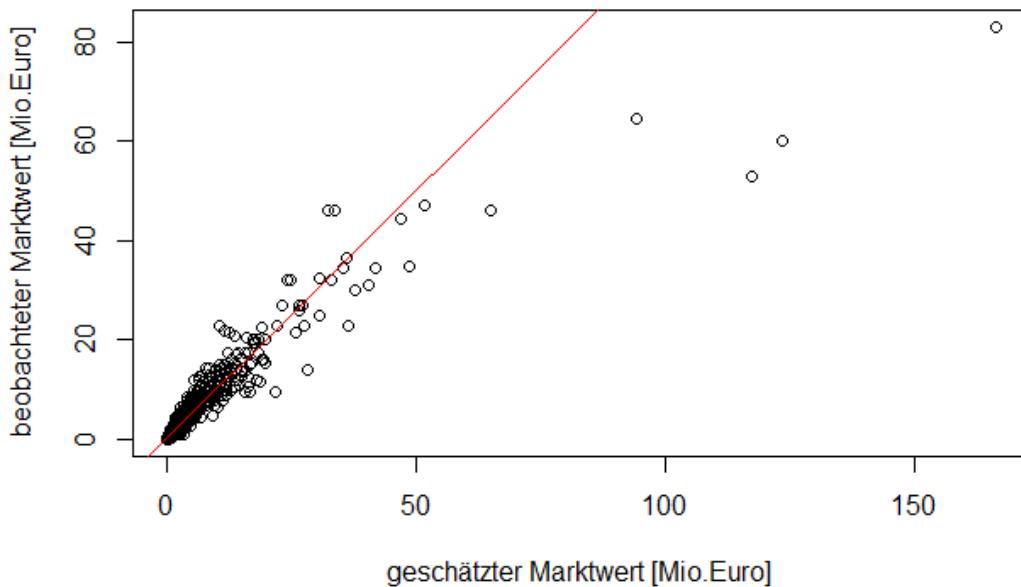


Abbildung 22: Gegenüberstellung von unseren geschätzten Marktwerten und den beobachteten Marktwerten.
Eigene Darstellung.

Unsere geschätzten Marktwerte liegen insbesondere für kleinere Marktwerte nah an den Beobachtungen (bei 100 % Übereinstimmung würden alle Punkte in Abbildung 22 auf der roten Linie liegen).

Wir können jedoch eine zunehmende Ungenauigkeit des Schätzers für große Marktwerte erkennen. Dies haben wir bereits in Kapitel 3.4 erwartet.

4.3 Z-Standardisierung

Anhand der Koeffizienten können wir allerdings nicht ablesen, welche der Variablen den größten Einfluss auf den Marktwert hat, da sie unterschiedlich skaliert sind. Während die Fähigkeitsvariablen einen Wert zwischen 1 und 100 annehmen, liegt der Wert für die internationale Reputation hingegen auf einer Skala von 1 bis 5. Damit wir trotzdem eine Aussage treffen können, müssen wir die Regression noch einmal mit Z-standardisierten Variablen durchführen.³⁴

$$Z \equiv \frac{X - \mu}{\sigma}$$

³⁴ Siehe Wooldridge, J., 2012, S. 189 ff.

Wenn wir alle Variablen standardisiert haben und die Regression erneut durchführen, kommen wir zu folgendem Ergebnis:

| Coefficients: | | Estimate | Std. Error | t value | Pr(> t) | |
|------------------|--|-----------|------------|---------|----------|-----|
| (Intercept) | | 0.091557 | 0.009580 | 9.557 | < 2e-16 | *** |
| Zshooting | | 0.717671 | 0.011486 | 62.480 | < 2e-16 | *** |
| Zpace | | 0.051770 | 0.006767 | 7.650 | 3.27e-14 | *** |
| Zpassing | | 0.034279 | 0.009990 | 3.431 | 0.000614 | *** |
| Zdribbling | | 0.241869 | 0.012003 | 20.151 | < 2e-16 | *** |
| Zphysic | | 0.183479 | 0.007164 | 25.611 | < 2e-16 | *** |
| Zrep | | 0.044516 | 0.006174 | 7.211 | 8.21e-13 | *** |
| Zdelta_potential | | 0.151365 | 0.018492 | 8.185 | 5.13e-16 | *** |
| Zage | | -0.110700 | 0.019068 | -5.805 | 7.59e-09 | *** |
| I(Zage^2) | | -0.091609 | 0.008080 | -11.337 | < 2e-16 | *** |

Abbildung 23 z-standardisierter Output der Regression II

Betrachtet man die geschätzten Steigungsparameter für „shooting“ und „reputation“ können wir die Auswirkung der z-Standardisierung erkennen. Während in der ursprünglichen nicht standardisierten Regression der Koeffizient für die internationale Reputation größer ist als der für die Schussfähigkeit (siehe Abbildung 21), wird mit der Standardisierung deutlich, dass tatsächlich „shooting“ einen größeren Einfluss auf den Marktwert hat (siehe Abbildung 23). Der Koeffizient der Reputation war nur so groß, da eine Einheit auf einer Skala [1-5], 20 Einheiten auf einer Skala [1-100] entspricht.

Den größten Effekt auf den Marktwert hat also die Fähigkeitenvariable Schussstärke, mit deutlichem Abstand folgen Dribbeln, Physische Stärke und das Potential. Die Passfähigkeit hat offenbar den geringsten Einfluss.

5 Fazit

5.1 Zusammenfassung

Das wesentliche Ziel unserer Forschungsarbeit ist es, den Marktwert eines Spielers anhand der von dem Videospiel FIFA vergebenen Fähigkeitsbewertungen (z.B. Schussstärke) und Allgemeingrößen (z.B. Alter) zu schätzen. Wir konnten unsere **erste Hypothese** bestätigen und feststellen, dass der Zusammenhang zwischen unseren erklärenden Variablen und dem Marktwert nicht linear ist, sondern vielmehr einer exponentiellen Funktion gleicht. (Siehe Kapitel 3.2). Um dennoch eine multiple lineare Regressionsanalyse durchführen zu können, haben wir eine Log-Level-Transformation durchgeführt.

Unsere **zweite Hypothese** lautete, dass die Anforderungen an einen Spieler und damit verbunden die Ausprägungen der geschätzten Steigungsparameter, von seiner jeweiligen Position abhängen. Dies konnten wir mit unserem ersten Regressionsmodell bestätigen. Die geschätzten Koeffizienten der 6 Fähigkeitsvariablen unterschieden sich eindeutig, abhängig davon ob es sich bei den beobachteten Spielern um Stürmer oder Verteidiger handelte. (Siehe Kapitel 4.2)

Hypothese III, nicht alle im von FIFA bereitgestellten Spielermerkmale haben tatsächlich einen Einfluss auf den Marktwert, konnten wir ebenfalls bestätigen. So konnten wir z.B. zeigen, dass die Nationalität eines Spielers keinen eindeutigen Effekt auf den Marktwert hat. (siehe Kapitel 2.3.3)

Hypothese IV, der Marktwert eines Spielers lässt sich nicht ausschließlich anhand seiner sichtbaren/messbaren Merkmale schätzen. Vielmehr nehmen (für uns) nicht ersichtliche Einflussgrößen oder solche die schlicht nicht im FIFA-Datensatz erfasst wurden, wie die Vermarktbarkeit eines Spielers oder das Budget des kaufenden Vereins, ebenfalls Einfluss auf den Marktwert bzw. letztlich den tatsächlichen Verkaufspreis. Auf die Beantwortung dieser Hypothese gehen wir im folgenden Kapitel 5.2 näher ein.

5.2 Reale Spielertransfers in der Saison 2019/2020

Um zu untersuchen, wie weit unsere geschätzten Marktwerte von den tatsächlich real gezahlten Transfersummen entfernt sind, haben wir uns eine Liste von 47

Stürmertransfers weltweit aus der Saison 2019/2020 erstellt und diese in unseren Datensatz eingearbeitet.³⁵

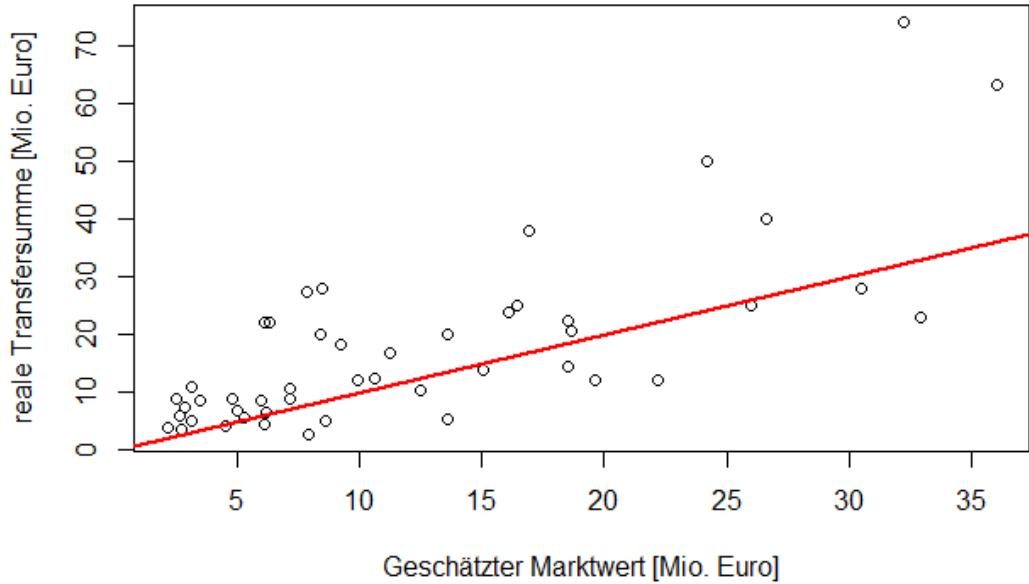


Abbildung 24: Vergleich unserer geschätzten Marktwerte (Regression II) mit real gezahlten Transfersummen (für 47 Spieler). Eigene Darstellung.

Unsere Schätzungen für den Marktwert liegen im Mittel 45,13 % von den real gezahlten Verkaufspreisen entfernt.

$$\text{Betrag der Abweichung} = |\widehat{\text{value}} - \text{Transfersumme}| \text{ für } i = 47 \text{ Spieler}$$

$$\text{Prozentuale Abweichung} = \frac{\text{Betrag der Abweichung}}{\text{Transfersumme}} \text{ für } i = 47 \text{ Spieler}$$

$$\text{gemittelte \% - Abweichung} = \frac{1}{47} * \sum_{i=1}^{47} \text{Prozentuale Abweichung} = 45,13 \%$$

Dies erscheint sehr viel, bestätigt aber auch unsere Hypothese IV, dass uns unbekannte Größen offensichtlich Einfluss auf den Markt- bzw. Verkaufspreis nehmen. Wir können in Abbildung 24 zusätzlich erkennen, dass unsere Schätzungen häufig unter den realen Verkaufspreisen liegen.

³⁵ Siehe Transfermarkt, 2019, transfermarkt.de, Zugriff v. 10.03.2021.

Nachfolgend möchten wir mögliche Erklärungsansätze für die beschriebenen Abweichungen vorstellen und erläutern.

5.2.1 Vermarktungswert

Wie im bereits anhand des einführenden Beispiels für Christiano Ronaldo erklärt (siehe Kapitel 1.3), können Spieler neben ihren sportlichen Leistungen, auch mittels Ihres Images einen enormen Wert zum Verein beitragen. Dies kann in Form von Erlösen aus Trikotverkäufen und Stadiontickets oder auch zusätzlicher Social-Media-Reichweite und damit besserer Werbefähigkeit des Klubs geschehen.³⁶ Diese Eigenschaft eines Spielers ist nicht in unserem Modell berücksichtigt – kann sich jedoch positiv auf den Verkaufswert und somit den realen Marktwert eines Spielers auswirken. Dies könnte insbesondere ein Grund für hohe Abweichungen bei hohen Transfersummen sein, da teure Spieler eher den Status eines Stars mit sich bringen.

5.2.2 Liquidität des kaufenden Vereins

Wir vermuten, dass Vereine, die über eine außerordentlich hohe Liquidität verfügen, häufiger bereit sind mehr Geld für einen Spieler auszugeben, als dessen Marktwert tatsächlich suggeriert. Ein Grund für höhere Spieler-Budgets können bessere TV-Verträge, externe Kapitalgeber oder auch hohe vorige Einnahmen aus Transfers sein.

In Sachen TV-Gelder-Ausschüttung liegt die weltweit meistgeschaute englische Premier League im internationalen Vergleich deutlich vorne. In der Saison 2018/2019 wurden insgesamt 2,6 Mrd. Euro an die 20 Ligavereine ausgeschüttet. Zum Vergleich: Die 18 Vereine der Bundesliga erhielten gerade einmal „nur“ 1,2 Mrd. Euro.^{37 38}

Hebt man alle Spieler-Transfers in die Premier League hervor, bestätigt Abbildung 25 unsere Vermutung.

³⁶ Siehe Williams, C., Erst.-Datum: unbekannt, ColossusBlog, Zugriff v. 10.03.2021.

³⁷ Siehe Süddeutsche Zeitung, 2020, sueddeutsche.de, Zugriff v. 10.03.2021.

³⁸ Siehe Fernsehgelder, 2019, fernsehgelder.de, Zugriff v. 10.03.2021.

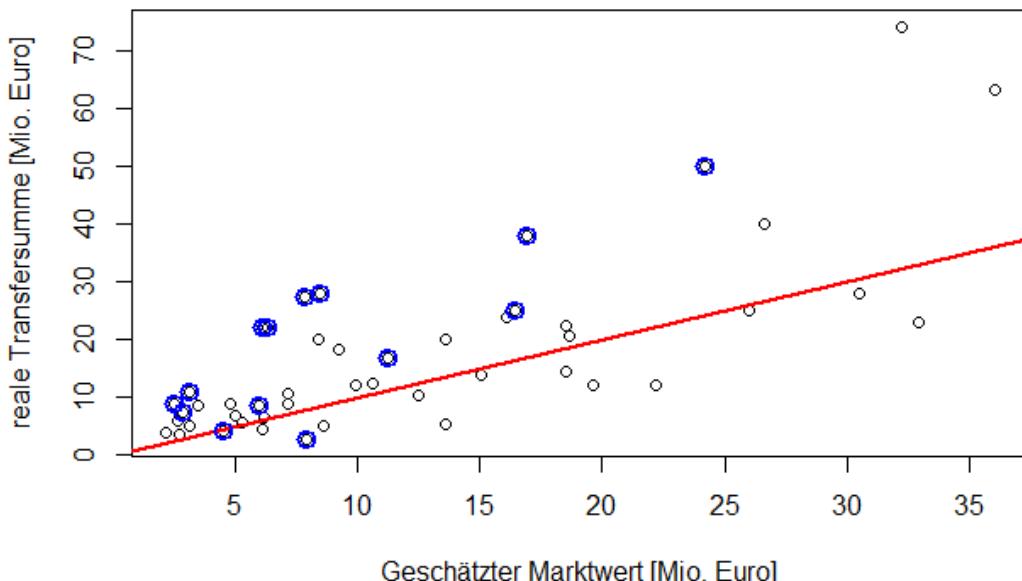


Abbildung 25: basierend auf Abbildung 24 – Transfers in die englische Premier League hervorgehoben. Eigene Darstellung.

Die Transfersummen für Spieler, die in die Premier League wechselten, liegen mit einer Ausnahme alle über unseren geschätzten Marktwerten (Siehe Abbildung 25).

Der teuerste Transfer in unserer Liste, der 74 Mio.-Euro-Wechsel von Romelu Lukaku von Manchester United zu Inter Mailand, wurde nicht zuletzt vom chinesischen Investor „Suning Commerce“ finanziert, der 2016 über 69 % des Mailänder Vereins kaufte.³⁹ Womöglich hat die zusätzliche Liquidität durch den Investor auch zu einer Erhöhung des Verkaufspreises geführt – der Verkäufer (Manchester United) wusste von dem breiten Budget und hat den Spieler letztlich über seinem Marktwert (unsere Schätzung liegt bei 32 Mio. Euro) an Inter Mailand verkauft.

5.2.3 Vertragsrestlaufzeit und Verletzungshistorie

Der FIFA-Datensatz enthält keine Informationen über vorangegangene Verletzungen oder die restliche Vertragslänge eines Spielers. Die Verletzungsanfälligkeit eines Spielers mindert dessen Marktwert vermutlich, da Verletzungspausen letztlich Opportunitätskosten für den jeweiligen Verein darstellen.⁴⁰ Ferner sinkt der

³⁹ Siehe Zeit Online, 2016, zeit.de, Zugriff v. 11.03.2021.

⁴⁰ Siehe Catapultsports, 2019, catapultsports.com, Zugriff v. 11.03.2021.

Verkaufswert eines Spielers mit abnehmender Vertragslaufzeit kontinuierlich – da dieser zu Vertragsende ablösefrei – also kostenlos – den Verein verlassen kann und somit keine Ablösezahlungen generiert.⁴¹

5.2.4 Angebot und Nachfrage

Nicht zuletzt erklären vermutlich auch Angebot und Nachfrage die Höhe der tatsächlich gezahlten Transfersumme. Die Covid-19-Pandemie und die damit verbundenen Fehleinnahmen durch Ticketerlöse haben viele Vereine in Geldnot gebracht.⁴² Die Klubs brauchen also dringend Geld und bieten mehr Spieler auf dem Markt an – das Angebot steigt, gleichzeitig sinkt die Kaufbereitschaft bei anderen Klubs aus Geldgründen – die Nachfrage sinkt. Dies führt zu weniger und günstigeren Transfers. Die Pandemie zeigt dieses Phänomen besonders stark, es kann jedoch in abgeschwächter Form auch im normalen Fußballbetrieb auftreten, wenn schlicht wenige Spieler einer Position und Qualität auf dem Markt verfügbar sind.

5.3 Abschluss

Die genannten Einflussfaktoren konnten wir nicht in unsere Schätzung einfließen lassen. Wir konnten dennoch, im Rahmen der uns zu Verfügung gestellten Daten, einen Einblick in die Marktwert-Entstehung eines Fußballspielers geben. Uns ist bewusst, dass insbesondere die Fähigkeitsbewertung innerhalb des Videospiels FIFA nie eine perfekte Wiedergabe der eigentlichen Spielerleistungen darstellen kann, da sie letztlich von Menschen (Siehe Kapitel 1.2) erstellt wurde. Vermutlich existiert ein besseres, womöglich exponentielles Regressionsmodell, zur Schätzung des Marktwertes. Im Rahmen unseres theoretischen Wissens war die Verwendung eines multiplen linearen Modells und inbegriffener Log-Level-Transformation die bestmögliche Wahl.

⁴¹ Siehe FIFA Transfer Regulations, 2019, Abschnitt 18.3, fifa.com, Zugriff v. 12.03.2021.

⁴² Siehe Tagesschau, 2020, tagesschau.de, Zugriff v. 12.03.2021

Literatur- und Quellenverzeichnis

Bücher

Auer, L. (2016): Ökonometrie – Eine Einführung, Springer Verlag, Berlin, Deutschland, 7. Auflage.

Wooldridge, J. (2012): Introductory Econometrics – A Modern Approach, South-Western, Cengage Learning, Mason, OH 45040USA, 5. Auflage.

Elektronische Quellen (Internet)

Ackermann, P. und Follert, F. (2018): Einige bewertungstheoretische Anmerkungen zur Marktwertanalyse der Plattform transfer-markt.de, https://zhb.tu-dortmund.de/wilkesmann/fussball_publi/Ackermann_Follert_Sciamus_2018.pdf, Abrufdatum: 15.01.2021 (leider nicht mehr verfügbar).

Catapultsports (2019): The Cost Of Injury, <https://www.catapultsports.com/the-cost-of-injury>, Abrufdatum: 11.03.2021.

Cullmann, A. (WS 2019): Ökonometrie Vorlesungsfolien 10, S. 3 ff. Abrufdatum: 10.03.2021

Fehr, R. (2018): Interview: So wird der Marktwert eines Fußballers berechnet, <https://www.watson.ch/sport/interview/996765299-so-wird-der-marktwert-eines-fussballers-berechnet>, Abrufdatum: 12.02.2021.

Fernsehgelder (2019): Ranking, <https://www.fernsehgelder.de/2018-19/ranking/>, Abrufdatum: 10.03.2021.

FIFA Transfer Regulations (2019): Regulations on the Status and Transfer of Players, Abschnitt 18.3, <https://resources.fifa.com/image/upload/regulations-on-the-status-and-transfer-of-players-june-2019.pdf?clouid=ao68trzk4bbaezlipx9u>, Abrufdatum: 12.03.2021.

Gerhards et al. (2014): Die Berechnung des Siegers: Marktwert, Ungleichheit, Diversität und Routine als Einflussfaktoren auf die Leistung professioneller Fußballteams, S. 234 ff., <https://www.polsoz.fu-berlin.de/soziologie/arbeitsbereiche/makrosoziologie/mitarbeiter/lehrstuhlinhaber/daten/Gerhards-et-al-2014-Die-Berechnung-des-Siegers.pdf>, Abrufdatum: 15.02.2021.

Manager Magazin (2018): Ronaldo und die 500.000 Trikots, <https://www.manager-magazin.de/unternehmen/handel/cristiano-ronaldo-juventus-turin-verkauft-in-24-h-520-000-trikots-a-1219296.html>, Abrufdatum: 15.02.2021.

Murphy, R. (2019): FIFA player ratings explained: How are the card number & stats decided?, <https://www.goal.com/en-ae/news/fifa-player-ratings-explained-how-are-the-card-number-stats/1hszd2fgr7wgfln2b2yjdpgynu>, Abrufdatum: 05.03.2021.

Premier League Squad List (2020): Premier League Squads confirmed, <https://www.premierleague.com/news/1869523>, Abrufdatum: 10.03.2021.

Rummenige, K.H (2017): The Neymar transfer triggers a chain reaction on the transfer market, <https://steemit.com/sports/@hoschitrooper/the-neymar-transfer-triggers-a-chain-reaction-on-the-transfer-market>, Abrufdatum: 10.03.2021.

Statistik am PC (2020): Heteroskedastizität in R erkennen (grafisch) - Daten analysieren in R (38), <https://www.youtube.com/watch?v=Di4mBCREQmk&t=174s>, Abrufdatum: 01.03.2021.

Süddeutsche Zeitung (2020): Wie die TV-Gelder in Spanien und England verteilt werden, <https://www.sueddeutsche.de/sport/bundesliga-tv-geld-verteilung-schlüssel-streit-1.4968073-2>, Abrufdatum: 10.03.2021.

Tagesschau (2020): Existenzkampf mit Ball, <https://www.tagesschau.de/wirtschaft/boerse/bundesliga-vereine-finanzen-corona-101.html>, Abrufdatum: 12.03.2021.

Transfermarkt (2021): Spielerprofil (Neymar, Kimmich und Kruse), <https://www.transfermarkt.de/neymar/profil/spieler/68290>, <https://www.transfermarkt.de/max-kruse/profil/spieler/36182>, <https://www.transfermarkt.de/joshua-kimmich/profil/spieler/161056>, Abrufdatum: 01.03.2021.

Williams, C. (unbekannt): Market value of football players: what is important?, <https://www.colossusbets.com/blog/market-value-of-football-players/#:~:text=Contract%20Duration&text=In%20other%20words%2C%20the%20market,the%20higher%20the%20transfer%20value>, Abrufdatum: 10.03.2021.

Zeit Online (2016): Chinesischer Investor übernimmt Inter Mailand,
<https://www.zeit.de/sport/2016-06/inter-mailand-suning-commerce-uebernahme-italien>, Abrufdatum: 11.03.2021.

Zeppenfeld, B. (2021): Ausgaben für internationale Transfers von Profifußballspielern weltweit in den Jahren 2012 bis 2020 (in Milliarden US-Dollar),
<https://de.statista.com/statistik/daten/studie/382052/umfrage/transfer-ausgaben-weltweit-profifussball/>, Abrufdatum: 15.02.2021