

Kursmaterial für Einführung in die Ökonometrie (Bachelor)

Rolf Tschernig & Harry Haupt

Universität Regensburg Universität Passau

August 2020 ¹

¹Dieses Kursmaterial basiert auf dem englischsprachigen "Intensive Course in Econometrics", der von Rolf Tschernig und Harry Haupt für das TEMPUS Project "New Curricula in Trade Theory and Econometrics" erarbeitet wurde. Florian Brezina hat die Übersetzung sowie das auf Deutschland ausgerichtete Außenhandelsbeispiel erstellt. Kathrin Kagerer, Joachim Schnurbus und Roland Jucknewitz, geb. Weigand, haben in wesentlichen Teilen das englischsprachige Kursmaterial verbessert und korrigiert. Patrick Kratzer hat unter Verwendung von Funktionen von Roland Jucknewitz das R-Programm für die empirischen Beispiele erstellt. Ihnen allen danken wir sehr herzlich. Eine englischsprachige und bzgl. der Folienseiten synchronisierte Version ist verfügbar unter "Introductory Econometrics — Slides — August 2020". Kommentare zu Änderungen und Korrekturen bzgl. früherer Versionen sind bei der Version vom Oktober 2019 zu finden. Wir bitten etwaige Fehler an rolf.tschernig@ur.de zu schicken.

© Die Folien dürfen für den individuellen Gebrauch und für Unterrichtszwecke, jedoch nicht für den kommerziellen Gebrauch gedruckt und reproduziert werden. Bitte zitieren als: Rolf Tschernig und Harry Haupt, Kursmaterial für Einführung in die Ökonometrie, Universität Regensburg, August 2020. Downloaded am [Tag Monat Jahr].

Inhaltsverzeichnis

1 Einführung: Was ist Ökonometrie?	4
1.1 Ein Außenhandelsbeispiel: Was bestimmt Handelsströme?	4
1.2 Ökonomische Modelle und der Bedarf für Ökonometrie	14
1.3 Kausalität und Experimente	22
1.4 Arten ökonomischer Daten	25
2 Das einfache Regressionsmodell	31
2.1 Das Regressionsmodell der Grundgesamtheit	32
2.2 Das Regressionsmodell der Stichprobe	47

2.3	Der KQ-Schätzer	50
2.4	Beste lineare Prognose, Korrelation und Kausalität	67
2.5	Algebraische Eigenschaften des KQ-Schätzers	74
2.6	Parameterinterpretation und funktionale Form	78
2.7	Statistische Eigenschaften: Erwartungswert und Varianz . .	90
2.8	Schätzung der Fehlervarianz	97
3	Das multiple lineare Regressionsmodell: Schätzung	100
3.1	Motivation: Fortführung des Außenhandelsbeispiels	100
3.2	Das multiple Regressionsmodell der Grundgesamtheit	105
3.3	Der KQ-Schätzer: Ableitung & algebraische Eigenschaften	119
3.4	Der KQ-Schätzer: Statistische Eigenschaften	132
3.5	Modellspezifikation I: Modellselektionskriterien	166

4 Multiple Regression: Hypothesentest	178
4.1 Grundlagen statistischer Tests	178
4.2 Wahrscheinlichkeitsverteilung des KQ-Schätzers	209
4.3 t -Tests im multiplen Regressionsmodell	216
4.4 Emp. Analyse einer vereinfachten Gravitationsgleichung ..	224
4.5 Konfidenzintervalle/Intervallschätzung	235
4.6 Testen einer <i>einzelnen</i> Linearkombination der Parameter ..	247
4.7 F -Test	253
4.8 Darstellung von Regressionsergebnissen	279
5 Multiples Regressionsmodell: Asymptotik	282
5.1 Verteilung des Mittelwertschätzers in großen Stichproben ..	283
5.2 Verteilung des KQ-Schätzers in großen Stichproben	298

6	Multiple Regression: Interpretation	303
6.1	Level- und Log-Modelle	303
6.2	Datenskalierung	304
6.3	Besonderheiten bei transformierten Regressoren	311
6.4	Qualitative Daten als Regressoren	322
7	Multiple Regression: Prognosen	340
7.1	Prognose und Prognosefehler	340
7.2	Statistische Eigenschaften linearer Prognosen	347
8	Multiple Regression: Heteroskedastie	348
8.1	Auswirkungen der Heteroskedastie auf die KQ-Schätzer	351
8.2	Heteroskedastie-robuste Standardfehler bei KQ-Schätzung	354
8.3	Der verallgemeinerte KQ-Schätzer (GLS-Schätzer)	357
8.4	Feasible Generalized Least Squares (FGLS)	366

9	Multiple Regression: Modelldiagnose	388
9.1	RESET-Test	388
9.2	Tests auf Heteroskedastie	391
9.3	Modellspezifikation II: Nützliche Tests	410
10	Anhang	I
10.1	Eine kurze Einführung in die Wahrscheinlichkeitstheorie	I
10.2	Wichtige Regeln der Matrix-Algebra	XXIII
10.3	Regeln zur Ableitung von Vektoren und Matrizen	XXX
10.4	Daten für die Schätzung der Gravitationsgleichung	XXXII
10.5	R-Programm für die empirischen Beispiele	XXXVIII

Organisation

Kontakt

Prof. Dr. Rolf Tschernig

Gebäude RW(L), 5. Stock, Raum 514

Universitätsstr. 31, 93040 Regensburg

Tel. (+49) 941/943 2737, Fax (+49) 941/943 4917

Email: rolf.tschernig@wiwi.uni-regensburg.de

<https://www.uni-regensburg.de/wirtschaftswissenschaften/vwl-tschernig/>

index.html

Zeiten, Räume und Kursleiter

siehe jeweilige Kurshomepage

<https://www.uni-regensburg.de/wirtschaftswissenschaften/vwl-tschnig/lehre/bachelor/einfuehrung-in-die-oekonometrie/index.html>

Prüfung

siehe jeweilige Kurshomepage

<https://www.uni-regensburg.de/wirtschaftswissenschaften/vwl-tschnig/lehre/bachelor/einfuehrung-in-die-oekonometrie/index.html>

Pflichtlektüre

Wooldridge, J.M. (2009). *Introductory Econometrics. A Modern Approach*, 4th ed., Thomson South-Western. Oder neuere Auflage.

Weiterführende Literatur

Stock, J.H. and Watson, M.W. (2007). *Introduction to Econometrics*, 2nd ed., Pearson, Addison-Wesley. (Oder neuere Auflage)

Software

Alle empirischen Beispiele werden in R (<https://www.r-project.org>) durchgeführt. Appendix 10.5 enthält alle R-Programme.

1 Einführung: Was ist Ökonometrie?

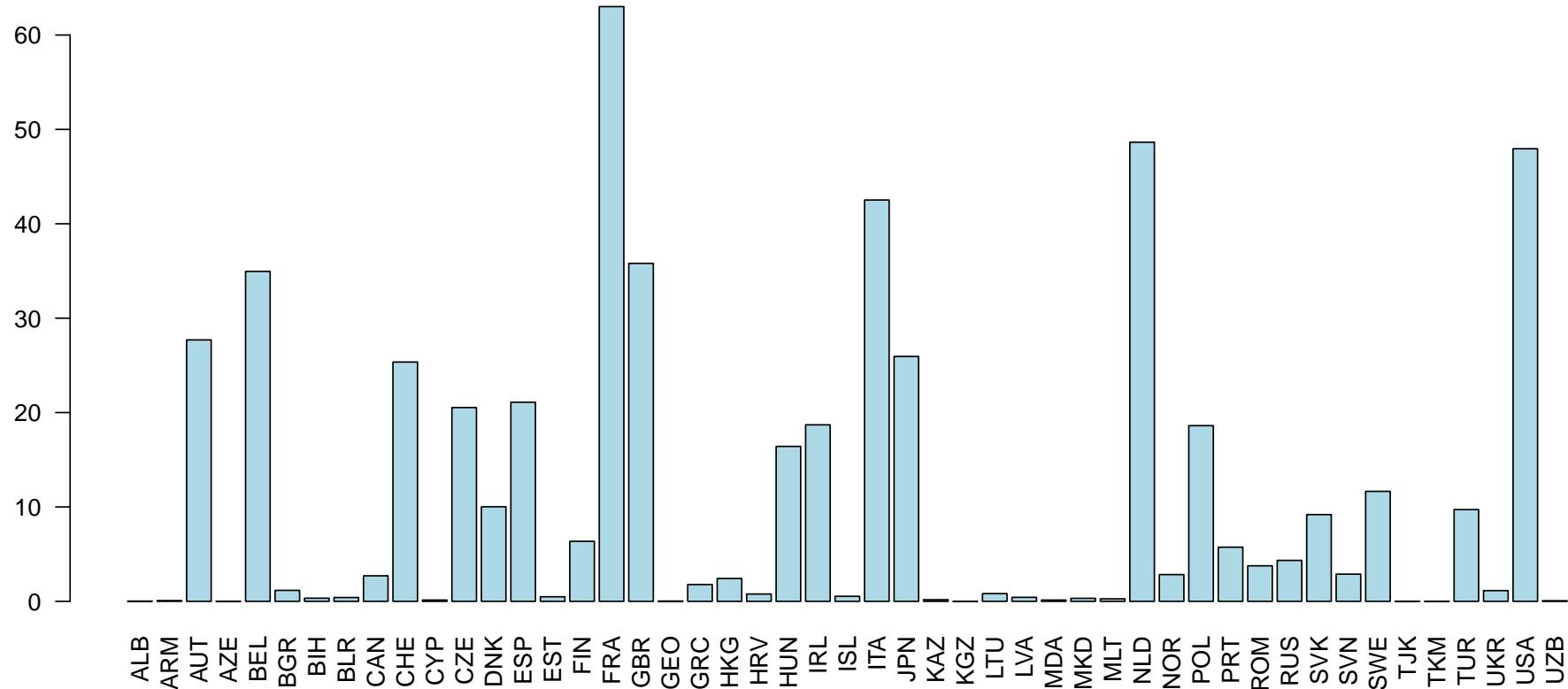
1.1 Ein Außenhandelsbeispiel: Was bestimmt Handelsströme?

Ziel/Wissenschaftliche Fragestellung: Ermittle die Faktoren, die die Importe nach Deutschland beeinflussen, und quantifiziere ihren Einfluss.

- Drei **Grundfragen**, die im Laufe einer Analyse beantwortet werden müssen:
 1. Welche (ökonomischen) Zusammenhänge sind (möglicherweise) relevant für die Fragestellung?
 2. Welche Daten könnten nützlich sein, um diese möglicherweise relevanten ökonomischen Vermutungen/Theorien zu überprüfen?
 3. Wie entscheidet man darüber, welche ökonomische Hypothese abzulehnen oder weiterhin zu verfolgen ist?
- Werfen wir erst einen Blick auf folgende interessante Daten: Die Importe nach Deutschland aus 54 Herkunftsländern im Jahr 2004 (in laufenden US-Dollars).

Importe nach Deutschland in 2004 in Mrd. US-Dollar

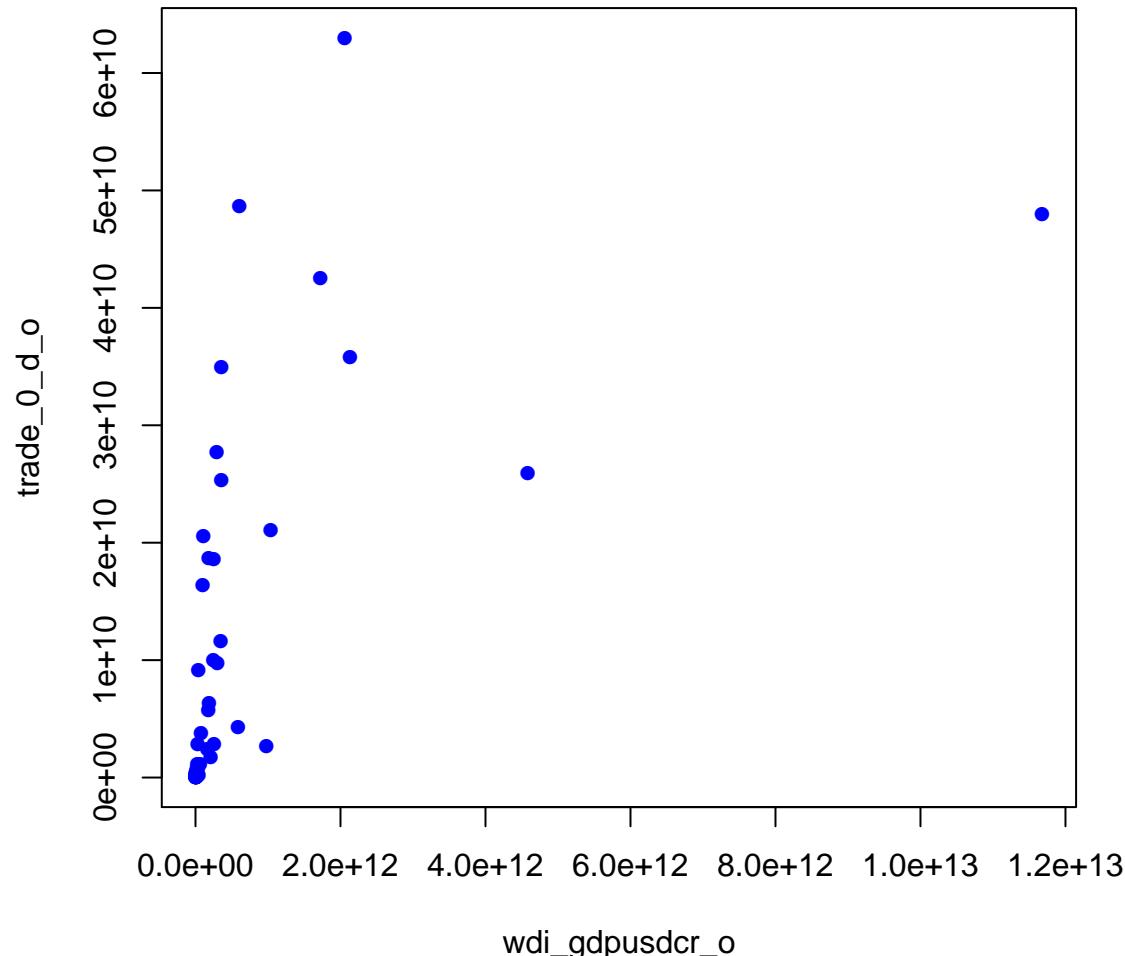
Imports to Germany in 2004 in Billions of US-Dollars



Die Originaldaten stammen von **UN Commodity Trade Statistics Database (UN COMTRADE)**

- Siehe Abschnitt 10.4 im Anhang für eine detailliertere Beschreibung der Daten. Die Daten finden Sie in der Text-Datei `importe_ger_2004.txt`. Wir danken Richard Frensch, Osteuropa-Institut, Regensburg, der alle in diesem Kurs verwendeten Daten zur Analyse von Handelsströmen bereitgestellt hat.
- Ein erster Versuch, die drei Grundfragen zu beantworten:
 1. Ignorieren wir fürs Erste die gesamte ökonomische Theorie und stellen die einfache Hypothese auf, dass die beobachteten Importe irgendwie mit dem BIP des Exportlandes zusammenhängen.
 2. Sammle BIP-Daten für die Herkunftsländer, z.B. vom **Internationalen Währungsfonds (IWF) – World Economic Outlook Database**
 3. Plotte die Daten, z.B. mittels eines Scatterplot.
Kann man entscheiden, ob es **einen Zusammenhang zwischen den Handelsströmen mit einem Exportland und dessen BIP gibt?**

Scatterplot (Streudiagramm)



Einige Fragen:

- Was sieht man?
- Gibt es einen Zusammenhang?
- Wenn ja, wie ist dieser zu quantifizieren?
- Existiert eine Kausalbeziehung - Welche Variable bestimmt welche?
- Wie verändern sich die Importe aus den USA, wenn sich das BIP der USA um 1% verändert?

- Gibt es andere relevante Faktoren, die die Importe bestimmen, z.B. die Entfernung?
- Ist es möglich, zukünftige Handelsströme zu prognostizieren?
- Was haben wir gemacht?
 - Wir haben versucht, die Realität zu vereinfachen,
 - indem wir eine Art ökonomisches Modell aufgestellt haben.

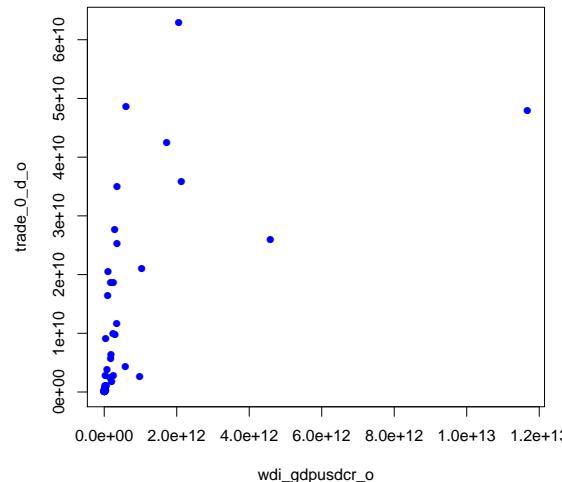
- **Ein (ökonomisches) Modell**

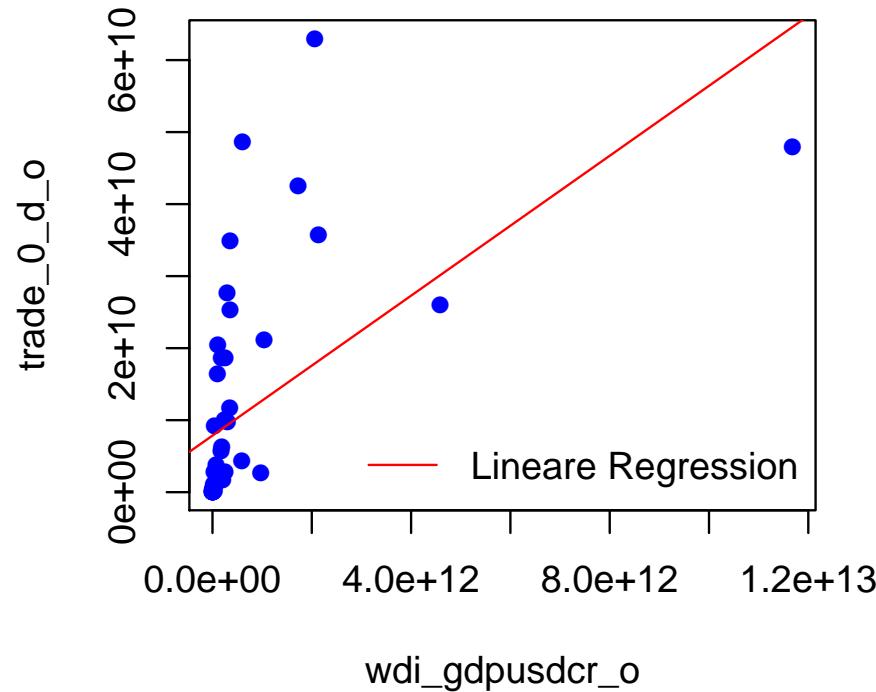
- soll die Komplexität der Realität reduzieren, sodass es genutzt werden kann, um die zu Anfangs formulierten Fragen zu beantworten;
- ist eine Kollektion clever gewählter Annahmen, aus denen (logische) Schlussfolgerungen gezogen werden können — Beispiel: Heckscher-Ohlin Modell;
- sollte so einfach wie möglich und so komplex wie nötig sein;
- kann ohne empirische Daten jedweder Art weder widerlegt noch “bestätigt” werden.

- Nehmen wir nun für den Zusammenhang zwischen Importen und dem BIP der Herkunftsländer ein einfaches formales Modell an:

$$Importe_i = \beta_0 + \beta_1 BIP_i, \quad i = 1, \dots, 49.$$

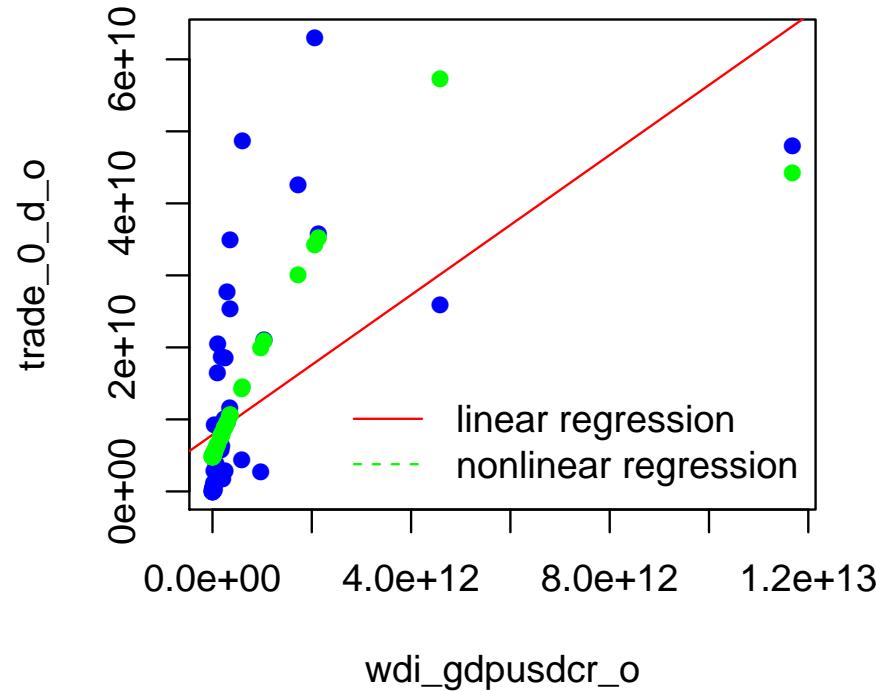
- Ist das sinnvoll?
- Wie bestimmt man die Werte der sogenannten Parameter β_0 und β_1 ?
- Legen wir eine Gerade durch die Punktewolke!





Weitere Fragen:

- Wie legen wir die Gerade durch die Punktewolke?
- Welche Eigenschaften hat die so angepasste Gerade?
- Was macht man mit den anderen relevanten Faktoren, die in der aktuellen Analyse vernachlässigt wurden?
- Welche Kriterien wählt man, um einen möglichen Zusammenhang zu ermitteln?



Noch mehr Fragen:

- Ist der mögliche Zusammenhang tatsächlich linear? Vergleichen Sie diesen mit den grünen Punkten eines nicht-linearen Zusammenhangs.
- Und: wie sehr dürfen die Ergebnisse für eine andere Stichprobe abweichen, z.B. für 2003?

1.2 Ökonomische Modelle und der Bedarf für Ökonometrie

- **Standardprobleme ökonomischer Modelle:**

- Das vermutete ökonomische Modell vernachlässigt vermutlich einige Faktoren.
- *Quantitative Ergebnisse* der Fragestellungen hängen i.d.R. von der Wahl des Datensatzes ab. Ein anderer Datensatz führt zu anderen numerischen Ergebnissen.

⇒ Quantitative Antworten beinhalten also immer **Unsicherheit**.

• Ökonometrie

- bietet Lösungen an, mit unbeobachteten Faktoren in ökonomischen Modellen umzugehen,
- bietet “both a numerical answer to the question and a measure how precise the answer is ([Stock und Watson, 2007](#), p. 7)”,
- bietet, wie wir später sehen werden, Werkzeuge zur Widerlegung ökonomischer Hypothesen an, indem mittels statistischer Methoden Theorien mit empirisch erhobenen Daten konfrontiert werden, *und* bietet Werkzeuge zur Quantifizierung der Wahrscheinlichkeiten an, mit denen solche Entscheidungen falsch sind,
- erlaubt, wie wir ebenfalls später sehen werden, die Quantifizierung der **Risiken** von Vorhersagen, Entscheidungen und sogar ihrer eigenen Analyse.

- Deshalb:

Ökonometrie kann auch nützlich sein, um Antworten auf Fragen zu geben wie beispielsweise:

- Wie zuverlässig sind prognostizierte Wachstumsraten oder Gewinne?
- Wie Wahrscheinlich ist es, dass der in der Zukunft tatsächlich realisierte Wert nahe des prognostizierten Wertes liegt? Mit anderen Worten, wie genau sind die Prognosen?

- **Hauptwerkzeug:** Das multiple Regressionsmodell

Es ermöglicht die Quantifizierung des Effekts einer Veränderung einer Variable auf eine andere Variable, während alle anderen Faktoren konstant gehalten werden (**ceteris paribus Analyse**).

- **Schritte einer ökonometrischen Analyse:**

1. Sorgfältige Formulierung der interessierenden Fragestellung/Aufgabe bzw. des Problems.
2. Spezifizierung eines ökonomischen Modells.
3. Sorgfältige Auswahl einer Klasse ökonometrischer Modelle.
4. Sammeln von Daten.
5. Auswahl und Schätzung eines ökonometrischen Modells.
6. Prüfen, ob Modellspezifikation korrekt.
7. Anwenden des Modells d.h. Interpretation und Prognose.

Beachte: Es existiert eine Vielzahl unterschiedlicher ökonometrischer Modelle und die Modellwahl hängt sehr stark ab von der wissenschaftlichen Fragestellung, der zugrunde liegenden ökonomischen

Theorie, der Verfügbarkeit von Daten und der Problemstruktur.

- **Ziele** dieses Kurses:

Ihnen grundlegende ökonometrische Werkzeuge an die Hand zu geben, damit Sie

- erfolgreich einfache empirisch-ökonometrische Analysen durchführen und quantitative Antworten auf quantitative Fragen geben können,
- fehlerhaft durchgeführte ökonometrische Studien und deren Konsequenzen erkennen können,
- erkennen, wann Sie einen erfahrenen Ökonometrie-Experten zu Rate ziehen sollten,
- an fortgeschrittenen empirischen (Ökonometrie-) Veranstaltungen teilnehmen können,
- sich anspruchsvollere ökonometrische Methoden aneignen können.

Einige Definitionen von Ökonometrie

- “... discover empirical relation between economic variables, provide forecast of various economic quantities of interest ... (First issue of volume 1, *Econometrica*, 1933).”
- “The science of model building consists of a set of quantitative tools which are used to construct and then test *mathematical representations of the real world*. The development and use of these tools are subsumed under the subject heading of econometrics ([Pindyck und Rubinfeld, 1998](#)).”

- “At a broad level, econometrics is the science and art of using economic theory and statistical techniques to analyze economic data. Econometric methods are used in many branches of economics, including finance, labor economics, macroeconomics, microeconomics, marketing, and economic policy. Econometric methods are also commonly used in other social sciences, including political science and sociology ([Stock und Watson, 2007](#), p. 3).”

Also, einige würden auch sagen: “Alchemy or Science?”, “Economic-tricks”, “Econo-mystiques”.

- “Econometrics is based upon the development of statistical methods for estimating economic relationships, testing economic theories, and evaluating and implementing government and business policy ([Wooldridge, 2009](#), p. 1).”

- **Überblick über Aufgaben ökonometrischer Methoden**
 - **kurz:** Ökonometrie kann nützlich sein, wenn man auf (ökonomische) Daten stößt und Zusammenhänge verstehen möchte.
 - **ausführlicher:**
 - * Sie bietet einen **formalen Rahmen** zur **Falsifizierung postulierter ökonomischer Zusammenhänge**, indem sie Wirtschaftstheorie mit Daten mittels statistischer Methoden konfrontiert: Ökonomische Hypothesen werden formuliert und auf der Basis passend (und wiederholt) gesammelter Daten getestet, sodass Testresultate die Hypothesen falsifizieren können.
 - * **Analyse der Auswirkungen von Wirtschaftspolitik.**
 - * **Prognosen.**

1.3 Kausalität und Experimente

- Gängiges Verständnis: “causality means that a specific action” (berühren einer heißen Herdplatte) “leads to a specific, measurable consequence” (sich verbrennen) (Stock und Watson, 2007, p. 8).
- Wie ermittelt man Kausalität? Beobachte **wiederholt** eine Handlung und ihre Konsequenzen! Allerdings lässt diese Vorgehensweise nur Rückschlüsse auf eine durchschnittliche Kausalität zu, da man für jede einzelne Handlung nicht gleichzeitig verschiedene Ergebnisse (Hand verbrannt, nicht verbrannt) beobachten kann.
- Man ist in der Wissenschaft also darauf aus, Handlungen und ihre Konsequenzen unter **identischen** Bedingungen zu wiederholen. Doch wie erzeugt man mehrere Wiederholungen einer Handlung?

- **Kontrollierte Zufallsexperimente:**

- es gibt eine **Kontrollgruppe**, die keine Behandlung erhält (z.B. Düngemittel) und eine **behandelte Gruppe**,
- wobei die Zuordnung zu den Gruppen **zufällig** erfolgt, um jeden möglichen systematischen Zusammenhang zwischen der Behandlung und anderen möglichen Einflüssen auszuschließen.

- **Kausaleffekt:**

Ein “**causal effect** is defined to be an effect on an outcome of a given action or treatment, as measured in an ideal randomized controlled experiment ([Stock und Watson, 2007, p. 9](#)).”

- In der Ökonomie sind kontrollierte Zufallsexperimente oft nur sehr schwer oder gar nicht durchführbar. Ein kontrolliertes Zufallsexperiment dient dann als theoretische Messlatte und die ökonometrische

Analyse hat das Ziel, so nahe wie möglich die Bedingungen eines kontrollierten Zufallsexperiments mit tatsächlich vorhandenen Daten nachzuahmen.

- Beachte: für **Prognosen** ist es nicht notwendig über Kausalitäten Bescheid zu wissen.
- Jedoch: i.A. lassen multiple Regressionsmodelle keine Schlüsse auf die Kausalitätsstruktur zu!
- Eine hervorragende, leicht lesbare Einführung in aktuelle Methoden der Kausalanalyse und ihrer Grenzen bietet Kapitel 4 in **Bauer, Fertig, und Schmidt (2009)** oder auch noch sehr unterhaltsam **Angrist und Pischke (2015)**.

1.4 Arten ökonomischer Daten

1. Querschnittsdaten (dieser Kurs)

- werden zu einem bestimmten Zeitpunkt von individuellen Einheiten erhoben.
- Einheiten: “economic agents”, z.B. Individuen, Haushalte, Investoren, Firmen, Branchen, Städte, Länder.
- I.A.: die Reihenfolge der Beobachtungen hat keine Bedeutung.
- Zum Indizieren wird meist der Index i verwendet.
- Optimal: Die Daten sind eine *Zufallsstichprobe* der zugrunde liegenden *Grundgesamtheit*, siehe Abschnitt 2.1.
- Querschnittsdaten erlauben es, Unterschiede zwischen den individuellen Einheiten zu erklären.

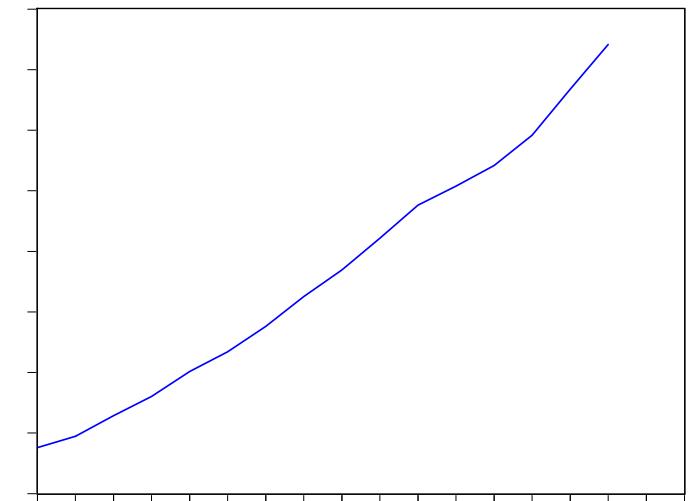
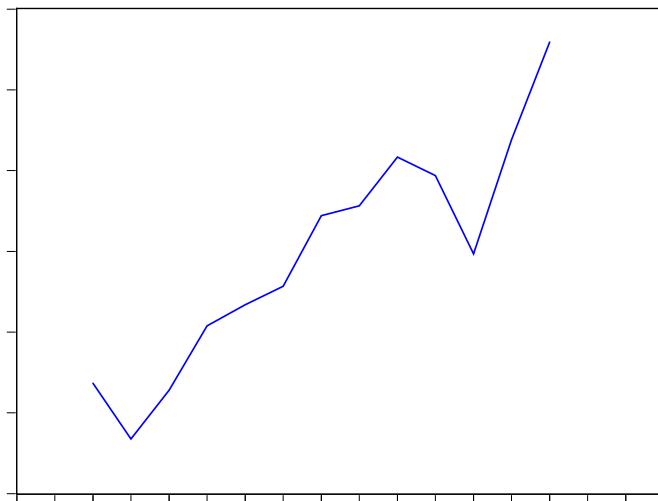
- Beispiel: Stichprobe der nach Deutschland exportierenden Länder des Jahres 2004 aus Abschnitt 1.1.

2. Zeitreihendaten (BA Zeitreihenökonometrie, Quantitative Wirtschaftsforschungs I; MA: Methoden der Ökonometrie, Applied Financial Econometrics, Quantitative Wirtschaftsforschung II)

- sequentielle Erhebung über verschiedene Zeitpunkte/-perioden.
- Zum Indizieren wird meist der Index t verwendet.
- Bei der Beobachtungsfrequenz ist zu beachten:
 - variabel vs. fest;
 - fest: jährlich, vierteljährlich, monatlich, wöchentlich, täglich;
 - variabel: Tickerdaten, Verweildauerdaten (z.B. Arbeitslosigkeit).

- Zeitreihendaten erlauben die Analyse dynamischer Effekte.
- Univariate versus multivariate Zeitreihendaten.

Beispiel: Handelsstrom von den USA nach Deutschland und das BIP der USA (in laufenden US-Dollars), 1990 - 2007, $T = 18$.



3. Paneldaten (BA: Weiterführende Fragen der Ökonometrie)

- eine Kollektion von Querschnittsdaten, die an mindestens zwei verschiedenen Zeitpunkten erhoben wurden.
- die individuellen Einheiten bleiben über die Zeit hinweg (in jeder Querschnittserhebung) identisch — außer durch Ausscheiden.
- Es wird ein doppelter Index benutzt: it wobei $i = 1, \dots, N$ und $t = 1, \dots, T$.
- Typisches Problem: fehlende Werte - für einige Zeitpunkte und/oder Einheiten sind keine Daten verfügbar.
- Beispiel: Wachstumsrate der Importe aus 54 Ländern nach Deutschland zwischen 1991 und 2008. Alle 54 Länder wurden für die Stichprobe des Jahres 1991 ausgewählt und sind für alle folgenden Jahre **gleich geblieben** ($T = 18, N = 54$).

4. Gepoolte Querschnittsdaten (BA: Weiterführende Fragen der Ökonometrie)

- auch eine Kollektion von Querschnittsdaten, die an verschiedenen Zeitpunkten erhoben wurden, jedoch mit sich möglicherweise verändernden individuellen Einheiten.
- Beispiel: 1995 sind die Herkunftsländer die Niederlande, Frankreich und Russland, 1996 sind es Polen, die USA und Italien.

Im Rahmen dieser Veranstaltung beschränken wir uns auf Querschnittsdaten:

- **einfaches Regressionsmodell** → Kapitel 2,
- **multiples Regressionsmodell** → Kapitel 3 bis 9.
- Zeitreihenanalyse erfordert fortgeschrittene ökonometrische Techniken, die über den Rahmen dieser Veranstaltung hinausgehen (bei gegebener zeitlicher Beschränkung).

Von Bedeutung ist noch die **arithmetische Qualität der Daten** (vgl. Statistik I):

- quantitative Variablen,
- qualitative (oder kategoriale) Variablen.

Zu Lesen: Abschnitte 1.1-1.3 in **Wooldridge (2009)**.

2 Das einfache Regressionsmodell

Unterschiede zwischen

- dem Regressionsmodell der Grundgesamtheit und
- dem Regressionsmodell der Stichprobe.

2.1 Das Regressionsmodell der Grundgesamtheit

- **Allgemein:**

y und x sind zwei Variablen, die Eigenschaften der Grundgesamtheit beschreiben. Für diese Grundgesamtheit möchte man “ y mittels x erklären” oder “erklären, wie y variiert, wenn x variiert” oder “für vorliegende Werte von x y prognostizieren”.

Beispiel: Wie verändert sich der Stundenlohn, wenn die Ausbildungsdauer um ein zusätzliches Jahr erhöht wird und alle anderen Einflussfaktoren konstant gehalten werden?

- **Wären wir allwissend**, ließe sich der Zusammenhang zwischen y und x formal wie folgt ausdrücken

$$y = m(x, z^1, \dots, z^s) \quad (2.1)$$

wobei z^1, \dots, z^s die s zusätzlichen Variablen bezeichnen, die neben

der Ausbildungsdauer x Einfluss auf den Stundenlohn y haben.

- **In der praktischen Arbeit** ist es aber möglich, dass
 - der Zusammenhang (2.1) zu kompliziert ist, um nützlich zu sein,
 - kein exakter Zusammenhang besteht,
 - zwar ein exakter Zusammenhang besteht, allerdings nicht alle s Einflussvariablen z^1, \dots, z^s beobachtbar sind,
 - man keine Ahnung über die funktionale Form von $m(\cdot)$ hat.
- **Unsere Lösung:**
 - **Aufstellen eines nützlichen Modells**, siehe Abschnitt 1.1,
 - für das eine Beziehung **im “Durchschnitt”** gilt. Was ist mit “im Durchschnitt” gemeint?
- **Wichtige Grundbausteine für unser Modell:**

– **Fassen Sie die Variable y als zufällig auf.** Man kann sich y etwa als den Wert der Variable vorstellen, der zufällig aus der Grundgesamtheit gezogen wird. Weiterhin kann für den Fall, dass die **Zufallsvariable** y diskrete Werte annimmt, jedem Wert von y eine bestimmte Wahrscheinlichkeit zugewiesen werden. (Ist die Zufallsvariable y stetig, kann eine Dichte zugeordnet werden.)

Mit anderen Worten: Man gebraucht **Wahrscheinlichkeitstheorie**. Siehe Appendices B und C in [Wooldridge \(2009\)](#).

Beispiele:

- * Die Grundgesamtheit besteht aus allen Wohnungen in Regensburg. Die Variable y bezeichnet die Miete für eine einzelne Wohnung, die zufällig aus der Menge aller Wohnungen in Regensburg ausgewählt wurde.

- * Die Grundgesamtheit besteht aus allen möglichen Werten für die Exporte eines bestimmten Landes nach Deutschland in einer bestimmten Periode.
- * Bei einem Würfel besteht die Grundgesamtheit aus allen Zahlen, die auf den Seiten des Würfels aufgemalt sind. Allerdings bevorzugen Statistiker in diesem Fall von sample space zu sprechen.
- In der Wahrscheinlichkeitstheorie bezeichnet der “Durchschnitt” einer Variablen y den **Erwartungswert** dieser Variable. Im Falle diskreter y erhält man

$$E[y] = \sum_{j \in \text{aller möglichen unterschiedlichen } y_j \text{ in der Grundgesamtheit}} y_j \text{Prob} (y = y_j)$$

- Manchmal möchte man aber nur eine **Teilmenge der Grundgesamtheit** betrachten, etwa alle y , für die eine andere Variable

x den selben Wert annimmt.

Beispiel: Man betrachtet nur die Mieten (y) der Wohnungen in Regensburg deren Wohnfläche $x = 75 \text{ m}^2$ beträgt.

- Ist der “Durchschnitt” bedingt auf bestimmte Werte einer anderen Variablen x , dann betrachtet man den **bedingten Erwartungswert** von y gegeben x : $E[y|x]$. Für diskrete Variablen y erhält man

$$E[y|x] = \sum_{j \in \text{aller möglichen unterschiedlichen } y_j \text{ in der Grundgesamtheit}} y_j \text{Prob} (y = y_j|x)$$

(Siehe Anhang 10.1 für eine kurze Einführung in die Wahrscheinlichkeitstheorie und die entsprechenden Definitionen für stetige Zufallsvariablen.)

Fortsetzung des Beispiels: Der bedingte Erwartungswert $E[y|x = 75]$ entspricht der durchschnittlichen Miete aller Woh-

nungen in Regensburg mit einer Wohnfläche von $x = 75 \text{ m}^2$.

- NB: Die Variable x kann ebenfalls zufällig sein. In diesem Fall ist der bedingte Erwartungswert $E[y|x]$ eine Funktion der Zufallsvariablen x

$$E[y|x] = g(x)$$

und damit selbst eine Zufallsvariable.

- Über die **Gleichung**

$$y = E[y|x] + (y - E[y|x]) \quad (2.2)$$

lässt sich der **Fehlerterm** oder **Störterm** definieren als

$$u \equiv y - E[y|x],$$

so dass man ein **einfaches Regressionsmodell der Grundgesamtheit** erhält

$$y = E[y|x] + u. \quad (2.3)$$

- **Interpretation:**

- Die Zufallsvariable y schwankt zufällig um den bedingten Erwartungswert $E[y|x]$:

$$y = E[y|x] + u.$$

- Der bedingte Erwartungswert $E[y|x]$ wird als der **systematische Teil** der Regression bezeichnet.
- Der Fehlerterm u wird als **unsystematischer Teil** der Regression bezeichnet.

- Statt das Unmögliche zu versuchen, etwa die Spezifizierung von $m(x, \dots)$ aus (2.1), beschränkt man die Untersuchung auf den “Durchschnitt” $E[y|x]$.

• Wie bestimmt man den bedingten Erwartungswert?

- Dieser Schritt erfordert Annahmen!
- Zur Vereinfachung treffen wir folgende **Annahme (A)**

$$E[y|x] = \beta_0 + \beta_1 x. \quad (2.4)$$

- **Erläuterung** der Annahme (A):
 - * Sie **beschränkt die Flexibilität von** $g(x) = E[y|x]$, so dass $g(x) = \beta_0 + \beta_1 x$ linear in x sein muss. Wenn jedoch $E[y|x] = \delta_0 + \delta_1 \log x$ gilt, ist Annahme (A) falsch.
 - * Sie kann erfüllt sein, wenn noch weitere Variablen y linear beeinflussen. Nehmen wir beispielsweise an, dass

$$E[y|x, z] = \delta_0 + \delta_1 x + \delta_2 z.$$

Wir erhalten dann über den Satz der iterierten Erwartungswerte

(LIE = law of iterated expectations), dass

$$E[y|x] = \delta_0 + \delta_1 x + \delta_2 E[z|x].$$

Ist $E[z|x]$ linear in x , erhält man

$$\begin{aligned} E[y|x] &= \delta_0 + \delta_1 x + \delta_2(\alpha_0 + \alpha_1 x) \\ &= \gamma_0 + \gamma_1 x \end{aligned} \tag{2.5}$$

mit $\gamma_0 = \delta_0 + \delta_2 \alpha_0$ und $\gamma_1 = \delta_1 + \delta_2 \alpha_1$. Beachte: I.A. gilt dann $E[y|x, z] \neq E[y|x]$. Die Modellwahl hängt in diesem Fall vom Ziel der Untersuchung ab: Das kleinere Modell kann für Voraussagen bevorzugt werden, das größere Modell wird benötigt, wenn die Berücksichtigung von z wichtig ist \Leftrightarrow kontrollierte Zufallsexperimente, siehe Abschnitt 1.3.

- * Gilt (2.5) nicht, ist **Annahme (A)** i.A. nicht gültig, wenn z.B. $E[z|x]$ nicht-linear in x ist. Das lineare Modell der Grundge-

samtheit bezeichnet man dann als **fehlspezifiziert**. Mehr dazu in Abschnitt **3.4**.

- **Eigenschaften des Fehlerterms u :** Aus **Annahme (A)** ergibt sich:
 1. $E[u|x] = 0$,
 2. $E[u] = 0$,
 3. $Cov(x, u) = 0$.

- Eine **alternative Annahmenkombination**:

Zusammen mit der Identität in (2.3) und dem obigen Resultat $E[u|x] = 0$ ist **Annahme (A)** äquivalent zu folgenden zwei Annahmen:

1. **Annahme SLR.1**

(Linearität in den Parametern)

$$y = \beta_0 + \beta_1 x + u, \quad (2.6)$$

2. **Annahme SLR.4**

(Bedingter Erwartungswert gleich Null)

$$E[u|x] = 0.$$

- **Lineares Regressionsmodell der Grundgesamtheit:**

Das einfache lineare Regressionsmodell der Grundgesamtheit ist gegeben durch Gleichung (2.6)

$$y = \beta_0 + \beta_1 x + u$$

und wird gewonnen, indem der bedingte Erwartungswert des Regressionsmodells (2.3) durch eine lineare Funktion (linear in den Parametern) spezifiziert wird.

Die Parameter β_0 und β_1 werden als **Achsenabschnitt/Konstante** bzw. **Steigungsparameter** bezeichnet.

• Einige Begriffsbestimmungen für Regressionen

y	x
Abhängige Variable	Unabhängige Variable
Erklärte Variable	Erklärende Variable
Antwortvariable	Kontrollvariable
Prognosevariable	Wirkungsvariable/Prediktorvariable
Regressand	Regressor
	Covariate

- Einfaches Beispiel: Würfelspiel

Es bezeichnen die Zufallszahlen x und u die Würfe von zwei fairen Würfeln, wobei $x, u = \{-2.5, -1.5, -0.5, 0.5, 1.5, 2.5\}$. Die Zufallszahl y ist auf Basis der beiden Würfe folgende Summe

$$y = \underbrace{2}_{\beta_0} + \underbrace{3}_{\beta_1} x + u.$$

Damit ist die Grundgesamtheit vollständig beschrieben!

- Bestimmen Sie den systematischen Zusammenhang zwischen x und y gegeben x .
- Interpretieren Sie den systematischen Zusammenhang.
- Wie erhält man $\beta_0 = 2$ und $\beta_1 = 3$, wenn diese Parameter nicht bekannt sind?

Im nächsten Abschnitt: Wie lassen sich β_0 und β_1 bestimmen?

2.2 Das Regressionsmodell der Stichprobe

Schätzer (estimators) und Schätzungen (estimates)

- In der Praxis müssen die unbekannten Parameter des Regressionsmodells der Grundgesamtheit β_0 und β_1 mit Hilfe einer Stichprobe geschätzt werden.
- Die Stichprobe muss **repräsentativ** sein und wird aus der Grundgesamtheit gezogen.
- Eine Stichprobe der Größe n der Zufallsvariablen x und y ist gegeben durch $\{(x_i, y_i) : i = 1, \dots, n\}$.
- Wir benötigen nun einen **Schätzer**, der es, gegeben die Stichprobenbeobachtungen $\{(x_i, y_i) : i = 1, \dots, n\}$, erlaubt, **Schätzungen** für die Parameter β_0 und β_1 der Grundgesamtheit zu berechnen.

- **Beachte:**
 - wenn wir einen Schätzer für die unbekannten Parameterwerte konstruieren wollen, dann liegen keine spezifischen Stichprobenwerte vor. Ein **Schätzer** ist eine Funktion, der die Stichprobenwerte als Argumente enthält.
 - Sobald wir einen Schätzer vorliegen *und* eine Stichprobe erhoben haben, können wir **Schätzungen** (= numerische Werte) für die unbekannten Parameter berechnen.
 - Für die Schätzung der unbekannten Parameter existieren häufig viele verschiedene Schätzer, die sich im Allgemeinen in ihren statistischen Eigenschaften (ihrer Qualität) unterscheiden!
- Beispiel:** Zwei verschiedene Schätzer für den Mittelwert: $\frac{1}{n} \sum_{i=1}^n y_i$ und $\frac{1}{2} (y_1 + y_n)$.

- Kennzeichnet man im Regressionsmodell der Grundgesamtheit

$$y = \beta_0 + \beta_1 x + u$$

die Schätzungen von β_0 und β_1 mit Tilden, also $\tilde{\beta}_0$ und $\tilde{\beta}_1$, dann ist das Stichprobenregressionsmodell gegeben durch

$$y_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_i + \tilde{u}_i, \quad i = 1, \dots, n,$$

mit folgenden Bezeichnungen:

- **Stichprobenregressionsfunktion (sample regression function)** oder **Regressionsgerade**

$$\tilde{y}_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_i$$

- **Schätzwerte/angepasste Werte (fitted values)** \tilde{y}_i
- **Residuen** $\tilde{u}_i = y_i - \tilde{y}_i, i = 1, \dots, n.$

Mit welcher Methode sollen wir schätzen?

2.3 Kleinst-Quadrat-Schätzer (OLS)

- Der Kleinst-Quadrat-Schätzer (KQ-Schätzer) wird häufig mit OLS (= ordinary least squares) abgekürzt. Der KQ-Schätzer geht auf C.F. Gauss (1777-1855) zurück.
- Man erhält ihn, indem man $\tilde{\beta}_0$ und $\tilde{\beta}_1$ so wählt, dass die **Summe der quadrierten Residuen**

$$\sum_{i=1}^n \tilde{u}_i^2 = \sum_{i=1}^n (y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_i)^2$$

minimal wird.

- Hierzu bildet man die ersten partiellen Ableitungen bezüglich $\hat{\beta}_0$ und $\hat{\beta}_1$ und setzt diese gleich Null:

$$\sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0, \quad (2.7)$$

$$\sum_{i=1}^n x_i \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i \right) = 0. \quad (2.8)$$

Die Gleichungen (2.7) und (2.8) werden als **Normalgleichungen** bezeichnet.

Es ist wichtig, sich die Interpretation der Normalgleichungen klar zu machen.

Aus (2.7) erhält man

$$\begin{aligned}\hat{\beta}_0 &= n^{-1} \sum_{i=1}^n y_i - \hat{\beta}_1 n^{-1} \sum_{i=1}^n x_i, \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x},\end{aligned}\tag{2.9}$$

wobei $\bar{z} = n^{-1} \sum_{i=1}^n z_i$ den geschätzten Mittelwert von z_i , $i = 1, \dots, n$ bezeichnet.

Durch Einsetzen von (2.9) in die Normalgleichung (2.8) erhält man

$$\sum_{i=1}^n x_i \left(y_i - \left(\bar{y} - \hat{\beta}_1 \bar{x} \right) - \hat{\beta}_1 x_i \right) = 0.$$

Umformen führt zu

$$\sum_{i=1}^n x_i (y_i - \bar{y}) = \hat{\beta}_1 \sum_{i=1}^n x_i (x_i - \bar{x}).$$

Man beachte, dass

$$\sum_{i=1}^n x_i(y_i - \bar{y}) = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}),$$

$$\sum_{i=1}^n x_i(x_i - \bar{x}) = \sum_{i=1}^n (x_i - \bar{x})^2,$$

so dass:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (2.10)$$

- **Terminologie:**

- Die Stichprobenfunktionen (2.9) und (2.10)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x},$$
$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

werden als **KQ-Schätzer** (bzw. KQ-Schätzvorschriften) für β_0 und β_1 bezeichnet.

- Für eine gegebene Stichprobe werden $\hat{\beta}_0$ und $\hat{\beta}_1$ als **KQ-Schätzungen** (bzw. KQ-Schätzwerte) von β_0 und β_1 bezeichnet.

- Die **KQ-Regressionsfunktion der Stichprobe** oder **KQ-Regressiongerade** für das einfache Regressionsmodell ist gegeben durch

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (2.11)$$

mit den **Residuen** $\hat{u}_i = y_i - \hat{y}_i$.

- Das **KQ-Regressionsmodell der Stichprobe** ist gegeben durch

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i \quad (2.12)$$

Beachte:

- Der KQ-Schätzer $\hat{\beta}_1$ existiert *nur dann*, wenn die Stichprobenbeobachtungen $x_i, i = 1, \dots, n$ eine Streuung aufweisen, da sonst der Nenner in (2.10) Null ist.

Annahme SLR.3

(Stichprobenvariation der unabhängigen Variablen):

In der Stichprobe sind nicht alle Werte der unabhängigen Variable $x_i, i = 1, 2, \dots, n$ gleich.

- Die Ableitung des KQ-Schätzers erfordert nur die gerade getroffene Annahme SLR.3, nicht jedoch die Annahmen SLR.1 und SLR.4 bezüglich der Grundgesamtheit!

- Möchte man jedoch Aussagen zur statistischen Qualität des KQ-Schätzers machen, sind weitere Annahmen notwendig, siehe Abschnitte 2.7, 3.4, und 4.2.
 - Der KQ-Schätzer lässt sich auch unter Berücksichtigung von Annahmen bezüglich der Grundgesamtheit ableiten.
- Der KQ-Schätzer als **Momentenschätzer**:
- Beachte, dass man aus Annahme SLR.4 $E[u|x] = 0$ zwei Momentenbedingungen erhält: $E[u] = 0$ und $Cov(x, u) = 0$. Über Einsetzen von Annahme SLR.1 $u = y - \beta_0 - \beta_1 x$, erhält man die Definition der **Momentenbedingungen** für die Modellparameter

$$E(y - \beta_0 - \beta_1 x) = 0 \quad (2.13)$$

$$E[x(y - \beta_0 - \beta_1 x)] = 0 \quad (2.14)$$

- Wie lassen sich die Momentenbedingungen durch Funktionen der Stichprobenwerte ersetzen?
- **Annahme SLR.2 (Zufallsstichprobe):**
Es liegt eine Zufallsstichprobe der Größe n vor, d.h. die Paare (x_i, y_i) und (x_j, y_j) , $i \neq j$, $i, j = 1, \dots, n$ sind paarweise unabhängig und identisch verteilt.
- Ein wichtiges Ergebnis aus der Statistik, siehe Abschnitt 5.1, besagt:
Bei Gültigkeit von Annahme SLR.2 ist der **Mittelwert einer Stichprobe** ein guter Schätzer des Erwartungswertes. (Annahme SLR.2 kann abgeschwächt werden, siehe z.B. Kapitel 11 in **Wooldridge (2009)**.)

- Ersetzt man die Erwartungswerte in (2.13) und (2.14) durch die Stichprobenmittelwerte, erhält man

$$n^{-1} \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0, \quad (2.15)$$

$$n^{-1} \sum_{i=1}^n x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0. \quad (2.16)$$

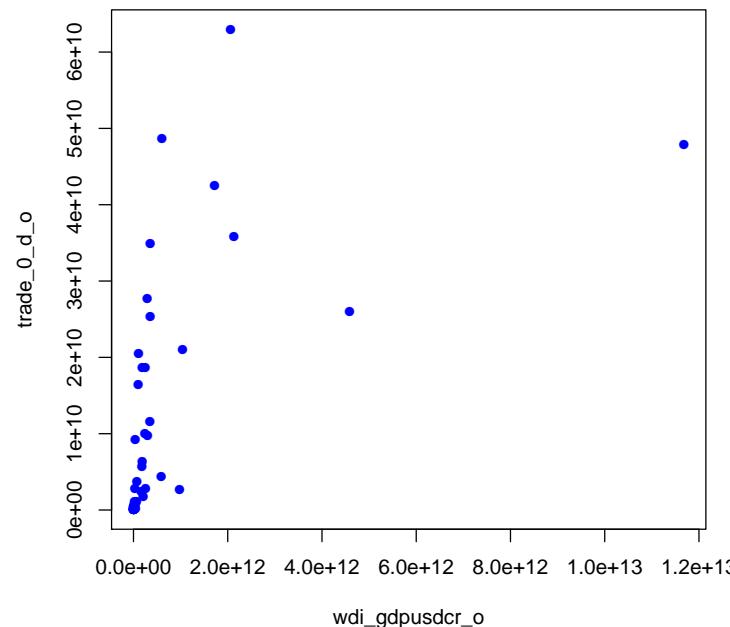
Werden (2.15) und (2.16) mit n multipliziert, erhält man wieder die Normalgleichungen (2.7) und (2.8).

Fortsetzung des Außenhandelsbeispiels

Fragestellung:

Steigen die Exporte nach Deutschland, wenn das BIP des exportierenden Landes steigt?

Scatterplot (aus Abschnitt 1.1)

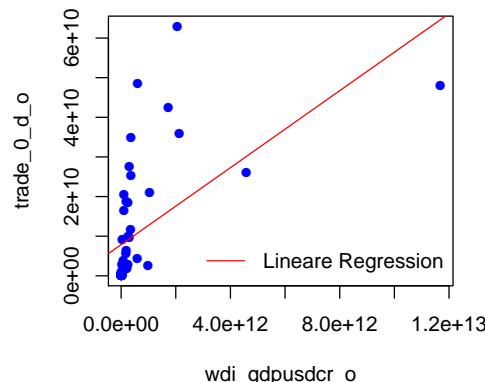


Die KQ-Regressionsgerade ist gegeben durch

$$\widehat{Importe}_i = 7.86 \cdot 10^{09} + 4.857 \cdot 10^{-03} BIP_i, \quad i = 1, \dots, 49,$$

und das Regressionsmodell der Stichprobe ist

$$Importe_i = 7.86 \cdot 10^{09} + 4.857 \cdot 10^{-03} BIP_i + \hat{u}_i, \quad i = 1, \dots, 49.$$



R-Output

Call:

```
lm(formula = trade_0_d_o ~ wdi_gdpusdcr_o)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.663e+10	-7.736e+09	-6.815e+09	2.094e+09	4.515e+10

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.858e+09	1.976e+09	3.977	0.000239 ***
wdi_gdpusdcr_o	4.857e-03	1.052e-03	4.617	3.03e-05 ***

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’
	0.1 ‘ ’	1		

Residual standard error: 1.31e+10 on 47 degrees of freedom

Multiple R-squared: 0.3121, Adjusted R-squared: 0.2974

F-statistic: 21.32 on 1 and 47 DF, p-value: 3.027e-05

- Eine nähere Beschreibung der Daten finden Sie im Anhang [10.4](#):

$Importe_i$ (aus Land i) TRADE_0_D_O

BIP_i (des Exportlandes i) WDI_GDPUSDCR_O

- Mögliche Interpretation des geschätzten Steigungsparameters:

$$\hat{\beta}_1 = \frac{\widehat{\Delta Importe}}{\Delta BIP}$$

zeigt an, um wie viel US-Dollar die durchschnittlichen Importe Deutschlands ansteigen, wenn das BIP des Exportlandes um 1 US-Dollar ansteigt.

- Ist das wirklich eine sinnvolle Interpretation? Fehlen nicht etwa andere wichtige Einflussfaktoren? Kann die Wirtschaftstheorie nicht auch mit einbezogen werden?
- Wie sieht es mit der Güte der Schätzungen aus?

Beispiel: Lohnregression

Fragestellung:

Wie verändert sich der Stundenlohn eines/einer Beschäftigten mit der Ausbildungsdauer?

- **Daten** (Quelle: Example 2.4 in [Wooldridge \(2009\)](#)): Stichprobe von U.S. Beschäftigten mit $n = 526$ Beobachtungen. Vorhandene Daten:
 - $wage$ in US-Dollar pro Stunde und
 - $educ$ Ausbildungsjahre eines Beschäftigten.
- Die KQ-Regressionsgerade und das Regressionsmodell der Stichprobe sind gegeben durch:

$$\widehat{wage}_i = -0.90 + 0.54 \text{ educ}_i, \quad i = 1, \dots, 526,$$

$$wage_i = -0.90 + 0.54 \text{ educ} + \hat{u}_i, \quad i = 1, \dots, 526.$$

Call:

```
lm(formula = wage ~ educ)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.3396	-2.1501	-0.9674	1.1921	16.6085

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.90485	0.68497	-1.321	0.187
educ	0.54136	0.05325	10.167	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.378 on 524 degrees of freedom

Multiple R-squared: 0.1648, Adjusted R-squared: 0.1632

F-statistic: 103.4 on 1 and 524 DF, p-value: < 2.2e-16

- Interpretation des geschätzten Steigungsparameters:

$$\hat{\beta}_1 = \frac{\Delta \widehat{wage}}{\Delta educ}$$

zeigt zum Beispiel an, um wie viel sich der durchschnittliche Studentenlohn im Durchschnitt ändert, wenn sich die Ausbildungsdauer um eine Einheit (also 1 Jahr) verändert:

- Ein zusätzliches Ausbildungsjahr erhöht den Stundenlohn im Durchschnitt um 54 Cent.
- Aber: Jemand gänzlich ohne Ausbildung erhält im Durchschnitt einen Stundenlohn von -90 Cent? Ist diese Interpretation sinnvoll?
- Ist es immer sinnvoll, den Steigungsparameter zu interpretieren? Geben Sie Acht, ob nicht eine scheinbare Kausalität vorliegt, siehe nächsten Abschnitt.
- Sind diese Schätzungen zuverlässig bzw. gut? Was bedeutet überhaupt “gut” im Rahmen der Ökonometrie und Statistik? Analysiere hierfür
 - die statistischen Eigenschaften des KQ-Schätzers und die Eigenschaften der KQ-Schätzungen, siehe Abschnitt 2.7 und
 - überprüfe die Wahl der funktionalen Form des bedingten Erwartungswertes $E[y|x]$, siehe Abschnitt 2.6.

2.4 Beste lineare Prognose, Korrelation und Kausalität

Beste lineare Prognose

- Was schätzt der KQ-Schätzer, wenn in der Grundgesamtheit, aus der die Stichprobe gezogen wurde, die Annahmen SLR.1 und SLR.4 (auch bekannt als Annahme (A)) **nicht gelten**?
- Beachte, dass $SSR(\gamma_0, \gamma_1)/n = \sum_{i=1}^n (y_i - \gamma_0 - \gamma_1 x_i)^2/n$ einen Stichprobenmittelwert angibt und somit eine Schätzung für den Erwartungswert

$$E \left[(y - \gamma_0 - \gamma_1 x)^2 \right], \quad (2.17)$$

darstellt, sofern Annahme SLR.2 (oder eine abgeschwächte Form davon) gilt. (Damit (2.17) existiert, müssen $0 < Var(x) < \infty$ und $Var(y) < \infty$ gelten.)

Gleichung (2.17) wird **mittlerer quadratischer Fehler des linearen Prädiktors**

$$\gamma_0 + \gamma_1 x.$$

genannt.

- Die theoretisch beste Anpassung des linearen Prädiktors $\gamma_0 + \gamma_1 x$ an y wird erreicht, indem man den mittleren quadratischen Fehler (2.17) bezüglich γ_0 und γ_1 minimiert. Dies ist analog zur Minimierung von $SSR(\gamma_0, \gamma_1)$. Dies führt zu (versuchen Sie dies selbst abzuleiten)

$$\gamma_0^* = E[y] - \gamma_1^* E[x], \quad (2.18)$$

$$\gamma_1^* = \frac{Cov(x, y)}{Var(x)} = Corr(x, y) \sqrt{\frac{Var(y)}{Var(x)}} \quad (2.19)$$

mit

$$\text{Corr}(x, y) = \frac{\text{Cov}(x, y)}{\sqrt{\text{Var}(x)\text{Var}(y)}}, \quad -1 \leq \text{Corr}(x, y) \leq 1$$

als Bezeichnung für die **Korrelation**, also dem Maß für die lineare Abhängigkeit zweier Variablen der Grundgesamtheit, hier x und y .

Der Ausdruck

$$\gamma_0^* + \gamma_1^* x \tag{2.20}$$

wird als **bester linearer Prädiktor** von y bezeichnet. Die Bezeichnung “Bester” wird hier definiert im Sinne von minimalem mittleren quadratischen Fehler.

- Beachte nun, dass im einfachen Regressionsmodell

$$y = \gamma_0^* + \gamma_1^* x + \varepsilon$$

gilt, dass $Cov(x, \varepsilon) = 0$ (eine schwächere Form von SLR.4) gilt, schließlich ist

$$Cov(x, y) = \frac{Cov(x, y)}{Var(x)}Var(x) + Cov(x, \varepsilon).$$

Dies deutet darauf hin, dass unter den Annahmen SLR.2 und SLR.3 gezeigt werden kann, dass der **KQ-Schätzer die Parameter γ_0^* und γ_1^* des besten linearen Prädiktors schätzt**. Der KQ-Schätzer (2.10) des Steigungsparameters lässt sich schreiben als Kombination der Stichprobenmittelwerte der Momente, die γ_1^* definieren, nämlich

$$\hat{\gamma}_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

- Indem man den **empirischen Korrelationskoeffizienten**

$$\widehat{Corr}(x, y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

verwendet, lässt sich $\hat{\gamma}_1$ umschreiben zu

$$\hat{\gamma}_1 = \widehat{Corr}(x, y) \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}.$$

Dies zeigt, dass der geschätzte Steigungsparameter **nicht null ist, wenn** in der Stichprobe Korrelation zwischen x und y vorliegt.

Kausalität

- Rufen Sie sich Abschnitt 1.3 wieder ins Gedächtnis.
- **Beachten** Sie, dass der **Steigungsparameter des besten linearen Prädiktors** γ_1^* **und sein KQ-Schätzwert** $\hat{\gamma}_1$ **nicht automatisch im Sinne eines kausalen Zusammenhangs interpretiert werden dürfen**, denn die Schätzung des besten linearen Prädiktors
 - berücksichtigt lediglich die Korrelation, nicht aber die Wirkungsrichtung,
 - schätzt u.U. nicht das interessierende Modell, z.B. wenn die Annahmen SLR.1 und SLR.4 verletzt sind und $\beta_1 \neq \gamma_1^*$,

- kann Unsinn produzieren, wenn
 - * im einfachen Regressionsmodell relevante Kontrollvariablen fehlen, sodass die Ergebnisse nicht die Resultate eines fiktiven kontrollierten Zufallsexperiments wiedergeben können (siehe Kapitel 3 und folgende Kapitel) oder
 - * $\widehat{Corr}(x, y)$ eine Scheinkorrelation schätzt, d.h. $Corr(x, y) = 0$ und Annahme SLR.2 (oder eine schwächere Form von ihr) sind verletzt.

Bevor man Kausalitäten interpretiert, sollten deshalb erst Techniken zur Spezifikation und Diagnose von Regressionsmodellen angewandt werden. Darüber hinaus ist es wichtig, die Bedeutung identifizierender Annahmen zu erkennen und sich der Grenzen bewusst zu werden, die für jede empirischen Kausalitätsanalyse gelten. Vgl. hierzu z.B. Kapitel 4 in **Bauer, Fertig, und Schmidt (2009)**.

2.5 Algebraische Eigenschaften des KQ-Schätzers

Grundlegende Eigenschaften:

- $\sum_{i=1}^n \hat{u}_i = 0$, wegen Normalgleichung (2.7),
- $\sum_{i=1}^n x_i \hat{u}_i = 0$, wegen Normalgleichung (2.8).
- Der Punkt (\bar{x}, \bar{y}) liegt auf der Regressionsgerade.

Können Sie diese Eigenschaften mit Ihrer Intuition erklären?

- **Total sum of squares (SST)**

$$SST \equiv \sum_{i=1}^n (y_i - \bar{y})^2.$$

- **Explained sum of squares (SSE)**

$$SSE \equiv \sum_{i=1}^n (\hat{y}_i - \bar{y})^2.$$

- **Sum of squared residuals (SSR)**

$$SSR \equiv \sum_{i=1}^n \hat{u}_i^2.$$

- Die **Zerlegung** $SST = SSE + SSR$ gilt immer dann, wenn die Regressionskonstante β_0 im Regressionsmodell enthalten ist.

- **Bestimmtheitsmaß R^2 (coefficient of determination oder R-squared)**

$$R^2 = \frac{SSE}{SST}.$$

- Interpretation: Anteil der Streuung der y_i , der durch die Streuung der x_i erklärt wird.
- Wenn die Regressionskonstante β_0 im Modell enthalten ist, dann gilt dank der Zerlegungsformel $SST = SSE + SSR$

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

und somit:

$$0 \leq R^2 \leq 1.$$

- Später sehen wir: Auswahl der Regressoren mit Hilfe von R^2 ist im Allgemeinen irreführend!

Zu Lesen:

- Abschnitte 1.4 und 2.1-2.3 in Wooldridge (2009) sowie bei Bedarf Appendix 10.1.
- 2.4 und 2.5 in Wooldridge (2009).

2.6 Parameterinterpretation, funktionale Form und Datentransformation

- Die Bezeichnung *linear* im einfachen linearen Regressionsmodell bedeutet nicht, dass eine lineare Beziehung zwischen den Variablen vorliegen muss, sondern dass die Parameter linear in das Modell eingehen.
- Beispiele für Modelle, die *linear* in den Parametern sind:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

$$y_i = \beta_0 + \beta_1 \ln x_i + u_i,$$

$$\ln y_i = \beta_0 + \beta_1 \ln x_i + u_i,$$

$$\ln y_i = \beta_0 + \beta_1 x_i + u_i,$$

$$y_i = \beta_0 + \beta_1 x_i^2 + u_i.$$

Der natürliche Logarithmus in der Ökonometrie

Die wohl in der empirischen Ökonomie am weitesten verbreitete Transformation ist der natürliche Logarithmus, kurz \ln . Die Interpretation des Steigungsparameters muss entsprechend angepasst werden.

Taylor-Approximation der logarithmischen Funktion:

$$\ln(1 + z) \approx z, \text{ falls } z \text{ nahe } 0.$$

Daraus lässt sich eine für die Berechnung von Wachstumsraten oder Renditen wichtige Approximation ableiten:

$$\begin{aligned} (\Delta x_t)/x_{t-1} &\equiv (x_t - x_{t-1})/x_{t-1} \\ &\approx \ln(1 + (x_t - x_{t-1})/x_{t-1}), \\ (\Delta x_t)/x_{t-1} &\approx \ln(x_t) - \ln(x_{t-1}). \end{aligned}$$

Für relative Veränderungen $\Delta x_t/x_{t-1}$ nahe Null ist dies eine gute Näherung. Prozentwerte erhält man durch Multiplikation mit 100:

$$100\Delta \ln(x_t) \approx \% \Delta x_t = 100(x_t - x_{t-1})/x_{t-1}.$$

Für kleine $\Delta x_t/x_{t-1}$ können demnach prozentuale Veränderungen gut über $100[\ln(x_t) - \ln(x_{t-1})]$ approximiert werden.

- Beispiele für Modelle, die *nichtlinear* in den Parametern $(\beta_0, \beta_1, \gamma, \lambda, \pi, \delta)$ sind:

$$y_i = \beta_0 + \beta_1 x_i^\gamma + u_i,$$

$$y_i^\gamma = \beta_0 + \beta_1 \ln x_i + u_i,$$

$$y_i = \beta_0 + \beta_1 x_i + \frac{1}{1 + \exp(\lambda(x_i - \pi))} (\gamma + \delta x_i) + u_i.$$

- Das letzte Beispiel ermöglicht das sanfte Umschalten zwischen zwei linearen Systemen/Regimen. Es sind natürlich nahezu unendlich vie-

le beliebige Formen von Nichtlinearität denkbar. Die Schätzung erfordert jedoch fortgeschrittene Methoden, wie etwa den nichtlinearen KQ-Schätzer. Dies geht jedoch über den Rahmen dieser Veranstaltung hinaus.

- Beachte, dass mit linearen Regressionsmodellen dennoch eine große Bandbreite nichtlinearer Zusammenhänge zwischen abhängiger und unabhängiger Variable abgebildet werden können, einige davon wurden am Anfang dieses Abschnitts aufgelistet.

Ökonomische Interpretation der KQ-Parameter

- Betrachten Sie das **Verhältnis der relativen Veränderungen** zweier **nicht-stochastischer** Variablen y und x

$$\frac{\frac{\Delta y}{y}}{\frac{\Delta x}{x}} = \frac{\% \text{-Veränderung von } y}{\% \text{-Veränderung von } x} = \frac{\% \Delta y}{\% \Delta x}.$$

Wenn $\Delta y \rightarrow 0$ und $\Delta x \rightarrow 0$, dann gilt $\frac{\Delta y}{\Delta x} \rightarrow \frac{dy}{dx}$.

- Dieses Ergebnis auf das obige Verhältnis angewendet, ergibt die **Elastizität**

$$\eta(x) = \frac{dy}{dx} \frac{x}{y}.$$

- **Interpretation:** Wenn die relative Veränderung von x 0,01 beträgt, dann ist die relative Veränderung von y genau $0,01\eta(x)$.

Bzw.: Wenn x sich um 1% ändert, dann ändert sich y um $\eta(x)\%$.

- Falls y, x **Zufallsvariablen** sind, wird die Elastizität auf Basis des bedingten Erwartungswertes von y gegeben x definiert:

$$\eta(x) = \frac{dE[y|x]}{dx} \frac{x}{E[y|x]}.$$

Dies lässt sich ableiten, indem man

$$\frac{\frac{E[y|x_1=x_0+\Delta x] - E[y|x_0]}{E[y|x_0]}}{\frac{\Delta x}{x_0}} =$$

$$\frac{E[y|x_1 = x_0 + \Delta x] - E[y|x_0]}{\Delta x} \frac{x_0}{E[y|x_0]}$$

betrachtet und dann Δx gegen 0 gehen lässt.

Verschiedene Modelle und Interpretationen von β_1

Für jedes Modell wird im Folgenden angenommen, dass die Annahmen SLR.1 und SLR.4 gültig sind.

- **Modelle, die linear in den Variablen sind (level-level oder Niveau)**

$$y = \beta_0 + \beta_1 x + u.$$

Es gilt

$$\frac{dE[y|x]}{dx} = \beta_1$$

und somit näherungsweise

$$\Delta E[y|x] = \beta_1 \Delta x.$$

In Worten: Der Steigungsparameter gibt die *absolute* Veränderung des bedingten Erwartungswertes der abhängigen Variable y an, wenn sich die unabhängige Variable x um eine *Einheit* ändert.

- **Linear-log Modelle (level-log)**

$$y = \beta_0 + \beta_1 \ln x + u.$$

Es gilt

$$\frac{dE[y|x]}{dx} = \beta_1 \frac{1}{x}$$

und somit näherungsweise

$$\Delta E[y|x] \approx \beta_1 \Delta \ln x = \frac{\beta_1}{100} 100 \Delta \ln x \approx \frac{\beta_1}{100} \% \Delta x.$$

In Worten:

Der bedingte Erwartungswert von y verändert sich um $\beta_1/100$ *Einheiten*, wenn sich x um 1% ändert.

- **Log-lineare Modelle (log-level)**

$$\ln y = \beta_0 + \beta_1 x + u$$

bzw.

$$y = e^{\ln y} = e^{\beta_0 + \beta_1 x + u} = e^{\beta_0 + \beta_1 x} e^u.$$

Somit gilt

$$E[y|x] = e^{\beta_0 + \beta_1 x} E[e^u|x].$$

Ist $E[e^u|x]$ konstant, gilt

$$\frac{dE[y|x]}{dx} = \beta_1 \underbrace{e^{\beta_0 + \beta_1 x} E[e^u|x]}_{E[y|x]} = \beta_1 E[y|x].$$

Man erhält näherungsweise

$$\frac{\Delta E[y|x]}{E[y|x]} \approx \beta_1 \Delta x, \quad \text{bzw.} \quad \% \Delta E[y|x] \approx 100 \beta_1 \Delta x$$

In Worten: Der bedingte Erwartungswert von y verändert sich um $100 \beta_1 \%$, wenn sich x um eine *Einheit* verändert.

• Log-log Modelle

werden oft als **loglineare** oder **constant-elasticity** Modelle bezeichnet und sind in der empirischen Praxis sehr populär

$$\ln y = \beta_0 + \beta_1 \ln x + u.$$

Ähnlich wie oben lässt sich zeigen, dass gilt:

$$\frac{dE[y|x]}{dx} = \beta_1 \frac{E[y|x]}{x}, \quad \text{und somit} \quad \beta_1 = \eta(x),$$

falls $E[e^u|x]$ konstant ist.

In diesem Modell entspricht der Steigungsparameter des log-log-Modells gerade der Elastizität für die ursprünglichen Niveauvariablen $E[y|x]$ und x .

In Worten: Der bedingte Erwartungswert von y verändert sich um $\beta_1\%$, wenn sich x um 1% verändert.

Fortführung des Außenhandelsbeispiels

R-Output

Call:

```
lm(formula = log(trade_0_d_o) ~ log(wdi_gdpusdcr_o))
```

Residuals:

Min	1Q	Median	3Q	Max
-2.6729	-1.0199	0.2792	1.0245	2.3754

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.77026	2.18493	-2.641	0.0112 *
log(wdi_gdpusdcr_o)	1.07762	0.08701	12.384	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.305 on 47 degrees of freedom

Multiple R-squared: 0.7654, Adjusted R-squared: 0.7604

F-statistic: 153.4 on 1 and 47 DF, p-value: < 2.2e-16

Beachte die unterschiedliche Interpretation des Steigungsparameters $\hat{\beta}_1$:

- Level-level Modell (Abschnitt 2.3): Ein Anstieg des BIPs des exportierenden Landes um 1 Mrd. US-Dollar geht einher mit einem durchschnittlichen Anstieg der Einfuhren nach Deutschland um 4,857 Mio. US-Dollar.
- Log-log Modell: Ein 1%iger Anstieg des BIPs des exportierenden Landes geht mit einem durchschnittlichen Anstieg der Importe um 1,077% einher.

Aber noch etwas Geduld, bevor diesen Ergebnissen Vertrauen geschenkt wird.

2.7 Statistische Eigenschaften des KQ-Schätzers: Erwartungswert und Varianz

- Vorbereitende Umformungen (alle Summern laufen über $i = 1, \dots, n$):

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} = \frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{\sum_{j=1}^n (x_j - \bar{x})^2} \\ &= \sum_{i=1}^n \underbrace{\left[\frac{(x_i - \bar{x})}{\sum_{j=1}^n (x_j - \bar{x})^2} \right]}_{w_i} y_i = \sum w_i y_i,\end{aligned}$$

wobei gilt (zur Übung verifizieren!):

$$\sum w_i = 0, \quad \sum w_i x_i = 1 \quad \text{und} \quad \sum w_i^2 = \frac{1}{\sum_{j=1}^n (x_j - \bar{x})^2}.$$

- **Unverzerrtheit (Erwartungstreue) der KQ-Schätzer:**

Gelten Annahmen SLR.1 bis SLR.4, dann gilt

$$E[\hat{\beta}_0] = \beta_0,$$

$$E[\hat{\beta}_1] = \beta_1.$$

Interpretation:

Würde man für sehr viele verschiedene Stichproben jeweils die KQ-Schätzungen berechnen, dann entspricht deren Durchschnitt ungefähr den Parameterwerten des Regressionsmodells der Grundgesamtheit.

Die Erwartungstreue ist eine Eigenschaft der Stichprobenverteilung der KQ-Schätzer für β_0 und β_1 . Sie besagt **nicht**, dass die Parameter für eine **konkrete** Stichprobe perfekt geschätzt werden!

Beweis für $\hat{\beta}_1$ (machen Sie sich klar, an welcher Stelle die einzelnen SLR-Annahmen zum Einsatz kommen):

1. $E \left[\hat{\beta}_1 \middle| x_1, \dots, x_n \right]$ lässt sich wie folgt umformen:

$$\begin{aligned}
 &= E \left[\sum w_i y_i \middle| x_1, \dots, x_n \right] \\
 &= E \left[\sum w_i (\beta_0 + \beta_1 x_i + u_i) \middle| x_1, \dots, x_n \right] \\
 &= \sum E [w_i (\beta_0 + \beta_1 x_i + u_i) \mid x_1, \dots, x_n] \\
 &= \beta_0 \sum w_i + \beta_1 \sum w_i x_i + \sum E [w_i u_i \mid x_1, \dots, x_n] \\
 &= \beta_1 + \sum w_i E [u_i \mid x_1, \dots, x_n] \\
 &= \beta_1 + \sum w_i E [u_i \mid x_i] \\
 &= \beta_1.
 \end{aligned}$$

2. Aus $E[\hat{\beta}_1] = E[E[\hat{\beta}_1 \mid x_1, \dots, x_n]]$ folgt die Erwartungstreue

$$E[\hat{\beta}_1] = \beta_1.$$

- **Varianzen der Parameterschätzer**

Zur Bestimmung der Varianzen der KQ-Parameterschätzer $\hat{\beta}_0$ und $\hat{\beta}_1$ benötigen wir eine weitere Annahme,

Annahme SLR.5 (Homoskedastie):

$$Var(u|x) = \sigma^2.$$

- **Varianzen der Parameterschätzer
bedingt auf die Stichprobenwerte**

Gelten Annahmen SLR.1 bis SLR.5, dann gilt

$$Var\left(\hat{\beta}_1 \middle| x_1, \dots, x_n\right) = \sigma^2 \frac{1}{\sum_{i=1}^n (x_i - \bar{x})^2},$$

$$Var\left(\hat{\beta}_0 \middle| x_1, \dots, x_n\right) = \sigma^2 \frac{n^{-1} \sum x_i^2}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Beweis (für die bedingte Varianz von $\hat{\beta}_1$):

$$\begin{aligned}
 & Var \left(\hat{\beta}_1 \middle| x_1, \dots, x_n \right) \\
 &= Var \left(\sum w_i u_i \middle| x_1, \dots, x_n \right) \\
 &= \sum Var (w_i u_i | x_1, \dots, x_n) \\
 &= \sum w_i^2 Var (u_i | x_1, \dots, x_n) \\
 &= \sum w_i^2 Var (u_i | x_i) \\
 &= \sum w_i^2 \sigma^2 \\
 &= \sigma^2 \sum w_i^2 \\
 &= \sigma^2 \frac{1}{\sum (x_i - \bar{x})^2}.
 \end{aligned}$$

- **Kovarianz zwischen den Schätzern für die Konstante und den Steigungsparameter:**

$$Cov(\hat{\beta}_0, \hat{\beta}_1 | x_1, \dots, x_n) = -\sigma^2 \frac{\bar{x}}{\sum_{i=1}^n (x_i - \bar{x})^2}.$$

Beweis: $Cov(\hat{\beta}_0, \hat{\beta}_1 | x_1, \dots, x_n)$ lässt sich wie folgt umformen:

$$\begin{aligned}
 &= Cov(\bar{y} - \hat{\beta}_1 \bar{x}, \hat{\beta}_1 | x_1, \dots, x_n) \\
 &= \underbrace{Cov(\bar{y}, \hat{\beta}_1 | x_1, \dots, x_n)}_{=0 \text{ siehe unten}} - Cov(\hat{\beta}_1 \bar{x}, \hat{\beta}_1 | x_1, \dots, x_n) \\
 &= -\bar{x} Cov(\hat{\beta}_1, \hat{\beta}_1 | x_1, \dots, x_n) \\
 &= -\bar{x} Var(\hat{\beta}_1 | x_1, \dots, x_n) \\
 &= -\sigma^2 \frac{\bar{x}}{\sum (x_i - \bar{x})^2}.
 \end{aligned}$$

$$\begin{aligned}
& Cov \left(\bar{y}, \hat{\beta}_1 \middle| x_1, \dots, x_n \right) \\
&= Cov \left(\beta_0 + \beta_1 \bar{x} + \bar{u}, \sum w_i y_i \middle| x_1, \dots, x_n \right) \\
&= Cov \left(\bar{u}, \sum w_i (\beta_0 + \beta_1 x_i + u_i) \middle| x_1, \dots, x_n \right) \\
&= Cov \left(\bar{u}, \sum w_i u_i \middle| x_1, \dots, x_n \right) \\
&= Cov (\bar{u}, w_1 u_1 \middle| x_1, \dots, x_n) + \dots + Cov (\bar{u}, w_n u_n \middle| x_1, \dots, x_n) \\
&= w_1 Cov (\bar{u}, u_1 \middle| x_1, \dots, x_n) + \dots + w_n Cov (\bar{u}, u_n \middle| x_1, \dots, x_n) \\
&= \sum w_i Cov (\bar{u}, u_i \middle| x_1, \dots, x_n) \\
&= Cov (\bar{u}, u_1 \middle| x_1, \dots, x_n) \sum w_i \\
&= 0.
\end{aligned}$$

2.8 Schätzung der Fehlervarianz

- Ein möglicher Schätzer der Fehlervarianz σ^2 ist

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n \hat{u}_i^2,$$

wobei die \hat{u}_i 's die Residuen des KQ-Schätzers bezeichnen.

Nachteil: Der Schätzer $\tilde{\sigma}^2$ berücksichtigt nicht, dass 2 Restriktionen auferlegt wurden, um die Residuen des KQ-Schätzers zu erhalten, nämlich: $\sum \hat{u}_i = 0$, $\sum \hat{u}_i x_i = 0$.

Dies führt zu verzerrten Schätzungen, $E[\tilde{\sigma}^2 | x_1, \dots, x_n] \neq \sigma^2$.

- **Unverzerrter Schätzer für die Fehlervarianz:**

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2.$$

- Unter den Annahmen SLR.1 bis SLR.5 gilt

$$E[\hat{\sigma}^2|x_1, \dots, x_n] = \sigma^2.$$

- **Standardfehler der Regression, standard error of the regression, standard error of the estimate oder root mean squared error:**

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2}$$

- $\hat{\sigma}^2$ kann in den Formeln für die Varianzen und Kovarianzen der Parameterschätzungen $\hat{\beta}_0$ und $\hat{\beta}_1$ zur Schätzung der unbekannten Fehlervarianz σ benutzt werden.

Beispiel:

$$\widehat{Var}(\hat{\beta}_1|x_1, \dots, x_n) = \frac{\hat{\sigma}^2}{\sum (x_i - \bar{x})^2}.$$

Bezeichnet man

$$sd(\hat{\beta}_1|x_1, \dots, x_n) = \sqrt{Var(\hat{\beta}_1|x_1, \dots, x_n)},$$

als die Standardabweichung, dann wird

$$\hat{sd}(\hat{\beta}_1|x_1, \dots, x_n) = \frac{\hat{\sigma}}{\left(\sum (x_i - \bar{x})^2\right)^{1/2}}$$

häufig als **Standardfehler von $\hat{\beta}_1$** bezeichnet und im Output von Softwarepaketen ausgegeben.

Zu Lesen: Abschnitte 2.4 und 2.5 in [Wooldridge \(2009\)](#) sowie bei Bedarf Appendix [10.1](#).

3 Das multiple lineare Regressionsmodell: Schätzung

3.1 Motivation für multiple Regression: Fortführung des Außenhandelsbeispiels

- In Abschnitt 2.6 wurden zur Erklärung der Importe Deutschlands zwei einfache lineare Regressionsmodelle geschätzt (und interpretiert): ein level-level Modell und ein log-log Modell.
- Es ist kaum glaubwürdig, dass die Importe Deutschlands nur vom BIP der Exportländer abhängen sollen. Was ist zum Beispiel mit der

Entfernung, gemeinsamen Staatsgrenzen oder anderen Kostenfaktoren des Außenhandels?

- Die empirische Literatur zu **Gravitationsgleichungen** ermittelte diese Größen als relevante Einflussfaktoren auf intra- und internationale Handelsströme. I.A. werden Handelsströme betrachtet, die in beide Richtungen fließen. Wir werden uns hier auf eine Richtung beschränken, nämlich die Exporte nach Deutschland im Jahr 2004. Eine so vereinfachte Gravitationsgleichung lautet

$$\ln(Importe_i) = \beta_0 + \beta_1 \ln(BIP_i) + \beta_2 \ln(Entfernung_i) + u_i. \quad (3.1)$$

Standard-Gravitationsgleichungen basieren auf bilateralen Importen und Exporten über mehrere Jahre hinweg und es werden hierfür Techniken für Paneldaten benötigt, die erst in **Weiterführende Fragen der Ökonometrie (BA)** besprochen werden.

- Eine kurze Einführung zu Gravitationsgleichungen finden Sie z.B. in [Fratianni \(2007\)](#). Eine theoretische Fundierung der Gravitationsgleichung findet sich in [Anderson und Wincoop \(2003\)](#).
- Bleiben relevante Variablen unberücksichtigt, kann Annahme SLR.1 und/oder SLR.4 verletzt sein, was zur Folge hätte, dass die Interpretation von Kausaleffekten in diesem Fall hochgradig irreführend sein kann, siehe Abschnitt [3.4](#). Um zu vermeiden, dass wir einen solchen Fehler begehen, kann das **multiple Regressionsmodell** verwendet werden.
- Um eine Vorstellung zu bekommen, wie sich der Elastizitätsparameter verändert, wenn eine zweite unabhängige Variable, z.B. Entfernung, eingefügt wird, betrachten wir die folgende Schätzung der einfachen Importgleichung [\(3.1\)](#):

R-Output

Call:

```
lm(formula = log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.99289	-0.58886	-0.00336	0.72470	1.61595

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.67611	2.17838	2.147	0.0371 *
log(wdi_gdpusdcr_o)	0.97598	0.06366	15.331	< 2e-16 ***
log(cepii_dist)	-1.07408	0.15691	-6.845	1.56e-08 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.9284 on 46 degrees of freedom

Multiple R-squared: 0.8838, Adjusted R-squared: 0.8787

F-statistic: 174.9 on 2 and 46 DF, p-value: < 2.2e-16

Statt der geschätzten Elastizität von 1,077, siehe Abschnitt 2.6, erhalten wir nun den Wert 0,976. Das R^2 steigt von 0,76 auf 0,88, was auf eine bessere statistische Anpassung schließen lässt. Weiterhin erfahren wir, dass ein Anstieg der Entfernung um 1% die Einfuhren um 1,074% reduziert. Ist dieses Modell also besser?

Oder ist es (ebenfalls) fehlspezifiziert?

Um diese Fragen beantworten zu können, müssen wir uns erstmal genauer mit dem multiplen Regressionsmodell befassen.

3.2 Das multiple Regressionsmodell der Grundgesamtheit

- **Annahmen:**

Die Annahmen SLR.1 und SLR.4 des einfachen linearen Regressionsmodells müssen für das multiple lineare Regressionsmodell (MLR) der Grundgesamtheit angepasst werden (siehe Abschnitt 3.3 in [Wooldridge \(2009\)](#)):

- **MLR.1 (Linearität in den Parametern)**

Das **multiple Regressionsmodell** gestattet mehr als nur eine, sagen wir k , erklärende Variable

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u \quad (3.2)$$

und das Modell ist linear in seinen Parametern.

Beispiel: Importgleichung (3.1).

– **MLR.4 (Bedingter Erwartungswert Null)**

$$E[u|x_1, \dots, x_k] = 0 \quad \text{für alle } x.$$

Beachte, dass auf alle erklärenden Variablen des multiplen Regressionsmodells (3.2) bedingt wird. Manchmal wird die Menge der Variablen auf die bedingt wird (conditioning set) auch als **Informationsmenge (information set)** bezeichnet.

- **Bemerkungen:**

- Um zu sehen, wofür man MLR.4 benötigt, nehmen wir in (3.2) auf alle k Regressoren bedingt den Erwartungswert von y :

$$E[y|x_1, x_2, \dots, x_k] = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + E[u|x_1, x_2, \dots, x_k].$$

Falls für einige x gilt, dass $E[u|x_1, x_2, \dots, x_k] \neq 0$, modelliert der strukturelle Teil $\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k$ den bedingten Erwartungswert $E[y|x_1, \dots, x_k]$ nicht korrekt.

- Sind MLR.1 und MLR.4 erfüllt, wird Gleichung (3.2)

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + u$$

als **lineares multiples Regressionsmodell der Grundgesamtheit** bezeichnet. Nicht selten wird es auch als **wahres Modell** bezeichnet (selbst wenn jedes Modell von der Wahrheit weit entfernt sein sollte). Alternativ kann man sich unter Gleichung (3.2) auch den **datengenerierenden Prozess** vorstellen (obwohl, streng genommen, ein datengenerierender Prozess zusätzlich noch die Spezifikation der Wahrscheinlichkeitsverteilungen aller Regressoren und der Fehler erfordert).

- Damit der KQ-Schätzer und das Regressionsmodell der Stichprobe auch die gewünschten Eigenschaften aufweisen, passen wir die Annahmen SLR.2 und SLR.3 entsprechend an:

– **MLR.2 (Zufallsstichprobe)**

Es wird eine Zufallsstichprobe der Größe n gezogen, d.h. $\{(x_{i1}, \dots, x_{ik}, y_i) : i = 1, \dots, n\}$ sind paarweise unabhängig und identisch verteilt.

– **MLR.3 (Keine perfekte Multikollinearität)**

(mehr zu MLR.3 in Abschnitt 3.3)

• **Interpretation:**

– Falls die Annahmen MLR.1 und MLR.4 korrekt sind *und* das Regressionsmodell der Grundgesamtheit sich kausal interpretieren lässt, dann ist das multiple Regressionsmodell ein gutes Werkzeug für eine **ceteris-paribus Untersuchung**. Damit ist es möglich die Werte aller erklärenden Variablen *außer einer* konstant zu halten und zu prüfen, wie sich der bedingte Erwartungswert der erklärten Variablen verändert. Dies gleicht dem Manipulieren ei-

ner Kontrollvariable in einem kontrollierten Zufallsexperiment. Sei nun x_j die Kontrollvariable, die von Interesse ist.

- Berechnen wir den bedingten Erwartungswert der Regressionsgleichung (3.2) und wenden Annahme MLR.4 an, erhalten wir

$$E[y|x_1, \dots, x_j, \dots, x_k] = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k.$$

- Betrachten wir nun eine Veränderung von x_j zu $x_j + \Delta x_j$:

$$E[y|x_1, \dots, x_j + \Delta x_j, \dots, x_k] = \beta_0 + \beta_1 x_1 + \dots + \beta_j(x_j + \Delta x_j) + \dots + \beta_k x_k.$$

*** Ceteris-paribus Effekt:**

Die Veränderung von x_j um Δx_j führt zu einer absoluten Veränderung des Erwartungswertes von (3.2):

$$\begin{aligned} \Delta E[y|x_1, \dots, x_j, \dots, x_k] &\equiv \\ &E[y|x_1, \dots, x_{j-1}, x_j + \Delta x_j, x_{j+1}, \dots, x_k] \\ &- E[y|x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k] = \beta_j \Delta x_j, \end{aligned}$$

wobei β_j mit der partiellen Ableitung erster Ordnung übereinstimmt

$$\frac{\partial E[y|x_1, \dots, x_{j-1}, x_j, x_{j+1}, \dots, x_k]}{\partial x_j} = \beta_j.$$

Der Parameter β_j gibt also den partiellen Effekt einer Veränderung von x_j auf den bedingten Erwartungswert von y an, während alle anderen Regressoren konstant gehalten werden.

* **Gesamteffekt:**

Man kann auch die gleichzeitige Veränderung mehrerer Regressoren betrachten, etwa für Δx_1 und Δx_k . Man erhält

$$\Delta E[y|x_1, \dots, x_k] = \beta_1 \Delta x_1 + \beta_k \Delta x_k.$$

- Beachte, dass die **Interpretation von** β_j davon abhängt, auf welche Weise die zugehörige Variable in die Gleichung eingeht, z.B. als logarithmierte Variable. Die Ergebnisse aus Abschnitt 2.6 bleiben für eine ceteris-paribus Untersuchung weiterhin gültig.

Fortsetzung des Außenhandelsbeispiels

- Betrachten wir das log-log Modell (3.1):

$$\ln(Importe_i) = \beta_0 + \beta_1 \ln(BIP_i) + \beta_2 \ln(Entfernung_i) + u_i.$$

Eine 1%ige Erhöhung der Entfernung führt zu einem Anstieg der Importe um $\beta_2\%$, wenn das **BIP konstant gehalten** wird. Mit anderen Worten heißt dies, dass man den Einfluss der Entfernung vom Einfluss des BIPs auf die Importe trennen kann. Ein Blick in die Output-Tabelle in Abschnitt 3.1 zeigt, dass ein 1%iger Anstieg der Entfernung die Importe um etwa 1,074% schrumpfen lässt.

- Man sollte aber bedenken, dass die Ermittlung von Entfernungen zwischen Ländern ein komplizierter Sachverhalt ist und die Ergebnisse der Schätzung variieren können, je nachdem, welche Methode man zur Berechnung der Distanzen angewandt hat. Unsere Daten

sind von **CEPII**, siehe auch Appendix **10.4**.

- Zudem können auch hier noch relevante Einflussgrößen vernachlässigt worden sein, siehe auch Abschnitt **4.4**.

Fortsetzung des Lohnbeispiels

- In Abschnitt **2.3** haben wir angenommen, dass sich der Stundenlohn ermitteln ließe über

$$wage = \beta_0 + \beta_1 \text{educ} + u.$$

Statt eines level-level Modells kann man auch ein log-level Modell erwägen:

$$\ln(wage) = \beta_0 + \beta_1 \text{educ} + u. \quad (3.3)$$

- Da wir erwarten, dass Erfahrung ebenfalls einen Einfluss auf den Stundenlohn hat, fügen wir die Variable *experience* ein. Dies ergibt

$$\ln(wage) = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + v. \quad (3.4)$$

Was erhalten wir dann für den erwarteten logarithmierten Lohn, gegeben die Variablen *educ* und *exper*?

$$E[\ln(wage)|\text{educ}, \text{exper}] = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper} + E[v|\text{educ}, \text{exper}]$$

$$E[\ln(wage)|\text{educ}, \text{exper}] = \beta_0 + \beta_1 \text{educ} + \beta_2 \text{exper},$$

wobei die zweite Gleichung nur gilt, wenn Annahme MLR.4 gilt, nämlich wenn

$$E[v|\text{educ}, \text{exper}] = 0.$$

- Beachte: Schätzt man statt (3.4) das einfache lineare log-level Modell (3.3), **obwohl das Modell der Grundgesamtheit *exper* enthält**, erhält man

$$E[\ln(wage)|\text{educ}] = \beta_0 + \beta_1 \text{educ} + \beta_2 E[\text{exper}|\text{educ}] + E[v|\text{educ}].$$

Dies weist darauf hin, dass das einfache Modell fehlspezifiziert ist, da es den Einfluss von *exper* über β_2 vernachlässigt. Demnach liegt im kleineren Modell Fehlspezifikation vor, wenn für einige Werte von *educ* oder *exper*

$$E[\ln(wage)|\text{educ}] \neq E[\ln(wage)|\text{educ}, \text{exper}]$$

gilt.

- **Empirische Ergebnisse:**

Siehe Example 2.10 in [Wooldridge \(2009\)](#), Daten: wage1.txt, Output aus R

- **Einfaches log-level Modell**

Call:

```
lm(formula = log(wage) ~ educ)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.21158	-0.36393	-0.07263	0.29712	1.52339

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.583773	0.097336	5.998	3.74e-09 ***
educ	0.082744	0.007567	10.935	< 2e-16 ***

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’
	0.1 ‘ ’			1

Residual standard error: 0.4801 on 524 degrees of freedom

Multiple R-squared: 0.1858, Adjusted R-squared: 0.1843

F-statistic: 119.6 on 1 and 524 DF, p-value: < 2.2e-16

$$\ln(wage_i) = 0.5838 + 0.0827 educ_i + \hat{u}_i, \quad i = 1, \dots, 526,$$
$$R^2 = 0.1858.$$

Gelten SLR.1 bis SLR.4, führt jedes zusätzliche Ausbildungsjahr im Durchschnitt zu einer Erhöhung des Stundenlohns um 8.3% und das Modell erklärt 18.6% der Variation der abhängigen Variable $\ln(wage)$.

– Multiples log-level Modell:

Call:

```
lm(formula = log(wage) ~ educ + exper)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.05800	-0.30136	-0.04539	0.30601	1.44425

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.216854	0.108595	1.997	0.0464 *
educ	0.097936	0.007622	12.848	< 2e-16 ***
exper	0.010347	0.001555	6.653	7.24e-11 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4614 on 523 degrees of freedom

Multiple R-squared: 0.2493, Adjusted R-squared: 0.2465

F-statistic: 86.86 on 2 and 523 DF, p-value: < 2.2e-16

$$\ln(wage_i) = 0.2169 + 0.0979 educ_i + 0.0103 exper_i + \hat{u}_i,$$

$$i = 1, \dots, 526,$$

$$R^2 = 0.2493.$$

- * **Ceteris-paribus Interpretation:** Bei Gültigkeit von MLR.1 bis MLR.4: Erhöht man die Zahl der Ausbildungsjahre um ein Jahr, erhöht sich der durchschnittliche Stundenlohn um 9.8%. Der Effekt ist etwas größer als im einfachen Regressionsmodell. Erhöht sich die Zahl der Jahre an Berufserfahrung um ein Jahr, erhöht sich der durchschnittliche Stundenlohn um 1%.
- * **Modellanpassung:**
Das Modell erklärt 24.9% der Stichprobenvariation der abhängigen Variable durch die Stichprobenvariation der Regressoren. Ist deshalb das multiple Regressionsmodell besser als das o.a. einfache Regressionsmodell mit Erklärungsgehalt 18.6%? Die Antwort erfolgt später, wenn wir Modellselektionsverfahren kennengelernt haben.

3.3 Der KQ-Schätzer: Ableitung & algebraische Eigenschaften

- Für einen beliebigen Schätzer ist das **Regressionsmodell der Stichprobe** für eine Stichprobe $(y_i, x_{i1}, \dots, x_{ik}), i = 1, \dots, n$ gegeben durch

$$y_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \tilde{\beta}_2 x_{i2} + \dots + \tilde{\beta}_k x_{ik} + \tilde{u}_i, \quad i = 1, \dots, n.$$

- Was war gleich noch mal die Idee des KQ-Schätzers? Wähle $\tilde{\beta}_0, \dots, \tilde{\beta}_k$ so, dass die **Summe der quadrierten Residuen (SSR)**

$$\text{SSR}(\tilde{\beta}_0, \dots, \tilde{\beta}_k) = \sum_{i=1}^n \tilde{u}_i^2 = \sum_{i=1}^n \left(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - \dots - \tilde{\beta}_k x_{ik} \right)^2$$

minimiert wird. Wir leiten nun $\text{SSR}(\tilde{\beta}_0, \dots, \tilde{\beta}_k)$ bezüglich aller $k+1$ Parameter ab und setzen diese Null. Man erhält die Bedingungen

erster Ordnung für ein Minimum:

$$\sum_{i=1}^n \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik} \right) = 0 \quad (3.5a)$$

$$\sum_{i=1}^n x_{i1} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik} \right) = 0 \quad (3.5b)$$

$$\sum_{i=1}^n x_{ik} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik} \right) = 0 \quad \begin{matrix} \vdots & \vdots \end{matrix} \quad (3.5c)$$

Dieses System von **Normalgleichungen** enthält $k + 1$ unbekannte Parameter und $k + 1$ Gleichungen. Unter einigen weiteren Bedingungen (siehe unten) existiert hierfür eine eindeutige Lösung.

Es ist mühsam, dieses Gleichungssystem für große k zu lösen. Dies können wir vermeiden, indem wir die Normalgleichungen und vorher das Regressionsmodell in Matrixschreibweise darstellen.

• Das multiple Regressionsmodell in Matrixschreibweise

Mithilfe der Matrixschreibweise kann das multiple Regressionsmodell umgeschrieben werden zu ([Wooldridge, 2009](#), Appendix E)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (3.6)$$

wobei

$$\underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} x_{10} & x_{11} & x_{12} & \cdots & x_{1k} \\ x_{20} & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}}_{\mathbf{u}}.$$

Die Matrix \mathbf{X} wird als Regressormatrix bezeichnet und hat n Zeilen und $k + 1$ Spalten. Die Spaltenvektoren \mathbf{y} und \mathbf{u} haben jeweils n Zeilen, der Spaltenvektor $\boldsymbol{\beta}$ hat $k + 1$ Zeilen.

• Herleitung: KQ-Schätzer in Matrixschreibweise

- Eine Möglichkeit, den KQ-Schätzer in Matrixschreibweise herzuleiten, ist, die Normalgleichungen (3.5) in Matrixschreibweise zu schreiben. Wir führen dies anhand der j -ten Gleichung

$$\sum_{i=1}^n x_{ij} \left(y_i - \hat{\beta}_0 x_{i0} - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik} \right) = 0$$

durch, die ausmultipliziert zu

$$\sum_{i=1}^n \left(x_{ij} y_i - \hat{\beta}_0 x_{ij} x_{i0} - \hat{\beta}_1 x_{ij} x_{i1} - \cdots - \hat{\beta}_k x_{ij} x_{ik} \right) = 0$$

wird und sich weiter umschreiben lässt zu

$$\sum_{i=1}^n \left(\hat{\beta}_0 x_{ij} x_{i0} + \hat{\beta}_1 x_{ij} x_{i1} + \cdots + \hat{\beta}_k x_{ij} x_{ik} \right) = \sum_{i=1}^n x_{ij} y_i.$$

Die linke Seite wird nun faktorisiert, man erhält

$$\left(\sum_{i=1}^n x_{ij} x_{i0} \right) \hat{\beta}_0 + \left(\sum_{i=1}^n x_{ij} x_{i1} \right) \hat{\beta}_1 + \cdots + \left(\sum_{i=1}^n x_{ij} x_{ik} \right) \hat{\beta}_k = \sum_{i=1}^n x_{ij} y_i.$$

Dies lässt sich analog auch für alle anderen Normalgleichungen durchführen. Die $k + 1$ Gleichungen, die man so erhält, werden in Vektoren geschrieben:

$$\begin{pmatrix} \left(\sum_{i=1}^n x_{i0} x_{i0} \right) \hat{\beta}_0 + \left(\sum_{i=1}^n x_{i0} x_{i1} \right) \hat{\beta}_1 + \cdots + \left(\sum_{i=1}^n x_{i0} x_{ik} \right) \hat{\beta}_k \\ \vdots \\ \left(\sum_{i=1}^n x_{ik} x_{i0} \right) \hat{\beta}_0 + \left(\sum_{i=1}^n x_{ik} x_{i1} \right) \hat{\beta}_1 + \cdots + \left(\sum_{i=1}^n x_{ik} x_{ik} \right) \hat{\beta}_k \end{pmatrix} = \begin{pmatrix} \sum_{i=1}^n x_{i0} y_i \\ \vdots \\ \sum_{i=1}^n x_{ik} y_i \end{pmatrix}.$$

Wendet man die Regeln für die Matrixmultiplikation an, erhält man

$$\underbrace{\begin{pmatrix} (\sum_{i=1}^n x_{i0}x_{i0}) & (\sum_{i=1}^n x_{i0}x_{i1}) & \cdots & (\sum_{i=1}^n x_{i0}x_{ik}) \\ \vdots & \vdots & \ddots & \vdots \\ (\sum_{i=1}^n x_{ik}x_{i0}) & (\sum_{i=1}^n x_{ik}x_{i1}) & \cdots & (\sum_{i=1}^n x_{ik}x_{ik}) \end{pmatrix}}_{\mathbf{X}'\mathbf{X}} \underbrace{\begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}}_{\hat{\boldsymbol{\beta}}} = \underbrace{\begin{pmatrix} \sum_{i=1}^n x_{i0}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{pmatrix}}_{\mathbf{X}'\mathbf{y}},$$

also die **Normalgleichungen in Matrixschreibweise**

$$(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}. \quad (3.7)$$

- Beachte: Die Matrix $\mathbf{X}'\mathbf{X}$ hat $k + 1$ Zeilen und Spalten und ist somit eine quadratische Matrix.

Die Inverse $(\mathbf{X}'\mathbf{X})^{-1}$ **existiert, falls** alle Spalten (und Zeilen) linear unabhängig sind. Es kann gezeigt werden, dass dies dann der Fall ist, wenn alle Spalten von \mathbf{X} linear unabhängig sind.

Damit dies erfüllt ist, benötigen wir folgende Annahme:

Annahme MLR.3 (Keine perfekte Kollinearität):

In der Stichprobe lässt sich kein Regressor als exakte Linearkombination von einem oder mehreren anderen Regressoren darstellen.

Ist das eine restriktive Annahme?

- Um schließlich den **KQ-Schätzer in Matrixschreibweise** zu erhalten, multiplizieren wir die Normalgleichung (3.7) von links mit $(\mathbf{X}'\mathbf{X})^{-1}$:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}. \quad (3.8)$$

Dies ist die Kurzschreibweise für

$$\underbrace{\begin{pmatrix} \hat{\beta}_0 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}}_{\hat{\boldsymbol{\beta}}} = \underbrace{\begin{pmatrix} (\sum_{i=1}^n x_{i0}x_{i0}) & (\sum_{i=1}^n x_{i0}x_{i1}) & \cdots & (\sum_{i=1}^n x_{i0}x_{ik}) \\ \vdots & \vdots & \ddots & \vdots \\ (\sum_{i=1}^n x_{ik}x_{i0}) & (\sum_{i=1}^n x_{ik}x_{i1}) & \cdots & (\sum_{i=1}^n x_{ik}x_{ik}) \end{pmatrix}}_{(\mathbf{X}'\mathbf{X})^{-1}} \cdot \underbrace{\begin{pmatrix} \sum_{i=1}^n x_{i0}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{pmatrix}}_{\mathbf{X}'\mathbf{y}}.$$

Algebraische Eigenschaften des KQ-Schätzers

- $\mathbf{X}'\hat{\mathbf{u}} = 0$, d.h. $\sum_{i=1}^n x_{ij}\hat{u}_i = 0$ für $j = 0, \dots, k$.

Beweis: Einsetzen von $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$ in die Normalgleichung ergibt $(\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})\hat{\boldsymbol{\beta}} + \mathbf{X}'\hat{\mathbf{u}}$ und folglich $\mathbf{X}'\hat{\mathbf{u}} = 0$.

- Falls $x_{i0} = 1$, $i = 1, \dots, n$, ergibt sich $\sum_{i=1}^n \hat{u}_i = 0$.
- Für den Spezialfall $k = 1$ ergeben sich die algebraischen Eigenschaften des einfachen Regressionsmodells.
- Falls eine Konstante im Modell enthalten ist, liegt der Punkt $(\bar{y}, \bar{x}_1, \dots, \bar{x}_k)$ immer in der Regressionshyperebene. Die Regressionshyperebene ist das Äquivalent zur Regressionsgeraden bei der einfachen Regression.
- Die Definitionen für SST, SSE und SSR bleiben unverändert.
- Falls eine Konstante im Modell enthalten ist, gilt auch hier

$$SST = SSE + SSR.$$

- **Bestimmtheitsmaß:**

R^2 ist genauso definiert wie im SLR-Fall

$$R^2 = \frac{\text{SSE}}{\text{SST}}$$

oder, falls eine Konstante im Modell enthalten ist,

$$R^2 = 1 - \frac{\text{SSR}}{\text{SST}}.$$

Es kann gezeigt werden, dass das R^2 das Quadrat des empirischen Korrelationskoeffizienten zwischen den beobachteten y_i 's und den geschätzten (erklärten) \hat{y}_i 's ist, also

$$\begin{aligned} R^2 &= \frac{\left(\sum_{i=1}^n (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}) \right)^2}{\left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \left(\sum_{i=1}^n (\hat{y}_i - \bar{\hat{y}})^2 \right)} \\ &= \left[\widehat{\text{Corr}}(y, \hat{y}) \right]^2. \end{aligned}$$

Beachte, dass $\left[\widehat{\text{Corr}}(y, \hat{y})\right]^2$ selbst dann genutzt werden kann, wenn R^2 nicht brauchbar ist. In diesem Fall wird der Ausdruck als Pseudo- R^2 bezeichnet.

- **Bereinigtes R^2 :**

Schreibt man R^2 um, indem man SSR/SST mit n erweitert

$$R^2 = 1 - \frac{\text{SSR}/n}{\text{SST}/n},$$

sieht man, dass sich SSR/n und SST/n als Schätzer für σ^2 bzw. σ_y^2 interpretieren lassen. Diese Schätzer sind allerdings verzerrt.

Verwendet man stattdessen unverzerrte Schätzer, erhält man das **“bereinigte” R^2** :

$$\bar{R}^2 = 1 - \frac{\text{SSR}/(n - k - 1)}{\text{SST}/(n - 1)}.$$

Alternative Schreibweisen:

$$\bar{R}^2 = 1 - \frac{n-1}{n-k-1} \cdot \frac{\text{SSR}}{\text{SST}}$$

$$\begin{aligned}\bar{R}^2 &= 1 - \frac{n-1}{n-k-1} (1 - R^2) \\ &= \frac{-k}{n-k-1} + \frac{n-1}{n-k-1} \cdot R^2\end{aligned}$$

Eigenschaften von \bar{R}^2 (siehe Abschnitt 6.3 in [Wooldridge \(2009\)](#)):

- \bar{R}^2 kann ansteigen *oder* fallen, wenn ein Regressor zusätzlich in die Gleichung eingefügt wird.
- \bar{R}^2 steigt immer an, wenn ein zusätzlicher Regressor die unverzerrte Schätzung der Fehlervarianz reduziert.

Vorsicht: Analog zu R^2 darf man auch \bar{R}^2 nicht für Regressionsmodelle mit unterschiedlichen y vergleichen, z.B. etwa wenn in einem Modell y der Regressand ist, im anderen $\ln(y)$.

- Die Größen R^2 , \bar{R}^2 oder $\left[\widehat{\text{Corr}}(y, \hat{y})\right]^2$ werden als **Gütemaße (goodness-of-fit measures)** bezeichnet.

3.4 Der KQ-Schätzer: Statistische Eigenschaften

Annahmen :

- **MLR.1 (Linearität in den Parametern)**
- **MLR.2 (Zufallsstichprobe)**
- **MLR.3 (Keine perfekte Kollinearität)**
- **MLR.4 (Bedingter Erwartungswert Null)**

3.4.1 Erwartungstreue der Parameterschätzungen

- Gelten MLR.1 bis MLR.4, dann gilt $E[\hat{\beta}] = \beta$.

Beweis:

$$\begin{aligned}
 \hat{\beta} &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} && \text{MLR.3} \\
 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\mathbf{X}\beta + \mathbf{u}) && \text{MLR.1} \\
 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
 &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}.
 \end{aligned}$$

Bildet man den bedingten Erwartungswert, erhält man

$$\begin{aligned}
 E[\hat{\beta}|\mathbf{X}] &= \beta + E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}] \\
 &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\mathbf{u}|\mathbf{X}] \\
 &= \beta. && \text{MLR.2 und MLR.4}
 \end{aligned}$$

Die letzte Gleichung gilt wegen

$$E[\mathbf{u}|\mathbf{X}] = \begin{pmatrix} E[u_1|\mathbf{X}] \\ E[u_2|\mathbf{X}] \\ \vdots \\ E[u_n|\mathbf{X}] \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

wobei Letzteres folgt, da für $i = 1, \dots, n$ gilt:

$$\begin{aligned} E[u_i|\mathbf{X}] &= E[u_i|x_{11}, \dots, x_{1k}, \dots, x_{nk}] \\ &= E[u_i|x_{i1}, \dots, x_{ik}] && \text{MLR.2} \\ &= 0 && \text{MLR.4.} \end{aligned}$$

- **Das Problem fehlender Regressoren (omitted variable bias)**

Wir partitionieren die $k + 1$ Regressoren in eine $(n \times k)$ Matrix \mathbf{X}_A und einen $(n \times 1)$ Vektor \mathbf{x}_a . Wir erhalten

$$\mathbf{y} = \mathbf{X}_A \boldsymbol{\beta}_A + \mathbf{x}_a \beta_a + \mathbf{u}. \quad (3.9)$$

Im Folgenden nehmen wir an, dass das Regressionsmodell der Grundgesamtheit dieselbe Struktur hat wie in (3.9).

Fortsetzung des Außenhandelsbeispiels (aus Abschnitt 3.2):

Wir nehmen an, dass die Importe in der Grundgesamtheit abhängen vom BIP, der Entfernung der Länder und der Offenheit (bezüglich der Waren- und Kapitalströme) des Herkunftslandes.

$$\begin{aligned}\ln(Importe_i) = & \beta_0 + \beta_1 \ln(BIP_i) + \beta_2 \ln(Entfernung_i) \\ & + \beta_3 \ln(Offenheit_i) + u_i,\end{aligned}\quad (3.10)$$

so dass \mathbf{X}_A die Konstante, BIP_i , und $Entfernung_i$ enthält und \mathbf{x}_a der Vektor für $Offenheit_i$ ist, jeweils $i = 1, \dots, n$.

Nehmen wir nun an, wir wären nur an den Parameterwerten $\hat{\beta}_A$ (den Parametern der Konstanten, von BIP , und $Entfernung$) interessiert und dass wir den Regressorvektor \mathbf{x}_a in der Schätzung nicht berücksichtigen können, etwa weil die Erhebung zu aufwändig ist und uns deshalb keine Daten vorliegen.

Wie wirkt es sich auf die Schätzung von $\hat{\beta}_A$ aus, wenn die **Variable** x_a in der Schätzung **weggelassen** wird? Das geschätzte Modell ist

demnach

$$\mathbf{y} = \mathbf{X}_A \boldsymbol{\beta}_A + \mathbf{w}. \quad (3.11)$$

Das Modell (3.11) wird häufig auch als kleineres Modell bezeichnet.

Mit anderen Worten: Welche Schätzeigenschaften hat der KQ-Schätzer für $\boldsymbol{\beta}_A$, wenn das kleinere Modell (3.11) zugrunde gelegt wird?

Herleitung:

- Wir bezeichnen den KQ-Schätzer für $\boldsymbol{\beta}_A$ des kleineren Modells mit $\tilde{\boldsymbol{\beta}}_A$. Der Beweis folgt analog dem Beweis der Erwartungstreue des kleineren Modells, allerdings wird nun für \mathbf{y} das wahre Modell der Grundgesamtheit (3.9) eingesetzt, nämlich

$$\begin{aligned}\tilde{\boldsymbol{\beta}}_A &= (\mathbf{X}'_A \mathbf{X}_A)^{-1} \mathbf{X}'_A \mathbf{y} \\ &= (\mathbf{X}'_A \mathbf{X}_A)^{-1} \mathbf{X}'_A (\mathbf{X}_A \boldsymbol{\beta}_A + \mathbf{x}_a \beta_a + \mathbf{u}) \\ &= \boldsymbol{\beta}_A + (\mathbf{X}'_A \mathbf{X}_A)^{-1} \mathbf{X}'_A \mathbf{x}_a \beta_a + (\mathbf{X}'_A \mathbf{X}_A)^{-1} \mathbf{X}'_A \mathbf{u}.\end{aligned}$$

- Aus dem Gesetz der iterierten Erwartungen lässt sich ableiten, dass $E[\mathbf{u}|\mathbf{X}_A] = E[E[\mathbf{u}|\mathbf{X}_A, \mathbf{x}_a]|\mathbf{X}_A] = E[0|\mathbf{X}_A] = 0$, falls Annahme MLR.4 für das Modell der Grundgesamtheit (3.9) gilt.
- Berechnen wir nun den bedingten Erwartungswert von $\tilde{\boldsymbol{\beta}}_A$. Dazu behandeln wir das (unbeobachtete) \mathbf{x}_a auf die gleiche Art wie auch \mathbf{X}_A und erhalten

$$E \left[\tilde{\boldsymbol{\beta}}_A | \mathbf{X}_A, \mathbf{x}_a \right] = \boldsymbol{\beta}_A + (\mathbf{X}'_A \mathbf{X}_A)^{-1} \mathbf{X}'_A \mathbf{x}_a \boldsymbol{\beta}_a.$$

Demzufolge ist der **Schätzer** $\tilde{\boldsymbol{\beta}}_A$ **nur dann unverzerrt**, wenn

$$(\mathbf{X}'_A \mathbf{X}_A)^{-1} \mathbf{X}'_A \mathbf{x}_a \boldsymbol{\beta}_a = \mathbf{0}. \quad (3.12)$$

Sehen wir uns den Term auf der linken Seite der Gleichung (3.12) genauer an, d.h.

$$(\mathbf{X}'_A \mathbf{X}_A)^{-1} \mathbf{X}'_A \mathbf{x}_a \boldsymbol{\beta}_a.$$

Der Teil

$$(\mathbf{X}'_A \mathbf{X}_A)^{-1} \mathbf{X}'_A \mathbf{x}_a = \tilde{\boldsymbol{\delta}}$$

ist nämlich der KQ-Schätzer für $\boldsymbol{\delta}$ der Regression von \mathbf{x}_a (sozusagen: das “neue” \mathbf{y}) auf \mathbf{X}_A (das “neue” \mathbf{X})

$$\mathbf{x}_a = \mathbf{X}_A \boldsymbol{\delta} + \boldsymbol{\varepsilon}.$$

Die Gültigkeit der Bedingung (3.12) (und somit keine Verzerrung) liegt dann vor, falls

- * $\tilde{\boldsymbol{\delta}} = \mathbf{0}$, also \mathbf{x}_a in der Stichprobe mit \mathbf{X}_A unkorreliert ist, oder
- * $\beta_a = 0$ gilt und damit das kleinere Modell dem Modell der Grundgesamtheit entspricht.

Gilt keine der beiden Bedingungen, ist $\tilde{\boldsymbol{\beta}}_A$ verzerrt

$$E[\tilde{\boldsymbol{\beta}}_A | \mathbf{X}_A, \mathbf{x}_a] = \boldsymbol{\beta}_A + \tilde{\boldsymbol{\delta}} \beta_a.$$

Dies bedeutet, dass der KQ-Schätzer $\tilde{\beta}_A$ i.A. für **jeden** Parameter des kleineren Modells verzerrt ist.

Da diese Verzerrung (= bias) dadurch verursacht wird, dass eine im Modell der Grundgesamtheit relevante Variable im benutzten Regressionsmodell weggelassen wurde (weglassen = to omit), wird diese Art von Verzerrung **omitted variable bias** genannt und man sagt, dass das kleinere Modell in diesem Fall **fehlspezifiziert** ist. (Siehe Appendix 3A.4 in [Wooldridge \(2009\)](#).)

- Was ist eigentlich mit dem unbedingten Erwartungswert und somit der unbedingten Verzerrung? Nach Anwenden des Gesetzes der iterierten Erwartungen erhält man

$$E \left[\tilde{\boldsymbol{\beta}}_A | \mathbf{X}_A \right] = \boldsymbol{\beta}_A + E \left[\tilde{\boldsymbol{\delta}} | \mathbf{X}_A \right] \beta_a,$$

$$E \left[\tilde{\boldsymbol{\beta}}_A \right] = \boldsymbol{\beta}_A + E \left[\tilde{\boldsymbol{\delta}} \right] \beta_a.$$

Interpretation: Der zweite Ausdruck gibt den Erwartungswert des KQ-Schätzers an, wenn immer neue Stichproben für \mathbf{y} **und** \mathbf{X}_A gezogen werden. Für wiederholte Stichproben liegt eine Verzerrung also nur dann vor, wenn zwischen den Variablen \mathbf{X}_A und \mathbf{x}_a in der Grundgesamtheit Korrelation vorliegt, denn ansonsten ist $E \left[\tilde{\boldsymbol{\delta}} \right] = 0$, siehe auch 2.4.

- **Fortsetzung des Lohnbeispiels** (aus Abschnitt 3.2):
 - Falls der beobachtete Regressor $educ$ mit der unbeobachteten Variable $ability$ korreliert ist, “fehlt” der Regressor $x_a = ability$ in der Regression und die KQ-Schätzer, z.B. für die Wirkung eines zusätzlichen Jahres an Ausbildung, sind verzerrt.
 - **Interpretation** der Informationsmengen bei der Berechnung des Erwartungswertes von $\hat{\beta}_{educ}$:
 - * Betrachten wir zuerst

$$E[\hat{\beta}_{educ}|educ, exper, ability] = \beta_{educ} + \tilde{\delta}\beta_{ability},$$

wobei

$$ability = (1 \ educ \ exper) \boldsymbol{\delta} + \varepsilon.$$

Der obige bedingte Erwartungswert gibt uns den über viele unterschiedliche Stichproben berechneten Durchschnittswert für

$\hat{\beta}_{educ}$ an, wenn jede Stichprobe (von Arbeitnehmern) wie folgt gezogen wird: Man stellt sicher, dass in jeder Stichprobe immer die gleiche Anzahl von Personen vorhanden ist mit z.B. 10 Jahren Ausbildung, 15 Jahren Erfahrung und 150 Einheiten Begabung sowie die gleiche Anzahl von Arbeitnehmern mit 11 Jahren Ausbildung, etc.. Wir brauchen also für jede Merkmalskombination ($educ$, $exper$, $ability$) dieselbe Anzahl an Arbeitnehmern in jeder Stichprobe, obwohl die Arbeitnehmer selbst natürlich nicht (komplett) identisch sein dürfen.

* Betrachten wir nun

$$E[\hat{\beta}_{educ}|educ, exper] = \beta_{educ} + E[\tilde{\delta}|educ, exper] \beta_{ability}.$$

Wenn wir jetzt neue Stichproben ziehen, müssen wir nur noch sicherstellen, dass die Anzahl der Arbeitnehmer mit bestimmten Werten für die Ausbildungsdauer und die Erfahrung gleich

bleibt. Im Vergleich zu oben wird nun die Variable Begabung nicht mehr kontrolliert.

* Zuletzt betrachten wir noch

$$E[\hat{\beta}_{educ}] = \beta_{educ} + E[\tilde{\delta}] \beta_{ability}.$$

Ziehen wir nun neue Stichproben darf alles variieren. Hatten wir in einer Stichprobe, sagen wir, 50 Arbeitnehmer mit 10 Jahren Ausbildung dürfen dies in einer anderen Stichprobe 73 Arbeitnehmer mit 10 Jahren Ausbildung sein. Dies war in den beiden vorgenannten Fällen nicht möglich.

- **Die Auswirkungen fehlender Variablen auf den bedingten Erwartungswert:**

- Allgemeine Terminologie:

- * Falls $E[y|X_A, x_a] \neq E[y|X_A]$ gilt,
dann ist das kleinere Modell ohne x_a **fehlspezifiziert** und die Schätzung unterliegt dem omitted variable bias (Verzerrung wg. fehlender Variablen).
 - * Falls $E[y|X_A, x_a] = E[y|X_A]$ gilt,
dann ist die Variable x_a im größeren Modell **überflüssig (redundant)** und sollte aus der Regressionsgleichung entfernt werden.
 - * **Fortsetzung des Außenhandelsbeispiels:** Nehmen wir an, dass das Modell der Grundgesamtheit nur die Variablen

BIP und *Entfernung* enthält. Ein einfaches Regressionsmodell mit *BIP* als einzigm Regressor ist fehlspezifiziert und ein multiples Regressionsmodell mit *BIP*, *Entfernung* und *Offenheit* enthält eine redundante Variable, nämlich *Offenheit*.

- **Es kann passieren, dass im fehlspezifizierten Modell die Annahmen MLR.1 bis MLR.4 erfüllt sind.**

Nehmen wir der Einfachheit halber an, dass in X_A nur eine Variable enthalten ist

$$E[y|x_A, x_a] = \beta_0 + \beta_A x_A + \beta_a x_a.$$

Über das Gesetz der iterierten Erwartungen erhalten wir dann

$$E[y|x_A] = \beta_0 + \beta_A x_A + \beta_a E[x_a|x_A].$$

Ist zudem $E[x_a|x_A]$ linear in x_A

$$x_a = \alpha_0 + \alpha_1 x_A + \varepsilon, \quad E[\varepsilon|x_A] = 0,$$

erhält man

$$\begin{aligned} E[y|x_A] &= \beta_0 + \beta_A x_A + \beta_a(\alpha_0 + \alpha_1 x_A) \\ &= \gamma_0 + \gamma_1 x \end{aligned}$$

mit $\gamma_0 = \beta_0 + \beta_a \alpha_0$ und $\gamma_1 = \beta_A + \beta_a \alpha_1$ als Parameter des besten linearen Prädiktors, siehe Abschnitt 2.4.

- Beachte, dass in diesem Fall SLR.1 und SLR.4 für das kleinere Modell erfüllt sind, obwohl es nicht das Modell der Grundgesamtheit ist. Trotzdem gilt

$$E[y|x_A, x_a] \neq E[y|x_A],$$

falls $\beta_a \neq 0$ und $\alpha_1 \neq 0$.

- Demnach ist die Modellwahl von großer Bedeutung, siehe Abschnitt 3.5. Falls es wichtig ist, dass x_a in der Regression berücksichtigt wird, dann nützt uns das kleinere Modell nicht viel, wenn

die Differenzen der Erwartungswerte für einige Werte der Regressoren groß sind.

Benötigt man das Modell aber zur Vorhersage kann man dem kleineren Modell den Vorzug geben, da es eine kleinere Schätzvarianz aufweist, siehe Abschnitte [3.4.3](#) und [3.5](#).

Zu Lesen: Abschnitt 3.3 in [Wooldridge \(2009\)](#).

3.4.2 Varianz der Parameterschätzungen

- **Annahme MLR.5 (Homoskedastie):**

$$Var(u_i|x_{i1}, \dots, x_{ik}) = \sigma^2, \quad i = 1, \dots, n.$$

- Die Annahmen MLR.1 bis MLR.5 werden häufig als **Gauss-Markov-Annahmen** bezeichnet.
- Beachte, dass wegen Annahme MLR.2 (Zufallsstichprobe) gilt, dass

$$Cov(u_i, u_j|x_{i1}, \dots, x_{ik}, x_{j1}, \dots, x_{jk}) = 0 \quad \text{für } i \neq j, 1 \leq i, j \leq n,$$

$$Cov(u_i, u_j) = 0 \quad \text{für } i \neq j, 1 \leq i, j \leq n,$$

wobei wir die zweite Gleichung aus der ersten erhalten, wenn wir das LIE anwenden. Wegen MLR.2 kann man auch schreiben

$$Var(u_i|x_{i1}, \dots, x_{ik}) = Var(u_i|\mathbf{X}), \quad Cov(u_i, u_j|\mathbf{X}) = 0, \quad i \neq j.$$

Man kann alle n Varianzen und alle Kovarianzen in eine Matrix

schreiben

$$Var(\mathbf{u}|\mathbf{X}) \equiv \begin{pmatrix} Var(u_1|\mathbf{X}) & Cov(u_1, u_2|\mathbf{X}) & \cdots & Cov(u_1, u_n|\mathbf{X}) \\ Cov(u_2, u_1|\mathbf{X}) & Var(u_2|\mathbf{X}) & \cdots & Cov(u_2, u_n|\mathbf{X}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(u_n, u_1|\mathbf{X}) & Cov(u_n, u_2|\mathbf{X}) & \cdots & Var(u_n|\mathbf{X}) \end{pmatrix} \quad (3.13)$$

$$= \sigma^2 \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

oder kürzer (MLR.2 und MLR.5 gemeinsam)

$$Var(\mathbf{u}|\mathbf{X}) = \sigma^2 \mathbf{I}. \quad (3.14)$$

• Varianz des KQ-Schätzers

Unter den Gauss-Markov Annahmen MLR.1 bis MLR.5 gilt

$$Var(\hat{\beta}_j | \mathbf{X}) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, \quad \text{falls } x_j \text{ nicht konstant,} \quad (3.15)$$

wobei SST_j die **gesamte Stichprobenstreuung** (total sum of squares) des j -ten Regressors bezeichnet, also

$$SST_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2,$$

und das Bestimmtheitsmaß R_j^2 aus der Regression des j -ten Regressors auf alle anderen Regressoren stammt

$$x_{ij} = \delta_0 x_{i0} + \cdots + \delta_{j-1} x_{i,j-1} + \delta_{j+1} x_{i,j+1} + \cdots + \delta_k x_{i,k} + v_i, \\ i = 1, \dots, n. \quad (3.16)$$

(Für den Beweis siehe Appendix 3A.5 in [Wooldridge \(2009\)](#).)

Interpretation der Varianz des KQ-Schätzers:

- Je größer die **Fehlervarianz** σ^2 ist, desto größer ist auch die Varianz von $\hat{\beta}_j$.

Beachte: Dies ist eine Eigenschaft der Grundgesamtheit, sodass dieser Bestandteil der Varianz nicht durch die Größe der Stichprobe beeinflusst wird. (Wie im einfachen Regressionsmodell.)

- Je größer die **gesamte Stichprobenstreuung** SST_j des j -ten **Regressors** x_j ist, desto kleiner ist $Var(\hat{\beta}_j | \mathbf{X})$.

Beachte: Die gesamte Stichprobenstreuung kann immer erhöht werden, indem man die Stichprobe vergrößert, schließlich erhöht jede neue Beobachtungen die Summe SST_j .

- Falls $SST_j = 0$, ist Annahme MLR.3 *verletzt*.

- Je größer das Bestimmtheitsmaß R_j^2 aus Regression (3.16) ist, desto größer ist die Varianz von $\hat{\beta}_j$.
- Je größer R_j^2 ist, desto besser lässt sich die Streuung in x_j durch die Streuung aller anderen Regressoren erklären, weil in diesem Fall eine starke lineare (aber eben nicht *perfekte*) Abhängigkeit zwischen x_j und den anderen erklärenden Variablen besteht.

Nur ein geringer Teil der Stichprobenstreuung von y ist somit *spezifisch* auf den j -ten Regressor selbst zurückzuführen (eben gerade die Fehlervarianz in (3.16)). Der andere Teil der Streuung lässt sich genauso “gut” auch durch die geschätzte Linearkombination aller anderen Regressoren erklären. Der Schätzer kann deshalb den Effekt auf den Regressanden y nur schwerlich der Variable x_j oder der Linearkombination der restlichen Regressoren zurechnen. Deswegen hat der Schätzer eine große Schätzvarianz.

– Spezialfälle:

- * $R_j^2 = 0$: Dann ist x_j mit allen anderen erklärenden Variablen empirisch (in der Stichprobe) unkorreliert. Dadurch ist der Parameterschätzer $\hat{\beta}_j$ von den anderen Regressoren unbeeinflusst.
- * $R_j^2 = 1$: Dann gilt Annahme MLR.3 nicht mehr.
- * R_j^2 nahe 1: Diese Situation wird **Multi- oder Fastkollinearität** bezeichnet. In diesem Fall ist $Var(\hat{\beta}_j | \mathbf{X})$ sehr groß.
- **Aber**: Das Problem der Multikollinearität nimmt in großen Stichproben ab, da dann für einen gegebenen Wert von R_j^2 die SST_j ansteigt und somit die Varianz sinkt. Demnach ist Multikollinearität immer auch ein Problem geringer Stichprobengrößen.

- **Schätzung der Fehlervarianz σ^2**

- Unverzerrte **Schätzung der Fehlervarianz σ^2** :

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n - (k + 1)}.$$

- **Eigenschaften** des KQ-Schätzers (Forts.):

Ist $sd(\hat{\beta}_j | \mathbf{X}) = \sqrt{Var(\hat{\beta}_j | \mathbf{X})}$ die Standardabweichung, dann ist

$$\hat{sd}(\hat{\beta}_j | \mathbf{X}) = \frac{\hat{\sigma}}{\left(SST_j (1 - R_j^2) \right)^{1/2}}$$

der **Standardfehler von $\hat{\beta}_j$** .

- **Varianz-Kovarianz-Matrix** des KQ-Schätzers:

Grundlagen: Die Kovarianz zweier gemeinsam geschätzter β_j und β_l — zwischen dem Schätzer des j -ten und des l -ten Parameters — lässt sich schreiben als

$$Cov(\hat{\beta}_j, \hat{\beta}_l | \mathbf{X}) = E[(\hat{\beta}_j - \beta_j)(\hat{\beta}_l - \beta_l) | \mathbf{X}], \quad j, l = 0, 1, \dots, k,$$

wobei unterstellt wird, dass beide Schätzer unverzerrt sind. Wir können eine $((k + 1) \times (k + 1))$ -Matrix aufstellen, die alle Varianzen und Kovarianzen enthält (nächste Folie):

$$\begin{aligned}
 Var(\hat{\boldsymbol{\beta}}|\mathbf{X}) &\equiv \\
 &= \begin{pmatrix} Cov(\hat{\beta}_0, \hat{\beta}_0|\mathbf{X}) & Cov(\hat{\beta}_0, \hat{\beta}_1|\mathbf{X}) & \cdots & Cov(\hat{\beta}_0, \hat{\beta}_k|\mathbf{X}) \\ Cov(\hat{\beta}_1, \hat{\beta}_0|\mathbf{X}) & Cov(\hat{\beta}_1, \hat{\beta}_1|\mathbf{X}) & \cdots & Cov(\hat{\beta}_1, \hat{\beta}_k|\mathbf{X}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\beta}_k, \hat{\beta}_0|\mathbf{X}) & Cov(\hat{\beta}_k, \hat{\beta}_1|\mathbf{X}) & \cdots & Cov(\hat{\beta}_k, \hat{\beta}_k|\mathbf{X}) \end{pmatrix} \\
 &= \begin{pmatrix} E[(\hat{\beta}_0 - \beta_0)(\hat{\beta}_0 - \beta_0)|\mathbf{X}] & \cdots & E[(\hat{\beta}_0 - \beta_0)(\hat{\beta}_k - \beta_k)|\mathbf{X}] \\ E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_0 - \beta_0)|\mathbf{X}] & \cdots & E[(\hat{\beta}_1 - \beta_1)(\hat{\beta}_k - \beta_k)|\mathbf{X}] \\ \vdots & \ddots & \vdots \\ E[(\hat{\beta}_k - \beta_k)(\hat{\beta}_0 - \beta_0)|\mathbf{X}] & \cdots & E[(\hat{\beta}_k - \beta_k)(\hat{\beta}_k - \beta_k)|\mathbf{X}] \end{pmatrix} \\
 &= E \left[\begin{pmatrix} (\hat{\beta}_0 - \beta_0) \\ \vdots \\ (\hat{\beta}_k - \beta_k) \end{pmatrix} \left| \begin{pmatrix} (\hat{\beta}_0 - \beta_0) & \cdots & (\hat{\beta}_k - \beta_k) \end{pmatrix} \right| \mathbf{X} \right] \\
 &= E \left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \right] .
 \end{aligned}$$

Als Nächstes zeigen wir, dass gilt:

$$Var(\hat{\beta}|\mathbf{X}) = E \left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)' | \mathbf{X} \right] = \sigma^2 (\mathbf{X}' \mathbf{X})^{-1}.$$

Beweis:

Wir erinnern uns daran, dass eine korrekte Modellspezifikation Folgendes bedeutet:

$$\hat{\beta} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' (\mathbf{X} \beta + \mathbf{u}) = \beta + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{u},$$

woraus sich $\hat{\beta} - \beta = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{u}$ ergibt. Dies können wir in $Var(\hat{\beta}|\mathbf{X})$ einsetzen und erhalten

$$\begin{aligned}
E \left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' | \mathbf{X} \right] &= E \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} \left((\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} \right)' | \mathbf{X} \right] \\
&= E \left[(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} \mathbf{u}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} | \mathbf{X} \right] \\
&= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' \underbrace{E[\mathbf{u}\mathbf{u}' | \mathbf{X}]}_{\sigma^2 \mathbf{I}_n} \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\
&= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1}.
\end{aligned}$$

Aus obiger Definition der $Var(\hat{\boldsymbol{\beta}} | \mathbf{X})$ kann man ersehen, dass die Diagonalelemente den Varianzen $Var(\hat{\beta}_j | \mathbf{X})$, $j = 0, \dots, k$, der Parameterschätzer entsprechen.

- **Effizienz der KQ-Methode**

Beachte: Der KQ-Schätzer ist ein **linearer Schätzer** bezüglich der abhängigen Variable, da für gegebenes \mathbf{X} gilt, dass

$$\hat{\beta}_j = \sum_{i=1}^n \left(\frac{\hat{v}_i}{\sum_{i=1}^n \hat{v}_i^2} \right) y_i,$$

wobei \hat{v}_i die Residuen aus Regression (3.16) bezeichnen. Der Schätzer ist also eine gewichtete Summe des Regressanden. Die Linearität *des Schätzers* sollte nicht mit der Linearität *des Modells in den Parametern* verwechselt werden. (Eine Ableitung ohne Matrixalgebra findet sich in Appendix 3A.2 in [Wooldridge \(2009\)](#).)

Weiterhin ist der KQ-Schätzer unverzerrt, also gilt $E[\hat{\beta}_j] = \beta_j$.

Gauss-Markov Theorem: Unter den Annahmen MLR.1 bis MLR.5 ist der KQ-Schätzer der **beste aller unverzerrten linearen Schätzer (BLUE = best linear unbiased estimator)**.

“Beste” bedeutet, dass der KQ-Schätzer, der wegen $E[\hat{\beta}_j] = \beta_j$ unverzerrt ist, unter allen unverzerrten linearen Schätzern die kleinstmögliche Varianz aufweist.

3.4.3 Trade-off zwischen Verzerrung und Multikollinearität

- **Beispiel:** Das Regressionsmodell der Grundgesamtheit sei

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u.$$

– Sei R_1^2 für eine gegebene Stichprobe nahe 1. Dann wird β_1 mit großer Varianz geschätzt. Die Varianz ergibt sich aus (3.15).

– Was wäre eine mögliche Lösung? Wir könnten x_2 aus der Regression entfernen und ein einfaches Regressionsmodell schätzen. Wie wir bereits gesehen haben, ist der Schätzer von β_1 dann verzerrt.

Also: Ist die Korrelation zwischen x_1 und x_2 nahe 1 oder -1, dann liegt — **für eine gegebene Stichprobengröße** — ein **Trade-off (Zielkonflikt) zwischen Varianz und Verzerrung** vor.

– Wir beobachten eine Art **statistischer Unschärferelation**: Die Stichprobe liefert nicht genügend Information, um die formulierte

Fragestellung genau zu beantworten.

- **Einzig wirklich gute Lösung:** Erhöhung der Stichprobengröße.
 - **Alternativ:** Man kombiniert die stark korrelierten Variablen.
- **Varianz der Parameterschätzungen bei fehlspezifizierten Modellen:**

Auch hier unterscheiden wir die unterschiedlichen Arten der Fehlspezifikation (siehe Abschnitt 3.4.1):

- Zu viele Variablen: Es werden auch Parameter geschätzt für Variablen, die im “wahren” datengenerierenden Prozess keine Rolle spielen (**überflüssige/redundante Variablen**).
- Zu wenige Variablen: Eine oder mehrere Variablen fehlen, obwohl diese im Regressionsmodell der Grundgesamtheit relevant sind (**fehlende Variablen**).

- Falsche Variablen: Eine Kombination aus beiden obigen Fällen.

Auswirkungen auf die Varianz der Parameterschätzer:

- **Fall 1 (Überflüssige Variablen):**

Das Modell der Grundgesamtheit sei $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$. Es wird folgende Spezifikation für die Stichprobe gewählt:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{z}\boldsymbol{\alpha} + \mathbf{w},$$

wobei der Vektor \mathbf{z} für jede Beobachtung den individuellen Wert der Variablen z enthält. Die Varianz des Parameterschätzers $\hat{\beta}_j$ ist

$$Var(\hat{\beta}_j | \mathbf{X}) = \frac{\sigma^2}{SST_j(1 - R_{j,\mathbf{X},\mathbf{z}}^2)},$$

wobei nun $R_{j,\mathbf{X},\mathbf{z}}^2$ das Bestimmtheitsmaß einer Regression von x_j auf alle anderen Variablen in \mathbf{X} und auf \mathbf{z} bezeichnet. Dass $R_{j,\mathbf{X},\mathbf{z}}^2 \geq R_j^2$ gilt, lässt sich leicht daraus ersehen, dass für die

Berechnung des zweiten R^2 einer Regression mit weniger Variablen verwendet wurde.

Daher: **Fügt man zusätzliche Variablen in das Regressionsmodell ein, erhöhen sich die Schätzvarianzen oder bleiben (im günstigsten Fall) gleich.**

– **Fall 2 (Fehlende Variablen):**

Hier gilt die Umkehrung von Fall 1: Es kann gezeigt werden, dass die **Schätzvarianz kleiner ist, wenn nicht das wahre Modell geschätzt wird, sondern eine Variable fehlt.**

– **Fall 3 (Überflüssige und fehlende Variablen):**

Sollte man wirklich vermeiden.

Also: Eine korrekte Modellspezifikation ist von entscheidender Bedeutung!

3.5 Modellspezifikation I: Modellselektionskriterien

- **Ziel der Modellselektion:**
 - Grundsätzlich: Suche nach dem Modell der Grundgesamtheit.
 - Praktisch: Suche nach dem “besten” Modell für die beabsichtigte Untersuchung.
 - Genauer: Unter der Annahme, das Modell der Grundgesamtheit sei ein multiples lineares Regressionsmodell, wollen wir alle Variablen finden, die im Regressionsmodell der Grundgesamtheit vorhanden sind und deren passende Transformationen (level oder log oder...). Es soll vermieden werden, dass relevante Variablen fehlen und überflüssige hinzugefügt werden.

- Kurze **Theorie der Modellselektion**:

- Es gibt zwei Aspekte:
 - a) die **Auswahl der Variablen (Wahl des Modells)**,
 - b) die **Schätzvarianz**.
- Betrachten wir Punkt a): Man wähle eine Zielfunktion, die unterschiedliche Modelle bewertet. Eine gängige Zielfunktion ist der **mittlere quadratische Fehler (MSE)**. Dieser ist für feste Parameter definiert als

$$MSE = E \left[(y - \beta_0 x_0 - \beta_1 x_1 - \cdots - \beta_k x_k)^2 \right], \quad (3.17)$$

siehe auch Gleichung (2.17) für den Fall eines einfachen Regressionsmodells.

Wir wählen das Modell mit dem kleinstmöglichen MSE.

Wichtige Fälle:

- * Falls x_0, \dots, x_k **alle relevanten Variablen** beinhaltet und das Modell der Grundgesamtheit ein multiples lineares Regressionsmodell ist sowie der MSE bezüglich der Parameter minimiert wird, dann gilt

$$MSE = E[u^2] = \sigma^2.$$

- * Es kann gezeigt werden, dass für den Fall, dass **relevante Variablen fehlen** sollten, der MSE zerlegt werden kann in Varianz und quadrierte Verzerrung. Der Einfachheit halber, “vergessen” wir alle Variablen außer x_1 und wählen das Modell

$$y = \gamma_0 + \gamma_1 x_1 + v.$$

Dann gilt

$$\begin{aligned} MSE_1 &= E \left[\{(y - E[y|x_1, \dots, x_k]) + (E[y|x_1, \dots, x_k] - E[y|x_1])\}^2 \right] \\ &= \sigma^2 + E \left[(E[y|x_1, \dots, x_k] - E[y|x_1])^2 \right]. \end{aligned}$$

Erste Gleichung: Der Term in der ersten runden Klammer entspricht der Abweichung des tatsächlichen Werts von y von seinem bedingten Erwartungswert aus der Regressionsgleichung der Grundgesamtheit (“wahres Modell”) (also u). Der Term in der zweiten runden Klammer ist die Abweichung des bedingten Erwartungswertes des “wahren” Modells vom bedingten Erwartungswert des von uns gewählten fehlspezifizierten Modells, also die Verzerrung für die Prognose von y bedingt auf x_1, \dots, x_k . Die zweite Gleichung erhält man mittels LIE. Da $E \left[(E[y|x_1, \dots, x_k] - E[y|x_1])^2 \right] > 0$ für jedes fehlspezifizierte Modell (vgl. Folie/Seite 145), gilt $MSE < MSE_1$.

- Betrachten wir Punkte a) und b): Falls die **Parameter geschätzt werden müssen**, muss ein weiterer Term dem mittleren quadratischen Fehler hinzugefügt werden, nämlich die Varianzen und Kovarianzen, die sich aus der Schätzung der Modellparameter ergeben. Man erhält

$$\begin{aligned} MSE &= \text{Varianz des Fehlers der Grundgesamtheit} \\ &+ \text{Verzerrung des gewählten Modells}^2 \\ &+ \text{Schätzvarianz,} \end{aligned}$$

wobei die Schätzvarianz i.A. ansteigt, wenn die Anzahl der Variablen ansteigt. Demnach kann es passieren, dass es für die Minimierung des MSE am Besten wäre, ein Modell zu wählen, dass eine oder mehrere relevante Variable vernachlässigt. Dies ist typischerweise dann der Fall, wenn wir Vorhersagen treffen wollen.

- Daher sind zuverlässige Methoden zum Schätzen des MSE nötig.

- **Was nicht funktioniert:**

- **Das Modell mit dem kleinsten Standardfehler der Regression $\hat{\sigma}$ zu wählen, funktioniert nicht.**
 - * Warum? Es ist immer möglich ein Modell auszuwählen, bei dem jedes Residuum den Wert 0 annimmt, d.h. $\hat{u}_i = 0$ für $i = 1, \dots, n$. Dann ist klarerweise $\hat{\sigma} = 0$, obwohl die Fehlervarianz des wahren Modells $\sigma^2 > 0$ ist.
 - * Wie macht man das? Wir nehmen einfach $k+1 = n$ Regressoren in das Modell der Stichprobe auf, die die Annahme MLR.3 erfüllen und lösen die Normalgleichungen (3.5). Wir erhalten dann eine perfekte Anpassung, da wir ein lineares Gleichungssystem mit n Gleichungen und n Unbekannten (den Parametern) lösen.

- * Beachte, dass man hierfür jeden beliebigen Regressor, der MLR.3 erfüllt, hinzufügen kann, selbst wenn dieser mit dem Modell der Grundgesamtheit nichts zu tun hat.
- * Es ist ebenfalls beachtenswert, dass das SSR konstant bleibt oder fällt, wenn für eine gegebene Stichprobengröße n ein weiterer Regressor hinzugefügt wird. Schließlich erhält das lineare Gleichungssystem durch das Hinzufügen des Regressors mehr Flexibilität, um die Parameter an die Stichprobenwerte anzupassen. Daher gilt auch für $\tilde{\sigma}^2 = \frac{\hat{u}'\hat{u}}{n} = \frac{SSR}{n}$, dass sie konstant bleibt oder fällt.
- * Für den Varianzschätzer $\hat{\sigma}^2 = \frac{SSR}{n-k-1}$ ergeben sich gegenläufige Effekte, denn die Abnahme von SSR könnte durch die Abnahme in $n - k - 1$ kompensiert werden.

Zusammengefasst lässt sich sagen, dass $\tilde{\sigma} = \sqrt{SSR/n}$ nicht geeignet ist, um diejenigen Variablen auszuwählen, die Bestandteil des Modells der Grundgesamtheit sind, weil der $\tilde{\sigma}$ immer gleich bleibt oder fällt, wenn zusätzliche Regressoren aufgenommen werden.

- Die Wahl des Modells anhand der größten R^2 **funktioniert ebenfalls nicht**.

Warum?

- Obwohl das bereinigte R^2 bei Hinzunahme neuer Regressoren sowohl fallen als auch steigen kann, versagt es ebenfalls im Falle $k + 1 = n$, denn dann gilt $\bar{R}^2 = 1$.

- **Lösung: Modellselektionskriterien**

- **Grundidee:**

$$\text{Selektionskriterium} = \ln \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n} + (k + 1) \cdot \text{Straffunktion}(n)$$

- * **Erster Term:** $\ln \tilde{\sigma}^2$ basiert auf dem **Varianzschätzer** $\tilde{\sigma}^2$ für das gewählte Modell.

Beachte, dass die geschätzte Varianz $\tilde{\sigma}^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}}/n$ mit jedem zusätzlich aufgenommenen Regressor sinkt oder gleich bleibt.

- * **Zweiter Term:** ist ein **Strafterm**, der die Anzahl der Parameter bestraft, um zu vermeiden, dass überflüssige Variablen mit ins Modell aufgenommen werden.

Benutzt man $\tilde{\sigma}^2$, wird die wahre Fehlervarianz typischerweise unterschätzt und deswegen bestraft der Strafterm die Hinzunahme von zusätzlichen Regressoren.

Der Strafterm steigt mit steigendem k und die Straffunktion muss so gewählt werden, dass sie mit steigendem n fällt, sodass eine größere Parameteranzahl in großen Stichproben weniger ausmacht. Warum?

- * Dies impliziert einen **Trade-off**: Regressoren werden dann in das Modell aufgenommen, wenn die Strafe geringer ausfällt als die Minderung des MSE.

Durch die Wahl der Straffunktion (und damit des Kriteriums) legt man fest, wie dieser Trade-off quantitativ vorgenommen wird.

- * **Regel**: Unter allen in Erwägung gezogenen Kandidaten wird die Spezifikation gewählt, für die das Kriterium den *kleinsten* Wert annimmt.

– Gängige Modellselektionskriterien:

* Akaike Criterion (AIC)

$$AIC = \ln \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n} + (k + 1) \frac{2}{n}, \quad (3.18)$$

* Hannan-Quinn Criterion (HQ)

$$HQ = \ln \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n} + (k + 1) \frac{2 \ln(\ln n)}{n}, \quad (3.19)$$

* Schwarz / Bayesian Information Criterion (SC/BIC)

$$SC = \ln \frac{\hat{\mathbf{u}}' \hat{\mathbf{u}}}{n} + (k + 1) \frac{\ln n}{n}. \quad (3.20)$$

Es ist ratsam, alle Kriterien zu prüfen. In günstigen Fällen liefern alle Kriterien dasselbe Resultat. Beachte, dass SC für Standard-Stichprobengrößen zusätzliche Parameter stärker bestraft als HQ, und HQ wiederum stärker als AIC.

- **Fortsetzung des Außenhandelsbeispiels:**

- Modell 1

$$\text{LOG(TRADE_0_D_0)} = -5.770261 + 1.077624 * \text{LOG(WDI_GDPUSDCR_0)}$$

$$\text{AIC} = 3.410063, \quad \text{HQ} = 3.439359, \quad \text{SC} = 3.487280$$

- Modell 2

$$\text{LOG(TRADE_0_D_0)} = 4.676112 + 0.975983 * \text{LOG(WDI_GDPUSDCR_0)} - 1.074076 * \text{LOG(CEPII_DIST)}$$

$$\text{AIC} = 2.748467, \quad \text{HQ} = 2.792411, \quad \text{SC} = 2.864293$$

- Modell 3

$$\text{LOG(TRADE_0_D_0)} = 2.741040 + 0.940664 * \text{LOG(WDI_GDPUSDCR_0)} - 0.970318 * \text{LOG(CEPII_DIST)} \\ 0.507250 * \text{EBRD_TFES_0}$$

$$\text{AIC} = 2.644544, \quad \text{HQ} = 2.703136, \quad \text{SC} = 2.798979$$

- Modell 4

$$\text{LOG(TRADE_0_D_0)} = 2.427777 + 1.025023 * \text{LOG(WDI_GDPUSDCR_0)} - 0.888646 * \text{LOG(CEPII_DIST)} \\ 0.353154 * \text{EBRD_TFES_0} - 0.151031 * \text{LOG(CEPII_AREA_0)}$$

$$\text{AIC} = 2.616427, \quad \text{HQ} = 2.689667, \quad \text{SC} = 2.809470$$

- Ein Vergleich aller 4 Modelle ergibt, dass das SC Modell 3 mit den Regressoren *BIP*, *Entfernung* und *Offenheit* auswählt, während AIC Modell 4 mit *Fläche* als zusätzlichem Regressor wählt. Für Details zu den Variablen siehe Appendix 10.4. Man kann gut erkennen, dass SC stärker als AIC zusätzliche Regressoren bestraft. Statistische Tests können weitere Information im Hinblick auf die Frage liefern, welches Modell zu wählen ist, siehe Abschnitt 4.3 und Folgende.

4 Multiple Regression: Hypothesentests und Konfidenzintervalle

4.1 Grundlagen statistischer Tests

Grundlagen statistischer Hypothesentests

- Allgemein: Ein statistischer Hypothesentest ermöglicht statistisch fundierte, eindeutige Antworten auf Ja-Nein-Fragen:
 - Verdienen in Deutschland Frauen und Männer gleich viel?

- Bewirken bestimmte wirtschaftspolitische Maßnahmen eine Reduzierung der Arbeitslosigkeit im Jahr 2020?
- Wird die Höhe der deutschen Importe durch das BIP der exportierenden Länder beeinflusst?

- **Drei Bestandteile eines statistischen Tests:**

1. Zwei disjunkte **Hypothesen** über einen oder mehrere Wert(e) des (der) Parameter(s) θ in einer Grundgesamtheit.

D.h. eine der beiden alternativen Hypothesen muss in der Grundgesamtheit gelten:

- **Nullhypothese** H_0
- **Alternativhypothese** H_1

Wäre θ bekannt, könnte man sofort entscheiden, ob H_0 gilt!

2. Eine **Teststatistik** t : Eine Teststatistik ist eine Funktion, die aus den Stichprobenwerten (\mathbf{X}, \mathbf{y}) berechnet wird. Vor Beobachten einer Stichprobe ist eine Teststatistik eine *Zufallsvariable*, nach dem Beobachten einer Teststatistik eine *Realisation* einer Zufallsvariable. Wir schreiben in beiden Fällen $t(\mathbf{X}, \mathbf{y})$.

3. Eine **Entscheidungsregel**, die festlegt, für welche Werte von $t(\mathbf{X}, \mathbf{y})$ die **Nullhypothese H_0 abgelehnt** und für welche Werte die **Nullhypothese nicht abgelehnt** wird. Der Wertebereich von $t(\mathbf{X}, \mathbf{y})$ wird in zwei disjunkte Teilbereiche unterteilt:

- **Ablehnungsbereich (rejection region), kritischer Bereich \mathcal{C}**

Liegt die Teststatistik $t(\mathbf{X}, \mathbf{y})$ innerhalb des kritischen Bereichs, wird H_0 abgelehnt:

Lehne H_0 ab, falls $t(\mathbf{X}, \mathbf{y}) \in \mathcal{C}$.

- **Nicht-Ablehnungsbereich**

Liegt die Teststatistik $t(\mathbf{X}, \mathbf{y})$ innerhalb des Nicht-Ablehnungsbereichs, wird H_0 *nicht abgelehnt*:

Lehne H_0 nicht ab, falls $t(\mathbf{X}, \mathbf{y}) \notin \mathcal{C}$.

- **Kritischer Wert** c : Eine oder mehrere Grenze(n) zwischen Ablehnungs- und Nicht-Ablehnungsbereich.
- **Eigenschaften eines Tests:**

- **Fehler 1. Art (Type I error bzw. α -Fehler):**

Der Fehler erster Art eines Tests bezeichnet die Wahrscheinlichkeit (vor Erheben einer Stichprobe), die Nullhypothese H_0 zu verwerfen, obwohl H_0 in der Grundgesamtheit korrekt ist,

$$\alpha(\theta) = P(\text{Lehne } H_0 \text{ ab} | H_0 \text{ ist wahr}) = P(T \in \mathcal{C} | H_0 \text{ ist wahr}).$$

Beachte: Der Fehler 1. Art hängt möglicherweise von θ ab.

- **Fehler 2. Art (Type II error bzw. β -Fehler):**

Der Fehler zweiter Art gibt an, mit welcher Wahrscheinlichkeit H_0 *nicht* abgelehnt wird, obwohl H_0 falsch ist,

$$\beta(\theta) = P(\text{Lehne } H_0 \text{ nicht ab} | H_1 \text{ ist wahr}).$$

- **Größe (size) eines Tests:** bezeichnet den größten Fehler 1. Art bzgl. aller zulässigen Parameter θ , genau genommen das Supremum der Fehler 1. Art über alle Parameter, die für das Modell der Grundgesamtheit berücksichtigt werden können:

$$\sup_{\theta} \alpha(\theta)$$

- **Signifikanzniveau (level of significance, level):** Das Signifikanzniveau α muss vor der Untersuchung festgelegt werden und gibt vor, wie groß der Fehler 1. Art maximal sein darf:

$$\alpha(\theta) \leq \alpha$$

Aus dieser Bedingung lässt sich der Ablehnungsbereich $\mathcal{C} = \mathcal{C}(\alpha)$ bestimmen.

- **Güte (power) eines Tests:** Die **Güte** $\pi(\theta)$ eines Tests gibt an, mit welcher Wahrscheinlichkeit eine falsche Nullhypothese abgelehnt wird:

$$\begin{aligned}\pi(\theta) &= 1 - \beta(\theta) = 1 - P(\text{Lehne } H_0 \text{ nicht ab} | H_1 \text{ ist wahr}) \\ &= P(\text{Lehne } H_0 \text{ ab} | H_1 \text{ ist wahr}).\end{aligned}$$

Um \mathcal{C} für ein gegebenes α zu berechnen, muss man die **Wahrscheinlichkeitsverteilung** der Teststatistik unter H_0 kennen.

Tests für den Mittelwert der Grundgesamtheit:

1. Man betrachtet zwei disjunkte **Hypothesen** über den Mittelwert der Grundgesamtheit (z.B. den Mittelwert μ des Stundenlohns in den USA des Jahres 1976).

a) Nullhypothese

$$H_0 : \mu = \mu_0$$

(In unserem Bsp: Der mittlere Stundenlohn beträgt 6 US-\$, also $H_0 : \mu = 6$)

b) Alternativhypothese

$$H_1 : \mu \neq \mu_0$$

(In unserem Bsp: Der mittlere Stundenlohn beträgt nicht 6 US-\$, also $H_1 : \mu \neq 6$)

2. Teststatistik:

- a) Man wählt einen Schätzer für den unbekannten Mittelwert μ , z.B. den KQ-Schätzer der Regression des Stundenlohns w auf eine Konstante:

Aus der Stichprobe w_1, \dots, w_n mit n Beobachtungen lässt sich der Stichprobenmittelwert

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n w_i.$$

berechnen.

- b) Ermittle die Wahrscheinlichkeitsverteilung des Schätzers: Der Einfachheit halber nehmen wir an, dass die individuellen Löhne w_i gemeinsam normalverteilt sind mit Erwartungswert μ und Varianz

σ_w^2 , also

$$w_i \sim N(\mu, \sigma_w^2).$$

Aus den Eigenschaften gemeinsam normalverteilter Zufallsvariablen folgt, dass

$$\hat{\mu} \sim N\left(\mu, \sigma_{\hat{\mu}}^2\right),$$

wobei $\sigma_{\hat{\mu}}^2 = \text{Var}(\hat{\mu}) = \text{Var}(n^{-1} \sum w_i) = n^{-1} \sigma_w^2$.

- c) Um die Teststatistik $t(w_1, \dots, w_n)$ zu erhalten, müssen alle unbekannten Parameter aus der Verteilung entfernt werden. In diesem einfachen Fall wird dies erreicht, indem $\hat{\mu}$ standardisiert wird; man erhält

$$t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu}{\sigma_{\hat{\mu}}} \sim N(0, 1).$$

- d) Die Teststatistik $t(w_1, \dots, w_n)$ kann dann berechnet werden,

wenn wir μ und $\sigma_{\hat{\mu}}$ kennen. Nehmen wir fürs Erste an $\sigma_{\hat{\mu}}$ sei uns bekannt.

Welchen Wert nimmt dann μ unter H_0 an?

$$H_0 : \mu = \mu_0.$$

Unter H_0 können wir die Teststatistik für eine gegebene Stichprobe berechnen über

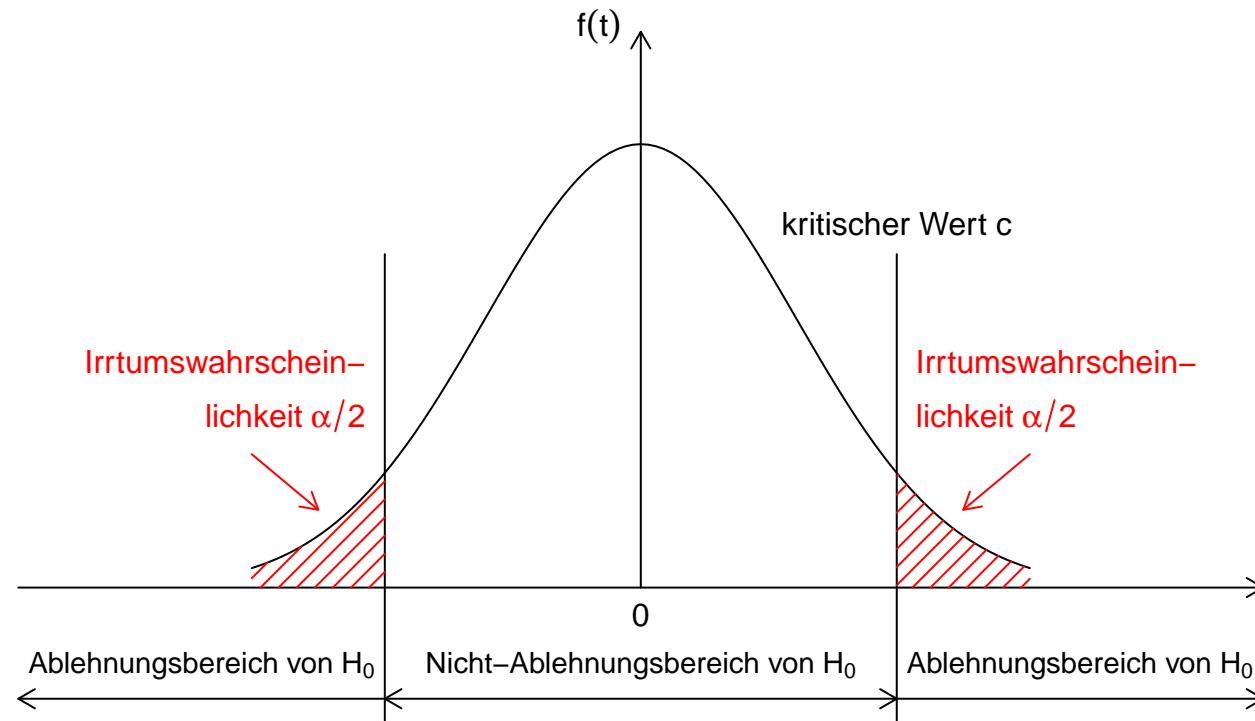
$$t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu_0}{\sigma_{\hat{\mu}}} \sim N(0, 1).$$

3. Entscheidungsregel:

Wann sollten wir H_0 ablehnen und wann nicht?

(Nun muss das **Signifikanzniveau** α gewählt werden!)

Ist die Abweichung von $\hat{\mu}$ vom Nullhypothesenwert μ_0 **groß genug**, sollten wir H_0 ablehnen.



Intuition: Falls t sehr groß (oder sehr klein) ist, dann ist

- a) der geschätzte Mittelwert $\hat{\mu}$ weit weg von μ_0 (unter H_0) und/oder
- b) die Standardabweichung $\sigma_{\hat{\mu}}$ der geschätzten Abweichung ist klein im Vergleich zur Differenz $\hat{\mu} - \mu_0$.

- Wann ist $|t|$ **groß genug** (um H_0 ablehnen zu können)?
- **Beachte:** Unter H_0 gilt, dass

$$t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu_0}{\sigma_{\hat{\mu}}} \sim N(0, 1)$$

und somit lässt sich der Ablehnungsbereich \mathcal{C} für ein gegebenes α ermitteln (siehe Grafik).

- Formal:

$$P(T < -c | H_0) + P(T > c | H_0) = \alpha$$

Dies lässt sich aufgrund der Symmetrie der Normalverteilung auch schreiben als

$$P(T < -c | H_0) = \frac{\alpha}{2} \quad \text{und} \quad P(T > c | H_0) = \frac{\alpha}{2}.$$

Die Werte von $-c$ und c lassen sich aus Tabellen ermitteln — es sind die $\alpha/2$ - und $1-\alpha/2$ -Quantile der Standardnormalverteilung.

- Unter H_1 gilt schließlich, dass

$$\frac{\hat{\mu} - \mu}{\sigma_{\hat{\mu}}} \sim N(0, 1).$$

Durch Erweitern erhält man

$$\frac{\hat{\mu} - \mu + \mu_0 - \mu_0}{\sigma_{\hat{\mu}}} = \frac{\hat{\mu} - \mu_0}{\sigma_{\hat{\mu}}} + \frac{\mu_0 - \mu}{\sigma_{\hat{\mu}}} = \underbrace{\frac{\hat{\mu} - \mu_0}{\sigma_{\hat{\mu}}}}_{t(w_1, \dots, w_n)} - \underbrace{\frac{\mu - \mu_0}{\sigma_{\hat{\mu}}}}_m$$

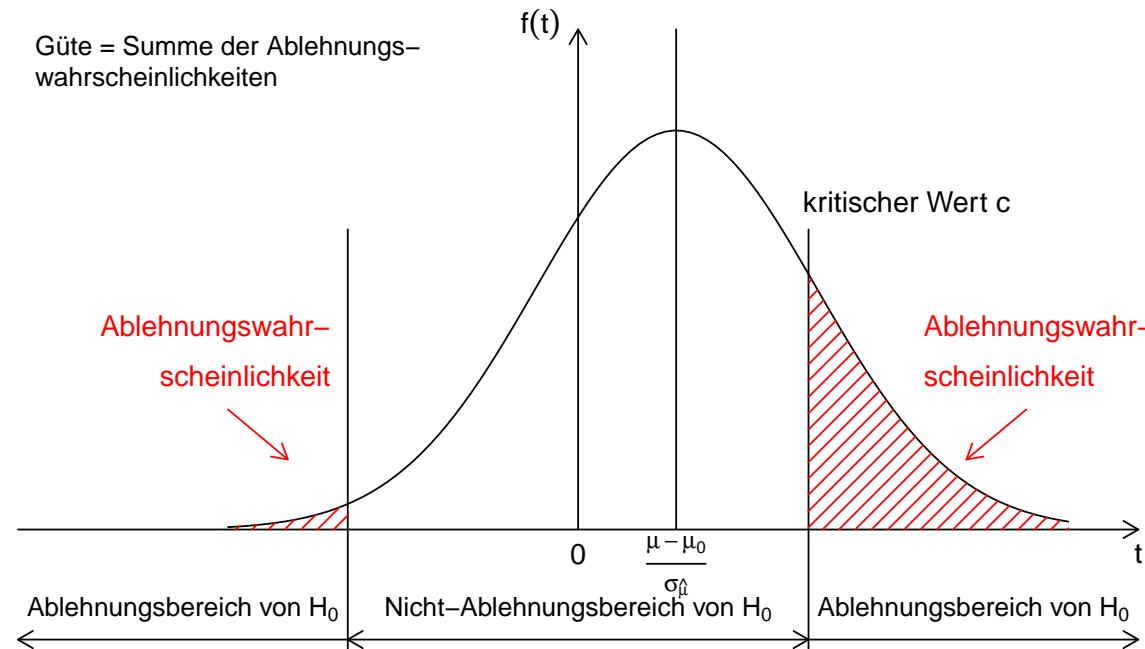
und somit erhalten wir unter H_1

$$t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu_0}{\sigma_{\hat{\mu}}} \sim N\left(\frac{\mu - \mu_0}{\sigma_{\hat{\mu}}}, 1\right),$$

da $X \sim N(m, 1)$ äquivalent zu $X - m \sim N(0, 1)$ ist.

- Fazit: Gilt H_1 , so ist die Teststatistik $t(w_1, \dots, w_n)$ um $(\mu - \mu_0)/\sigma_{\hat{\mu}}$ verschoben.

- In der Abbildung der Verteilung unter H_1 (für ein konkretes $\mu \neq \mu_0$) ergibt sich die Güte bzw. power aus der Summe der beiden schraffierten Flächen: $\pi(\mu) = P(t < -c|H_1) + P(t > c|H_1)$.



- Für ein gegebenes $\sigma_{\hat{\mu}}$ steigt die Güte des Tests mit steigender Differenz zwischen der Nullhypothese μ_0 und dem wahren Wert μ . Es ist dann “einfacher”, eine falsche Nullhypothese abzulehnen.

- Für den Fall, dass H_0 wahr ist, ist $(\mu - \mu_0)/\sigma_{\hat{\mu}} = 0$ und man erhält wieder die Verteilung unter H_0 .
- Man sieht auch, dass der Fehler 2ter Art — gegeben durch $\beta(\mu) == 1 - (1 - \beta(\mu)) = 1 - \pi(\mu)$ — *nicht* Null ist!

4. Bleibt noch ein Problem: In der Praxis kennen wir die Streuung des Mittelwertschätzers $\sigma_{\hat{\mu}} = \sigma_w/\sqrt{n}$ nicht.

Ausweg: Wir schätzen diese durch

$$\hat{\sigma}_{\hat{\mu}} = \frac{\hat{\sigma}_w}{\sqrt{n}}.$$

Man erhält letztlich die bekannte ***t*-Statistik**

$$t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu_0}{\hat{\sigma}_{\hat{\mu}}},$$

doch hier ist Vorsicht geboten!

Die neue Teststatistik ist nicht mehr normalverteilt, sondern folgt der t -Verteilung mit $n - 1$ Freiheitsgraden (kurz t_{n-1}). Damit gilt

$$t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu_0}{\hat{\sigma}_{\hat{\mu}}} \sim t_{n-1}.$$

Um die kritischen Werte

$$P(T < -c | H_0) = \frac{\alpha}{2} \quad \text{und} \quad P(T > c | H_0) = \frac{\alpha}{2},$$

zu erhalten, muss man in die Tabelle der t -Verteilung sehen (siehe Appendix G, Table G.2 in [Wooldridge \(2009\)](#)).

Fortsetzung des Lohnbeispiels:

Stundenlöhne w_i , $i = 1, \dots, 526$ von U.S.-amerik. Angestellten:

1. Hypothesen:

- a) Nullhypothese: $H_0 : \mu = 6$
- b) Alternativhypothese: $H_1 : \mu \neq 6$

2. Schätzung und Berechnung der t -Statistik in R:

Call:

```
lm(formula = wage ~ 1)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.3661	-2.5661	-1.2461	0.9839	19.0839

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.896	0.161	36.62	<2e-16 ***

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’
	0.1 ‘ ’			1

Residual standard error: 3.693 on 525 degrees of freedom

Also sind (unter Verwendung der gerundeten Werten im Output)

$$\hat{\mu} = 5.896, \quad \hat{\sigma}_{\hat{\mu}} = 0.161$$

und

$$t(w_1, \dots, w_{526}) = \frac{5.896 - 6}{0.161} = -0.6459627, \quad \text{exakt: } -0.6452201.$$

3. Bestimmung der kritischen Werte:

Nehmen wir ein Signifikanzniveau von $\alpha = 5\%$ an. Der kritische Wert $c = t_{525,0.05}$ kann dann aus der Tabelle der t -Verteilung mit $n - 1 = 525$ Freiheitsgraden ermittelt werden: $c = t_{525,0.05} = 1.96$.

4. Testentscheidung: Wir lehnen $H_0 : \mu = 6$ nicht ab, da

$$-c = -1.96 < t = -0.645 < c = 1.96,$$

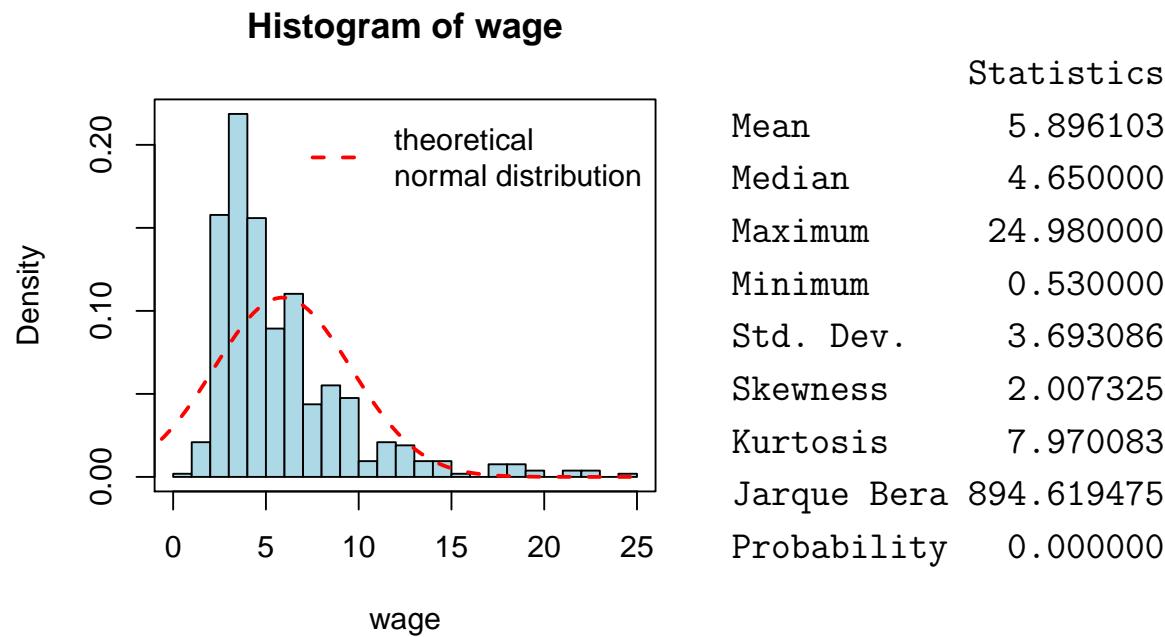
und somit $t \notin \mathcal{C}$ (die Teststatistik liegt nicht im Ablehnungsbereich).

5. Aber:

Wir nahmen an, dass die Stundenlöhne w_i normalverteilt sind. Stimmt das?

Betrachten wir das Histogram der Stichprobenbeobachtungen w_i :

Ergebnis:



- Es scheint, dass die Normalverteilungsannahme für unseren Test nicht erfüllt ist. Dann könnten die Testergebnisse irreführend sein!
- Es gibt aber auch Tests, die keine Normalverteilungsannahme erfordern, siehe Abschnitt 5.1.

Ein- und zweiseitige Hypothesentests

- **Zweiseitige Tests**

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

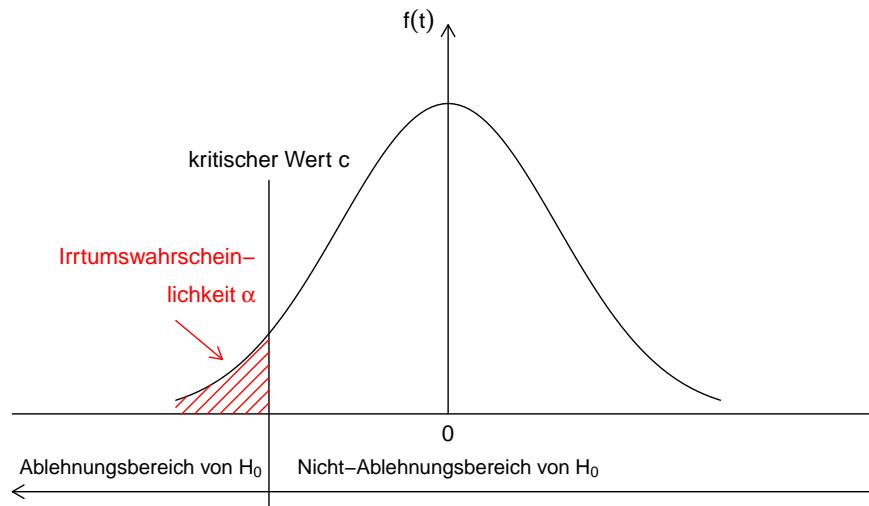
- **Einseitige Tests**

- **Test mit linksseitiger Alternativhypothese**

$$H_0 : \theta \geq \theta_0 \quad \text{versus} \quad H_1 : \theta < \theta_0.$$

Beachte: Häufig, so auch in [Wooldridge \(2009\)](#), liest man $H_0 : \theta = \theta_0$ versus $H_1 : \theta < \theta_0$. Diese Schreibweise ist nicht ganz präzise, da ja jeder mögliche Parameterwert entweder zu H_0 oder zu H_1 gehören muss. Das wird bei dieser Schreibweise aber nicht deutlich.

$$H_0 : \theta \geq \theta_0 \quad \text{versus} \quad H_1 : \theta < \theta_0$$



* **Entscheidungsregel:**

$$t < c \quad \Rightarrow \text{Lehne } H_0 \text{ ab.}$$

* Wir benötigen keinen Ablehnbereich auf der rechten Seite, da alle $\theta > \theta_0$ Elemente von H_0 sind und somit in den Nicht-Ablehnbereich fallen.

- * Den kritischen Wert berechnet man über die Dichte unter $\theta = \theta_0$, da dann für einen gegebenen kritischen Wert c der schraffierte Bereich größer ist als für ein beliebiges $\theta > \theta_0$. Man bevorzugt deshalb einen kritischen Wert, der den maximalen Fehler 1. Art, also die Größe kontrolliert, also die Größe des Tests durch das vorgegebene Signifikanzniveau begrenzt wird.

Fortsetzung des Lohnbeispiels:

(Im Folgenden ignorieren wir, dass die Stundenlöhne nicht normalverteilt sind.)

- * Die Nullhypothese besagt, dass der mittlere Stundenlohn 6 US-\$ oder mehr beträgt (H_1 sagt, dass er kleiner ist als 6 US-\$):

$$H_0 : \mu \geq 6 \quad \text{versus} \quad H_1 : \mu < 6.$$

- * Berechnung der Teststatistik: Wie im Fall des zweiseitigen Tests, schließlich ist μ_0 wieder die Grenze zwischen Null- und Alternativhypothese:

$$t(w_1, \dots, w_{526}) = \frac{5.896 - 6}{0.161} = -0.6459627, \quad \text{exakt: } -0.6452201.$$

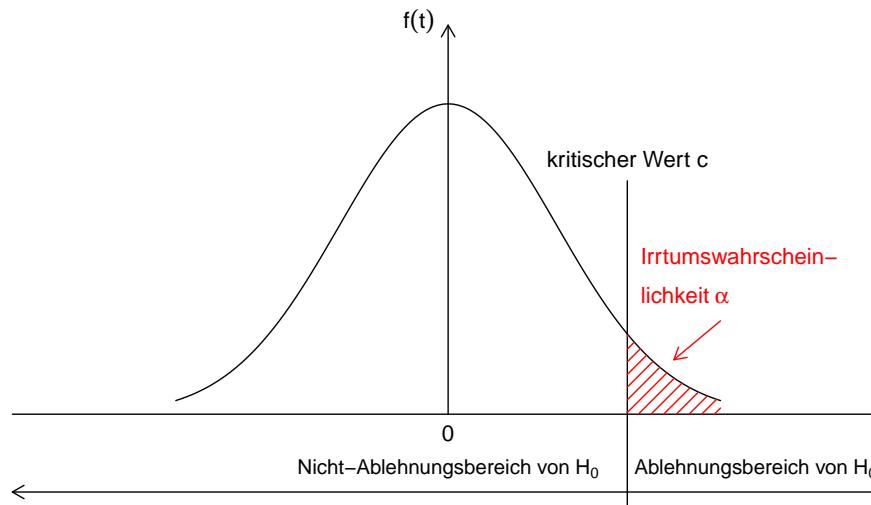
- * Berechnung des kritischen Werts: Für $\alpha = 0.05$ lautet der kritische Wert (beachte nun: einseitiger Test) aus der t -Verteilung mit 525 Freiheitsgraden (df) 1.645, also $c = -1.645$, da der linksseitige kritische Wert benötigt wird.
- * Entscheidung: Da

$$t = -0.6459627 > c = -1.645$$

wird die Nullhypothese nicht abgelehnt. (An Grafik verdeutlichen!)

– Test mit rechtsseitiger Alternative

$$H_0 : \theta \leq \theta_0 \quad \text{versus} \quad H_1 : \theta > \theta_0$$



Wie bei linksseitiger Alternativhypothese, nur spiegelverkehrt.

- **Warum machen wir einseitige Tests?** Betrachten wir folgende Aufgabe: Wir wollen untersuchen, ob es statistische Evidenz für die Behauptung gibt, der Stundenlohn läge über 5,60 \$.

– Da wir mittels statistischer Tests Hypothesen nicht bestätigen,

sondern nur ablehnen können, müssen wir die Alternativhypothese so wählen, dass sie unsere Vermutung widerspiegelt. In diesem Fall ist dies, dass der mittlere Stundenlohn größer als 5,60 \$ ist. Lehnen wir die Nullhypothese ab, erhalten wir statistische Evidenz für die Alternativhypothese. Es gibt jedoch Ausnahmen von dieser Regel, siehe z.B. Abschnitte [4.6](#) und [4.7](#).

- Wir müssen also testen, ob der mittlere Stundenlohn *statistisch signifikant größer* als 5,60 \$ ist.

Deshalb benötigen wir einen einseitigen Test. Unser Hypothesenpaar lautet

$$H_0 : \mu \leq 5.60 \quad \text{versus} \quad H_1 : \mu > 5.60.$$

- Für $\alpha = P(T > c | H_0) = 0,05$ ist der kritische Wert $c = 1.645$.

- Entscheidung:

$$t = \frac{5.896 - 5.60}{0.161} = 1.838509 > c = 1.645$$

⇒ Lehne H_0 (bei Größe 5%) ab, d.h. die Daten bestätigen, dass der mittlere Stundenlohn **statistisch signifikant** über 5,60 \$ liegt.

- Wollten wir im Gegensatz dazu untersuchen, ob der mittlere Stundenlohn **egal in welche Richtung** vom Wert 5,60 \$ abweicht, wäre unser Hypothesenpaar

$$H_0 : \mu = 5.60 \quad \text{versus} \quad H_1 : \mu \neq 5.60.$$

Für gegebenes Signifikanzniveau, $\alpha = 0.05$, ergeben sich die kritischen Werte -1,96 bzw. 1,96 und es gilt

$$-1.96 < 1.84 < 1.96.$$

Demnach kann die Nullhypothese *nicht* abgelehnt werden.

- Also, wenn man die Lage der Alternative genauer kennt, dann ist es “leichter”, die Nullhypothese abzulehnen, da der Test bezüglich der vermuteten Alternativhypothese mehr power/Güte hat.

p-Werte (*p*-values)

- Für jede Teststatistik lässt sich das *größte* Signifikanzniveau berechnen, bei dem — für eine gegebene Stichprobe — die berechnete Teststatistik *gerade noch nicht* zu einer Ablehnung der Nullhypothese geführt hätte. Diese Wahrscheinlichkeit nennt man ***p*-value (probability value)**.

Im Falle eines einseitigen Tests mit rechtsseitiger Alternative erhält man ([Wooldridge, 2009](#), Appendix C.6, p. 776)

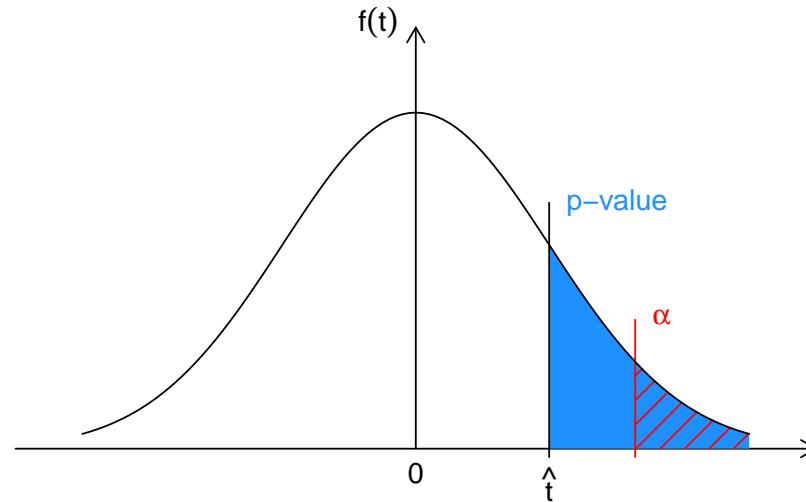
$$P(T \leq t(\mathbf{y}) | \mathcal{H}_0) \equiv 1 - p.$$

- Da $P(T > t(\mathbf{y})|\mathcal{H}_0) = 1 - P(T \leq t(\mathbf{y})|\mathcal{H}_0)$, gilt auch

$$P(T > t(\mathbf{y})|\mathcal{H}_0) = p.$$

Deshalb sagt man auch, dass der *p*-value das *kleinste* Signifikanzniveau angibt, bei dem die Nullhypothese *gerade noch* abgelehnt werden *kann*. Siehe Abschnitt 4.2, p. 133 in [Wooldridge \(2009\)](#).

- Die **Entscheidungsregel** lässt sich auch für ***p*-values** aufstellen: **Lehne \mathcal{H}_0 ab**, falls der ***p*-value kleiner** als das Signifikanzn. α ist.



Beachte: In der Grafik ist \hat{t} die Kurzschreibweise von $t(\mathbf{y})$.

Linksseitiger Test: $p = P(T < t(\mathbf{X}, \mathbf{y}))$

Rechtsseitiger Test: $p = P(T > t(\mathbf{X}, \mathbf{y}))$

Zweiseitiger Test: $p = P(T < -|t(\mathbf{X}, \mathbf{y})|) + P(T > |t(\mathbf{X}, \mathbf{y})|)$

- Viele Computerprogramme (so auch R) geben routinemäßig den p -value an für

$$H_0 : \theta = 0 \quad \text{versus} \quad \theta \neq 0.$$

Zu Lesen: Appendix C.6 in [Wooldridge \(2009\)](#).

4.2 Wahrscheinlichkeitsverteilung des KQ-Schätzers

Für das multiple Regressionsmodell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

nehmen wir weiterhin MLR.1 bis MLR.5 an, wie wir dies auch schon in den Abschnitten 3.2 und 3.4 getan haben.

- In Abschnitt 3.4.1 sahen wir, dass der KQ-Schätzer

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

unter Annahme MLR.1 auch geschrieben werden kann als

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + \underbrace{(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'}_{\mathbf{W}}\mathbf{u}. \quad (4.1)$$

- Um die Wahrscheinlichkeitsverteilung einer Teststatistik ableiten zu können, benötigt man die Wahrscheinlichkeitsverteilung des zu grunde liegenden Schätzers, da die Teststatistik eine Funktion des Schätzers ist. Zudem wird die Wahrscheinlichkeitsverteilung gebraucht, um Intervallschätzer zu berechnen, siehe Abschnitt 4.5.

Aus Gleichung (4.1) folgt, dass die Wahrscheinlichkeitsverteilung des KQ-Schätzers lediglich vom Fehlervektor \mathbf{u} abhängt, wenn wir auf die Regressormatrix \mathbf{X} bedingen. Analog zu dem Fall, in dem wir den Mittelwert getestet haben, machen wir die Annahme, dass die relevanten Zufallsvariablen normalverteilt sind.

- **Annahme MLR.6 (Normalverteilung der Fehler):**

Bedingt auf die Regressormatrix \mathbf{X} , ist der Vektor der Stichprobenfehler \mathbf{u} stochastisch unabhängig und identisch normalverteilt, also

$$u_i | x_{i1}, \dots, x_{ik} \sim i.i.d. N(0, \sigma^2), \quad i = 1, \dots, n.$$

Gemeinsam mit MLR.2 kann auch geschrieben werden, dass \mathbf{u} multivariat normalverteilt ist mit Mittelwert $\mathbf{0}$ und Varianz-Kovarianzmatrix $\sigma^2 \mathbf{I}$

$$\mathbf{u} | \mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I}).$$

- Natürlich könnte man für die Fehler \mathbf{u} auch jede andere Wahrscheinlichkeitsverteilung annehmen. Allerdings hat es zwei Vorteile, wenn man die Normalverteilung annimmt:

1. Die Wahrscheinlichkeitsverteilung des KQ-Schätzers und der abgeleiteten Teststatistiken lassen sich recht einfach herleiten, siehe die folgenden Abschnitte.

2. Unter bestimmten Bedingungen gilt die so ermittelte Wahrscheinlichkeitsverteilung des KQ-Schätzers auch dann, wenn die Fehler in Wirklichkeit nicht normalverteilt sind. Man nennt dies dann die **asymptotische Verteilung**, siehe Kapitel 5.

Zu den Eigenschaften normalverteilter Zufallsvariablen und Vektoren sei auf Appendix B und D in [Wooldridge \(2009\)](#) verwiesen.

- **Eigenschaften der multivariaten Normalverteilung:**

- Falls $Z \sim N(\mu, \sigma^2)$, dann $aZ + b \sim N(a\mu + b, a^2\sigma^2)$.
- Sind die Zufallsvariablen Z und V gemeinsam normalverteilt, dann sind Z und V dann und nur dann stochastisch unabhängig, wenn $Cov(Z, V) = 0$. (Beachte, dass nur bei der Normalverteilung aus $Cov(Z, V) = 0$ die stochastische Unabhängigkeit folgt!)
- Jede Linearkombination unabhängig und identisch normalverteilter Zufallsvariablen $\mathbf{z} \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ ist ebenfalls normalverteilt.

Sei

$$\mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix}, \quad \mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix},$$

dann gilt $\sum_{j=1}^n w_j z_j | \mathbf{w} = \mathbf{w}' \mathbf{z} | \mathbf{w} \sim N(\mathbf{w}' \boldsymbol{\mu}, \sigma^2 \mathbf{w}' \mathbf{w})$.

Allgemeiner gilt für $\mathbf{z} = (z_1, \dots, z_n)' \sim N(\boldsymbol{\mu}, \sigma^2 \mathbf{I})$ und

$$\mathbf{W} = \begin{pmatrix} w_{01} & w_{02} & \cdots & w_{0n} \\ w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & & \vdots \\ w_{k1} & w_{k2} & \cdots & w_{kn} \end{pmatrix} \quad |\mathbf{W} = \mathbf{Wz}| \mathbf{W} \sim N \left(\mathbf{W}\boldsymbol{\mu}, \sigma^2 \mathbf{W}\mathbf{W}' \right) . \quad (4.2)$$

- Die Eigenschaft (4.2) einer Linearkombination normalverteilter Zufallsvariablen kommt uns sehr gelegen, schließlich ist der KQ-

Schätzer (4.1) eine solche Linearkombination.

Also erhalten wir

$$\hat{\beta} - \beta | \mathbf{W} = \mathbf{W}\mathbf{u} | \mathbf{W} \sim N\left(0, \sigma^2 \mathbf{W}\mathbf{W}'\right).$$

Da $\mathbf{W}\mathbf{W}' = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{X})^{-1}$, erhält man

$$\hat{\beta} | \mathbf{X} \sim N\left(\beta, \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\right).$$

Auf ähnliche Art lässt sich zeigen, dass

$$\hat{\beta}_j | \mathbf{X} \sim N\left(\beta_j, \sigma_{\hat{\beta}_j}^2\right) \quad (4.3)$$

mit $\sigma_{\hat{\beta}_j}^2 = \frac{\sigma^2}{SST_j(1-R_j^2)}$ (siehe (3.15) in Abschnitt 3.4).

- Beachte, dass (4.3) eine Verallgemeinerung des Beispiels 4.1 zum Hypothesentest auf den Mittelwert ist. Falls \mathbf{X} ein Spaltenvektor mit lauter Einsen ist, gilt $\hat{\beta}_0 = \hat{\mu}$. (Zur Übung verifizieren)

4.3 t -Tests im multiplen Regressionsmodell

- **Ableitung der Teststatistik und deren Verteilung**

- Aus (4.3) wissen wir, dass $\hat{\beta}_j | \mathbf{X} \sim N\left(\beta_j, \sigma_{\hat{\beta}_j}^2\right)$.
- Standardisieren führt zu
$$\frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N(0, 1), \quad \text{keine Bed., da } \mathbf{X} \text{ nur in } \sigma_{\hat{\beta}_j} \text{ enthalten.}$$

(Ohne Beweis:) Müssen wir σ^2 schätzen, ist die Teststatistik **t -verteilt** mit $n - k - 1$ Freiheitsgraden. Indem wir $k + 1$ Regressionsparameter schätzen, ergeben sich aus den Normalgleichungen $k + 1$ Restriktionen, also gilt

$$t(\mathbf{X}, \mathbf{y}) = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t_{n-k-1}.$$

• Kritischer Bereich und Entscheidungsregel

– Zweiseitiger Test

* Hypothesen:

$$H_0 : \beta_j = \beta_{j0} \quad \text{versus} \quad H_1 : \beta_j \neq \beta_{j0}.$$

Für ein zuvor gewähltes Signifikanzniveau erhält man die kritischen Werte aus der Tabelle der *t*-Verteilung, wobei $P(T < -c|H_0) = \alpha/2$ und $P(T > c|H_0) = \alpha/2$ bzw. gleichbedeutend $2 \cdot P(T > c|H_0) = \alpha$ gelten muss.

* Entscheidungsregel:

- Lehne H_0 ab, falls $|t(\mathbf{X}, \mathbf{y})| > c$. Sonst lehne H_0 nicht ab.
- Alternativ: Berechne *p*-values

$$p = P(|T| > |t(\mathbf{X}, \mathbf{y})| | H_0) = 2 \cdot P(T > |t(\mathbf{X}, \mathbf{y})| | H_0)$$

und lehne H_0 ab, falls $p < \alpha$; ansonsten lehne H_0 nicht ab.

– Einseitiger Test mit linksseitiger Alternative

* Hypothesen:

$$H_0 : \beta_j \geq \beta_{j0} \quad \text{versus} \quad H_1 : \beta_j < \beta_{j0}.$$

Für ein gegebenes Signifikanzniveau lassen sich die kritischen Werte aus der Tabelle der *t*-Verteilung ablesen, wobei gelten muss

$$P(T < c | H_0) = \alpha.$$

* Entscheidungsregel:

- Lehne H_0 ab, falls $t(\mathbf{X}, \mathbf{y}) < c$; ansonsten lehne H_0 nicht ab.
- Alternativ: Berechne *p*-values

$$p = P(T < t(\mathbf{X}, \mathbf{y}) | H_0)$$

und lehne H_0 ab, falls $p < \alpha$; ansonsten lehne H_0 nicht ab.

– Einseitiger Test mit rechtsseitiger Alternative

* Hypothesen:

$$H_0 : \beta_j \leq \beta_{j0} \quad \text{versus} \quad H_1 : \beta_j > \beta_{j0}.$$

Für ein gegebenes Signifikanzniveau lassen sich die kritischen Werte aus der Tabelle der *t*-Verteilung ablesen, wobei gelten muss

$$P(T > c | H_0) = \alpha.$$

* Entscheidungsregel:

- Lehne H_0 ab, falls $t(\mathbf{X}, \mathbf{y}) > c$; ansonsten lehne H_0 nicht ab.
- Alternativ: Berechne *p*-values

$$p = P(T > t(\mathbf{X}, \mathbf{y}) | H_0)$$

und lehne H_0 ab, falls $p < \alpha$; ansonsten lehne H_0 nicht ab.

• Ökonomische versus statistische Signifikanz

- Für ein gegebenes (statistisches) Signifikanzniveau α steigt die Güte eines Test mit zunehmender Stichprobengröße an, weil $\sigma_{\hat{\beta}_j}$ im Nenner der Teststatistik in diesem Fall abnimmt.
- Lässt sich die Nullhypothese nicht ablehnen, kann dies auch daran liegen, dass die Stichprobe schlicht zu klein ist (falls die Nullhypothese in der Grundgesamtheit tatsächlich falsch sein sollte).
- Andererseits kann der Parameter einer Variablen bei einer großen Stichprobe auch dann signifikant sein, wenn der Einfluss dieser Variablen in der Grundgesamtheit nur sehr gering ist. Selbst wenn $\beta_j x_j$ **nur geringen ökonomischen** Einfluss auf die abhängige Variable ausübt, wäre der Parameter der Variablen (statistisch) signifikant.

- **Aber Vorsicht:** Um Verzerrungen wegen eines zu kleinen Modells zu vermeiden (omitted variable bias), müssen statistisch signifikante Variablen im Modell verbleiben, siehe Abschnitt 3.4.1.
- **Wahl des Signifikanzniveaus**
 - Es gibt zwei Gründe, warum das Signifikanzniveau α mit zunehmender Stichprobengröße n gesenkt werden sollte:
 - * Eine größere Stichprobe lässt die Güte des Modells ansteigen. Ändert man nichts am Signifikanzniveau führt der Anstieg der Stichprobengröße zu einer Reduktion des Fehlers 2. Art, $\beta(\theta) = 1 - \pi(\theta)$. Es könnte aber von Interesse sein, auch den Fehler 1. Art zu verringern. Er bezeichnet die Wahrscheinlichkeit, eine Variable in das Modell aufzunehmen, obwohl diese im Modell der Grundgesamtheit nicht relevant ist. Eine Reduktion dieses Fehlers ist hier durchaus sinnvoll.

- * Im Allgemeinen wählt man die relevanten Variablen aus einer großen Zahl potentiell einflussreicher Variablen. Da für jede statistisch signifikante Variable das Signifikanzniveau α gilt, werden im Durchschnitt αK überflüssige Variablen fälschlicherweise in das Modell aufgenommen (K steht für die Anzahl der insgesamt in Erwägung gezogener Variablen). Gewöhnlich lässt man K mit zunehmender Stichprobengröße n ansteigen, sodass das Signifikanzniveau α fallen sollte, wenn man vermeiden möchte, dass αK ansteigt.
- Falls man das Hannan-Quinn-(HQ) oder Schwarz-(SC) Modellselektionskriterium anwendet ((3.19) bzw. (3.20)), sinkt das Signifikanzniveau mit zunehmender Stichprobengröße. Beim AIC-Kriterium (3.18) ist dies nicht der Fall.

- **Insignifikanz, Multikollinearität und Stichprobengröße**

- Die Teststatistik $t(\mathbf{X}, \mathbf{y})$ ist (wie wir gesehen haben) klein, wenn
 - * die Abweichung der Nullhypothese vom wahren Wert klein ist, etwa zwischen β_j und β_{j0}
 - * oder der geschätzte Standardfehler $\hat{\sigma}_{\hat{\beta}_j}$ von β_j groß ist.

Letzteres kann auch von Multikollinearität in \mathbf{X} herrühren. Also: Ein hoher Grad an Multikollinearität macht die Ablehnung der Nullhypothese unwahrscheinlicher (weil dann $|t(\mathbf{X}, \mathbf{y})|$ im Durchschnitt klein ist).

- Dies ist etwa ein Grund, warum man insignifikante Variablen in der Regression beibehalten möchte. Die zugehörigen Parameterwerte sind aber mit Bedacht zu interpretieren.

Zu Lesen: Bei Bedarf Appendices C.5, E.3 in **Wooldridge (2009)**.

4.4 Empirische Analyse einer vereinfachten Gravitationsgleichung

Fortsetzung des Außenhandelsbeispiels (aus Abschnitt 3.5):

Vergleiche hierzu "Schritte einer ökonometrischen Untersuchung" aus Abschnitt 1.2.

1. Fragestellung:

Quantifizierung des Einflusses von BIP-Veränderungen in den Exportländern auf die Importe nach Deutschland.

2. Ökonomisches Modell:

Unter idealisierten Annahmen wie vollständige Spezialisierung der Produktion, identischen Konsumpräferenzen in den Ländern, keinen Transport- und Handelskosten, und einer ausschließlichen Beschränkung auf die Importe, unterstellt die Wirtschaftstheorie (siehe

Section II, equation (5) in **Fratianni (2007)**)

$$Importe_i = A \cdot BIP_i \cdot Entfernung_i^{\beta_2}, \quad \beta_2 < 0.$$

Dies impliziert eine Elastizität von *BIP* auf *Importe* in Höhe von 1. Eine 1%ige Veränderung des BIPs des exportierenden Landes führt demnach zu einer Steigerung der Importe um ebenfalls 1%.

Diese Hypothese lässt sich statistisch testen.

3. Ökonometrisches Modell:

Das einfachste ökonometrische Modell erhält man, wenn man logarithmiert und einen Fehlerterm hinzufügt. Dies ergibt

$$\ln(Importe_i) = \beta_0 + \beta_1 \ln(BIP_i) + \beta_2 \ln(Entfernung_i) + u_i.$$

4. Daten sammeln: siehe Appendix 10.4.

5. Wahl und Schätzung eines ökonometrischen Modells:

In der Realität dürften weitere Variablen Einfluss auf die Importe haben. Es müssen also weitere Kontrollvariablen hinzugefügt werden. Die Übung zur Modellwahl in Abschnitt 3.5 wies darauf hin, dass auf Basis des Schwarz-Kriteriums die Variable *Offenheit* hinzugenommen werden sollte:

(Modell 3)

$$\begin{aligned}\ln(\text{Importe}_i) = & \beta_0 + \beta_1 \ln(\text{BIP}_i) + \beta_2 \ln(\text{Entfernung}_i) \\ & + \beta_3 \text{Offenheit}_i + u_i.\end{aligned}$$

Call:

```
lm(formula = log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o)
```

Residuals:

Min	1Q	Median	3Q	Max
-2.1999	-0.5587	0.1009	0.5866	1.5220

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.74104	2.17518	1.260	0.2141
log(wdi_gdpusdcr_o)	0.94066	0.06134	15.335	< 2e-16 ***
log(cepii_dist)	-0.97032	0.15268	-6.355	9.26e-08 ***
ebrd_tfes_o	0.50725	0.19161	2.647	0.0111 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

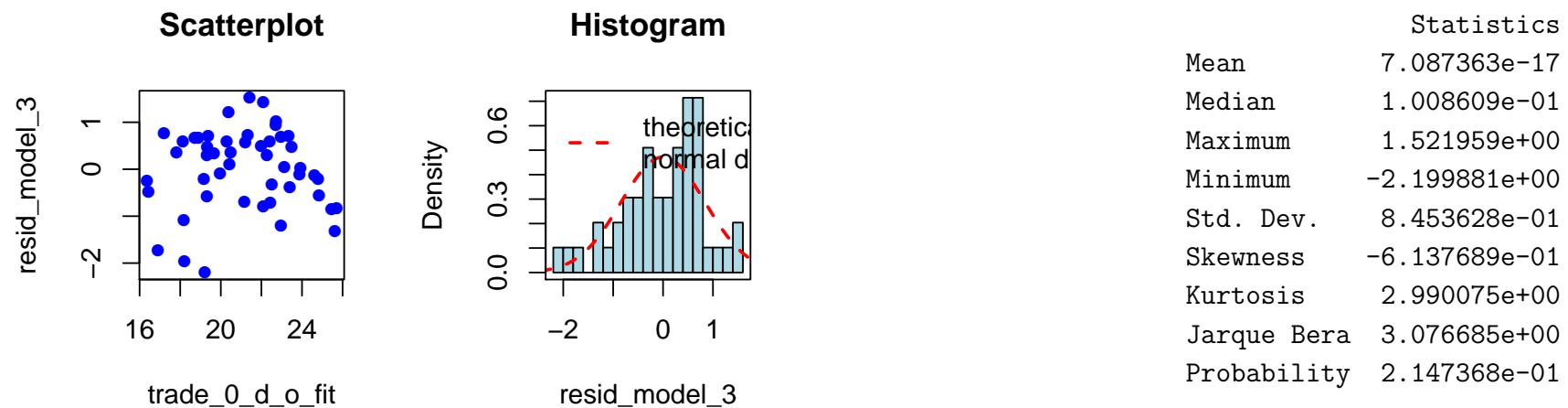
Residual standard error: 0.8731 on 45 degrees of freedom

Multiple R-squared: 0.8995, Adjusted R-squared: 0.8928

F-statistic: 134.2 on 3 and 45 DF, p-value: < 2.2e-16

6. Modelldiagnose:

- Prüfen einer evtl. Verletzung von MLR.5 (Homoskedastie). Dazu plotten wir die Residuen gegen die gefitteten Werte.
- Prüfen einer möglichen Verletzung von MLR.6 (Normalverteilte Fehler.) Dazu lassen wir uns ein Histogramm der Residuen anzeigen.



Der Scatterplot zeigt keine Verletzung von MLR.5 an. Warum?
 Statistische Tests hierfür werden in Abschnitt 9.2 besprochen.

Das Histogramm deutet allerdings auf eine asymmetrische Verteilung hin. Wäre dies der Fall, wären die Fehler nicht normalverteilt. Die **Asymmetrie einer Verteilung** wird u.a. durch das dritte Moment, die **Schiefe (Skewness)** gemessen. Die symmetrische Normalverteilung hat eine Schiefe von 0. Ein Blick in den Kasten rechts vom Histogramm zeigt, dass die geschätzte Schiefe -0.6 beträgt.

Das **vierte Moment, die Kurtosis (Wölbung)** wird nahe 3 geschätzt, dem Wert der Kurtosis einer Standardnormalverteilung.

Für Spezialisten: Ob das 3. und/oder 4. Moment (Schiefe und Wölbung) der Normalverteilung widersprechen, lässt sich mit dem **Lomnicki-Jarque-Bera-Test** überprüfen. Der p -value wird im Kasten in der letzten Zeile aufgelistet. Die Nullhypothese normalverteilter Fehler ist also auf keinem vernünftigen Signifikanzniveau abzulehnen.

Wir können mit unserem Modell also weiterarbeiten.

7. Benutzen des Modells: Durchführen von Tests: Zweiseitiger Test

- Wir können nun das statistische Hypothesenpaar aufstellen:

H_0 : Die BIP-Elastizität der Importe ist 1. versus H_1 : Die Elastizität ist ungleich 1.

$H_0 : \beta_1 = 1$ versus $H_1 : \beta_1 \neq 1$.

- Berechne die t -Statistik aus der passenden Zeile des Outputs

	Estimate	Std. Error	t value	Pr(> t)
log(wdi_gdpusdcr_o)	0.94066	0.06134	15.335	< 2e-16 ***

$$t(\mathbf{X}, \mathbf{y}) = \frac{\hat{\beta}_1 - \beta_{10}}{\hat{\sigma}_{\hat{\beta}_1}} = \frac{0.94066 - 1}{0.06134} = -0.9673948$$

- Wähle Signifikanzniveau, z.B. $\alpha = 0.05$.

Berechne kritische Werte: Wir haben $n - k - 1 = 49 - 3 - 1 = 45$

Freiheitsgrade. Aus Table G.2 in [Wooldridge \(2009\)](#) lässt sich ein ungefährer kritischer Wert ermitteln, einen präzisen kritischen Wert erhält man z.B. mit

- R: `(crit <- qt(0.975, df = 49 - 3 -1))` im Kommandofenster ergibt 2.014103 (Dezimaltrennzeichen ist ".") oder
- Excel: Man wendet die Formel $c = \text{TINV}(\text{alpha}; n-k-1) = 2.0141$ an. (Beachte, dass Excel stillschweigend bereits einen zweiseitigen Test annimmt.)

- Da

$$-c < t(\mathbf{X}, \mathbf{y}) < c$$

$$-2.014103 < -0.9673948 < 2.014103$$

können wir die Nullhypothese nicht ablehnen.

- p -values lassen sich in R berechnen über

```
pval <- 2 * pt(teststat, df = 49-3-1) = 0.3385174.
```

Demnach lässt sich H_0 selbst auf dem 10% Signifikanzniveau nicht ablehnen. Der p -value besagt, dass wir unter H_0 in etwa 34 von 100 Stichproben eine t -Statistik erhalten würden, deren Absolutbetrag mindestens 0.9673948 beträgt.

Einseitiger Test

- Wir können auch eine Hypothese bezüglich des Vorzeichens von β_2 aufstellen, z.B. dem Einfluss von *Entfernung* auf *Importe*. Da wir an Evidenz für $\beta_2 < 0$ interessiert sind, packen wir dies in H_1 :

$$H_0 : \beta_2 \geq 0 \quad \text{versus} \quad H_1 : \beta_2 < 0.$$

- Berechne die t -Statistik aus der passenden Zeile des Outputs

Estimate	Std. Error	t value	Pr(> t)
-9.703183e-01	1.526847e-01	-6.355048e+00	9.262691e-08

$$t(\mathbf{X}, \mathbf{y}) = \frac{\hat{\beta}_2 - \beta_{20}}{\hat{\sigma}_{\hat{\beta}_2}} = \frac{-0.9703183 - 0}{0.1526847} = -6.355046.$$

- Wir wählen wieder $\alpha = 0.05$ und berechnen den kritischen Wert über die R-Funktion

$$qt(1-0.05, \text{ df}=49-3-1) = 1.679427.$$

- Wegen

$$t(\mathbf{X}, \mathbf{y}) = -6.3550 < -1.6794 = c,$$

lehnen wir die Nullhypothese ab. Somit hat beim gegebenen Signifikanzniveau die logarithmierte Entfernung statistisch signifikant negativen Einfluss auf die Importe.

- Als zugehörigen *p*-value erhalten wir mittels R

$pt(\text{teststat}, \text{ df}=49-3-1) = 4.631369\text{e-08}$. Die Entfernung hat also selbst auf dem 1% Signifikanzniveau negativen Einfluss.

Beachte, dass wir in Abschnitt 3.5 bereits andere Modellspezifikationen in Betracht gezogen haben. Es könnte interessant sein, zu prüfen, ob die Testergebnisse auch dann robust sind, wenn andere Modellspezifikationen verwendet werden wie etwa Modell 2 oder Modell 4.

4.5 Konfidenzintervalle/Intervallschätzung

- Wie groß ist eigentlich die Wahrscheinlichkeit, dass der geschätzte Parameterwert tatsächlich dem “wahren” Parameterwert entspricht?
- Ein Parameterschätzer — oder genauer, ein Punktschätzer — lässt keine Rückschlüsse darüber zu, wie “nahe” dieser Wert am wahren Wert der Grundgesamtheit liegt.
- Wissenschaftstheoretische Bedenken gegen Punktschätzungen liefert die Position des **kritischen Rationalismus** (Sir Karl Popper): Eine empirische Aussage ist nur dann wissenschaftlich, wenn sie prinzipiell **falsifizierbar** ist.
- Beispiel: Angenommen, wir machen im Rahmen eines Modells eine Schätzung für einen Preisindex und erhalten den Wert 5.12. Tatsächlich liegt aber ein Wert von 5.24 vor. → Wir machen dann

eine **falsche** Vorhersage, weil die Vorhersage nicht exakt eingetreten ist.

Dieser “Fehler” kann drei Gründe haben:

- Zufallsfehler des Regressionsmodells der Grundgesamtheit.
- Schätzfehler des Regressionsmodells der Stichprobe.
- Das Regressionsmodell ist nicht korrekt oder (realistischer) ist keine gute Annäherung. Mindestens eine unserer Annahmen ist unbegründet.

Problem:

Von einem subjektiven Standpunkt aus kann man unterschiedlicher Meinung bezüglicher dieser “Erklärungen” sein:

- Der eine glaubt, dass sich die Abweichung auf den Zufallsfehler zurückführen lässt.
- Eine andere meint, das Modell sei falsch.

Lösung:

Es sollten also **objektive Kriterien** festgelegt werden und zwar bevor der vorherzusagende Wert eintritt, damit eine objektive (wissenschaftliche) Entscheidung getroffen werden kann.

Damit kann man sich nicht nachträglich einer Falsifikation seiner Aussage entziehen, eine Aussage wird daher prinzipiell falsifizierbar und im Sinne von Popper wissenschaftlich.

- Wir präzisieren unsere Eingangsfrage:

Wie groß ist eigentlich die Wahrscheinlichkeit, dass der geschätzte Parameterwert $\hat{\beta}_j$ tatsächlich dem wahren Parameterwert β_j entspricht, wenn (wie in Abschnitt 4.3 gezeigt wurde) ,

$$\hat{\beta}_j \sim N\left(\beta_j, \sigma_{\hat{\beta}_j}^2\right)$$

und $(\hat{\beta}_j - \beta_j)/\sigma_{\hat{\beta}_j} \sim N(0, 1)$, bzw. falls $\sigma_{\hat{\beta}_j}$ geschätzt wurde,

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t_{n-k-1}?$$

- **Alternative Frage:**

Wie groß ist die Wahrscheinlichkeit, dass für ein gegebenes c der wahre Wert β_j vor *Beobachtung einer Stichprobe* in folgendem Intervall liegt

$$[\hat{\beta}_j - c \cdot \hat{\sigma}_{\hat{\beta}_j}, \hat{\beta}_j + c \cdot \hat{\sigma}_{\hat{\beta}_j}]?$$

Man beachte, dass die beiden Grenzen des Intervalls vor *Erhebung einer Stichprobe* Zufallszahlen sind. Die **Lage ist zufällig**, weil $\hat{\beta}_j$ eine Zufallsvariable ist, und seine **Länge ist zufällig** wegen $\hat{\sigma}_{\hat{\beta}_j}$.

Dieses Intervall ist das bekannteste Beispiel für einen **Intervallschätzer**.

- **Antwort für ein gegebenes $\sigma_{\hat{\beta}_j}$:**

Wie groß ist die Wahrscheinlichkeit, dass der wahre Wert β_j im Intervall $[\hat{\beta}_j - c \cdot \sigma_{\hat{\beta}_j}, \hat{\beta}_j + c \cdot \sigma_{\hat{\beta}_j}]$ liegt, das vor Ziehung einer

Stichprobe zufällig ist und wobei der Wert c von uns gewählt wird?

– Sie ist $2\Phi(c) - 1$, da

$$\begin{aligned}
 P\left(\hat{\beta}_j - c\sigma_{\hat{\beta}_j} \leq \beta_j \leq \hat{\beta}_j + c\sigma_{\hat{\beta}_j}\right) &= P\left(-c\sigma_{\hat{\beta}_j} \leq \beta_j - \hat{\beta}_j \leq c\sigma_{\hat{\beta}_j}\right) \\
 &= P\left(-c \leq \frac{\beta_j - \hat{\beta}_j}{\sigma_{\hat{\beta}_j}} \leq c\right) \\
 &= P\left(-c \leq \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \leq c\right) \\
 &= \Phi(c) - \Phi(-c) \\
 &= \Phi(c) - (1 - \Phi(c)) \\
 &= 2\Phi(c) - 1.
 \end{aligned}$$

- Beispiel: Für $c = 1.96$ erhält man $\Phi(1.96) - \Phi(-1.96) = 0.975 - 0.025 = 0.95$: Der wahre Wert β_j liegt mit 95%iger Wahrscheinlichkeit innerhalb des Intervalls $\hat{\beta}_j \pm c \cdot \sigma_{\hat{\beta}_j}$. Diese Wahrscheinlichkeit kann man auch mit α in Beziehung setzen, indem man $0.95 = 1 - \alpha$ schreibt. Hier ist also $\alpha = 0.05$.
- **Antwort für ein geschätztes $\sigma_{\hat{\beta}_j}$:** Der wahre Wert β_j liegt mit Wahrscheinlichkeit $1 - \alpha$ im Intervall $\hat{\beta}_j \pm c \cdot \hat{\sigma}_{\hat{\beta}_j}$. Beachte, dass man für die Berechnung der Wahrscheinlichkeit die t_{n-k-1} -Verteilung verwenden muss, da

$$P\left(\hat{\beta}_j - c\hat{\sigma}_{\hat{\beta}_j} \leq \beta_j \leq \hat{\beta}_j + c\hat{\sigma}_{\hat{\beta}_j}\right) = P\left(-c \leq \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \leq c\right).$$

- Das Intervall

$$[\hat{\beta}_j - c \cdot \hat{\sigma}_{\hat{\beta}_j}, \hat{\beta}_j + c \cdot \hat{\sigma}_{\hat{\beta}_j}]$$

wird als **Konfidenzintervall** bezeichnet.

Man sagt, dass das Konfidenzintervall den wahren Wert mit einer Vertrauenswahrscheinlichkeit von $(1 - \alpha)100\%$ enthält. Der Wert $(1 - \alpha)$ wird auch **Konfidenzniveau** oder **coverage probability** des Konfidenzintervalls genannt.

- **In der Praxis** gibt man sich einen Wert $1 - \alpha$ für das Konfidenzniveau vor und berechnet dann c anhand der passenden Verteilung: $N(0, 1)$ oder t_{n-k-1} .
- **Interpretation:** Würde man R mal jeweils neue Stichproben aus derselben Grundgesamtheit ziehen und für jede Stichprobe ein Konfidenzintervall für gegebenes Konfidenzniveau $1 - \alpha$ berechnen, dann würde in ungefähr $(1 - \alpha)R$ aller Stichproben der wahre Wert im

jeweiligen Konfidenzintervall liegen.

- **Beachte:**

- Liegt bereits eine Stichprobe vor, dann ist der wahre Parameter entweder in dem auf Basis der beobachteten Stichprobe berechneten Konfidenzintervall enthalten oder nicht. Mit anderen Worten, es macht keinen Sinn, bei einer bereits vorliegenden Stichprobe von einer Überdeckungswahrscheinlichkeit bezüglich der vorliegenden Stichprobe zu sprechen.
- Die Konstante c entspricht dem (oberen) kritischen Wert eines zweiseitigen Tests mit Signifikanzniveau α .
- Das Konfidenzintervall ist ein Zufallsintervall und demnach sind seine Lage und Länge i.A. für jede Stichprobe unterschiedlich.
- Je größer $(1-\alpha)$ (je kleiner also α), desto größer ist das Konfiden-

zintervall. Mit anderen Worten: Je sicherer man gehen möchte, desto “ungenauer” wird das Intervall. Warum?

- Ein zweiseitiger t -Test und ein Konfidenzintervall beinhalten dieselbe Information. Die Nullhypothese eines zweiseitigen t -Tests wird dann und nur dann abgelehnt, wenn der Wert der Nullhypothese außerhalb des Konfidenzintervalls liegt. Um sich dies klar zu machen, fertigen Sie dazu eine Zeichnung an.
- Ein Konfidenzintervall für eine Stichprobe enthält alle Nullhypotesen eines zweiseitigen t -Tests, die nicht zum Signifikanzniveau α abgelehnt werden können.
- Wenn man immer neue Stichproben aus der Grundgesamtheit zieht, wie viele Konfidenzintervalle werden im Durchschnitt den wahren Wert nicht enthalten?

- **Fortsetzung des Außenhandelsbeispiels** (aus Abschnitt 4.4):

- Berechne ein 95%-Konfidenzintervall für die BIP-Elastizität der Importe β_{BIP} .
- Aus Abschnitt 4.4 kann geschlossen werden, dass MLR.1 bis MLR.6 gelten und die logarithmierten Importe somit normalverteilt sind.
- Da $\sigma_{\hat{\beta}_{BIP}}$ geschätzt werden muss, muss die t -Verteilung mit $n - k - 1 = 45$ Freiheitsgraden verwendet werden. Für ein Konfidenzniveau von 0.95 erhält man $\alpha = 0.05$ und somit $c = 2.014103$ (z.B. in R via `qt(1-0.05/2, df = 49 - 3 - 1)`).
- Die relevante Zeile des Outputs sah wie folgt aus, siehe Abschnitt 4.4:

	Estimate	Std. Error	t value	Pr(> t)
<code>log(wdi_gdpusdcr_o)</code>	0.94066	0.06134	15.335	< 2e-16 ***

- Somit ist das 95%-Konfidenzintervall gegeben durch

$$[\hat{\beta}_{BIP} - c \cdot \hat{\sigma}_{\hat{\beta}_{BIP}}, \hat{\beta}_{BIP} + c \cdot \hat{\sigma}_{\hat{\beta}_{BIP}}]$$
$$[0.94066 - 2.014103 \cdot 0.06134, 0.94066 + 2.014103 \cdot 0.06134]$$
$$[0.81712, 1.06421].$$

- Alle Nullhypotesen für die BIP-Elastizität im Konfidenzintervall $[0.81712, 1.06421]$ können gegeben ein Konfidenzniveau von 95% nicht abgelehnt werden. Beachte, dass der Wert 1 im Konfidenzintervall enthalten ist. Dies bestätigt das Testergebnis aus Abschnitt 4.4, wo wir $H_0 : \beta_{BIP} = 1$ nicht ablehnen konnten.

4.6 Testen einer einzelnen Linearkombination der Parameter

- **Beispiel:** Cobb-Douglas Produktionsfunktion

$$\log Y = \beta_0 + \beta_1 \log K + \beta_2 \log L + u,$$

wobei Y den Output bezeichnet; K und L bezeichnen die Produktionsfaktoren Kapital bzw. Arbeit. Beachte, dass β_1 und β_2 in diesem Fall Elastizitäten darstellen.

Falls die Restriktion $\beta_1 + \beta_2 = 1$ gilt, hat die Produktionsfunktion konstante Skalenerträge, d.h. ein 1%iger Anstieg der Arbeit und des Kapitals führen im Schnitt zu einer 1%igen Steigerung des Outputs.

Um einen empirischen Test auf konstante Skalenerträge durchzuführen, verwenden wir das folgende Hypothesenpaar:

$$H_0 : \beta_1 + \beta_2 = 1 \quad \text{versus} \quad H_1 : \beta_1 + \beta_2 \neq 1.$$

- **Wie konstruiert man die Teststatistik?**

1. Zuerst definieren wir die Hilfsparameter θ und θ_0 , wobei

$$\theta = \beta_1 + \beta_2, \quad \theta_0 = 1,$$

oder entsprechend

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0.$$

2. Anschließend lösen wir θ nach einem der Parameter β_i auf, hier β_1

$$\beta_1 = \theta - \beta_2$$

und setzen dies in die Gleichung der Ausgangsregression ein. Nun werden noch einige Umformungen vorgenommen, also

$$\begin{aligned} \log Y &= \beta_0 + (\theta - \beta_2) \log K + \beta_2 \log L + u \\ \log Y &= \beta_0 + \theta \log K + \beta_2 \underbrace{(\log L - \log K)}_{\text{neue Variable}} + u. \end{aligned} \quad (4.4)$$

Daraufhin kann (4.4) geschätzt werden und man erhält die Teststatistik

$$t_\theta = \frac{\hat{\theta} - \theta_0}{\hat{\sigma}_{\hat{\theta}}},$$

die direkt mit den Schätzungen aus (4.4) berechnet werden kann.

Beispiel:

In einem klassischen Modell aus dem Marketing werden die Verkäufe (S) (bzw. deren natürlicher Logarithmus) eines Konsumgutes auf den (natürlichen Logarithmus) des Preises dieses Gutes (P) sowie auf den (natürlichen Logarithmus) der Kreuzpreise (P_{K1}, P_{K2}) konkurrierender Güter regressiert. Aus den zur Verfügung stehenden Daten erhält man für diese Regression folgenden Output:

Call:

```
lm(formula = log(S) ~ log(P) + log(P_K1) + log(P_K2))
```

Residuals:

Min	1Q	Median	3Q	Max
-4.8760	-0.6421	-0.0098	0.6352	3.7577

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.40779	0.07956	55.40	<2e-16 ***
log(P)	-3.95528	0.06809	-58.09	<2e-16 ***
log(P_K1)	0.71027	0.07391	9.61	<2e-16 ***
log(P_K2)	1.15416	0.07982	14.46	<2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.022 on 6913 degrees of freedom

Multiple R-squared: 0.3323, Adjusted R-squared: 0.332

F-statistic: 1147 on 3 and 6913 DF, p-value: < 2.2e-16

Wir wollen folgende Behauptung prüfen: Die Elastizitäten der Kreuzpreise sind identisch (obwohl die konkurrierenden Güter aus unterschiedlichen Marktsegmenten stammen), alles andere konstant gehalten (*ceteris paribus*).

- Die Ausgangshypothesen sind demnach

$$H_0 : \beta_{K1} = \beta_{K2} \quad \text{versus} \quad H_1 : \beta_{K1} \neq \beta_{K2}.$$

Die Reparameterisierung führt dann zu

$$\theta = \beta_{K1} - \beta_{K2}, \quad \theta_0 = 0$$

$$H_0 : \theta = 0 \quad \text{versus} \quad H_1 : \theta \neq 0.$$

- Wegen $\beta_{K1} = \theta + \beta_{K2}$, lässt sich das Ausgangsmodell

$$\ln(S) = \beta_1 + \beta_2 \ln(P) + \beta_{K1} \ln(P_{K1}) + \beta_{K2} \ln(P_{K2}) + u$$

überführen in

$$\ln(S) = \beta_1 + \beta_2 \ln(P) + \theta \ln(P_{K1}) + \beta_{K2}(\ln(P_{K2}) + \ln(P_{K1})) + u.$$

- Aus den Schätzungen der letzten Regressionsgleichung

```
lm(formula = log(S) ~ log(P) + log(P_K1) + I(log(P_K1) + log(P_K2)))
```

Residuals:

Min	1Q	Median	3Q	Max
-4.8760	-0.6421	-0.0098	0.6352	3.7577

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.40779	0.07956	55.403	< 2e-16 ***
log(P)	-3.95528	0.06809	-58.085	< 2e-16 ***
log(P_K1)	-0.44389	0.11254	-3.944	8.09e-05 ***
I(log(P_K1) + log(P_K2))	1.15416	0.07982	14.460	< 2e-16 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.022 on 6913 degrees of freedom

Multiple R-squared: 0.3323, Adjusted R-squared: 0.332

F-statistic: 1147 on 3 and 6913 DF, p-value: < 2.2e-16

berechnet man die *t*-Statistik

$$t = \frac{-0.44389 - 0}{0.112544} \approx -3.94, \text{ Exakter Wert: } -3.944165.$$

Für ein gegebenes Signifikanzniveau $\alpha = 0.05$ sind die kritischen Werte -1,96 und 1,96. Demnach ist H_0 abzulehnen.

Zu Lesen: Abschnitte 4.3-4.4 in [Wooldridge \(2009\)](#).

4.7 Gemeinsames Testen mehrerer Linearkombinationen der Parameter: *F*-Test

Beispiel möglicher Restriktionen im Rahmen des MLR:

1. $H_0 : \beta_1 = 3$
2. $H_0 : \beta_2 = \beta_k$
3. $H_0 : \beta_1 = 1, \beta_k = 0$
4. $H_0 : \beta_1 = \beta_3, \beta_2 = \beta_4$
5. $H_0 : \beta_j = 0, j = 1, \dots, k$
6. $H_0 : \beta_j + 2\beta_l = 1, \beta_k = 2$

Fall 1. und 2. können wir bereits mithilfe des *t*-Tests prüfen. In allen anderen Fällen wird der *F*-Test benötigt.

4.7.1 Testen mehrerer Null- bzw. Ausschlussrestriktionen

Fortsetzung des Außenhandelsbeispiels (aus Abschnitt 4.5):

Wir betrachten Modell 4 aus Abschnitt 3.5:

```
lm(formula = log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o + log(cepii_area_o))
```

Residuals:

Min	1Q	Median	3Q	Max
-2.1825	-0.6344	0.1613	0.6301	1.5243

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.42778	2.13258	1.138	0.2611
log(wdi_gdpusdcr_o)	1.02502	0.07654	13.392	< 2e-16 ***
log(cepii_dist)	-0.88865	0.15614	-5.691	9.57e-07 ***
ebrd_tfes_o	0.35315	0.20642	1.711	0.0942 .
log(cepii_area_o)	-0.15103	0.08523	-1.772	0.0833 .

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

Residual standard error: 0.853 on 44 degrees of freedom

Multiple R-squared: 0.9062, Adjusted R-squared: 0.8976

F-statistic: 106.2 on 4 and 44 DF, p-value: < 2.2e-16

Werden die Kontrollvariablen *Offenheit* (`EBRD_TFES_0`) und *Fläche* (`LOG(CEPII_AREA_0)`) in der Spezifikation des Modells 4 tatsächlich gebraucht?

Genauer: Sind die beiden Parameter der angesprochenen Variablen gemeinsam signifikant von Null verschieden?

$$H_0 : \beta_{Offenheit} = 0 \text{ und } \beta_{Flaeche} = 0$$

versus

$$H_1 : \beta_{Offenheit} \neq 0 \text{ und/oder } \beta_{Flaeche} \neq 0$$

Wie lassen sich mehrere Hypothesen gemeinsam testen?

- Wir wissen, dass die SSR sinkt (oder konstant bleibt), wenn wir zusätzliche Regressoren aufnehmen.
⇒ **Idee:** Vergleiche die SSR des Modells unter Berücksichtigung der Nullrestriktionen (restringiertes Modell) mit der SSR des Ausgangsmodells (unrestringiertes Modell).
- Die Schätzung unter H_0 ist simpel : Man entfernt alle Regressoren aus dem Modell deren Parameter unter H_0 Null sein sollen und schätzt dieses restringierte Modell.
Im Falle des Modells 4 des Außenhandelsbeispiels erhält man folgende Schätzungen (diese entsprechen denen des Modells 2 aus Abschnitt 3.5):

Call:

```
lm(formula = log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.99289	-0.58886	-0.00336	0.72470	1.61595

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.67611	2.17838	2.147	0.0371 *
log(wdi_gdpusdcr_o)	0.97598	0.06366	15.331	< 2e-16 ***
log(cepii_dist)	-1.07408	0.15691	-6.845	1.56e-08 ***

Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Residual standard error: 0.9284 on 46 degrees of freedom

Multiple R-squared: 0.8838, Adjusted R-squared: 0.8787

F-statistic: 174.9 on 2 and 46 DF, p-value: < 2.2e-16

Ergebnisse:

- Das R^2 des unrestringierten Modells beträgt 0.9062, während das R^2 des restringierten Modells 0.8838 beträgt.
- Der Standardfehler der Regression $\hat{\sigma}$ steigt dementsprechend von 0.853 auf 0.9284.
- Sind diese Unterschiede groß? Es sieht zwar nicht unbedingt so aus, aber was bedeutet “groß” in diesem Fall eigentlich?
- Beachte, dass alle drei Modellselektionskriterien, AIC, HQ, und SC, das unrestringierte Modell “bevorzugen”, siehe Abschnitt 3.5. Wird dies durch den Test bestätigt?

- Um eine Statistik (eine Funktion, die aus Stichprobenwerten berechnet werden kann) als Teststatistik benutzen zu können, muss ihre Wahrscheinlichkeitsverteilung unter H_0 bekannt sein.

Man kann zeigen (→ Masterveranstaltung Methoden der Ökonometrie oder Abschnitt 4.4 in [Davidson und MacKinnon \(2004\)](#)), dass die folgende Teststatistik einer F -Verteilung folgt

$$F = \frac{(\text{SSR}_{H_0} - \text{SSR}_{H_1})/q}{\text{SSR}_{H_1}/(n - k - 1)} \sim F_{q, n - k - 1}.$$

Deswegen wird dieser Test auch **F -Test** genannt und die Teststatistik wird kurz als **F -Statistik** bezeichnet.

- Beachte, dass die **F -Verteilung** zwei unterschiedliche Freiheitsgrade besitzt, zum einen die q Freiheitsgrade der Zufallsvariable im Zähler, zum anderen die $n - k - 1$ Freiheitsgrade der Zufallsvariable im Nenner.

Der Wert q bezeichnet die **Anzahl der gemeinsam getesteten Restriktionen**.

- **Details** der F -Statistik:
 - Ihr kleinstmöglicher Wert ist 0, da $\text{SSR}_{H_0} \geq \text{SSR}_{H_1}$ und $\text{SSR}_{H_1} > 0$. (Die F -Statistik kann demnach also nicht normalverteilt sein!)
 - Eine obere Schranke existiert nicht.
- Wann sollte die gemeinsame Nullhypothese abgelehnt werden?
 - Je größer der Absolutbetrag der Differenz der SSRs des restri- gierten und unrestringierten Modells, $\text{SSR}_{H_0} - \text{SSR}_{H_1}$, desto wahr- scheinlicher ist es, dass die Ausschlussrestriktionen ungültig sind, denn dann tragen die ausgeschlossenen Variablen dazu bei, dass die SSR im unrestringierten Modell deutlich kleiner ausfällt, was auf einen möglichen Erklärungsgehalt dieser Variablen hinweist.

- Jedoch haben absolute Differenzen alleine wenig Aussagekraft. Warum?
- Es hat deutlich mehr Sinn, die relativen Unterschiede der SSRs zu betrachten. Genau dies macht die F -Statistik. Sie skaliert die SSRs auf Basis der SSR des unrestringierten Modells. Ist die relative Differenz groß, ist die gemeinsame Nullhypothese wahrscheinlich verletzt.
- Ist der relative Unterschied klein, dann ist es wahrscheinlich, dass die ausgeschlossenen Variablen keinen relevanten Einfluss im unrestringierten Modell auf den Regressanden haben und somit weggelassen werden können.

- **Entscheidungsregel:**

Lehne H_0 ab, falls die Teststatistik größer ist als der **kritische Wert**:

Lehne H_0 ab, falls $F > c$.

Der **kritische Bereich** ist also (c, ∞) .

Berechnung des kritischen Bereichs:

Für ein gegebenes Signifikanzniveau α ist der kritische Wert c über die Wahrscheinlichkeit

$$P(F > c | H_0) = \alpha$$

definiert. Den entsprechenden Wert für c gegeben α findet man in den Tabellen für die F -Verteilung, z.B. Table G.3 in Appendix G in [Wooldridge \(2009\)](#) oder berechnet man ihn mittels R (`qf(1-alpha, df1=q, df2= n-k-1)`) oder Excel (`Finv(0,05;q;n-k-1)` für $\alpha=0,05$).

Fortsetzung des Außenhandelsbeispiels (vom Beginn dieses Abschnitts):

- Da die gemeinsame Nullhypothese zwei Ausschlussrestriktionen enthält, besitzt der Zähler $q = 2$ Freiheitsgrade. Die Freiheitsgrade des Nenners entsprechen den Freiheitsgraden des Modells 4, $n - k - 1 = 49 - 4 - 1 = 44$. Wir wählen ein Signifikanzniveau $\alpha = 0.05$ und schlagen den passenden kritischen Wert in Table G.3 in Appendix G in [Wooldridge \(2009\)](#) nach. Die aufgeführten Werte sind $F_{2,40} = 3,23$ und $F_{2,60} = 3,15$. Dem ersten Wert entspricht ein in Wahrheit kleineres Signifikanzniveau als 0,05, dem zweiten Wert eines über 0,05. Ist man an einem exakten Wert interessiert, erhält man diesen aus R, nämlich $qf(1-0.05, 2, 44) = 3.209278$.
- Aus den Standardfehlern und Freiheitsgraden der Modelle 4 und 2 lassen sich jeweils die SSRs (SSR = (Residual standard

$\text{error})^2 * \text{df}$) berechnen und damit die F -Statistik als

$$F = \frac{(39,64485 - 32,01770)/2}{32,01770/44} = 5,240768.$$

Da

$$F = 5,240768 > 3,20928 = c,$$

lehnen wir H_0 auf einem Signifikanzniveau von 5% ab.

- Überprüfen Sie, dass diese Entscheidung auch auf einem Signifikanzniveau von 1% bestehen bleibt.

Die beiden Variablen Offenheit ($EBRD_TFES_O$) und Fläche ($LOG(CEPII_AREA_O)$) sind auf dem 1% Signifikanzniveau statistisch signifikant und somit hat mindestens eine der beiden Variablen sowohl auf dem 5% als auch auf dem 1%-Niveau Einfluss auf die Importe.

Berechnung von p -values der F -Statistik:

- In der Praxis ist man häufig am *größten* Signifikanzniveau interessiert, bei dem die Nullhypothese bei **gegebener** Teststatistik gerade noch nicht abgelehnt werden kann.

Wie in Abschnitt 4.1 bereits erläutert wurde, liefert uns der p -value genau diese Information. Alternativ gibt dieser das *kleinste* Signifikanzniveau an, bei dem die Nullhypothese gerade noch abgelehnt werden kann.

Die Nullhypothese wird dann abgelehnt, wenn der p -value *kleiner* ist als das Signifikanzniveau α , dass man sich vorgegeben hatte, bevor man die Berechnungen angestellt hat.

- **Fortsetzung des Außenhandelsbeispiels:** Der p -value kann in Excel berechnet werden ($=FVERT(5,24077;2;44) = 0,00909$). Der p -value kann auch in R berechnet werden:

$$1 - pf(5.24077, df1 = 2, df2 = 44) = 0.00908809.$$

Es liegt also starke statistische Evidenz gegen die Nullhypothese vor.

Direkte Berechnung der F -Statistik in R:

- Hierzu verwendet man das R-Paket car, das beim ersten Mal mit `install.packages("car")` installiert werden muss. In jedem Fall muss das Paket mit dem Befehl `library(car)` geladen werden.
- Um den F-Test durchzuführen, verwendet man den Befehl `linearHypothesis(model, ...)`. Im vorliegenden Fall wird die Restriktion folgendermaßen eingegeben:

```
linearHypothesis(model_4, c("ebrd_tfes_o = 0",
```

"log(cepii_area_o) = 0"). Man erhält

Linear hypothesis test

Hypothesis:

ebrd_tfes_o = 0

log(cepii_area_o) = 0

Model 1: restricted model

Model 2: $\log(\text{trade_0_d_o}) \sim \log(\text{wdi_gdpusdcr_o}) + \log(\text{cepii_dist}) + \text{ebrd_tfes_o} + \log(\text{cepii_area_o})$

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	46	39.645				
2	44	32.018	2	7.6272	5.2408	0.009088 **

						Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Bemerkungen:

- Man kann natürlich auch eine einfache Hypothese mit zweiseitiger Alternative

$$H_0 : \beta_j = 0 \quad \text{versus} \quad H_1 : \beta_j \neq 0$$

mittels eines F -Tests überprüfen.

Es gilt, dass das Quadrat einer Zufallsvariablen X , die einer t -Verteilung mit $n - k - 1$ Freiheitsgraden folgt, gerade einer Zufallsvariablen entspricht, die der F -Verteilung mit $(1, n - k - 1)$ Freiheitsgraden folgt

$$X \sim t_{n-k-1} \implies X^2 \sim F_{1, n-k-1}.$$

Für das oben angegebene Hypothesenpaar führen ein zweiseitiger t -Test und ein F -Test zu genau den gleichen Ergebnissen.

- Es kann passieren, dass die Regressoren jeweils einzeln getestet nicht statistisch signifikant sind, beim gemeinsamen Test (auf dem gleichen Signifikanzniveau) aber doch statistisch signifikant sind. Dies ist ein Zeichen für Multikollinearität zwischen den Regressoren. Die gegebene Stichprobengröße reicht dann nur aus, um die gemeinsame statistische Signifikanz der beiden Regressoren zu ermitteln. Sie ist aber nicht groß genug, um jedem Regressor einzeln Evidenz für deren Signifikanz nachzuweisen. In diesem Fall kann man die Kovarianz der dem Test unterworfenen Parameterschätzungen überprüfen (in R: `vcov(model)` liefert die Kovarianzmatrix der Parameterschätzer)
- Es kann auch passieren, dass eine Variable für sich genommen statistisch signifikant ist, beim gemeinsamen Test mit anderen Variablen aber nicht-signifikant wird. Dies passiert dann, wenn in der gemeinsamen Hypothese Variablen vorhanden sind, die für das Regressions-

modell der Grundgesamtheit überflüssig sind. In diesem Fall wird die Güte der einzelnen Hypothese durch die Irrelevanz der anderen Variablen abgeschwächt.

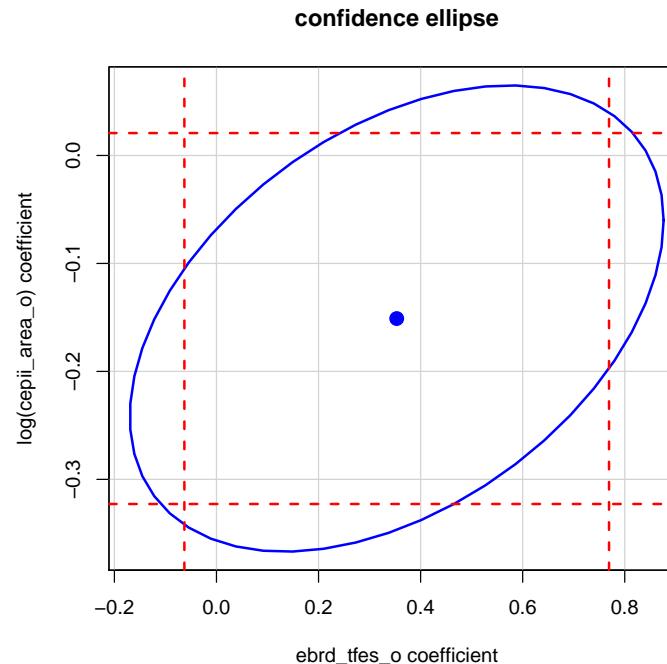
- Somit gibt es keine allgemeingültige Regel, ob einzelne oder gemeinsame Tests bevorzugt werden sollten.
- **Fortsetzung des Außenhandelsbeispiels** (von Mitte dieses Abschnitts): Vergleicht man die vier verschiedenen Modellspezifikationen mittels Modellselektionskriterien miteinander, siehe Abschnitt 3.5, bevorzugt das AIC das Modell 4 (SC bevorzugt Modell 3). Sieht man sich die Parameterschätzungen aus Modell 4 genauer an, stellt man fest, dass zwei Parameter selbst auf dem 5% Niveau statistisch insignifikant sind: $\beta_{Offenheit}$ und $\beta_{Flaeche}$.

Warum wurde Modell 4 bei Verwendung des AIC als das beste ausgewiesen und nicht etwa Modell 2, das die beiden insignifikanten Variablen nicht enthält?

Antwort:

Die Parameterschätzungen für $\beta_{Offenheit}$ und $\beta_{Flaeche}$ sind möglicherweise in hohem Maße korreliert, sodass nur deren gemeinsamer Einfluss signifikant ist. Ein Grund hierfür könnte sein, dass, neben anderen Faktoren, ein großer Teil der Streuung in *Fläche* durch die Streuung in *Offenheit* erklärt werden kann. Obiger Test hatte uns ja gezeigt, dass beide gemeinsam selbst auf dem 1%-Niveau signifikant sind.

Die Auswirkungen von Multikollinearität werden deutlich in



Die Ellipse ist eine Verallgemeinerung des Konfidenzintervalls für den Fall mit zwei Dimensionen. Alle Punkte außerhalb der Ellipse liegen im Ablehnungsbereich der Nullhypothese. Der Nullpunkt liegt außerhalb der Ellipse, während der Nullpunkt in jedem der beiden ein-dimensionalen Konfidenzintervalle liegt. (Den Graphen erhält man mit R-Befehl `confidenceEllipse(...)`. Siehe R-Programm in Appendix 10.5, Folie 270, für Details.)

- **R^2 -Version der F -Statistik:**

Enthält das Modell eine Konstante, gilt die Zerlegung $\text{SSR} = \text{SST}(1 - R^2)$. Die entsprechenden SSRs in die F -Statistik eingesetzt liefert dann

$$F = \frac{(R_{H_1}^2 - R_{H_0}^2)/q}{(1 - R_{H_1}^2)/(n - (k + 1))} \sim F_{q, n - k - 1}.$$

Beachte:

- SST kürzt sich nur dann heraus, wenn die abhängige Variable y unter H_0 und H_1 gleich ist, wie dies zum Beispiel bei Ausschlussrestriktionen der Fall ist. Dies gilt aber nicht immer, wenn man lineare Restriktionen testet.
- Aufgrund von Rundungsfehlern können beide Versionen der F -Statistik leicht unterschiedliche Ergebnisse liefern.

Overall F -Test

Die meisten Standardsoftwarepakete (wie auch R) liefern uns beim KQ-Output für $y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + u$ auch gleich die F -Statistik und ihren p -value für folgendes Hypothesenpaar mit:

“Keiner der Regressoren (außer der Konstante) hat Einfluss auf die abhängige Variable. Somit sind die entsprechenden Parameter alle Null.”

$$H_0 : \beta_1 = \dots = \beta_k = 0 \quad (\text{und } y = \beta_0 + u)$$

$$H_1 : \beta_j \neq 0 \text{ für mindestens ein } j = 1, \dots, k.$$

Falls H_0 nicht abgelehnt werden kann, bedeutet dies u.U., dass

- evtl. die Auswahl der Regressoren falsch ist,
 - zumindest ein großer Teil der Regressoren keinen Einfluss auf y hat,
 - für die gegebene Stichprobengröße n die Variablenanzahl zu groß ist.
- Dieser Test ist ein erster grober Check auf die Gültigkeit des Modells.

4.7.2 Testen mehrerer allgemeiner linearer Restriktionen

- Verallgemeinerung des F -Tests für Ausschlussrestriktionen.
- Funktioniert ebenfalls über die Berechnung der relativen Unterschiede der SSRs.
- R^2 -Version kann im allgemeinen Fall aber nicht angewendet werden!

Beispiel möglicher Hypothesenpaare:

$$H_0 : \beta_2 = \beta_3 = 1 \quad \text{versus} \quad H_1 : \beta_2 \neq 1 \text{ und/oder } \beta_3 \neq 1,$$

$$H_0 : \beta_1 = 1, \beta_j = 2\beta_l \quad \text{versus} \quad H_1 : \beta_1 \neq 1 \text{ und/oder } \beta_j \neq 2\beta_l.$$

Fortsetzung des Außenhandelsbeispiels (aus vorherigem Unterabschnitt):

- Man könnte vermuten, dass aufgrund der Multikollinearität zwischen den Schätzungen für *Offenheit* und *Fläche* der Einfluss

von *Offenheit* im Absolutbetrag unterschätzt werden könnte (in Modell 3 war die Parameterschätzung ja 0,507250), während der Einfluss von *Fläche* Null ist. Das betrachtete Hypothesenpaar lautet also:

$$H_0 : \beta_{Offenheit} = 0,5 \quad \text{und} \quad \beta_{Flaeche} = 0$$

$$H_1 : \beta_{Offenheit} \neq 0,5 \quad \text{und/oder} \quad \beta_{Flaeche} \neq 0$$

Um die SSR unter H_0 berechnen zu können, müssen die Restriktionen in die Regression wie folgt integriert werden:

$$\log(Importe) - (0,5)Offenheit = \beta_0 + \beta_{BIP} \log(BIP) + \beta_{Entfernung} Entfernung + u$$

Der R-Output lautet:

Call:

```
lm(formula = log(trade_0_d_o) - 0.5 * ebrd_tfes_o ~ log(wdi_gdpusdcr_o) +
  log(cepii_dist))
```

Residuals:

Min	1Q	Median	3Q	Max
-2.1968	-0.5605	0.1032	0.5904	1.5233

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.76870	2.02633	1.366	0.178
log(wdi_gdpusdcr_o)	0.94117	0.05922	15.893	< 2e-16 ***
log(cepii_dist)	-0.97180	0.14596	-6.658	2.97e-08 ***

Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Residual standard error: 0.8636 on 46 degrees of freedom

Multiple R-squared: 0.8884, Adjusted R-squared: 0.8836

F-statistic: 183.1 on 2 and 46 DF, p-value: < 2.2e-16

Dies ermöglicht es uns, die F -Statistik zu berechnen

$$F = \frac{(\text{SSR}_{H_0} - \text{SSR}_{H_1}) / q}{\text{SSR}_{H_1} / (n - k - 1)}$$

$$= \frac{(34.30373 - 32.01770) / 2}{32.01770 / 44} = 1.570776 < c = 3.20928.$$

Direkt in R: `linearHypothesis(model_4, c("ebrd_tfes_o = 0.5", "log(cepii_area_o) = 0")):`

Linear hypothesis test

Hypothesis:

`ebrd_tfes_o = 0.5`

`log(cepii_area_o) = 0`

Model 1: restricted model

Model 2: `log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o)`

	Res.Df	RSS	Df	Sum of Sq	F	Pr(>F)
1	46	34.304				
2	44	32.018	2	2.286	1.5708	0.2193

→ *Die Behauptung, dass die “Fläche des Landes keinen Einfluss hat und die Offenheit des Marktes einen Einfluss von 0.5 hat”, kann auf keinem vernünftigen Signifikanzniveau abgelehnt werden, da der p-value etwa 22% beträgt.*

4.8 Darstellung von Regressionsergebnissen

Im Allgemeinen untersuchen empirische Wissenschaftler mehrere unterschiedliche Spezifikationen der Regressionsgleichung.

Um sichtbar zu machen, wie robust die Schlussfolgerungen bezüglich der Modellwahl sind, werden in der Regel die Ergebnisse für die wichtigsten Spezifikationen angegeben, sodass jeder Leser die Resultate selbst beurteilen kann.

Am Einfachsten geht dies, indem man die relevanten Ergebnisse in einer Tabelle zusammenfasst, siehe Beispiel weiter unten.

Für jede Spezifikation sollte man zumindest Folgendes angeben:

- KQ-Parameterschätzungen $\hat{\beta}_j$ der Regressionsparameter β_j , $j = 0, 1, \dots, k$ (und die Variablennamen),
- Standardfehler von $\hat{\beta}_j$, $\hat{\sigma}_{\hat{\beta}_j}$,
- Stichprobengröße n ,
- R^2 und bereinigtes R^2 ,
- Standardfehler der Regression oder geschätzte Varianz der Regressionsfehler $\hat{\sigma}^2$.

Wenn möglich sollten zudem angegeben werden

- Modellselektionskriterien wie AIC, HQ oder SC,
- Residuenquadratsumme (SSR).

Auf Grundlage der SSRs lassen sich einfach F -Tests durchführen.

Für das Außenhandelsbeispiel ergibt sich z.B.:

Abhängige Variable: $\ln(\text{Importe nach Deutschland})$				
Unabhängige Variable/Modell	(1)	(2)	(3)	(4)
Konstante	-5.77	4.676	2.741	2.427
	(2.184)	(2.178)	(2.175)	(2.132)
$\ln(BIP)$	1.077	0.975	0.940	1.025
	(0.087)	(0.063)	(0.0613)	(0.076)
$\ln(\text{Entfernung})$	—	-1.074	-0.970	-0.888
		(0.156)	(0.152)	(0.156)
$Offenheit$	—	—	0.507	0.353
			(0.191)	(0.206)
$\ln(\text{Flaeche})$	—	—	—	-0.151
				(0.085)
Stichprobengröße	49	49	49	49
R^2	0.765	0.883	0.899	0.906
Standardfehler der Regression	1.304	0.928	0.873	0.853
Residuenquadratsumme	80.027	39.644	34.302	32.017
AIC	3.4100	2.7484	2.6445	2.6164
HQ	3.4393	2.7924	2.7031	2.6896
SC	3.4872	2.8642	2.7989	2.8094

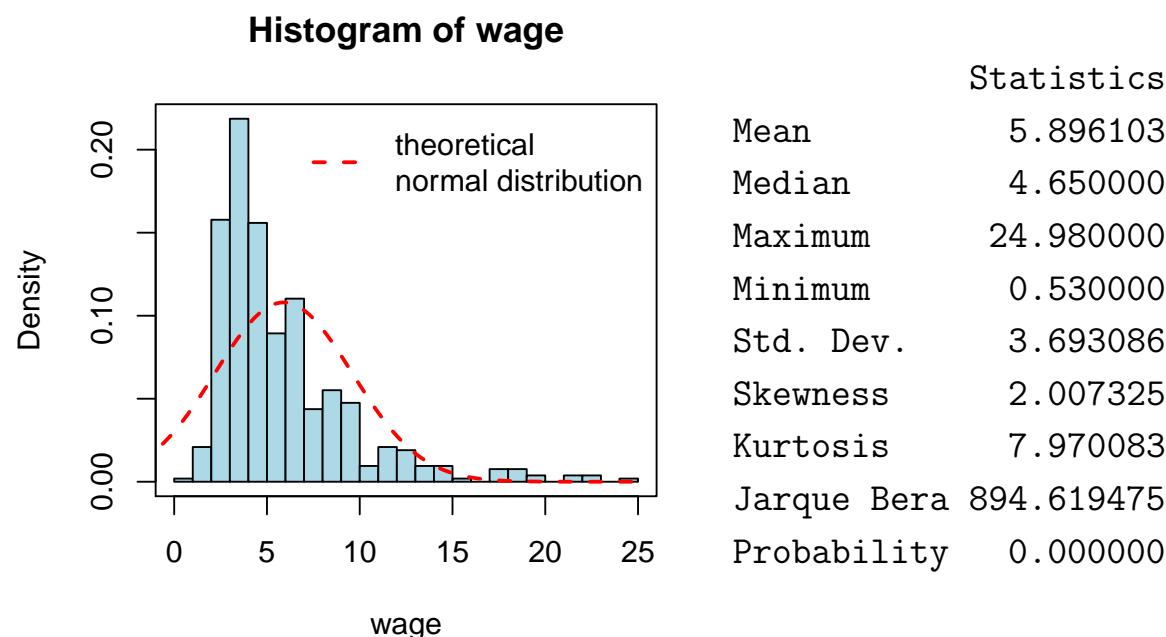
Zu Lesen: Abschnitte
4.5-4.6 in **Wooldridge**
(2009).

5 Multiples Regressionsmodell: Asymptotik

Die Annahme normalverteilter Fehler (MLR.6) ist in der Praxis häufig verletzt. Wie lassen sich dann Testentscheidungen und Konfidenzintervalle bestimmen?

5.1 Verteilung des Mittelwertschätzers in großen Stichproben

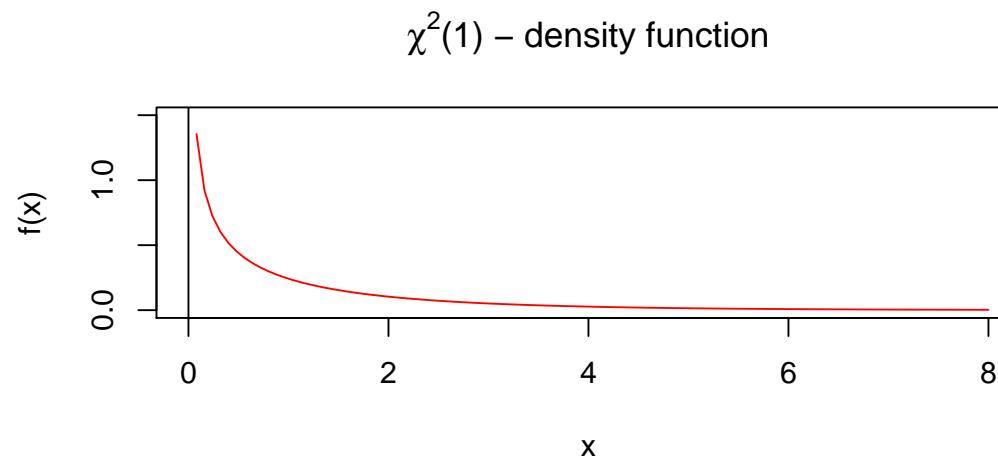
- Beispiel: Testen des Mittelwerts des Stundenlohns: Die empirische Verteilung ist linkssteil und rechtsschief (wie das für Preise und Löhne typisch ist, da diese nicht additiv generiert werden).



- Beispiele für Zufallsvariablen mit rechtsschiefer Verteilung:
 - Eine $\chi^2(m)$ -verteilte Zufallsvariable X ist definiert als die Summe von m quadrierten i.i.d. standardnormalverteilten Zufallsvariablen

$$X = \sum_{j=1}^m u_j^2, \quad u_j \sim i.i.d. N(0, 1).$$

(Details zur χ^2 -Verteilung finden Sie in Appendix B in [Wooldridge \(2009\)](#).)



Momente einer $\chi^2(1)$ -verteilten Zufallsvariable:

$$E[X] = E[u^2] = Var(u) + E[u]^2 = 1,$$

$$Var(X) = E[X^2] - E[X]^2 = E[u^4] - 1^2 = 2,$$

$$\frac{u^2 - 1}{\sqrt{2}} = \frac{X - 1}{\sqrt{2}} \sim (0, 1).$$

Beachte, dass man für eine standardnormalverteilte Zufallsvariable $E[u^4] = 3$ (=Wölbung) erhält.

- lineare Funktionen einer $\chi^2(1)$ -verteilten Zufallsvariable, z.B.

$$y_i = \nu + \sigma_y \frac{u_i^2 - 1}{\sqrt{2}}, \quad u_i \sim i.i.d.N(0, 1). \quad (5.1)$$

Momente:

$$E[y_i] = \nu,$$

$$Var(y_i) = Var\left(\sigma_y \frac{u_i^2 - 1}{\sqrt{2}}\right) = \sigma_y^2 Var\left(\frac{u_i^2}{\sqrt{2}}\right) = \sigma_y^2.$$

- **Erwartungswert und Varianz des Mittelwertschätzers**

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n y_i.$$

$$E[\hat{\mu}_n] = \frac{1}{n} \sum_{i=1}^n E[y_i] = \nu,$$

$$Var(\hat{\mu}_n) = \frac{1}{n^2} \sum_{i=1}^n Var(y_i) = \frac{Var(y_i)}{n} = \frac{\sigma_y^2}{n},$$

$$sd(\hat{\mu}_n) = \frac{\sigma_y}{\sqrt{n}}.$$

In diesem Beispiel ist der Schätzer unverzerrt und die Varianz nimmt mit zunehmender Stichprobengröße mit Rate n ab.

- **Konsistenz eines Schätzers $\hat{\theta}_n$:**

Für jedes $\epsilon > 0$ und $\delta > 0$ existiert ein N , sodass

$$P \left(|\hat{\theta}_n - \theta| < \epsilon \right) > 1 - \delta \quad \text{für alle } n > N.$$

Alternativ:

- $\lim_{n \rightarrow \infty} P \left(|\hat{\theta}_n - \theta| < \epsilon \right) = 1,$
- $\text{plim } \hat{\theta}_n = \theta,$
- $\hat{\theta}_n \xrightarrow{p} \theta.$

Die “plim”-Schreibweise steht für **probability limit**. Dieses Konvergenzkonzept wird gewöhnlich auch als Konvergenz in Wahrscheinlichkeit oder (schwache) Konsistenz bezeichnet. Hinweise zu den Rechenregeln für “plims” finden Sie in Appendix C.3 in **Wooldridge (2009)**.

Ein konsistenter Schätzer $\hat{\theta}_n$ hat die Eigenschaften

- $\lim_{n \rightarrow \infty} E[\hat{\theta}_n] = \theta$ **und**
- $\lim_{n \rightarrow \infty} \text{Var}(\hat{\theta}_n) = 0$.

Gilt eine dieser Eigenschaften nicht, bezeichnet man den Schätzer als **inkonsistent**. Allgemein:

- **Schwaches Gesetz der großen Zahlen (Weak law of large numbers (WLLN)):**

Für $y_i \sim \text{i.i.d.}$ mit $-\infty < E[y_i] = \mu < \infty$ ist der Mittelwertschätzer $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n y_i$ schwach konsistent, d.h.

$$\hat{\mu}_n \xrightarrow{p} \mu.$$

- Damit schätzen wir auch die Varianz einer i.i.d. Zufallsvariable $w_i \sim \text{i.i.d.}(\mu_w, \sigma_w^2)$ mit $\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (w_i - \mu_w)^2$ konsistent. Warum?

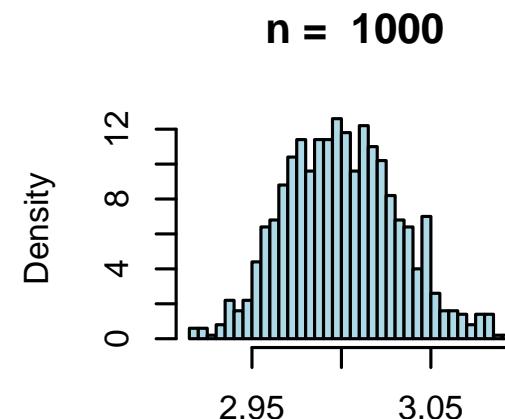
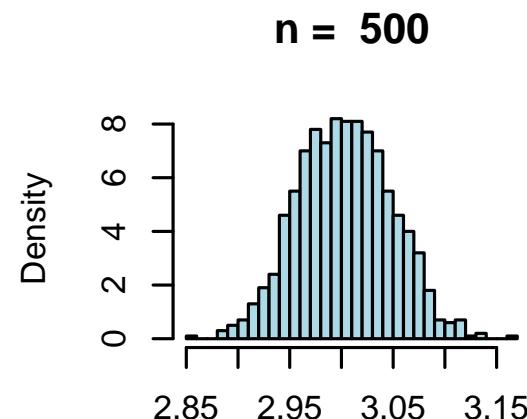
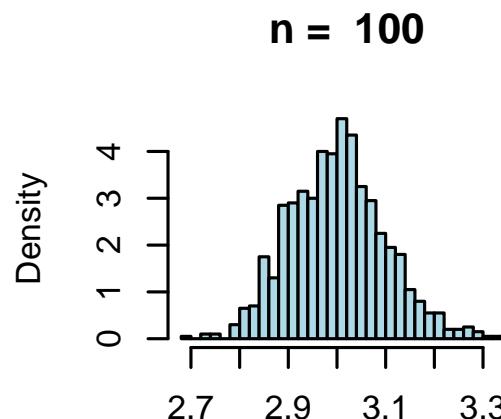
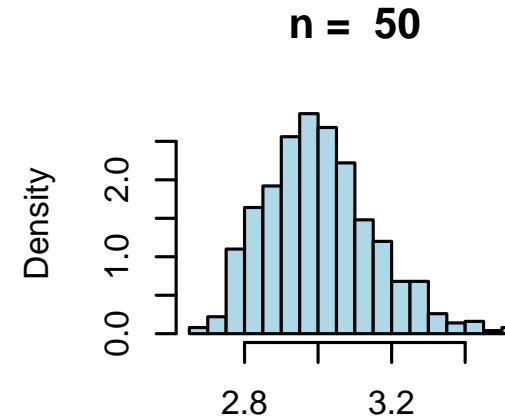
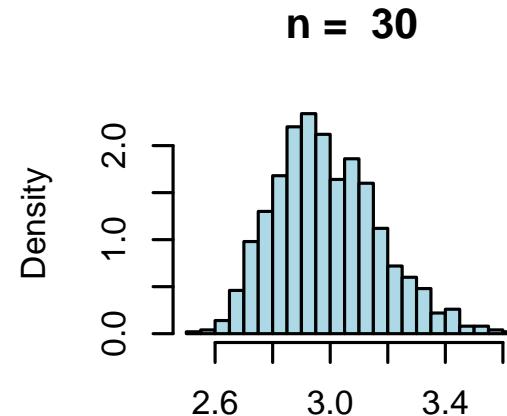
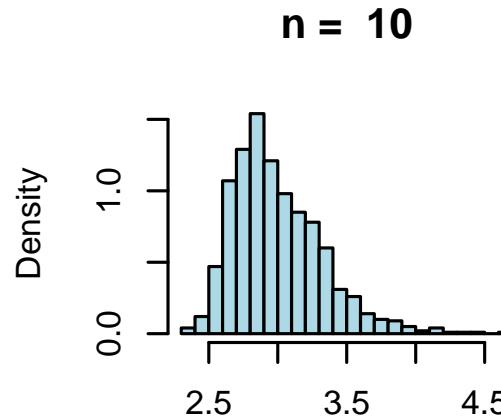
- Wie aber lässt sich die asymptotische Wahrscheinlichkeitsverteilung des Mittelwertschätzers $\hat{\mu}_n$ ableiten?
- Monte-Carlo Simulation (MC):

Das R-Programm `EOE_ws19_Emp_Beispiele.R`, ab Zeile 559, ermöglicht es uns, nacheinander $R = 1000$ Stichproben der Größe n mit den Elementen $\{y_1, \dots, y_n\}$ zu ziehen, wobei $y_i \sim i.i.d.(\nu, \sigma_y^2)$ mit $\nu = 3$ und $\sigma_y^2 = 1$ aus (5.1) generiert wird. Gleichung (5.1) wird häufig auch als **datengenerierender Prozess (DGP)** bezeichnet. Jede der $r = 1, \dots, 1000$ Stichproben $\{y_1^r, y_2^r, \dots, y_n^r\}$ wird auf diese Weise generiert. Für jede generierte Stichprobe wird der Mittelwertschätzer $\hat{\mu}^r = \frac{1}{n} \sum_{i=1}^n y_i^r$ berechnet und gespeichert. Hat man alle R Durchgänge hinter sich, berechnet man auf Basis dieser R Schätzungen $\hat{\mu}^1, \hat{\mu}^2, \dots, \hat{\mu}^R$ ein Histogramm.

Hier erst mal die Ergebnisse der simulierten Momente:

	Durchschnitte der geschätzten Mittelwerte	Standardabweichung wahre Standardabw.	des MC DGP
$n = 10$	2.999717	0.323645	0.316228
$n = 30$	2.988812	0.180521	0.182574
$n = 50$	3.005385	0.148377	0.141421
$n = 100$	3.001922	0.098153	0.100000
$n = 500$	3.003529	0.045176	0.044721
$n = 1000$	3.000575	0.031675	0.031623

- Die wahren Momente werden sehr genau geschätzt
- und wir sehen, wie das Gesetz der großen Zahlen funktioniert.



- Ergebnisse der simulierten Verteilungen:
 - Rechtsschiefe nimmt mit zunehmender Stichprobengröße n ab.
 - Test auf Normalität (Lomnicki-Jarque-Bera-Test): Die Nullhypothese einer Normalverteilung kann für große n nicht mehr abgelehnt werden.

Theoretische Erklärung dieses Phänomens: Eines der wichtigsten Instrumente der Statistik ist, dass unter bestimmten (ziemlich schwachen) Bedingungen ein zentraler Grenzwertsatz gilt!

- **Zentraler Grenzwertsatz (central limit theorem (CLT)):**
Für $y_i \sim i.i.d.(\mu, \sigma^2)$ mit $0 < \sigma^2 < \infty$ ist $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^n y_i$ asymptotisch normalverteilt:

$$\sqrt{n}(\hat{\mu}_n - \mu) \xrightarrow{d} N(0, \sigma^2).$$

- **Interpretation:** Je größer die Anzahl der Stichprobenelemente n , desto genauer wird die exakte tatsächliche Verteilung von $\hat{\mu}_n$ (siehe die MC-Ergebnisse) von einer exakt spezifizierten Normalverteilung approximiert. Deswegen auch die Bezeichnung **Verteilung für große Stichproben**.
- Aber wie gut ist die asymptotische Näherung für ein festes n ?
 - * Das CLT gibt uns hierzu keine Auskunft, allerdings können wir für bestimmte Fälle MC-Simulationen durchführen oder etwas aufwändigere Statistiken für endliche Stichproben anwenden, um Antworten zu erhalten.
 - * Erfahrung: Je näher die Verteilung von y_i an der Normalverteilung ist, umso kleiner wird das n , dass für eine sehr gute Annäherung erforderlich ist. In manchen Fällen ist $n = 30$ bereits ausreichend.

- Alternative Schreibweisen ($\Phi(z)$ bezeichnet die kumulative Verteilungsfunktion der Standardnormalverteilung):

$$\sqrt{n} \left(\frac{\hat{\mu}_n - \mu}{\sigma} \right) \xrightarrow{d} N(0, 1) \quad (5.2)$$

$$P \left(\sqrt{n} \left(\frac{\hat{\mu}_n - \mu}{\sigma} \right) \leq z \right) \longrightarrow \Phi(z) \quad (5.3)$$

$$\frac{\hat{\mu}_n - \mu}{\sigma/\sqrt{n}} \stackrel{\text{approx}}{\sim} N(0, 1) \quad (5.4)$$

$$\hat{\mu}_n \stackrel{\text{approx}}{\sim} N \left(\mu, \frac{\sigma^2}{n} \right) \quad (5.5)$$

→ Der Mittelwertschätzer ist **asymptotisch normalverteilt**.

- In großen Stichproben ist der standardisierte Mittelwertschätzer annähernd standardnormalverteilt. Aus (5.4) folgt also

$$w_i \sim i.i.d. N(\mu, \sigma^2) \quad t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu}{\sigma_{\hat{\mu}}} \sim N(0, 1)$$

$$w_i \sim i.i.d. (\mu, \sigma^2) \quad t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu}{\sigma_{\hat{\mu}}} \stackrel{\text{approx}}{\sim} N(0, 1)$$

und es kann gezeigt werden, dass (bei geschätzter Varianz)

$$w_i \sim i.i.d. N(\mu, \sigma^2) \quad t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu}{\hat{\sigma}_{\hat{\mu}}} \sim t_{n-k-1}$$

$$w_i \sim i.i.d. (\mu, \sigma^2) \quad t(w_1, \dots, w_n) = \frac{\hat{\mu} - \mu}{\hat{\sigma}_{\hat{\mu}}} \stackrel{\text{approx}}{\sim} N(0, 1)$$

und wir erhalten das Ergebnis: **Die Theorie der t -Tests und Konfidenzintervalle für den Mittelwertschätzer von i.i.d. Variablen (bei kleinen Stichproben) gilt in guter Annäherung auch für (genügend) große Stichproben.**

- Somit sind die Test-Ergebnisse unserer empirischen Übungen immer noch approximativ gültig!
- Wie sieht es mit dieser "approximativen Gültigkeit" im Kontext der Regression aus?

5.2 Verteilung des KQ-Schätzers in großen Stichproben

- Der KQ-Schätzer

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{u} = \boldsymbol{\beta} + \mathbf{W}\mathbf{u}$$

hängt von \mathbf{X} bzw. \mathbf{W} ab. Damit der KQ-Schätzer konsistent und asymptotisch normalverteilt ist, müssen für die Regressorvariablen bestimmte Bedingungen gelten, wenn $n \rightarrow \infty$. Eine dieser Bedingungen ist, dass für alle $i, l = 0, 1, \dots, k$ der $\text{plim} \frac{1}{n} \sum_{i=1}^n x_{ij}x_{il} = E[x_jx_l] = a_{ij}$ oder

$$\frac{1}{n} \mathbf{X}'\mathbf{X} \xrightarrow{p} \mathbf{A}. \quad (5.6)$$

- **Asymptotische Normalverteilung des KQ-Schätzers**

Gelten unsere Standardannahmen MLR.1-MLR.5, so sind alle notwendigen Bedingungen für asymptotische Normalverteilung erfüllt. Dann gilt (für eine Beweisskizze siehe Appendix E.4 in [Wooldridge \(2009\)](#)):

$$\sqrt{n} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N \left(0, \sigma^2 \mathbf{A}^{-1} \right). \quad (5.7)$$

Für die (**asymptotische**) **Verteilung der t -Statistiken** erhalten wir:

$$\text{MLR.1-MLR.6 } t(\mathbf{X}, \mathbf{y}) = \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \sim N(0, 1)$$

$$\text{MLR.1-MLR.5 } t(\mathbf{X}, \mathbf{y}) = \frac{\hat{\beta}_j - \beta_j}{\sigma_{\hat{\beta}_j}} \underset{\text{approx}}{\sim} N(0, 1)$$

und es kann gezeigt werden, dass

$$\text{MLR.1-MLR.6 } t(\mathbf{X}, \mathbf{y}) = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}/(\text{SST}_j(1-R_j^2))} \sim t_{n-k-1}$$

$$\text{MLR.1-MLR.5 } t(\mathbf{X}, \mathbf{y}) = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}/(\text{SST}_j(1-R_j^2))} \underset{\text{approx}}{\sim} N(0, 1)$$

Aus sehr vielen MC-Simulationen und der empirischen **Praxis** lässt sich folgern, dass

- man für kleine n so verfährt als wären die Fehler normalverteilt und man die kritischen Werte der t -Verteilung benutzt:

$$\text{MLR.1-MLR.5 } t(\mathbf{X}, \mathbf{y}) = \frac{\hat{\beta}_j - \beta_j}{\hat{\sigma} / (\text{SST}_j(1 - R_j^2))} \stackrel{\text{approx}}{\sim} t_{n-k-1}$$

- man analog für die F -Statistik die kritischen Werte aus der F -Verteilung ermittelt.
- Beachte auch hier: Die kritischen Werte sind nur approximativ gültig, nicht exakt. Analog dazu sind auch die p -values (, die von R berechnet werden) nur approximativ gültig!

- **Fazit:**

- Bei der Berechnung von Teststatistiken und Konfidenzintervallen (nicht: Prognoseintervallen) verfahren wir wie bisher. Jedoch gelten alle statistischen Ergebnisse nur approximativ.
- Ist die Annahme homoskedastisch verteilter Fehler verletzt, gelten selbst die asymptotischen Resultate nicht mehr und wir benötigen für diese Fälle Modelle für heteroskedastische Fehler (mit strengeren Annahmen für das LLN und das CLT), siehe Kapitel 8.

Zu Lesen: Kapitel 5 und Appendix C.3 in [Wooldridge \(2009\)](#).

6 Multiple Regression: Interpretation

6.1 Level- und Log-Modelle

Wiederholen Sie Abschnitt 2.6 zu level-level-, level-log-, log-level- und log-log-Modellen. Alle Ergebnisse sind bei einer ceteris-paribus-Untersuchung auch beim multiplen Regressionsmodell gültig.

6.2 Datenskalierung

- **Skalierung der abhängigen Variable:**

- Ausgangsmodell:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}.$$

- Variablentransformation: $y_i^* = a \cdot y_i$ mit Skalierungsfaktor a .

→ Neue, transformierte Regressionsgleichung:

$$\underbrace{ay}_{{\mathbf{y}}^*} = \mathbf{X} \underbrace{a\boldsymbol{\beta}}_{\boldsymbol{\beta}^*} + \underbrace{a\mathbf{u}}_{\mathbf{u}^*}$$

$$\mathbf{y}^* = \mathbf{X}\boldsymbol{\beta}^* + \mathbf{u}^* \tag{6.1}$$

- KQ-Schätzer für β^* aus (6.1):

$$\begin{aligned}\hat{\beta}^* &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}^* \\ &= a (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = a\hat{\beta}.\end{aligned}$$

- Fehlervarianz:

$$Var(\mathbf{u}^*) = Var(a\mathbf{u}) = a^2 Var(\mathbf{u}) = a^2 \sigma^2 \mathbf{I}.$$

- Varianz-Kovarianz-Matrix:

$$Var(\hat{\beta}^*) = \sigma^{*2} (\mathbf{X}'\mathbf{X})^{-1} = a^2 \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = a^2 Var(\hat{\beta})$$

- t -Statistik:

$$t^* = \frac{\hat{\beta}_j^* - 0}{\sigma_{\hat{\beta}_j^*}} = \frac{a\hat{\beta}_j}{a\sigma_{\hat{\beta}_j}} = t.$$

- **Skalierung erklärender Variablen:**

- Variablentransformation: $\mathbf{X}^* = \mathbf{X} \cdot a$. Neue Regressionsgleichung:

$$\mathbf{y} = \mathbf{X}a \cdot a^{-1}\boldsymbol{\beta} + \mathbf{u} = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{u}. \quad (6.2)$$

- KQ-Schätzer für $\boldsymbol{\beta}^*$ aus (6.2):

$$\begin{aligned}\hat{\boldsymbol{\beta}}^* &= (\mathbf{X}^{*\prime} \mathbf{X}^*)^{-1} \mathbf{X}^{*\prime} \mathbf{y} = (a^2 \mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' a \mathbf{y} \\ &= a^{-2} a (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = a^{-1} \hat{\boldsymbol{\beta}}.\end{aligned}$$

- Ergebnis: Die bloße Größe von β_j zeigt nicht an wie relevant der Einfluss des j ten Regressors ist. Man muss immer auch die Skalierung der Variable berücksichtigen.
- **Beispiel:** In Abschnitt 2.3 wurde ein einfaches level-level-Modell für den Einfluss des BIPs auf die Importe geschätzt. Die Parameterschätzung $\hat{\beta}_{BIP} = 4.857 \cdot 10^{-3}$ scheint ziemlich klein zu

sein. Wenn man jedoch berücksichtigt, dass das BIP in Dollar gemessen wird, ist dieser Parameterwert keineswegs klein. Wenn wir das BIP in Millionen Dollar reskalieren (mittels $a = 10^{-6}$) erhalten wir $\hat{\beta}^*_{BIP} = 10^6 \cdot 4.857 \cdot 10^{-03} = 4857$.

- **Skalierung von Variablen in logarithmischer Form**

verändert lediglich die Konstante β_0 , da $\ln y^* = \ln ay = \ln a + \ln y$.

- **Standardisierte Koeffizienten:**

Wir haben gerade gezeigt, dass die Relevanz der erklärenden Variablen nicht aus der Größe der zugehörigen Koeffizienten ersichtlich wird. Dies wäre dann möglich, wenn die Regression passend standardisiert würde.

Herleitung: Zuerst betrachten wir das folgende Regressionsmodell der Stichprobe

$$y_i = \hat{\beta}_0 + x_{i1}\hat{\beta}_1 + \dots + x_{ik}\hat{\beta}_k + \hat{u}_i, \quad (6.3)$$

und seine Darstellung, nachdem die Mittelwerte aller n Beobachtungen eingesetzt wurden

$$\bar{y} = \hat{\beta}_0 + \bar{x}_1\hat{\beta}_1 + \dots + \bar{x}_k\hat{\beta}_k. \quad (6.4)$$

Nun berechnen wir die Differenz von (6.4) und (6.3)

$$(y_i - \bar{y}) = (x_{i1} - \bar{x}_1)\hat{\beta}_1 + \dots + (x_{ik} - \bar{x}_k)\hat{\beta}_k + \hat{u}_i. \quad (6.5)$$

Schließlich wird Gleichung (6.5) durch die geschätzte Standardabweichung von y , also $\hat{\sigma}_y$, geteilt und jeder Term auf der rechten Seite wird noch mit den geschätzten Standardabweichungen des zu-

gehörigen Regressors, also $\hat{\sigma}_{x_j}$, $j = 1, \dots, k$ erweitert

$$\frac{(y_i - \bar{y})}{\hat{\sigma}_y} = \frac{(x_{i1} - \bar{x}_1)}{\hat{\sigma}_y} \cdot \frac{\hat{\sigma}_{x_1}}{\hat{\sigma}_{x_1}} \hat{\beta}_1 + \dots + \frac{(x_{ik} - \bar{x}_k)}{\hat{\sigma}_y} \cdot \frac{\hat{\sigma}_{x_k}}{\hat{\sigma}_{x_k}} \hat{\beta}_k + \frac{\hat{u}_i}{\hat{\sigma}_y}.$$

Einfaches Umstellen einiger Terme führt zu

$$\underbrace{\frac{(y_i - \bar{y})}{\hat{\sigma}_y}}_{z_{i,y}} = \underbrace{\frac{(x_{i1} - \bar{x}_1)}{\hat{\sigma}_{x_1}}}_{z_{i,x_1}} \underbrace{\frac{\hat{\sigma}_{x_1}}{\hat{\sigma}_y} \hat{\beta}_1}_{\hat{b}_1} + \dots + \underbrace{\frac{(x_{ik} - \bar{x}_k)}{\hat{\sigma}_{x_k}}}_{z_{i,x_k}} \underbrace{\frac{\hat{\sigma}_{x_k}}{\hat{\sigma}_y} \hat{\beta}_k}_{\hat{b}_k} + \underbrace{\frac{\hat{u}_i}{\hat{\sigma}_y}}_{\xi_i}.$$

Gewöhnlich werden in der Literatur die transformierten Variablen $z_{i,y}$ und $z_{i,x_1}, \dots, z_{i,x_k}$ als ***z-scores*** bezeichnet.

In **Kurzschriftweise** haben wir also

$$z_{i,y} = z_{i,x_1} \hat{b}_1 + \dots + z_{i,x_k} \hat{b}_k + \xi_i,$$

wobei die \hat{b}_j als **standardisierte Koeffizienten** (oder einfach **beta-Koeffizienten**) bezeichnet werden.

Die Größen der standardisierten Koeffizienten können nun miteinander verglichen werden. Die erklärende Variable mit dem größten Parameterwert \hat{b}_j hat demnach den relativ größten Einfluss auf die abhängige Variable.

Interpretation: Eine Erhöhung von x_j um eine Standardabweichung, verändert y um \hat{b}_j Standardabweichungen.

Standardisierte Koeffizienten lassen sich etwa in SPSS berechnen (siehe Example 6.1 in [Wooldridge \(2009\)](#)).

6.3 Besonderheiten bei transformierten Regressoren

- **Weitere Details zu logarithmierten Variablen:**

Wir betrachten das log-level-Regressionsmodell

$$\ln y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \quad (6.6)$$

wobei x_2 eine Dummyvariable ist, die entweder den Wert 0 oder 1 annimmt.

- Wie können wir den **exakten Einfluss** von x_2 ermitteln, d.h. wie interpretieren wir β_2 ? Aus (6.6) folgt

$$y = e^{\ln y} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + u} = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2} \cdot e^u$$

und für den bedingten Erwartungswert

$$E[y|x_1, x_2] = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2} \cdot E[e^u|x_1, x_2]. \quad (6.7)$$

Nun setzt man die beiden möglichen Werte von x_2 in (6.7) ein und erhält

$$\begin{aligned} E[y|x_1, x_2 = 0] &= e^{\beta_0 + \beta_1 x_1} \cdot E[e^u|x_1, x_2] \\ E[y|x_1, x_2 = 1] &= e^{\beta_0 + \beta_1 x_1} \cdot E[e^u|x_1, x_2] \cdot e^{\beta_2} \\ &= E[y|x_1, x_2 = 0] \cdot e^{\beta_2}. \end{aligned}$$

– Falls also $E[e^u|x_1, x_2]$ (für x_2) konstant ist, ist die **relative mittlere Veränderung der abhängigen Variable aufgrund einer Veränderung von x_2 um eine Einheit**

$$\begin{aligned} \frac{\Delta E[y|x_1, x_2]}{E[y|x_1, x_2 = 0]} &= \frac{E[y|x_1, x_2 = 1] - E[y|x_1, x_2 = 0]}{E[y|x_1, x_2 = 0]} \\ &= \frac{E[y|x_1, x_2 = 0] \cdot e^{\beta_2} - E[y|x_1, x_2 = 0]}{E[y|x_1, x_2 = 0]} \\ &= e^{\beta_2} - 1. \end{aligned}$$

Dies ist gleichbedeutend mit

$$\% \Delta E[y|x_1, x_2] = 100 \left(e^{\beta_2} - 1 \right).$$

– Im allgemeinen Fall mit k Regressoren:

$$\% \Delta E[y|x_1, x_2, \dots, x_k] = 100 \left(e^{\beta_j \Delta x_j} - 1 \right). \quad (6.8)$$

Offensichtlich stellt (6.8) den **exakten Partialeffekt** (Teileffekt) dar, während die Standardinterpretation als approximative semi-Elastizität in einigen Fällen ziemlich grob sein kann.

– **Fortsetzung des Außenhandelsbeispiels** (aus Abschnitt 4.8 und insbesondere aus Abschnitt 4.4):

Für Modell 3 erhielten wir die Regressionsgleichung

$$\begin{aligned} \text{LOG(TRADE}_0\text{D}_0) &= 2.74104 + 0.9406645 * \text{LOG(WDI_GDPUSDCR}_0) \\ &\quad - 0.9703183 * \text{LOG(CEPII_DIST)} + 0.5072497 * \text{EBRD_TFES}_0 + \text{RESIDUAL} \end{aligned}$$

- * Die approximative Interpretation von $\hat{\beta}_{Offenheit}$ ist, dass eine Veränderung um eine Einheit die *Importe* durchschnittlich um $100\hat{\beta}_{Offenheit} = 50,7\%$ verändert.
- * Der exakte Partialeffekt beträgt $100 \left(e^{\hat{\beta}_{Offenheit}} - 1 \right) = 66.1\%$, er ist also wesentlich größer!
- * Die Unterschiede des approximativen und des exakten Einflusses sind natürlich noch größer, wenn $\hat{\beta}$ weiter weg von der Null liegt.

- **Modelle mit quadratischen Regressoren:**

- Als Beispiel sei folgendes multiples Regressionsmodell

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2^2 + u$$

angenommen. Der marginale Einfluss einer Veränderung von x_2 auf den bedingten Erwartungswert y ist

$$\frac{\partial E[y|x_1, x_2]}{\partial x_2} = \beta_2 + 2\beta_3 x_2.$$

Somit beeinflusst eine Veränderung von x_2 um Δx_2 ceteris paribus die unabhängige Variable y im Durchschnitt um

$$(\beta_2 + 2\beta_3 x_2)\Delta x_2.$$

Der Einfluss hängt also offensichtlich vom Niveau von x_2 ab (und somit ist eine Interpretation von β_2 allein nicht sinnvoll!).

- In einigen empirischen Anwendungen verwendet man quadratische oder logarithmierte Regressoren, um eine **nicht-lineare Regressionsfunktion** zu approximieren.

Beispiel: Nicht-konstante Elastizitäten lassen sich folgendermaßen approximieren

$$\ln y = \beta_0 + \beta_1 x_1 + \beta_2 \ln x_2 + \beta_3 (\ln x_2)^2 + u.$$

Die Elastizität von y bezüglich x_2 ist demnach

$$\beta_2 + 2\beta_3 \ln x_2$$

und ist dann und nur dann konstant, wenn $\beta_3 = 0$.

– Fortsetzung des Außenhandelsbeispiels:

Bisher haben wir nur multiple Regressionsmodelle betrachtet, die in den Ausgangsvariablen log-log- oder log-level-spezifiziert waren.

Nun wollen wir eine weitere Spezifikation für die Modellierung von Importen betrachten, in der ein logarithmierter Regressor auch quadratisch in die Gleichung eingeht.

Modell 5:

$$\begin{aligned} \ln(\text{Importe}) = & \beta_0 + \beta_1 \ln(\text{BIP}) + \beta_2 (\ln(\text{BIP}))^2 + \beta_3 \ln(\text{Entfernung}) \\ & + \beta_4 \text{Offenheit} + \beta_5 \ln \text{Flaeche} + u. \end{aligned}$$

Eben wurde gezeigt, dass dann für die Elastizität der *Importe* bezüglich *BIP* gilt:

$$\beta_1 + 2\beta_2 \ln(\text{BIP}). \quad (6.9)$$

Die Schätzung des Modells 5 liefert:

Call:

```
lm(formula = log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + I(log(wdi_gdpusdcr_o)^2) +
  log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))
```

Residuals:

Min	1Q	Median	3Q	Max
-2.0672	-0.5451	0.1153	0.5317	1.3870

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-35.23314	17.44175	-2.020	0.04964 *
log(wdi_gdpusdcr_o)	3.90881	1.32836	2.943	0.00523 **
I(log(wdi_gdpusdcr_o)^2)	-0.05711	0.02627	-2.174	0.03523 *
log(cepii_dist)	-0.74856	0.16317	-4.587	3.86e-05 ***
ebrd_tfes_o	0.41988	0.20056	2.094	0.04223 *
log(cepii_area_o)	-0.13238	0.08228	-1.609	0.11497

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8191 on 43 degrees of freedom

Multiple R-squared: 0.9155, Adjusted R-squared: 0.9056

F-statistic: 93.12 on 5 and 43 DF, p-value: < 2.2e-16

Vergleichen wir AIC, HQ und SC des Modells 5 mit den Werten aus den Modellen 1 bis 4, siehe Abschnitt 4.4, erkennen wir, dass Modell 5 durchgängig die geringsten Werte erzielt. Zudem ist der (approximative) p -value von β_2 gleich 0.03523 und der quadrierte Term somit auf dem 5% Niveau statistisch signifikant.

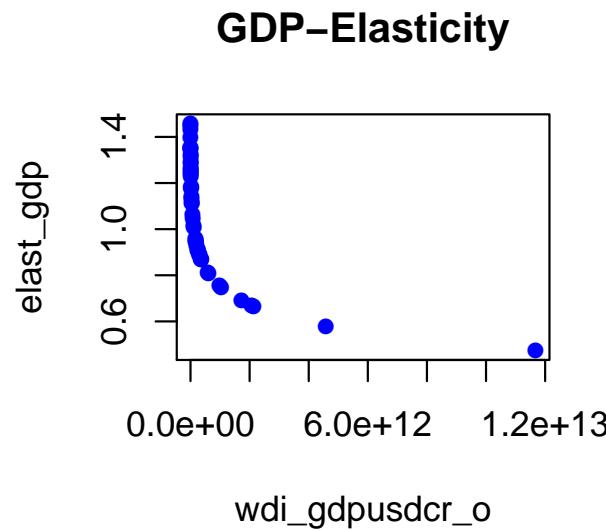
Damit ist der statistische Nachweis für eine nicht-lineare Elastizität erbracht. Setzen wir die Parameterschätzungen in (6.9) ein, erhalten wir

$$\eta(BIP) = 3.90881 - 0.05711 \ln(BIP).$$

Man kann die Elastizität von $\eta(BIP)$ für jeden beobachteten Wert von BIP gegen BIP plotten. In R wird dies mit einem kleinen Programm bewerkstelligt

R-Code

```
# Modell 5:
model_5 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + I(log(wdi_gdpusdcr_o)^2)
               + log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))
# Generiere die Elastizitäten für verschiedene BIPs
elast_gdp <- model_5$coef[2] + 2* model_5$coef[3]*log(wdi_gdpusdcr_o)
# Erstelle Scatterplot
plot(wdi_gdpusdcr_o, elast_gdp, pch = 16, col = "blue", main = "GDP-Elasticity")
```



Die BIP-Elastizität der Importe ist für kleine Volkswirtschaften (bezüglich des BIPs) viel größer als für große Volkswirtschaften.

Vorsicht: Nichtlinearitäten ergeben sich manchmal daraus, dass relevante Variablen fehlen. Können Sie sich denken, welche Kontrollvariable zu Modell 5 hinzugefügt werden sollte?

- **Interaktionen:**

Beispiel:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_2 x_1 + u.$$

Der marginale Effekt einer Veränderung von x_2 ist gegeben durch

$$\Delta E[y|x_1, x_2] = (\beta_2 + \beta_3 x_1) \Delta x_2.$$

Somit hängt der marginale Effekt auch vom Niveau von x_1 ab!

6.4 Qualitative Daten als Regressoren

Dummy-Variable oder Binärvariable

Eine **Binärvariable** kann genau **zwei** unterschiedliche Werte annehmen und ermöglicht es, **zwei** qualitativ unterschiedliche Zustände zu beschreiben.

Beispiele: weiblich vs. männlich, angestellt vs. arbeitslos, etc.

- Im Allgemeinen werden diese Werte mit 0 und 1 kodiert. Dies ermöglicht eine sehr einfache Interpretation. Beispiel:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_{k-1} x_{k-1} + \delta D + u,$$

wobei D entweder 0 oder 1 ist.

- **Interpretation** (bereits bekannt):

$$\begin{aligned}
 E[y|x_1, \dots, x_{k-1}, D = 1] - E[y|x_1, \dots, x_{k-1}, D = 0] &= \\
 \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1} + \delta & \\
 - (\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1}) &= \delta
 \end{aligned}$$

Der Koeffizient δ einer Dummyvariablen gibt somit an, inwiefern sich der Achsenabschnitt im Falle, dass $D = 1$ ist, verschiebt. Alle Steigungsparameter β_i , $i = 1, \dots, k - 1$ bleiben *unverändert*.

- **Fortsetzung des Lohnbeispiels:**

- Ausgangsfrage: Ist das Einkommen von Frauen signifikant niedriger als das von Männern?
- Daten: Stichprobe von $n = 526$ ArbeitnehmerInnen in den U.S.A. aus dem Jahre 1976. (Quelle: Examples 2.4, 7.1 in **Wooldridge (2009)**).

- * *wage*: Stundenlohn in US-\$,
- * *educ*: Dauer der Ausbildung,
- * *exper*: Berufserfahrung in Jahren,
- * *tenure*: Beschäftigungsdauer bei aktueller Firma,
- * *female*: dummy=1 falls weiblich, dummy=0 andernfalls.

```
lm(formula = log(wage) ~ female + educ + exper + I(exper^2) +
  tenure + I(tenure^2))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.83160	-0.256558	-0.02126	0.25500	1.13370

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.4166910	0.0989279	4.212	2.98e-05 ***
female	-0.2965110	0.0358054	-8.281	1.04e-15 ***
educ	0.0801966	0.0067573	11.868	< 2e-16 ***
exper	0.0294324	0.0049752	5.916	6.00e-09 ***
I(exper^2)	-0.0005827	0.0001073	-5.431	8.65e-08 ***
tenure	0.0317139	0.0068452	4.633	4.56e-06 ***
I(tenure^2)	-0.0005852	0.0002347	-2.493	0.013 *

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.3998 on 519 degrees of freedom

Multiple R-squared: 0.4408, Adjusted R-squared: 0.4343

F-statistic: 68.18 on 6 and 519 DF, p-value: < 2.2e-16

- **Beachte:** Um den Koeffizienten einer Dummyvariablen interpretieren zu können, muss man die Referenzgruppe kennen. Die Referenzgruppe ist diejenige Gruppe für die der Dummy den Wert Null annimmt.
- **Prognose:** Wie viel verdient ein Frau mit 12 Jahren Ausbildung, 10 Jahren Erfahrung und einem Jahr Beschäftigungsduer? (Man kann natürlich auch andere Werte einsetzen)

$$\begin{aligned}E[\ln(wage)|female = 1, educ = 12, exper = 10, tenure = 1] \\= 0.4167 - 0.2965 \cdot 1 + 0.0802 \cdot 12 + 0.0294 \cdot 10 \\- 0.0006 \cdot (10^2) + 0.0317 \cdot 1 - 0.0006 \cdot (1^2) \\= 1.35\end{aligned}$$

Demnach ist der erwartete Stundenlohn **approximativ** $\exp(1.35) = 3.86$ US-\$.

- Wir wissen bereits, dass im Falle eines log-level-Modells der Erwartungswert von y gegeben die Regressoren x_1, x_2 gegeben ist durch

$$E[y|x_1, x_2] = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2} \cdot E[e^u|x_1, x_2].$$

Der wahre Wert von $E[e^u|x_1, x_2]$ hängt von der Wahrscheinlichkeitsverteilung von u ab.

Es gilt: Falls u normalverteilt ist mit Varianz σ^2 , dann ist

$$E[e^u|x_1, x_2] = e^{E[u|x_1, x_2] + \sigma^2/2}.$$

Die genaue Prognose ist somit

$$E[y|x_1, x_2] = e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + E[u|x_1, x_2] + \sigma^2/2}.$$

Die exakte Vorhersage des Stundenlohns ist dann

$$\begin{aligned}
 E[wage | female = 1, educ = 12, exper = 10, tenure = 1] \\
 &= \exp(0.4167 - 0.2965 \cdot 1 + 0.0802 \cdot 12 + 0.02943 \cdot 10 \\
 &\quad - 0.0006 \cdot (10^2) + 0.0317 \cdot 1 - 0.0006 \cdot (1^2) + 0.3998^2 / 2) \\
 &= 4.18.
 \end{aligned}$$

Der genaue Wert des mittleren Stundenlohns der beschriebenen Person ist also etwa 4.18\$ und damit 30 Cent größer als der approximierte Wert.

- Der Parameter δ entspricht der Differenz des logarithmierten Einkommens zwischen weiblichen und männlichen Arbeitnehmern, wobei *alles andere konstant gehalten wird* (z.B. Ausbildungsdauer, Erfahrung, etc.).

Frage: Wie groß ist der exakte Lohnunterschied?

Antwort: $100(e^{-0.2965} - 1) = 34.51\%$.

Beachte, dass die ceteris-paribus-Untersuchung sehr viel mehr Aussagekraft hat als der Vergleich unbedingter Mittelwerte der Löhne von Männern und Frauen. Nimmt man normalverteilte Fehler an, hätte man

$$\frac{E[wage_f] - E[wage_m]}{E[wage_m]} = \frac{e^{E[\ln(wage_f)] + \sigma_f^2/2} - e^{E[\ln(wage_m)] + \sigma_m^2/2}}{e^{E[\ln(wage_m)] + \sigma_m^2/2}}.$$

Setzt man die Schätzungen ein, erhält man

$$\frac{e^{1.416 + 0.44^2/2} - e^{1.814 + 0.53^2/2}}{e^{1.814 + 0.53^2/2}} = -0.3570,$$

was, nebenbei, sehr ähnlich dem ist, was wir erhalten, wenn wir die Schätzungen in $(E[wage_f] - E[wage_m])/E[wage_m]$ einsetzen

und daraus einen Wert von -0.3538 ermitteln.

Frauen verdienen 36% weniger als Männer, wenn man keine anderen Einflüsse berücksichtigt.

Mehrere Untergruppen

- **Beispiel:** Ein Arbeitnehmer ist weiblich oder männlich und verheiratet oder ledig \Rightarrow 4 Untergruppen
 1. weiblich und ledig
 2. weiblich und verheiratet
 3. männlich und ledig
 4. männlich und verheiratet

Wie gehen wir vor?

- Wir bestimmen eine Gruppe als **Referenzgruppe**, etwa: weiblich und ledig
- Definiere Dummyvariablen für die anderen Untergruppen, in R:

```
* femmarr <- female * married

* malesing <- (1 - female) * (1 - married)

* malemarr <- (1 - female) * married

lm(formula = log(wage) ~ femmarr + malesing + malemarr + educ +
  exper + I(exper^2) + tenure + I(tenure^2))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.89697	-0.24060	-0.02689	0.23144	1.09197

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.2110279	0.0966445	2.184	0.0294 *
femmarr	-0.0879174	0.0523481	-1.679	0.0937 .
malesing	0.1103502	0.0557421	1.980	0.0483 *
malemarr	0.3230259	0.0501145	6.446	2.64e-10 ***
educ	0.0789103	0.0066945	11.787	< 2e-16 ***
exper	0.0268006	0.0052428	5.112	4.50e-07 ***
I(exper^2)	-0.0005352	0.0001104	-4.847	1.66e-06 ***

```
tenure      0.0290875  0.0067620   4.302 2.03e-05 ***
I(tenure^2) -0.0005331  0.0002312  -2.306   0.0215 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Residual standard error: 0.3933 on 517 degrees of freedom

Multiple R-squared: 0.4609, Adjusted R-squared: 0.4525

F-statistic: 55.25 on 8 and 517 DF, p-value: < 2.2e-16

Interpretationsbeispiele:

- Verheiratete Frauen verdienen 8.8% weniger als ledige Frauen. Dieser Effekt ist jedoch nur auf dem 10% Niveau signifikant (bei einem zweiseitigen Test).
- Der Lohnunterschied zwischen verheirateten Männern und Frauen beträgt ungefähr $32.3 - (-8.8) = 41.1\%$. Hierfür kann keine t -Statistik direkt berechnet werden. (Lösung: Führe die Schätzung nochmal durch mit einer der beiden Untergruppen als Referenzgruppe.)

Bemerkungen:

- Es ist nicht empfehlenswert für *alle* Untergruppen eine Dummyvariable zu erstellen, weil dann die Unterschiede bezüglich der Referenzgruppe nicht direkt getestet werden können.
- Falls man für alle Untergruppen eine Dummyvariable verwendet, darf keine Konstante mehr im Modell enthalten sein, sonst wäre MLR.3 verletzt. Warum?

• Ordinale Daten in der Regression

Beispiel: Universitäten-Ranking

Die Qualitätsunterschiede zwischen Rang 1 und 2 bzw. den Rängen 11 und 12, können gewaltig voneinander abweichen. Deshalb sind Rangfolgen **nicht** als Regressoren geeignet. Stattdessen weisen wir jeder Universität außer einer (der “Referenzkategorie”) eine Dum-

myvariable D_j zu, was zur Folge hat, dass wir einige neue Parameter zu schätzen haben (Deshalb sollten wir im Außenhandelsbeispiel evtl. die Variable *Offenheit* in mehrere Dummys aufspalten...).

Beachte: Der Koeffizient einer Dummyvariablen D_j gibt nun die Verschiebung des Achsenabschnitts zwischen Universität j und der Referenzuni an.

Gelegentlich ist die Rangliste zu lang, sodass zu viele Parameter zu schätzen wären. Es ist dann meist hilfreich, die Daten in Gruppen zusammenzufassen, z.B. Ränge 1-10, 11-20, etc..

Interaktionen und Dummyvariablen

- **Interaktionen zwischen Dummyvariablen:**
 - z.B. zum Definieren von Untergruppen (z.B. verheiratete Männer).
 - Beachte, dass eine sinnvolle Interpretation und ein Vergleich der Einflüsse der Untergruppen entscheidend von einer korrekten Verwendung der Dummies abhängt. Wir fügen z.B. die Dummies *male* und *married* und deren Interaktion einer Lohngleichung

$$y = \beta_0 + \delta_1 \text{male} + \delta_2 \text{married} + \delta_3 \text{male} \cdot \text{married} + \dots$$

hinzu. Ein Vergleich zwischen verheirateten und ledigen Männern

ist somit gegeben durch

$$\begin{aligned} E[y|male = 1, married = 1] - E[y|male = 1, married = 0] \\ = \beta_0 + \delta_1 + \delta_2 + \delta_3 + \dots - (\beta_0 + \delta_1 + \dots) = \delta_2 + \delta_3. \end{aligned}$$

- **Interaktionen zwischen Dummies und quantitativen Variablen:**

- Dies ermöglicht für unterschiedliche Gruppen unterschiedliche Steigungsparameter

$$y = \beta_0 + \beta_1 D + \beta_2 x_1 + \beta_3 (x_1 \cdot D) + u.$$

Beachte: β_1 bezeichnet hier *lediglich* für den Fall $x_1 = 0$ die Unterschiede zwischen beiden Gruppen.

Falls $x_1 \neq 0$, ist diese Differenz

$$\begin{aligned} & E[y|D = 1, x_1] - E[y|D = 0, x_1] \\ &= \beta_0 + \beta_1 \cdot 1 + \beta_2 x_1 + \beta_3(x_1 \cdot 1) - (\beta_0 + \beta_2 x_1) \\ &= \beta_1 + \beta_3 x_1 \end{aligned}$$

Selbst, wenn β_1 negativ ist, ist der Gesamteffekt möglicherweise positiv!

– Fortsetzung des Lohnbeispiels:

```
lm(formula = log(wage) ~ female + educ + exper + I(exper^2) +
  tenure + I(tenure^2) + I(female * educ))
```

Residuals:

Min	1Q	Median	3Q	Max
-1.83264	-0.25261	-0.02374	0.25396	1.13584

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.3888060	0.1186871	3.276	0.00112 **
female	-0.2267886	0.1675394	-1.354	0.17644
educ	0.0823692	0.0084699	9.725	< 2e-16 ***
exper	0.0293366	0.0049842	5.886	7.11e-09 ***
I(exper^2)	-0.0005804	0.0001075	-5.398	1.03e-07 ***
tenure	0.0318967	0.0068640	4.647	4.28e-06 ***
I(tenure^2)	-0.0005900	0.0002352	-2.509	0.01242 *
I(female * educ)	-0.0055645	0.0130618	-0.426	0.67028

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’
	0.1 ‘ ’	1		

Residual standard error: 0.4001 on 518 degrees of freedom

Multiple R-squared: 0.441, Adjusted R-squared: 0.4334

F-statistic: 58.37 on 7 and 518 DF, p-value: < 2.2e-16

Ist die *Rendite der Ausbildung* geschlechtsspezifisch?

• Tests auf Gruppenunterschiede

- werden mittels F -Tests durchgeführt.
- **Chow Test:** Ermöglicht es zu testen, ob Gruppenunterschiede im Sinne gruppenspezifischer Achsenabschnitte und/oder (mindestens einem) gruppenspezifischen Steigungsparameter vorliegen.

Beispiel:

$$y = \beta_0 + \beta_1 D + \beta_2 x_1 + \beta_3(x_1 \cdot D) + \beta_4 x_2 + \beta_5(x_2 \cdot D) + u. \quad (6.10)$$

Hypothesenpaar:

$$H_0 : \beta_1 = \beta_3 = \beta_5 = 0 \quad \text{vs.}$$

$$H_1 : \beta_1 \neq 0 \text{ und/oder } \beta_3 \neq 0 \text{ und/oder } \beta_5 \neq 0.$$

Anwenden des F -Tests:

- * Schätze die Regressionsgleichung für jede Gruppe l

$$y = \beta_{0l} + \beta_{2l}x_1 + \beta_{4l}x_2 + u, \quad l = 1, 2,$$

und berechne SSR_1 und SSR_2 .

- * Schätze dann die Regression für beide Gruppen zusammen und berechne die SSR .
- * Berechne die F -Statistik

$$F = \frac{SSR - (SSR_1 + SSR_2)}{SSR_1 + SSR_2} \frac{n - 2(k + 1)}{(k + 1)},$$

wobei die Freiheitsgrade der F -Verteilung $k+1$ und $n-2(k+1)$ sind.

Zu Lesen: Kapitel 6 (ohne Abschnitt 6.4) und Kapitel 7 (ohne die Abschnitte 7.5 und 7.6) in [Wooldridge \(2009\)](#).

7 Multiple Regression: Prognosen

7.1 Prognose und Prognosefehler

- Betrachten wir das multiple Regressionsmodell $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, d.h.

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i, \quad 1 \leq i \leq n.$$

- Wir suchen eine **Prognose** \hat{y}_0 von y_0 für gegebene x_{01}, \dots, x_{0k} .
- Den **Prognosefehler** definieren wir als

$$y_0 - \hat{y}_0.$$

- Wir nehmen an, dass die Annahmen MLR.1 bis MLR.5 für die **Prognosestichprobe** (\mathbf{x}_0, y_0) gelten. Dann gilt

$$y_0 = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k} + u_0 \quad (7.1)$$

sowie

$$E[u_0 | x_{01}, \dots, x_{0k}] = 0,$$

sodass

$$E[y_0 | x_{01}, \dots, x_{0k}] = \beta_0 + \beta_1 x_{01} + \cdots + \beta_k x_{0k} = \mathbf{x}'_0 \boldsymbol{\beta},$$

wobei $\mathbf{x}'_0 = (1, x_{01}, \dots, x_{0k})$.

MLR.4 stellt sicher, dass für *bekannte* Parameter die Prognosen unverzerrt sind. Die Prognose ist dann, salopp formuliert, im Durchschnitt korrekt (im Durchschnitt über viele Stichproben).

Man kann zeigen, dass der bedingte Erwartungswert bezüglich der Minimierung des mittleren quadrierten Prognosefehlers optimal ist.

- In der Praxis sind uns die wahren Regressionsparameter β_j , $j = 0, \dots, k$ nicht bekannt. Setzen wir die KQ-Schätzer $\hat{\beta}_j$ ein, so erhalten wir

$$\hat{y}_0 = \hat{E}[y_0 | x_{01}, \dots, x_{0k}] = \hat{\beta}_0 + \hat{\beta}_1 x_{01} + \dots + \hat{\beta}_k x_{0k}.$$

In kompakter Schreibweise lautet die **Prognosevorschrift**:

$$\hat{y}_0 = \mathbf{x}'_0 \hat{\beta}. \quad (7.2)$$

- Diese Prognosevorschrift hat nur dann Sinn, wenn (y_0, \mathbf{x}'_0) auch Bestandteil der Grundgesamtheit ist. Andernfalls ist das Regressionsmodell der Grundgesamtheit für (y_0, \mathbf{x}'_0) nicht gültig und damit wäre die auf der Schätzung basierende Prognose möglicherweise völlig irreführend.

- Allgemeine **Zerlegung** des **Prognosefehlers**

$$\begin{aligned}
 \hat{u}_0 &= y_0 - \hat{y}_0 & (7.3) \\
 &= \underbrace{(y_0 - E[y_0 | \mathbf{x}_0])}_{\text{unvermeidbarer Fehler } v_0} \\
 &\quad + \underbrace{(E[y_0 | \mathbf{x}_0] - \mathbf{x}'_0 \boldsymbol{\beta})}_{\text{mögl. Spezifikationsfehler}} \\
 &\quad + \underbrace{(\mathbf{x}'_0 \boldsymbol{\beta} - \mathbf{x}'_0 \hat{\boldsymbol{\beta}})}_{\text{Schätzfehler}}
 \end{aligned}$$

- Falls MLR.1 und MLR.4 für die Grundgesamtheit gelten und die Prognosestichprobe zur Grundgesamtheit gehört, dann ist der Spezifikationsfehler Null. In (7.1) gilt dann $v_0 = u_0$.
- Ist der Schätzer konsistent, also $\text{plim } \hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$, dann ist der Schätzfehler in großen Stichproben vernachlässigbar klein.

- Verwenden wir den KQ-Schätzer, ist der Schätzfehler

$$\begin{aligned}
 \mathbf{x}'_0 \boldsymbol{\beta} - \mathbf{x}'_0 \hat{\boldsymbol{\beta}} &= \mathbf{x}'_0 (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\
 &= \mathbf{x}'_0 \boldsymbol{\beta} - \mathbf{x}'_0 \left((\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} \right) \\
 &= \mathbf{x}'_0 \boldsymbol{\beta} - \mathbf{x}'_0 \left(\boldsymbol{\beta} + (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{u} \right) \\
 &= -\mathbf{x}'_0 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{u}.
 \end{aligned} \tag{7.4}$$

Somit hängt der Schätzfehler nur von der Stichprobe der Schätzung ab.

- Unter MLR.1 bis MLR.5 ist der KQ-Prognosefehler (aus (7.3) und (7.4)) gegeben durch:

$$\begin{aligned}
 \hat{u}_0 &= u_0 + \mathbf{x}'_0 (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\
 &= u_0 - \mathbf{x}'_0 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{u}.
 \end{aligned} \tag{7.5}$$

- **Prognosefehlervarianz:**

- **Erweiterung der Annahme MLR.2** (Zufallsstichprobe):
 u_0 und \mathbf{u} sind unkorreliert.
- Die auf \mathbf{X} und \mathbf{x}_0 bedingte Varianz von (7.5) lautet:

$$\begin{aligned}
 \text{Var}(\hat{u}_0|\mathbf{X}, \mathbf{x}_0) &= \text{Var}(u_0|\mathbf{X}, \mathbf{x}_0) + \text{Var}\left(\mathbf{x}'_0(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})|\mathbf{X}, \mathbf{x}_0\right) \\
 &= \sigma^2 + \mathbf{x}'_0 \text{Var}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}|\mathbf{X}) \mathbf{x}_0 \\
 &= \sigma^2 + \mathbf{x}'_0 \sigma^2 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0
 \end{aligned}$$

oder

$$\text{Var}(\hat{u}_0|\mathbf{X}, \mathbf{x}_0) = \sigma^2 \left(1 + \mathbf{x}'_0 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\right). \quad (7.6)$$

- In der Praxis eher relevant: **Geschätzte Prognosefehlervarianz**

$$\widehat{\text{Var}}(\hat{u}_0|\mathbf{X}, \mathbf{x}_0) = \hat{\sigma}^2 \left(1 + \mathbf{x}'_0 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{x}_0\right).$$

- **Prognoseintervall:** Das Prognoseintervall für ein multiples Regressionsmodell ist (für eine zuvor gewählte Konfidenzwahrscheinlichkeit $1 - \alpha$) gegeben als

$$\left[\hat{y}_0 - t_{n-k-1} \sqrt{\widehat{Var}(\hat{u}_0 | \mathbf{X}, \mathbf{x}_0)}, \hat{y}_0 + t_{n-k-1} \sqrt{\widehat{Var}(\hat{u}_0 | \mathbf{X}, \mathbf{x}_0)} \right].$$

Beachte:

- Ableitung und Struktur folgt analog zum Fall der Konfidenzintervalle für Parameterschätzungen.
- Im Gegensatz zu Konfidenzintervallen sind Prognoseintervalle selbst in großen Stichproben nur dann gültig, wenn die Prognosefehler normalverteilt sind. Dies folgt aus der Tatsache, dass der wahre Prognosefehler u_0 nicht gemittelt wird, wie dies bei $\hat{\beta} - \beta = \mathbf{Wu}$ wegen des Zentralen Grenzwertsatzes der Fall war.

7.2 Statistische Eigenschaften linearer Prognosen

Die Prognosevorschrift ist offensichtlich linear (in y), da

$$\hat{y}_0 = \mathbf{x}'_0 \hat{\boldsymbol{\beta}} = \mathbf{x}'_0 (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}.$$

Gauss-Markov Eigenschaft linearer Prognosen

Falls $\hat{\boldsymbol{\beta}}$ der BLU-Schätzer für $\boldsymbol{\beta}$ ist, dann ist

$$\hat{y}_0 = \mathbf{x}'_0 \hat{\boldsymbol{\beta}}$$

die BLU-Prognosevorschrift. Unter allen linearen Prognosevorschriften mit mittlerem Prognosefehler von Null weist diese die geringste Prognosefehlervarianz auf.

Zu Lesen: Abschnitt 6.4 in [Wooldridge \(2009\)](#).

8 Multiple Regression: Heteroskedastie

- In diesem Kapitel nehmen wir weiterhin an, dass MLR.1 bis MLR.4 gelten.
- Falls aber MLR.5 nicht gilt, d.h.

$$Var(u_i|x_{i1}, \dots, x_{ik}) = \sigma_i^2 \neq \sigma^2, \quad i = 1, \dots, n,$$

weisen die Fehler des Regressionsmodells Heteroskedastie auf. Genauer gesagt: Es gilt nun (statt MLR.5)

– Annahme GLS.5: Heteroskedastie

$$\begin{aligned} \text{Var}(u_i | x_{i1}, \dots, x_{ik}) &= \sigma_i^2(x_{i1}, \dots, x_{ik}) \\ &= \sigma^2 h(x_{i1}, \dots, x_{ik}) = \sigma^2 h_i, \quad i = 1, \dots, n. \end{aligned}$$

Die Fehlervarianz der i -ten Stichprobenbeobachtung σ_i^2 ist eine Funktion $h(\cdot)$ der Regressoren.

- **Beispiele:**

- Die Varianz der Nettomieten hängt von der Größe der Wohnung ab.
- Die Varianz der Konsumausgaben hängt von der Höhe des Einkommens ab.
- Die Varianz der logarithmierten Stundenlöhne hängt von der Ausbildungsdauer ab.

- Die **Kovarianzmatrix der Fehler** der Regression ist:

$$Var(\mathbf{u}|\mathbf{X}) = E[\mathbf{u}\mathbf{u}'|\mathbf{X}] = \begin{pmatrix} \sigma^2 h_1 & 0 & \cdots & 0 \\ 0 & \sigma^2 h_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 h_n \end{pmatrix} = \sigma^2 \underbrace{\begin{pmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_n \end{pmatrix}}_{\Psi}.$$

Somit erhalten wir

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad Var(\mathbf{u}|\mathbf{X}) = \sigma^2 \Psi, \quad (8.1)$$

was wir als das **Ausgangsmodell in Matrixschreibweise** bezeichnen werden.

- Werden Modelle mit heteroskedastischen Fehlern geschätzt, sind drei Fälle zu unterscheiden:
 1. Die Funktion $h(\cdot)$ ist bekannt, siehe Abschnitt 8.3.
 2. Die Funktion $h(\cdot)$ ist nur teilweise bekannt, siehe Abschnitt 8.4.
 3. Die Funktion $h(\cdot)$ ist vollständig unbekannt, siehe Abschnitt 8.2.

8.1 Auswirkungen der Heteroskedastie auf die KQ-Schätzer

- Der KQ-Schätzer ist weiterhin **unverzerrt** und **konsistent**.
- **Varianz des KQ-Schätzers** bei heteroskedastischen Fehlern (vgl. Abschnitt 3.4.2):

Aus $\hat{\beta} - \beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$ kann man ableiten, dass

$$\begin{aligned}
 Var(\hat{\beta}|\mathbf{X}) &= E \left[(\hat{\beta} - \beta) (\hat{\beta} - \beta)' | \mathbf{X} \right] \\
 &= E \left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} | \mathbf{X} \right] \\
 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \underbrace{E [\mathbf{u}\mathbf{u}' | \mathbf{X}]}_{\sigma^2 \Psi} \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\
 &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\sigma^2 \Psi \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{8.2}
 \end{aligned}$$

- Beachte, dass man für homoskedastische Fehler $\Psi = \mathbf{I}$ erhält. (8.2) führt dann zur bereits bekannten KQ-Kovarianzmatrix, nämlich $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$.
- **Verwendet man die gewöhnliche Kovarianzmatrix** $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$, **obwohl Heteroskedastie vorliegt**, ist dies irreführend und führt zu fehlerhaften Schlussfolgerungen.

- Wir können (8.2) aber nicht direkt verwenden, da Ψ unbekannt ist. Im nächsten Abschnitt wird ein passender Schätzer vorgestellt.
- Selbst wenn Ψ bekannt sein sollte, ist der KQ- nicht der BLU-Schätzer und somit auch **nicht effizient**. Stattdessen muss der verallgemeinerte KQ-Schätzer (= GLS = generalized least squares estimator) verwendet werden, siehe Abschnitt 8.3.

8.2 Heteroskedastie-robuste Standardfehler bei KQ-Schätzung

- **Ableitung heteroskedastie-robuster Standardfehler**

Sei $\mathbf{x}'_i = (1, x_{i1}, \dots, x_{ik})$. Beachte, dass der mittlere Term der Kovarianzmatrix (8.2) mit Dimension $(k+1) \times (k+1)$ auch geschrieben werden kann als

$$\mathbf{X}' \sigma^2 \Psi \mathbf{X} = \sum_{i=1}^n \sigma^2 h_i \mathbf{x}'_i \mathbf{x}_i.$$

Da $E[u_i^2 | \mathbf{X}] = \sigma^2 h_i$, kann man $\sigma^2 h_i$ durch den “Durchschnitt auf Basis von einer Beobachtung” u_i^2 schätzen. Dies ist natürlich kein besonders guter Schätzer aber für unseren Zweck tut er's. Da u_i unbekannt ist, nimmt man das Residuum \hat{u}_i .

Demnach kann man die Kovarianzmatrix (8.2) des KQ-Schätzers bei Heteroskedastie mittels

$$\widehat{Var}(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{i=1}^n \hat{u}_i^2 \mathbf{x}_i \mathbf{x}_i' \right) (\mathbf{X}'\mathbf{X})^{-1} \quad (8.3)$$

schätzen.

- Anmerkungen:

- Die Standardfehler, die man aus (8.3) erhält, bezeichnet man als **heteroskedastie-robuste Standardfehler** oder auch als **White-Standardfehler**. Letztere Bezeichnung führt auf Halbert White zurück, einen Ökonometriker an der University of California in San Diego.
- Für ein einzelnes $\hat{\beta}_j$ kann der heteroskedastie-robuste Standardfehler kleiner oder größer sein als der gewöhnliche KQ-

Standardfehler.

- Es kann gezeigt werden, dass der KQ-Schätzer $\hat{\beta}$ keine bekannte endliche Stichprobenverteilung mehr besitzt, wenn man heteroskedastie-robuste Standardfehler verwendet. Er ist jedoch **asymptotisch normalverteilt**. Es bleiben also die kritischen Werte und die p -values approximativ gültig, falls man (8.3) verwendet.
- Der KQ-Schätzer mit White-Standardfehlern ist **unverzerrt** und **konsistent**, da MLR.1 bis MLR.4 von Heteroskedastie unberührt bleiben .
- Der KQ-Schätzer ist aber **nicht effizient**. Effiziente Schätzer werden im nächsten Abschnitt behandelt.

8.3 Der verallgemeinerte KQ-Schätzer (GLS-Schätzer)

- Ausgangsmodell (8.1):

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + u_i, \quad (8.4)$$

$$Var(u_i|x_{i1}, \dots, x_{ik}) = \sigma^2 h(x_{i1}, \dots, x_{ik}) = \sigma^2 h_i.$$

- **Grundidee:** Gewichtete Schätzung von (8.4):

Man transformiert das Ausgangsmodell in ein Modell, dass alle Annahmen, einschließlich MLR.5, erfüllt. Dies geschieht durch eine Art Standardisierung der Regressionsfehler u_i . Erreicht wird dies, indem die u_i und mit ihnen die gesamte Regressionsgleichung (8.4) durch die Quadratwurzel von h_i dividiert werden:

$$\underbrace{\frac{y_i}{\sqrt{h_i}}}_{y_i^*} = \beta_0 \underbrace{\frac{1}{\sqrt{h_i}}}_{x_{i0}^*} + \beta_1 \underbrace{\frac{x_{i1}}{\sqrt{h_i}}}_{x_{i1}^*} + \dots + \beta_k \underbrace{\frac{x_{ik}}{\sqrt{h_i}}}_{x_{ik}^*} + \underbrace{\frac{u_i}{\sqrt{h_i}}}_{u_i^*}.$$

Das sich daraus ergebende Modell ist

$$y_i^* = \beta_0 x_{i0}^* + \beta_1 x_{i1}^* + \dots + \beta_k x_{ik}^* + u_i^*. \quad (8.5)$$

Beachte: Der transformierte Fehler u_i^* hat die Varianz

$$\begin{aligned} \text{Var}(u_i^* | x_{i1}, \dots, x_{ik}) &= \text{Var} \left(\frac{u_i}{\sqrt{h_i}} \middle| x_{i1}, \dots, x_{ik} \right) \\ &= E \left[\frac{u_i^2}{h_i} \middle| x_{i1}, \dots, x_{ik} \right] \\ &= \frac{1}{h_i} E[u_i^2 | x_{i1}, \dots, x_{ik}] = \frac{1}{h_i} \sigma^2 h_i = \sigma^2. \end{aligned}$$

Ergebnis: Wir haben das Ausgangsmodell (8.4) dahingehend transformiert, dass nun für das Regressionsmodell (8.5) die Annahme homoskedastisch verteilter Fehler MLR.5 gilt.

- Deshalb besitzt der KQ-Schätzer auf Grundlage des **transformierten Modells** (8.5) alle gewünschten Eigenschaften: **BLU** (best linear unbiased).
- Dem KQ-Schätzer des transformierten Modells (8.5) liegt die Minimierung einer *gewichteten* Residuenquadratsumme (SSR) zugrunde

$$\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_k x_{ik})^2 / h_i.$$

Er wird deswegen auch **gewichteter KQ-Schätzer** (= WLS = weighted LS) genannt. Beachte, dass in der vorliegenden Form $h(\cdot)$ bekannt sein muss.

- Falls $\sqrt{h_i}$ nicht identisch mit einem Regressor in Modell (8.4) ist, besitzt das transformierte Modell keine Konstante.
- Es folgt das transformierte Modell in Matrixschreibweise.

- Detaillierte Darstellung von \mathbf{y}^* , \mathbf{X}^* , und \mathbf{u}^* in Matrixschreibweise:

$$\underbrace{\begin{pmatrix} y_1^* \\ y_2^* \\ \vdots \\ y_n^* \end{pmatrix}}_{\mathbf{y}^*} = \underbrace{\begin{pmatrix} h_1^{-1/2} & 0 & \cdots & 0 \\ 0 & h_2^{-1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_n^{-1/2} \end{pmatrix}}_{\mathbf{P}} \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{\mathbf{y}}$$

$$\underbrace{\begin{pmatrix} x_{10}^* & x_{11}^* & \cdots & x_{1k}^* \\ x_{20}^* & x_{21}^* & \cdots & x_{2k}^* \\ \vdots & \vdots & & \vdots \\ x_{n0}^* & x_{n1}^* & \cdots & x_{nk}^* \end{pmatrix}}_{\mathbf{X}^*} = \mathbf{P} \cdot \underbrace{\begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ \vdots & \vdots & & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix}}_{\mathbf{X}}, \quad \underbrace{\begin{pmatrix} u_1^* \\ u_2^* \\ \vdots \\ u_n^* \end{pmatrix}}_{\mathbf{u}^*} = \mathbf{P} \cdot \underbrace{\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}}_{\mathbf{u}}$$

- Für die Transformationsmatrix \mathbf{P} gilt, dass

$$\mathbf{P}'\mathbf{P} = \Psi^{-1}$$

und somit

$$E[\mathbf{u}\mathbf{u}'|\mathbf{X}] = \sigma^2 \Psi = \sigma^2 (\mathbf{P}'\mathbf{P})^{-1}.$$

- Damit lässt sich das **transformierte Modell (8.5) in Matrixschreibweise** darstellen als

$$\mathbf{P}\mathbf{y} = \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}\mathbf{u},$$

oder

$$\mathbf{y}^* = \mathbf{X}^*\boldsymbol{\beta} + \mathbf{u}^*, \quad E[\mathbf{u}^*(\mathbf{u}^*)'|\mathbf{X}^*] = \sigma^2 \mathbf{I}. \quad (8.6)$$

- Man erhält (8.6), indem man das Ausgangsmodell (8.1) $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ von links mit der Transformationsmatrix \mathbf{P} multipliziert.
- Wie sieht nun die Formel für den KQ-Schätzer des transformierten Modells (8.6) aus?

Verallgemeinerter KQ-Schätzer

- KQ-Schätzung von (8.6) ergibt

$$\begin{aligned}
 \hat{\beta}_{GLS} &= (\mathbf{X}^{*'} \mathbf{X}^*)^{-1} \mathbf{X}^{*'} \mathbf{y}^* \\
 &= ((\mathbf{P}\mathbf{X})' \mathbf{P}\mathbf{X})^{-1} (\mathbf{P}\mathbf{X})' \mathbf{P}\mathbf{y} \\
 &= (\mathbf{X}' \mathbf{P}' \mathbf{P}\mathbf{X})^{-1} \mathbf{X}' \mathbf{P}' \mathbf{P}\mathbf{y}
 \end{aligned}$$

und demnach

$$\hat{\beta}_{GLS} = (\mathbf{X}' \Psi^{-1} \mathbf{X})^{-1} \mathbf{X}' \Psi^{-1} \mathbf{y}. \quad (8.7)$$

$\hat{\beta}_{GLS}$ in (8.7) wird als **verallgemeinerter KQ-Schätzer** oder als **GLS-Schätzer** — generalized least squares estimator (GLS estimator) — bezeichnet.

Bei Heteroskedastie ist Ψ eine Diagonalmatrix und jede der n Beobachtungen wird mit $1/\sqrt{h_i}$ gewichtet.

- **Eigenschaften** bei bekannter Funktion $h(\cdot)$:

Unter MLR.1 bis MLR.4 sowie GLS.5 ist der GLS-Schätzer $\hat{\beta}_{GLS}$

- **erwartungstreu** und **konsistent**,
- **BLUE** (best linear unbiased) und somit **effizient**,
- besitzt die Kovarianzmatrix $Var(\hat{\beta}_{GLS} | \mathbf{X}) = \sigma^2 (\mathbf{X}' \boldsymbol{\Psi}^{-1} \mathbf{X})^{-1}$
- und ist selbst dann unverzerrt und konsistent, falls $\boldsymbol{\Psi}$ fehlspezifiziert sein sollte, da $\boldsymbol{\Psi}$ eine Funktion in \mathbf{X} und nicht in \mathbf{u} ist und somit

$$E[\hat{\beta}_{GLS} - \beta | \mathbf{X}] = (\mathbf{X}' \boldsymbol{\Psi}^{-1} \mathbf{X})^{-1} \mathbf{X}' \boldsymbol{\Psi}^{-1} E[\mathbf{u} | \mathbf{X}] = \mathbf{0}.$$

Aus all dem folgt, dass der KQ-Schätzer ineffizient ist, da KQ- und verallgemeinerter KQ-Schätzer lineare Schätzer sind. Die KQ-Varianzen sind größer oder gleich den Varianzen des GLS-Schätzers. Dies kann mittels Matrixalgebra gezeigt werden.

- Analog zu Annahme MLR.6 aus Abschnitt 4.2, fordern wir nun

- **Annahme GLS.6: Normalverteilung**

$$u_i | \mathbf{x}_i \sim N(0, \sigma^2 h_i), \quad i = 1, \dots, n,$$

welche, zusammen mit MLR.2 (Zufallsstichprobe) die **multivariate Normalverteilung**

$$\mathbf{u} | \mathbf{X} \sim N(\mathbf{0}, \sigma^2 \Psi)$$

garantiert. Beachte, dass GLS.6 voraussetzt, dass die u_i gegeben \mathbf{x}_i unabhängig aber nicht identisch verteilt sind, da die Varianz für alle i unterschiedlich sein kann (An den Kovarianzen hat sich ja nichts geändert, sie sind weiterhin alle Null wegen MLR.2).

Alle **Teststatistiken**, die auf Grundlage des transformierten Modells (8.6) oder entsprechend angepasst auf Grundlage des Ausgangsmodells (8.1) berechnet werden, weisen die **exakten Verteilungen** aus Kapitel 4 auf (normal, t , F).

- **Häufiges Problem in der Praxis:** h_i ist unbekannt. In diesem Fall muss der durchführbare (feasible) GLS-Schätzer verwendet werden
→ Fall 2.

8.4 Feasible Generalized Least Squares (FGLS)

- In der Regel ist h_i unbekannt und muss daher geschätzt werden. Häufig sind weder die relevanten Einflüsse noch der Funktionszusammenhang bekannt.
- Man benötigt deshalb eine Spezifikation, die so flexibel ist, dass sie eine große Spanne an Möglichkeiten erfassen kann, z.B.

$$h_i = h(x_{i1}, \dots, x_{ik}) = \exp(\delta_1 x_{i1} + \dots + \delta_k x_{ik})$$

und somit

$$Var(u_i|x_{i1}, x_{i2}, \dots, x_{ik}) = \sigma^2 h_i = \sigma^2 \exp(\delta_1 x_{i1} + \dots + \delta_k x_{ik}).$$

Bemerkung: Auf pp. 282, fügt [Wooldridge \(2009\)](#) der Spezifikation von h_i noch den Faktor $\exp \delta_0$ hinzu. Da dieser Einfluss konstant ist, kann er ebenfalls in σ^2 erfasst werden.

- Wie lassen sich die unbekannten Parameter $\delta_1, \dots, \delta_k$ schätzen?

Standardisierung von u_i führt zu $v_i = u_i/(\sigma\sqrt{h_i})$ mit $E[v_i|\mathbf{X}] = 0$ und $Var(v_i|\mathbf{X}) = 1$. Somit ist $u_i = \sigma\sqrt{h_i}v_i$ und

$$u_i^2 = \sigma^2 h_i v_i^2, \quad i = 1, \dots, n.$$

Logarithmieren führt zu

$$\begin{aligned} \ln u_i^2 &= \ln \sigma^2 + \ln h_i + \ln v_i^2 \\ &= \ln \sigma^2 + \ln \exp(\delta_1 x_{i1} + \dots + \delta_k x_{ik}) + \ln v_i^2 \\ &= \underbrace{\ln \sigma^2 + E[\ln v_i^2]}_{\alpha_0} + \delta_1 x_{i1} + \dots + \delta_k x_{ik} + \underbrace{\ln v_i^2 - E[\ln v_i^2]}_{e_i} \\ \ln u_i^2 &= \alpha_0 + \delta_1 x_{i1} + \dots + \delta_k x_{ik} + e_i. \end{aligned} \tag{8.8}$$

Für die Regressionsgleichung (8.8) sind die Annahmen MLR.1-MLR.4 erfüllt. Somit ist der KQ-Schätzer für δ_j unverzerrt und konsistent.

In der Praxis werden die u_i^2 's der Varianzregression (8.8) durch die \hat{u}_i^2 's aus der Stichprobenregression $\mathbf{y} = \mathbf{X}\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}$ aus (8.1) ersetzt. Man erhält so die $\hat{\delta}_j$'s für die Berechnung der gefitteten Werte \hat{h}_i , die im zweiten Schritt in den GLS-Schätzer (8.7) eingesetzt werden.

- Leitfaden zur **FGLS**-Methode:

Erster Schritt

- Regressiere \mathbf{y} auf \mathbf{X} und berechne den Residuenvektor $\hat{\mathbf{u}}$ mittels KQ-Schätzung des Ausgangsmodells (8.1).
- Berechne die $\ln \hat{u}_i^2$'s, $i = 1, \dots, n$, die als abhängige Variable in der Varianzregression (8.8) dienen.
- Schätze die Varianzregression (8.8) mit KQ.
- Berechne $\hat{h}_i = \exp \left(\hat{\delta}_1 x_{i1} + \dots + \hat{\delta}_k x_{ik} \right)$, $i = 1, \dots, n$.

Zweiter Schritt

Den FGLS-Schätzer $\hat{\beta}_{FGLS}$ erhält man **analog** zum Vorgehen bei GLS. Das Ausgangsmodell (8.1) wird von links mit der Matrix

$$\hat{\mathbf{P}} = \begin{pmatrix} \hat{h}_1^{-1/2} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{h}_n^{-1/2} \end{pmatrix}$$

multipliziert. Man erhält eine Variante des transformierten Regressionsmodells

$$\mathbf{y}^\# = \mathbf{X}^\# \boldsymbol{\beta} + \mathbf{u}^\#. \quad (8.9)$$

Die KQ-Schätzung von (8.9) führt also zum **FGLS-Schätzer**

$$\hat{\boldsymbol{\beta}}_{FGLS} = \left(\mathbf{X}' \hat{\boldsymbol{\Psi}}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}' \hat{\boldsymbol{\Psi}}^{-1} \mathbf{y}, \quad (8.10)$$

mit $\hat{\boldsymbol{\Psi}}^{-1} = \hat{\mathbf{P}}' \hat{\mathbf{P}}$.

- **Schätzeigenschaften der FGLS-Schätzer:**

- Sie sind **konsistent**, d.h. sie konvergieren in Wahrscheinlichkeit gegen die wahren Werte, wenn $n \rightarrow \infty$

$$\text{plim } \hat{\beta}_{FGLS} = \beta.$$

- Der FGLS-Schätzer ist **asymptotisch effizient**: Für ein korrekt spezifiziertes h_i und eine hinreichend große Stichprobe ist der FGLS-Schätzer dem KQ-Schätzer vorzuziehen, da Ersterer eine kleinere Schätzvarianz hat. (Dies ist plausibel, schließlich nutzt der FGLS-Schätzer – anders als der KQ-Schätzer mit heteroskedastie-robusten Standardfehlern – auch die Information bezüglich der funktionalen Form der Heteroskedastie.)
- Ist die Varianzfunktion h_i **fehlspezifiziert**, ist der FGLS-Schätzer **ineffizient**.

- Seien Sie sich bewusst, dass die FGLS- und KQ-Schätzungen erheblich voneinander abweichen können.
- **Vergleich: KQ mit heteroskedastie-robusten Standardfehlern versus FGLS**
 - Hat man eine Idee, wie die Varianzfunktion h_i aussehen könnte, ist FGLS zu bevorzugen. Falls nicht, könnte KQ mit heteroskedastie-robusten Standardfehlern besser sein.
 - Es ist immer eine **gute Idee, die KQ-Regression auch mit heteroskedastie-robusten Standardfehlern durchzuführen**, um zu sehen, ob die Signifikanz der Parameter von möglicher Heteroskedastie abhängt.
 - Da jeder Schätzer, der Heteroskedastie berücksichtigt, vermieden werden soll, wenn gar keine Heteroskedastie vorliegt, sollte man

auf Heteroskedastie testen, siehe Abschnitt 9.2.

- **Fortsetzung des Außenhandelsbeispiels**

- Wir betrachten Modell 5 aus Abschnitt 6.3 und vergleichen KQ-, FGLS-, und KQ-Schätzungen mit heteroskedastie-robusten Standardfehlern miteinander.
- Ein R-Programm zur Durchführung von KQ, FGLS mit beiden Schritten, sowie KQ mit White-Standardfehlern, und zur Anzeige von Scatterplots der Residuen gegen die gefitteten Werte sieht wie folgt aus (Ausschnitt aus EOE_ws19_Emp_Beispiele.R, Zeilen 672 folgende):

```
# definiere log der abhängigen Variable
log_imp <- log(trade_0_d_o)

### Erster Schritt a) KQ-Regression und Berechnung der Residuen

# KQ-Regression
```

```
eq_ols_model5 <- lm(log_imp ~ log(wdi_gdpusdcr_o) + I((log(wdi_gdpusdcr_o))^2) +  
                      log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))  
  
# Berechne Residuen  
res_ols_model5 <- eq_ols_model5$resid  
  
# Berechne gefittete/angepasste Werte  
fit_ols_model5 <- fitted.values(eq_ols_model5)  
  
# Plotte die Residuen gegen die gefitteten Werte, um zu untersuchen,  
# ob Heteroskedastie vorliegen könnte  
dev.off()  
plot(fit_ols_model5, res_ols_model5, pch = 16)  
  
### Erster Schritt b) bis d)  
  
# Quadriere die Residuen und logarithmiere sie anschließend  
ln_u_hat_sq <- log(res_ols_model5^2)  
  
# Schätze die Varianzgleichung  
eq_h_model5 <- lm(ln_u_hat_sq ~ log(wdi_gdpusdcr_o) + I((log(wdi_gdpusdcr_o))^2) +  
                      log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))  
  
# Berechne die gefitteten Werte der logarithmierten Residuenanalyse
```

```
ln_u_hat_sq_hat <- fitted.values(eq_h_model5)

# Berechne die h's aus den gefitteten Werten der Varianzregression
h_hat <- exp(ln_u_hat_sq_hat)

### Zweiter Schritt: FGLS-Schätzung

# Schätze FGLS mit den gewichteten weights = 1/h_hat
eq_fgls_model5 <- lm(log_imp ~ log(wdi_gdpusdcr_o) + I((log(wdi_gdpusdcr_o))^2) +
                        log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o),
                        weights = 1/h_hat)
summary(eq_fgls_model5)

# Berechne die gefitteten Werte aus FGLS
fit_fgls_model5 <- fitted.values(eq_fgls_model5)

# Berechne die Residuen aus FGLS
res_fgls_model5 <- resid(eq_fgls_model5)

# Standardisierung der Residuen mittels der Gewichte
res_fgls_model5_star <- res_fgls_model5*h_hat^(-1/2)

# Plotte die Residuen gegen die gefitteten Werte
```

```
plot(fit_fgls_model5, res_fgls_model5_star, pch = 16)

### KQ-Regression mit heteroskedastie-robusten Standardfehlern
library(lmtest)
eq_white_model5 <- coeftest(eq_ols_model5, vcov=hccm(eq_ols_model5, type="hc1"))

# Graphiken/Outputs für Skript
summary(eq_ols_model5)
summary(eq_h_model5)
summary(eq_fgls_model5)
eq_white_model5
```

– KQ-Output mit gewöhnlichen Standardfehlern

Call:

```
lm(formula = log_imp ~ log(wdi_gdpusdcr_o) + I((log(wdi_gdpusdcr_o))^2) +
  log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))
```

Residuals:

Min	1Q	Median	3Q	Max
-2.0672	-0.5451	0.1153	0.5317	1.3870

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-35.23314	17.44175	-2.020	0.04964 *
log(wdi_gdpusdcr_o)	3.90881	1.32836	2.943	0.00523 **
I((log(wdi_gdpusdcr_o))^2)	-0.05711	0.02627	-2.174	0.03523 *
log(cepii_dist)	-0.74856	0.16317	-4.587	3.86e-05 ***
ebrd_tfes_o	0.41988	0.20056	2.094	0.04223 *
log(cepii_area_o)	-0.13238	0.08228	-1.609	0.11497

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.8191 on 43 degrees of freedom

Multiple R-squared: 0.9155, Adjusted R-squared: 0.9056

F-statistic: 93.12 on 5 and 43 DF, p-value: < 2.2e-16

– FGLS - 1ter Schritt: Schätze Varianzregression (8.8)

Call:

```
lm(formula = ln_u_hat_sq ~ log(wdi_gdpusdcr_o) + I((log(wdi_gdpusdcr_o))^2) +
  log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))
```

Residuals:

Min	1Q	Median	3Q	Max
-5.6970	-0.6885	0.4991	1.4881	2.8326

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	63.57453	48.98487	1.298	0.201
log(wdi_gdpusdcr_o)	-4.79105	3.73067	-1.284	0.206
I((log(wdi_gdpusdcr_o))^2)	0.08839	0.07377	1.198	0.237
log(cepii_dist)	-0.36408	0.45827	-0.794	0.431
ebrd_tfes_o	0.23452	0.56327	0.416	0.679
log(cepii_area_o)	0.03706	0.23109	0.160	0.873

Residual standard error: 2.3 on 43 degrees of freedom

Multiple R-squared: 0.09998, Adjusted R-squared: -0.004677

F-statistic: 0.9553 on 5 and 43 DF, p-value: 0.4557

– FGLS - 2ter Schritt: schätze (8.10)

Call:

```
lm(formula = log_imp ~ log(wdi_gdpusdcr_o) + I((log(wdi_gdpusdcr_o))^2) +
  log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o), weights = 1/h_hat)
```

Weighted Residuals:

Min	1Q	Median	3Q	Max
-4.1788	-1.3479	0.2645	1.2478	3.6620

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-30.66686	16.80239	-1.825	0.0749 .
log(wdi_gdpusdcr_o)	3.55935	1.28177	2.777	0.0081 **
I((log(wdi_gdpusdcr_o))^2)	-0.05016	0.02482	-2.021	0.0495 *
log(cepii_dist)	-0.74852	0.11358	-6.590	5.06e-08 ***
ebrd_tfes_o	0.39046	0.18441	2.117	0.0401 *
log(cepii_area_o)	-0.13856	0.05551	-2.496	0.0165 *

Signif. codes:	0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1			

Residual standard error: 1.973 on 43 degrees of freedom

Multiple R-squared: 0.9055, Adjusted R-squared: 0.8945

F-statistic: 82.41 on 5 and 43 DF, p-value: < 2.2e-16

Im Gegensatz zu EViews erhält man mit dem R-Befehl `eq_fgls_model5 <- lm(..., weights=...)` nur Statistiken für das gewichtete Modell (das “Sternchen”-Modell ((8.6) or (8.9)). Die entsprechende Residuenquadratsumme und weitere Statistiken für das “Sternchen”-Modell, die EViews anzeigt, erhält man mit

```
(SSR <- sum(w_scaled*(log_imp_star - regressor_star%*%coef(eq_fgls_model5))^2)) # SSR
mean(log_imp * (w_scaled)) # Mean dependent var
sd(log_imp * (w_scaled)) # S.D. dependent var
sqrt(SSR/(n-k-1)) # S.E. of regression
```

Entsprechende Statistiken für das ursprüngliche Modell (in EViews “Unweighted Statistics”) erhält man in R mit

```
# R-squared
(r_squared <- 1 - sum(residuals(eq_fgls_model5)^2) /
  sum((log_imp - mean(log_imp))^2))
# Adjusted R-squared
-k/(n-k-1) + (n-1)/(n-k-1)*r_squared
# Mean dependent var
mean(log_imp)
# S.D. dependent var
sd(log_imp)
# S.E. of regression
sqrt(sum(residuals(eq_fgls_model5)^2)/(n-k-1))
# Sum squared resid
sum(residuals(eq_fgls_model5)^2)
```

- KQ mit heteroskedastie-robusten Standardfehlern

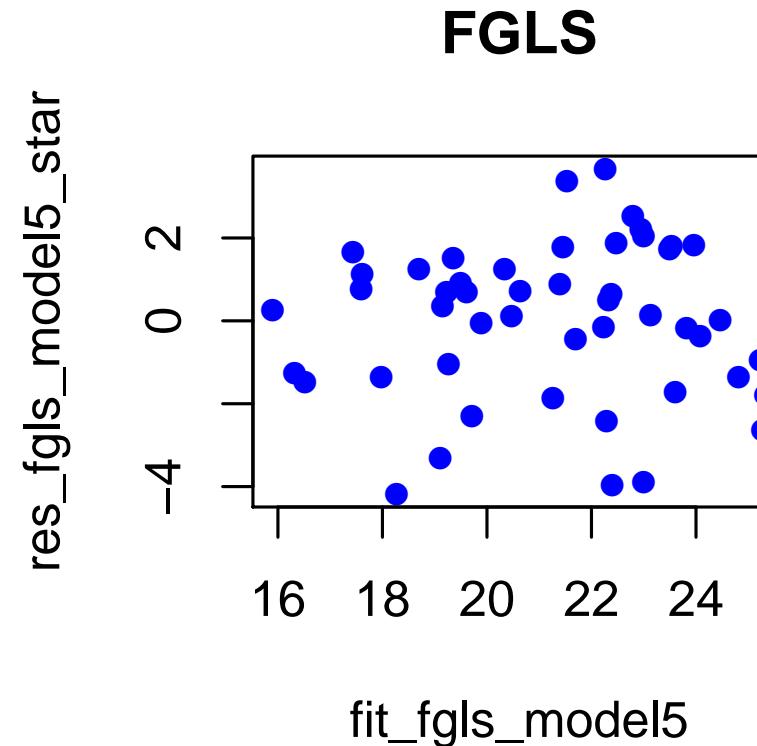
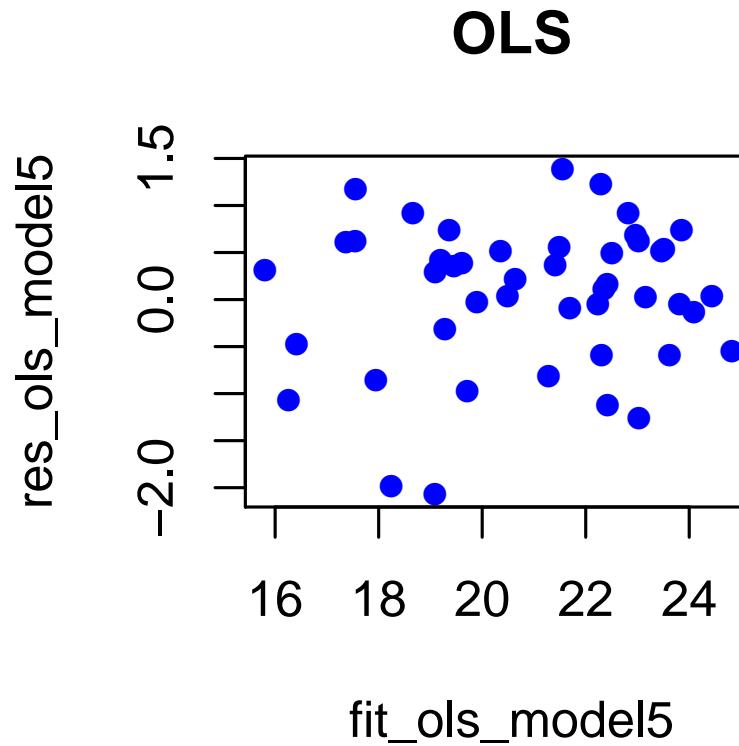
Diese erhält man in R mit dem Befehl `coeftest()` aus dem R-Paket `lmtest`

t test of coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-35.233143	16.148517	-2.1818	0.034635 *
log(wdi_gdpusdcr_o)	3.908811	1.244314	3.1413	0.003041 **
I((log(wdi_gdpusdcr_o))^2)	-0.057108	0.024340	-2.3462	0.023644 *
log(cepii_dist)	-0.748559	0.124427	-6.0160	3.465e-07 ***
ebrd_tfes_o	0.419883	0.155896	2.6934	0.010045 *
log(cepii_area_o)	-0.132380	0.046455	-2.8496	0.006693 **

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’
	0.1 ‘ ’			1

- Diagnoseplots: (Standardisierte) Residuen gegen gefittete Werte



- **Outputtabelle für Modell 4 und Modell 5 für unterschiedliche Schätzer** (vergleiche Abschnitt 4.8):

Abhängige Variable: $\ln(\text{Importe nach Deutschland})$			
Unabhängige Variablen/Modell	(4)-OLS	(5)-OLS	(5)-FGLS
Konstante	2.427 (2.132) [1.337]	-35.233 (17.441) [16.148]	-30.666 (16.802)
$\ln(BIP)$	1.025 (0.076) [0.070]	3.908 (1.328) [1.244]	3.559 (1.281)
$(\ln(BIP))^2$	— — (0.026) [0.024]	-0.057 (0.024)	-0.050
$\ln(\text{Entfernung})$	-0.888 (0.156) [0.120]	-0.748 (0.163) [0.124]	-0.748 (0.113)
$Offenheit$	0.353 (0.206) [0.180]	0.419 (0.200) [0.155]	0.390 (0.184)
$\ln(\text{Flaeche})$	-0.151 (0.085) [0.050]	-0.132 (0.082) [0.046]	-0.138 (0.055)
Stichprobengröße	49	49	49
R^2	0.906	0.915	0.905
Standardfehler der Regression	0.853	0.819	0.736
Residuenquadratsumme	32.017	28.846	23.330
AIC	2.6164	2.5529	
HQ	2.6896	2.6408	
SC	2.8094	2.7845	

Anmerkungen: KQ- bzw. FGLS-Standardfehler in runden, White-Standardfehler in eckigen Klammern

– Ergebnisse und Interpretation:

- * KQ- und FGLS-Parameterschätzungen für sind für alle Parameter sehr ähnlich: Die Auswirkungen möglicher Heteroskedastie sind nur schwach. Es sollte daher überprüft werden, ob überhaupt Heteroskedastie vorliegt. Läge keine Heteroskedastie vor, wäre der FGLS-Schätzer nicht mehr effizient und wir sollten daher beim KQ-Schätzer bleiben.
- * Berücksichtigt man die Heteroskedastie, ist auf Basis der FGLS-Schätzung kein Parameter auf dem 5% Signifikanzniveau insignifikant ebenso wie auf Basis heteroskedastie-robuster KQ-Standardfehler.

* Sieht man sich die Scatterplots der KQ- und der standardisierten FGLS-Residuen gegen die gefitteten Werte an, so schließt man daraus nicht unbedingt auf Heteroskedastie. Auch dies ist ein Hinweis darauf, dass man einen Heteroskedastietest anwenden sollte, siehe Abschnitt 9.2.

- **Zigarettenbeispiel** ([Wooldridge, 2009](#), Example 8.7) mit R

Erster Schritt

1. KQ-Schätzung

```
lm(formula = cigs ~ lincome + lcigpric + educ + age + I(age^2) +
  restaurn, data = smoke_all)

Residuals:
    Min      1Q  Median      3Q      Max
-15.819 -9.381 -5.975  7.922 70.221

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.639855  24.078660 -0.151  0.87988
lincome      0.880268  0.727783  1.210  0.22682
lcigpric     -0.750855  5.773343 -0.130  0.89656
educ        -0.501498  0.167077 -3.002  0.00277 ***
age          0.770694  0.160122  4.813 1.78e-06 ***
I(age^2)     -0.009023  0.001743 -5.176 2.86e-07 ***
restaurn     -2.825085  1.111794 -2.541  0.01124 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 13.4 on 800 degrees of freedom
Multiple R-squared:  0.05274, Adjusted R-squared:  0.04563
F-statistic: 7.423 on 6 and 800 DF,  p-value: 9.499e-08
```

2. Speichere die Residuen mit `u_hat_cig <- resid(ols_1)`

3. Logarithmiere die quadrierten Residuen:

```
ln_u_sq <- log(u_hat_cig^2)
```

4. Schätzung der Varianzregression (8.8) mittels KQ führt zu

```
lm(formula = ln_u_sq ~ lincome + lcigpric + educ + age + I(age^2) +
    restaurn, data = smoke_all)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.2186	-0.2237	-0.0227	0.2951	4.9588

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-1.9207040	2.5630344	-0.749	0.45384
lincome	0.2915405	0.0774683	3.763	0.00018 ***
lcigpric	0.1954209	0.6145390	0.318	0.75057
educ	-0.0797036	0.0177844	-4.482	8.49e-06 ***
age	0.2040054	0.0170441	11.969	< 2e-16 ***
I(age^2)	-0.0023921	0.0001855	-12.893	< 2e-16 ***
restaurn	-0.6270116	0.1183440	-5.298	1.51e-07 ***

Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.427 on 800 degrees of freedom

Multiple R-squared: 0.2474, Adjusted R-squared: 0.2417

F-statistic: 43.82 on 6 and 800 DF, p-value: < 2.2e-16

– Speichere die \hat{h}_i 's mit `h_hat_cig <- exp(ln_u_sq - resid(ols_2))`

Zweiter Schritt

Gewichtete KQ-Schätzung mit den Gewichten $h_{\text{hat_cig}}^{-1}$

Call:

```
lm(formula = cigs ~ lincome + lcigpric + educ + age + I(age^2) +
  restaurn, data = smoke_all, weights = h_hat_cig^(-1))
```

Weighted Residuals:

Min	1Q	Median	3Q	Max
-1.9036	-0.9532	-0.8099	0.8415	9.8556

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.6354329	17.8031409	0.317	0.751674
lincome	1.2952396	0.4370117	2.964	0.003128 **
lcigpric	-2.9403048	4.4601450	-0.659	0.509932
educ	-0.4634464	0.1201587	-3.857	0.000124 ***
age	0.4819480	0.0968082	4.978	7.86e-07 ***
I(age^2)	-0.0056272	0.0009395	-5.990	3.17e-09 ***
restaurn	-3.4610642	0.7955050	-4.351	1.53e-05 ***

Signif. codes:	0 ‘***’	0.001 ‘**’	0.01 ‘*’	0.05 ‘.’
	0.1 ‘ ’	1		

Residual standard error: 1.579 on 800 degrees of freedom

Multiple R-squared: 0.1134, Adjusted R-squared: 0.1068

F-statistic: 17.06 on 6 and 800 DF, p-value: < 2.2e-16

- Vergleiche diese mit den KQ-Schätzungen mit White-Standardfehlern.

9 Multiple Regression: Modelldiagnose

9.1 RESET-Test

RESET Test (REgression Specification Error Test)

Idee und Durchführung:

- Falls im Ausgangsmodell

$$y = x_0\beta_0 + \dots + x_k\beta_k + u = \mathbf{x}'\boldsymbol{\beta} + u$$

Annahme MLR.4 $E[u|x_0, \dots, x_k] = 0$ erfüllt ist, gilt

$$E[y|x_0, \dots, x_k] = x_0\beta_0 + \dots + x_k\beta_k = \mathbf{x}'\boldsymbol{\beta}.$$

- Jeder Term, der dem Modell hinzugefügt wird, sollte demnach insignifikant sein. Somit sollte auch jede nichtlineare Funktion unabhängiger Variablen insignifikant sein.
- Deswegen ist die Nullhypothese des RESET-Tests so formuliert, dass die Signifikanz nichtlinearer Funktionen der gefitteten Werte $\hat{y} = \mathbf{x}'\hat{\boldsymbol{\beta}}$, die dem Modell hinzugefügt wurden, getestet werden kann. Beachte, dass die gefitteten Werte eine lineare Funktion der Regressoren des Ausgangsmodells darstellen.

- In der Praxis erwiesen sich die zweite und dritte Potenz der \hat{y} als hinreichend, um den RESET-Test durchführen zu können:

$$y = \mathbf{x}'\boldsymbol{\beta} + \alpha\hat{y}^2 + \gamma\hat{y}^3 + \varepsilon.$$

Das Hypothesenpaar lautet

$$\begin{aligned} H_0 : \alpha &= 0, \gamma = 0 & (\text{lineares Modell ist korrekt spezifiziert}) \\ H_1 : \alpha &\neq 0 \text{ und/oder } \gamma \neq 0. \end{aligned}$$

Getestet wird die Nullhypothese mittels eines F -Tests mit 2 Freiheitsgraden im Zähler und $n - k - 3$ im Nenner.

- **Beachte**, dass die Nullhypothese auch wegen fehlender relevanter Regressoren abgelehnt werden könnte.
- In R steht der Befehl `resettest()` im R-Paket `lmtest` zur Verfügung.

9.2 Tests auf Heteroskedastie

- Wie bereits erwähnt, ist es nicht sinnvoll “automatisch” den FGLS-Schätzer zu verwenden. Sind die Fehler homoskedastisch, sollte der KQ-Schätzer mit den gewöhnlichen KQ-Fehlern verwendet werden.
- Man sollte also vorher testen, ob statistische Evidenz für Heteroskedastie vorliegt.
- Im Folgenden werden zwei unterschiedliche Tests vorgestellt: Der Breusch-Pagan-Test und der White-Test. Beide haben “homoskedastische Fehler” als Nullhypothese .
- Beide Tests sind in R verfügbar. Der Breusch-Pagan-Test `bptest` im R-Paket `lmtest`. Der White-Test `white_lm` im R-Paket `skedastic`. Letzterer ist auch in `EOE_ws19_Emp_Beispiele.R`, Zeilen 848 folgende, programmiert.

Es wird angenommen, dass für das multiple lineare Regressionsmodell

$$y = \beta_0 + x_1\beta_1 + \dots + x_k\beta_k + u$$

die Annahmen MLR.1 bis MLR.4 erfüllt sind.

Das zu testende **Hypothesenpaar** lautet

$$H_0 : \text{Var}(u_i|\mathbf{x}_i) = \sigma^2 \quad (\text{Homoskedastie}),$$

$$H_1 : \text{Var}(u_i|\mathbf{x}_i) = \sigma_i^2 \neq \sigma^2 \quad (\text{Heteroskedastie}).$$

Die Grundidee der Heteroskedastie-Tests ist, dass unter der Nullhypothese kein Regressor Erklärungsgehalt für $\text{Var}(u_i|\mathbf{x}_i)$ haben sollte. Gilt die Nullhypothese nicht, kann die bedingte Varianz $\text{Var}(u_i|\mathbf{x}_i)$ durch (beinahe) jede beliebige Funktion der Regressoren x_j , ($1 \leq j \leq k$) bestimmt sein.

Beachte: Der Breusch-Pagan- und der White-Test unterscheiden sich bezüglich ihrer Alternativhypothese.

Breusch-Pagan-Test

- Idee: Betrachten wir die Regression

$$u_i^2 = \delta_0 + \delta_1 x_{i1} + \cdots + \delta_k x_{ik} + v_i, \quad i = 1, \dots, n. \quad (9.1)$$

Unter den Annahmen MLR.1 bis MLR.4 ist der KQ-Schätzer für die δ_j 's unverzerrt.

Das Hypothesenpaar lautet somit:

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_k = 0 \quad \text{versus}$$

$$H_1 : \delta_1 \neq 0 \text{ und/oder } \delta_2 \neq 0 \text{ und/oder } \dots,$$

da unter H_0 gilt, dass $E[u_i^2 | \mathbf{X}] = \delta_0$.

• Abweichungen von der bisherigen Anwendung des F -Tests:

- Die quadrierten Fehler u_i^2 sind auf keinen Fall normalverteilt, weil sie quadrierte Größen sind und somit keine negativen Werte annehmen können. Somit können auch die v_i nicht normalverteilt sein und die F -Verteilung der F -Statistik ist bei endlichen Stichproben nicht exakt gültig. Jedoch lässt sich auch hier der Zentrale Grenzwertsatz (CLT) anwenden, siehe Abschnitt 5.2, und die F -Statistik folgt in großen Stichproben approximativ der F -Verteilung.
- Die Fehler u_i sind unbekannt. Sie können aber durch die Residuen \hat{u}_i der KQ-Schätzung ersetzt werden, ohne dass die asymptotische Gültigkeit des F -Tests dadurch beeinflusst würde (der Beweis ist formal recht aufwändig und unterbleibt hier).

- Man kann auch die R^2 -Version der Teststatistik verwenden. Beachte, dass das R^2 wegen $\text{SSR} = \text{SST}$ gleich Null ist, falls nur auf eine Konstante regressiert wird (es ist dann ja gar kein Regressor vorhanden, der Streuung aufweist). Wir bezeichnen das Bestimmtheitsmaß der KQ-Schätzung aus (9.1) mit $R_{\hat{u}^2}^2$ und erhalten

$$F = \frac{R_{\hat{u}^2}^2/k}{(1 - R_{\hat{u}^2}^2)/(n - k - 1)}.$$

Die Teststatistik des Overall- F -Tests, der auf die gemeinsame Signifikanz aller Regressoren testet, wird von den meisten Softwareprogrammen standardmäßig ausgegeben.

- H_0 wird dann abgelehnt, wenn F den kritischen Wert für ein gewähltes Signifikanzniveau übersteigt (oder wenn der p -value kleiner ist als das Signifikanzniveau).

- **Fortsetzung des Zigarettenbeispiels: (aus Abschnitt 8.4):**

```
lm(formula = u_hat_sq ~ lincome + lcigpric + educ + age + I(age^2) +
  restaurn, data = smoke_all)
```

Residuals:

Min	1Q	Median	3Q	Max
-270.1	-127.5	-94.0	-39.1	4667.8

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-636.30311	652.49456	-0.975	0.3298
lincome	24.63849	19.72180	1.249	0.2119
lcigpric	60.97656	156.44869	0.390	0.6968
educ	-2.38423	4.52753	-0.527	0.5986
age	19.41748	4.33907	4.475	8.75e-06 ***
I(age^2)	-0.21479	0.04723	-4.547	6.27e-06 ***
restaurn	-71.18138	30.12789	-2.363	0.0184 *

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

Residual standard error: 363.2 on 800 degrees of freedom

Multiple R-squared: 0.03997, Adjusted R-squared: 0.03277

F-statistic: 5.552 on 6 and 800 DF, p-value: 1.189e-05

Die F -Statistik für obige H_0 beträgt 5.55 und der zugehörige p -value ist kleiner als 1%. Die Nullhypothese homoskedastischer Fehler wird somit auf dem 1%-Niveau abgelehnt.

- **Beachte:**

- Vermutet man, dass die Heteroskedastie von speziellen Variablen verursacht wird, die zuvor nicht in der Regression berücksichtigt wurden, können diese in die Regression (9.1) eingefügt werden.
- Falls H_0 nicht abgelehnt wird, bedeutet dies nicht automatisch, dass die u_i 's homoskedastisch sind. Sollte die Spezifikation (9.1) nicht alle relevanten Variablen enthalten, die Heteroskedastie verursachen könnten, kann es passieren, dass alle δ_j , $j = 1, \dots, k$ gemeinsam insignifikant sind.
- Eine Variante des Breusch-Pagan-Tests ist ein Test auf multiplikative Heteroskedastie, d.h. die Varianz hat die Form $\sigma_i^2 = \sigma^2 \cdot h(\mathbf{x}'_i \boldsymbol{\beta})$. Wird etwa der Fall $h(\cdot) = \exp(\cdot)$ angenommen, erhält man die Testgleichung $\ln(\hat{u}_i^2) = \ln(\sigma^2) + \mathbf{x}'_i \boldsymbol{\beta} + v$.

White-Test

- **Hintergrund:**

Um die asymptotische Verteilung des KQ-Schätzers abzuleiten, wird die Annahme homoskedastischer Fehler MLR.5 *nicht* benötigt.

Es reicht bereits aus, dass die quadrierten Fehler u_i^2 mit allen Regressoren, deren Quadraten und deren Kreuzprodukten unkorreliert sind.

Dies lässt sich recht einfach mit folgender Regression testen, wobei

die unbekannten Fehler bereits durch die Residuen ersetzt wurden:

$$\begin{aligned}
 \hat{u}_i^2 &= \delta_0 + \delta_1 x_{i1} + \cdots + \delta_k x_{ik} \\
 &\quad + \delta_{k+1} x_{i1}^2 + \cdots + \delta_{J_1} x_{ik}^2 \\
 &\quad + \delta_{J_1+1} x_{i1} x_{i2} + \cdots + \delta_{J_2} x_{ik-1} x_{ik} \\
 &\quad + v_i, \quad i = 1, \dots, n.
 \end{aligned} \tag{9.2}$$

- Das Hypothesenpaar lautet:

$$H_0 : \delta_j = 0 \text{ für } j = 1, 2, \dots, J_2,$$

$$H_1 : \delta_j \neq 0 \text{ für mindestens ein } j.$$

Es kann wieder ein F -Test verwendet werden, dessen Verteilung approximativ die F -Verteilung ist (asymptotische Verteilung).

Besser bekannt ist die LM Version des Tests. Die LM-Statistik ist gegeben durch $LM = n R^2$ wobei man R^2 aus der Schätzung von (9.2) erhält. Die LM-Statistik ist asymptotisch $\chi^2(J_2)$ -verteilt.

- Hat man viele Regressoren, ist es mühsam den F -Test für (9.2) per Hand durchzuführen. Die meisten Softwareprogramme liefern den White-Test bereits mit.
- Bei großem k muss bei der Durchführung des White-Tests eine große Anzahl an Parametern geschätzt werden. In kleinen Stichproben ist dies kaum zu verwirklichen. Man nimmt dann nur die Quadrate x_{ij}^2 in die Regression auf und vernachlässigt alle Kreuzprodukte.
- **Beachte:** Sollte die Nullhypothese abgelehnt werden, kann dies auch daran liegen, dass MLR.1 oder MLR.4 verletzt ist. Dann ist die Ausgangsgleichung **fehlspezifiziert!**
- **Fortsetzung des Zigarettenbeispiels:**

Verwendung einer selbsterstellten R-Funktion `whitetest()`, siehe Anhang 10.5, Folie LXVII. Nicht alle Ergebniszeilen angegeben:

F Statistic	df1	df2	p Value
2.159257e+00	2.500000e+01	7.810000e+02	9.047555e-04

LM Statistic	df	p Value
52.172439390	25.000000000	0.001139947

Call:

```
lm(formula = form, data = dat)
```

Residuals:

Min	1Q	Median	3Q	Max
-326.8	-138.2	-81.2	-10.4	4620.0

Coefficients: (1 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.937e+04	2.056e+04	1.429	0.1535
lincome	-1.050e+03	9.634e+02	-1.089	0.2763
lcigpric	-1.034e+04	9.755e+03	-1.060	0.2894
educ	-1.175e+02	2.513e+02	-0.467	0.6403
age	-2.641e+02	2.358e+02	-1.120	0.2629
I(age^2)	3.469e+00	3.195e+00	1.086	0.2779
restaurn	-2.868e+03	2.987e+03	-0.960	0.3372
I(lincome^2)	-3.941e+00	1.707e+01	-0.231	0.8175
I(lcigpric^2)	6.685e+02	1.204e+03	0.555	0.5790
I(educ^2)	-2.903e-01	1.288e+00	-0.225	0.8217
I(I(age^2)^2)	1.178e-04	1.458e-04	0.808	0.4196
I(restaurn^2)	NA	NA	NA	NA
lincome:lcigpric	3.299e+02	2.392e+02	1.379	0.1683
lincome:educ	-9.592e+00	8.047e+00	-1.192	0.2336
lincome:age	-3.355e+00	6.682e+00	-0.502	0.6158
lincome:I(age^2)	2.670e-02	7.302e-02	0.366	0.7147
lincome:restaurn	-5.989e+01	4.969e+01	-1.205	0.2285

```
lcigpric:educ      3.291e+01  5.906e+01  0.557  0.5775
lcigpric:age       6.288e+01  5.529e+01  1.137  0.2558
lcigpric:I(age^2) -6.224e-01  5.947e-01 -1.046  0.2957
lcigpric:restaurn  8.622e+02  7.206e+02  1.196  0.2319
educ:age           3.617e+00  1.725e+00  2.097  0.0363 *
educ:I(age^2)     -3.556e-02  1.766e-02 -2.013  0.0445 *
educ:restaurn     -2.896e+00  1.066e+01 -0.272  0.7859
age:I(age^2)       -1.911e-02  2.866e-02 -0.667  0.5050
age:restaurn       -4.933e+00  1.084e+01 -0.455  0.6492
I(age^2):restaurn 3.845e-02  1.205e-01  0.319  0.7497
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Residual standard error: 362.9 on 781 degrees of freedom

Multiple R-squared: 0.06465, Adjusted R-squared: 0.03471

F-statistic: 2.159 on 25 and 781 DF, p-value: 0.0009048

Ergebnis: Auch mit dem White-Test wird H_0 abgelehnt.

Fortsetzung des Außenhandelsbeispiels (aus Abschnitt 8.4)

- Breusch-Pagan-Test auf Heteroskedastie mit den KQ-Residuen mit dem R-Befehl `bptest()` im R-Paket `lmtest`

studentized Breusch-Pagan test

```
data: eq_ols_model5
BP = 5.3378, df = 5, p-value = 0.3761
```

- White-Test (ohne Kreuzprodukte) auf Heteroskedastie mit den KQ-Residuen

```
# führe White-Test durch, Funktion whitetest() auf Folie 399 definiert
ols_model5_white <- whitetest(eq_ols_model5, crossterms=0)
```

Ergebnis:

F Statistic	df1	df2	p Value
1.0337453	5.0000000	43.0000000	0.4101294

LM Statistic	df	p Value
5.2579260	5.0000000	0.3852202

Call:

```
lm(formula = form, data = dat)
```

Residuals:

Min	1Q	Median	3Q	Max
-0.8842	-0.3981	-0.1658	0.1013	3.2860

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	4.879e+00	4.939e+00	0.988	0.329

I(log(wdi_gdpusdcr_o)^2)	-1.269e-02	1.400e-02	-0.906	0.370
I(I((log(wdi_gdpusdcr_o))^2)^2)	7.637e-06	1.070e-05	0.714	0.479
I(log(cepii_dist)^2)	-4.135e-03	1.213e-02	-0.341	0.735
I(ebrd_tfes_o^2)	3.897e-02	3.575e-02	1.090	0.282
I(log(cepii_area_o)^2)	1.065e-03	3.938e-03	0.270	0.788

Residual standard error: 0.871 on 43 degrees of freedom

Multiple R-squared: 0.1073, Adjusted R-squared: 0.003503

F-statistic: 1.034 on 5 and 43 DF, p-value: 0.4101

- Breusch-Pagan-Test auf Heteroskedastie mit den standardisierten FGLS-Residuen

LM-Teststatistik	p-Wert
2.5984906	0.7615946

```
lm(formula = data.frame(cbind(u_star_sq, regressor_star)))
```

Residuals:

Min	1Q	Median	3Q	Max
-0.6089	-0.3920	-0.1971	0.2204	1.9828

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.069617	0.388161	0.179	0.859
log.wdi_gdpusdcr_o.	0.035974	0.105189	0.342	0.734
I..log.wdi_gdpusdcr_o...2.	-0.002430	0.002957	-0.822	0.416
log.cepii_dist.	-0.040875	0.095084	-0.430	0.669
ebrd_tfes_o	0.224651	0.168832	1.331	0.190
log.cepii_area_o.	0.039700	0.049151	0.808	0.424

Residual standard error: 0.6394 on 43 degrees of freedom

Multiple R-squared: 0.05303, Adjusted R-squared: -0.05708

F-statistic: 0.4816 on 5 and 43 DF, p-value: 0.788

- White-Test (ohne Kreuzprodukte) auf Heteroskedastie mit den FGLS-Residuen

LM-Teststatistik	p-Wert
5.5752453	0.4724093

Call:

```
lm(formula = cbind(u_star_sq, regressor_white))
```

Residuals:

Min	1Q	Median	3Q	Max
-0.68248	-0.40380	-0.13190	0.07897	1.91210

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-2.577e-01	4.313e-01	-0.598	0.5533
w_scaled_sq	1.358e+01	8.538e+00	1.590	0.1193
log.wdi_gdpusdcr_o.	-3.678e-02	2.292e-02	-1.605	0.1160
I..log.wdi_gdpusdcr_o...2.	2.384e-05	1.550e-05	1.538	0.1315
log.cepII_dist.	-1.139e-02	7.836e-03	-1.453	0.1536
ebrd_tfes_o	7.024e-02	3.207e-02	2.191	0.0341 *
log.cepII_area_o.	1.983e-03	1.516e-03	1.308	0.1981

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

Residual standard error: 0.6259 on 42 degrees of freedom
Multiple R-squared: 0.1138, Adjusted R-squared: -0.01282
F-statistic: 0.8987 on 6 and 42 DF, p-value: 0.5049

Ergebnisse:

- Beachte, dass die Spezifikation des White-Tests ohne Kreuzprodukte den Voreinstellungen in EViews 6.0 folgt und somit (im Gegensatz zu (9.2)) keine Niveauvariablen beinhaltet.
- Sowohl der Breusch-Pagan- als auch der White-Test lehnen die Nullhypothese homoskedastischer KQ-Residuen auf jedem brauchbaren Signifikanzniveau *nicht* ab. Somit war also die Verwendung Heteroskedastie-robuster Standardfehler oder von FGLS in Abschnitt 8.4 nicht effizient.
- Breusch-Pagan- und White-Test lehnen die Nullhypothese homoskedastischer standardisierter FGLS-Fehler ebenfalls nicht ab.

- Zusammengefasst lässt sich sagen, dass unter allen betrachteten Modellen und Schätzverfahren die OLS-Schätzungen des Modells 5 am verlässlichsten erscheinen.

Zu Lesen: Kapitel 8 in [Wooldridge \(2009\)](#) (ohne Abschnitt 8.5, der sich mit linearen Wahrscheinlichkeitsmodellen beschäftigt).

9.3 Modellspezifikation II: Nützliche Tests

9.3.1 Vergleich von Modellen mit identischem Regressanden

Definition: *Nicht-geschachtelt* (non-nested) bedeutet, dass ein Modell nicht als ein Spezialfall des anderen Modells geschrieben werden kann.

Ausgangssituation: Zwei nicht-geschachtelte Modelle

$$(M1) \quad y = x_0\beta_0 + \dots + x_k\beta_k + u = \mathbf{x}'\boldsymbol{\beta} + u,$$

$$(M2) \quad y = z_0\gamma_0 + \dots + z_m\gamma_m + v = \mathbf{z}'\boldsymbol{\gamma} + v,$$

wobei $k = m$ nicht zu gelten braucht.

Entscheidung zwischen (M1) und (M2) mittels

- Informationskriterien (AIC, SC, HQ, ...),
- Encompassing-Test,
- nicht-geschachtelter F -Test,
- J -Test.

Alle drei Testverfahren lassen sich aus dem Encompassing-Prinzip begründen.

Encompassing-Prinzip

Es seien zwei nicht-geschachtelte Modelle gegeben:

$$(M1) \quad y = \mathbf{x}' \boldsymbol{\beta} + u,$$

$$(M2) \quad y = \mathbf{z}' \boldsymbol{\gamma} + v.$$

Um zu verdeutlichen, dass (M1) und (M2) nicht-geschachtelt sind, definieren wir

$$\mathbf{x}' = \begin{pmatrix} \mathbf{w}' & \mathbf{x}'_B \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_A & \boldsymbol{\beta}_B \end{pmatrix},$$

$$\mathbf{z}' = \begin{pmatrix} \mathbf{w}' & \mathbf{z}'_B \end{pmatrix}, \quad \boldsymbol{\gamma} = \begin{pmatrix} \boldsymbol{\gamma}_A & \boldsymbol{\gamma}_B \end{pmatrix},$$

sodass \mathbf{w} alle gemeinsamen Regressoren enthält

$$(M1) \quad y = \mathbf{w}' \boldsymbol{\beta}_A + \mathbf{x}'_B \boldsymbol{\beta}_B + u,$$

$$(M2) \quad y = \mathbf{w}' \boldsymbol{\gamma}_A + \mathbf{z}'_B \boldsymbol{\gamma}_B + v.$$

Idee des Encompassing-Prinzips:

- Falls (M1) korrekt spezifiziert ist, muss es möglich sein, die Ergebnisse einer Schätzung von (M2) zu erklären (und vice versa).
- Falls nicht, muss (M1) abgelehnt werden (und vice versa).

Ableitung:

Betrachten wir das “**künstlich geschachtelte Modell**”

$$(KGM) \quad y = \mathbf{w}'\mathbf{a} + \mathbf{x}'_B \mathbf{b}_x + \mathbf{z}'_B \mathbf{b}_z + \varepsilon, \quad E[\varepsilon | \mathbf{w}, \mathbf{x}_B, \mathbf{z}_B] = 0.$$

Unterschiedliche Szenarien:

- (KGM) ist das korrekt spezifizierte Modell, sodass (M1) und (M2) fehlspezifiziert sind. Es wird aber Modell (M2) geschätzt.
- Anstelle des korrekt spezifizierten Modells (M1) wird Mod. (M2) geschätzt.
- Anstelle des korrekt spezifizierten Modells (M2) wird Mod. (M1) geschätzt.

Allgemein liegt in allen Fällen ein **omitted variable bias** vor.

Details:

- (KGM) ist das korrekt spezifizierte Modell, sodass (M1) und (M2) fehlspezifiziert sind. Modell (M2) wird geschätzt. $\Rightarrow \mathbf{x}_B$ vergessen.

$$\begin{aligned}
 E[y|\mathbf{w}, \mathbf{z}_B] &= E[\mathbf{w}'\mathbf{a} + \mathbf{x}'_B \mathbf{b}_x + \mathbf{z}'_B \mathbf{b}_z + \varepsilon|\mathbf{w}, \mathbf{z}_B] \\
 &= E[\mathbf{w}'\mathbf{a}|\mathbf{w}, \mathbf{z}_B] + E[\mathbf{x}'_B \mathbf{b}_x|\mathbf{w}, \mathbf{z}_B] \\
 &\quad + E[\mathbf{z}'_B \mathbf{b}_z|\mathbf{w}, \mathbf{z}_B] + E[\varepsilon|\mathbf{w}, \mathbf{z}_B] \\
 &= \mathbf{w}'\mathbf{a} + \textcolor{red}{E[\mathbf{x}'_B|\mathbf{w}, \mathbf{z}_B]\mathbf{b}_x} + \mathbf{z}'_B \mathbf{b}_z + E[\varepsilon|\mathbf{w}, \mathbf{z}_B].
 \end{aligned}$$

Der Einfachheit halber wird angenommen, dass x_B ein Skalar und linear in \mathbf{w} und \mathbf{z}_B ist. Es gilt dann, dass

$$\begin{aligned}
 x_B &= \mathbf{w}'\mathbf{q} + \mathbf{z}'_B \mathbf{p} + \nu, \\
 E[x_B|\mathbf{w}, \mathbf{z}_B] &= \mathbf{w}'\mathbf{q} + \mathbf{z}'_B \mathbf{p}.
 \end{aligned}$$

Auch gilt, dass

$$E [E[\varepsilon | \mathbf{w}, \mathbf{x}_B, \mathbf{z}_B] | \mathbf{w}, \mathbf{z}_B] = E[\varepsilon | \mathbf{w}, \mathbf{z}_B].$$

Da (KGM) korrekt ist, gilt $E[\varepsilon | \mathbf{w}, \mathbf{x}_B, \mathbf{z}_B] = 0$ und somit

$$E[0] = 0 = E[\varepsilon | \mathbf{w}, \mathbf{z}_B].$$

Schätzt man (M2) statt (KGM), erhält man

$$\begin{aligned} E[y | \mathbf{w}, \mathbf{z}_B] &= \mathbf{w}' \mathbf{a} + [\mathbf{w}' \mathbf{q} + \mathbf{z}'_B \mathbf{p}] b_x + \mathbf{z}'_B \mathbf{b}_z \\ &= \mathbf{w}' \underbrace{[\mathbf{a} + \mathbf{q} b_x]}_{\gamma_A} + \mathbf{z}'_B \underbrace{[\mathbf{b}_z + \mathbf{p} b_x]}_{\gamma_B}. \end{aligned} \quad (9.3)$$

Beachte, dass die Verzerrungen $\mathbf{q} b_x$ und $\mathbf{p} b_x$ daraus resultieren, dass die Variable \mathbf{x}_B vergessen wurde. Diese Auswirkungen verzerren den direkten Einfluss von \mathbf{w} via \mathbf{a} und von \mathbf{z}_B via \mathbf{b}_z auf y .

- $(M1)$ ist das korrekt spezifizierte Modell. Modell $(M2)$ wird geschätzt.

Dann ist $b_z = 0$ und aus (9.3) ergibt sich folgende Restriktion:

$$\mathbf{p}b_x = \gamma_B.$$

Man sieht nun, dass die Kenntnis des korrekt spezifizierten Modells $(M1)$ genügt, um das Modell $(M2)$ abzuleiten, und somit auch um γ_B oder den Erwartungswert des KQ-Schätzers zu prognostizieren. Mit anderen Worten: Da $(M2)$ bezüglich der relevanten Variablen “kleiner” ist als $(M1)$, kann das Verhalten von $(M2)$ mithilfe von $(M1)$ prognostiziert werden, wenn für Letzteres ein unverzerrter Schätzer verwendet wird. Man sagt dann, dass **“Modell $(M1)$ umfasst Modell $(M2)$ ”**. (Die Kenntnis von $(M1)$ reicht hier nicht aus, falls (KGM) das richtige Modell sein sollte, $b_z \neq 0$.)

- $(M2)$ ist das korrekt spezifizierte Modell. Modell $(M1)$ wird geschätzt.

Kann analog zum obigen Fall abgeleitet werden.

Somit lassen sich für die Nullhypothese “ $(M1)$ umfasst $(M2)$ ” zwei äquivalente Hypothesen testen:

- $H_0 : pb - \gamma_B = 0$ - kompliziert, hier nicht detailliert dargestellt.
(Diese Version wird häufig **encompassing test** genannt und hat seine Vorteile in allgemeineren Modellen.)
- $H_0 : b_z = 0$ in (KGM) - einfach: mittels eines **nicht-geschachtelten F-Tests**.

Vorgehensweise bei mehr als zwei Alternativen:

- Das verbleibende Modell tritt nach dem selben Prinzip solange gegen weitere, alternative Modelle an, so lange es nicht abgelehnt wird.

- Problem dieses Prinzips : Es kann passieren, dass beide Nullhypothesen abgelehnt werden müssen.

Nicht-geschachtelter F -Test

Idee und Durchführung:

- Hypothesen: “ H_0 : Modell (M1) ist korrekt” versus “ H_1 : Modell (M1) ist inkorrekt”.
- Wir zerlegen wieder $\mathbf{z}' = (\mathbf{w}', \mathbf{z}'_B)$, wobei die k_A Regressoren aus \mathbf{w} in \mathbf{x} enthalten sind, nicht aber die k_B Regressoren aus \mathbf{z}_B .
- Formulierung des **künstlich geschachtelten Modells (KGM)**

$$y = \mathbf{x}'\boldsymbol{\beta} + \mathbf{z}'_B \mathbf{b}_z + \varepsilon.$$

- Auf Grundlage dieses KGM testen wir

$$H_0 : \mathbf{b}_z = \mathbf{0}$$

mittels eines F -Tests mit k_B Freiheitsgraden im Zähler und $n - m - k_B$ im Nenner.

- Für den Test (M2) vs. (M1) verfahren wir analog mit der Zerlegung $\mathbf{x}' = (\mathbf{w}', \mathbf{x}'_B) \dots$

***J*-Test (Davidson-MacKinnon-Test)**

Idee und Durchführung:

- Für den *J*-Test wird das KGM so formuliert, dass sowohl (M1) als auch (M2) im KGM geschachtelt sind:

$$y = (1 - \lambda)\mathbf{x}'\boldsymbol{\beta} + \lambda\mathbf{z}'\boldsymbol{\gamma} + \varepsilon.$$

Falls $\lambda = 0$ erhält man Modell (M1), falls $\lambda = 1$ Modell (M2).

- Problem: λ , $\boldsymbol{\beta}$ und $\boldsymbol{\gamma}$ sind im obigen Ansatz nicht identifiziert.
- Lösung: Ersetze $\boldsymbol{\gamma}$ durch den KQ-Schätzer aus (M2) $\hat{\boldsymbol{\gamma}}$.
D.h. teste $H_0 : \lambda = 0$ mittels der Testgleichung $y = \mathbf{x}'\boldsymbol{\beta}^* + \lambda\hat{y}_{M2} + \eta$, wobei $\boldsymbol{\beta}^* = (1 - \lambda)\boldsymbol{\beta}$ und $\hat{y}_{M2} = \mathbf{z}'\hat{\boldsymbol{\gamma}}$, also der gefittete Wert der KQ-Schätzung von (M2).

- Um zu testen, ob (M2) gültig ist, kann man analog vorgehen ...
- Interpretation der Logik des Tests:

Um das Modell (M1) zu testen, wird es um die gefitteten Werte des Modells (M2) erweitert; diese (d.h. der von den Regressoren in (M2) erklärte Teil von y) werden in der Testgleichung auf ihre Signifikanz hin überprüft.
- Vorteile des J -Tests gegenüber einem nicht-geschachtelten F -Test:
 - es muss nur eine Restriktion getestet werden,
 - höhere Güte, falls k_B bzw. m_B sehr groß sind,
 - im Falle von $k_B = 1$ bzw. $m_B = 1$ sind die Tests gleich.

9.3.2 Modellvergleich bei unterschiedlichen Regressanden

Idee und Durchführung (des P -Tests):

Beispiel: **lineares Modell vs. log-log Alternative**

- Schritt 1: Führe für beide Modelle jeweils eine KQ-Schätzung durch.
- Schritt 2: Berechne die zugehörigen gefitteten Werte \hat{y}_{lin} (lineares Modell) und $\widehat{\ln(y_{log})}$ (log-log Modell).
- Schritt 3a: Teste den linearen Ansatz gegen die log-log Alternative indem im KGM

$$y = \sum x_j \beta_{j,lin} + \delta_{lin} [\ln(\hat{y}_{lin}) - \widehat{\ln(y_{log})}] + u,$$

mittels eines t -Tests die Nullhypothese

$$H_0 : \delta_{lin} = 0 \quad (\text{lineares Modell ist korrekt})$$

überprüft wird.

- Schritt 3b: Teste den log-log Ansatz gegen die lineare Alternative, indem im KGM

$$\ln(y) = \sum \ln(x_j) \beta_{j,log} + \delta_{log} [\hat{y}_{lin} - \exp(\widehat{\ln y_{log}})] + v,$$

mittels eines t -Tests die Nullhypothese

$$H_0 : \delta_{log} = 0 \quad (\text{log-log Modell ist korrekt})$$

überprüft wird.

Problem: Es ist sowohl möglich, dass beide Hypothesen abgelehnt werden können (d.h. eine andere Funktionsform liegt vor) als auch beide nicht abgelehnt werden können (d.h. das Problem zu geringer Macht oder irgend ein anderes Problem).

Beachte: In diesem Fall (unterschiedliche Regressanden) ist ein Vergleich anhand der Informationskriterien **nicht möglich**.

Zu Lesen: Kapitel 9 in [Wooldridge \(2009\)](#).

10 Anhang

10.1 Eine kurze Einführung in die Wahrscheinlichkeitstheorie

Vorab: Die folgenden Seiten sind nicht zur Abschreckung gedacht, sondern als Ergänzung zu den Darstellungen, die in einführenden Lehrbüchern zur Ökonometrie enthalten sind. Es ist der Versuch, die Intuition hinter den vielen Definitionen und Konzepten in der Wahrscheinlichkeitstheorie zu erklären. Deshalb wird nicht überall auf For-

meldn verzichtet, auch wenn dadurch vielleicht nicht alles beim ersten oder zweiten Lesen klar wird. Korrekturen, Kommentare und Kritik sind willkommen!

Eine sehr ausführliche Einführung in die Wahrscheinlichkeitstheorie bietet z.B. [Casella und Berger \(2002\)](#) (in Englisch).

- **Ergebnismenge (sample space, outcome space):**

Die Menge Ω enthält alle möglichen Ergebnisse (outcomes) eines Zufallsexperiments. Die Menge kann abzählbar viele oder überabzählbar viele Ergebnisse enthalten.

Beispiele:

- Urne mit 4 farblich unterschiedlichen Kugeln: $\Omega = \{\text{gelb, rot, blau, grün}\}$
- zukünftiges Monatseinkommen eines Haushalts: $\Omega = [0, \infty)$

Anmerkungen:

- Sind die Ergebnisse endlich viele, dann bezeichnet man die einzelnen Ergebnisse häufig mit ω_i . Für S Ergebnisse, ist Ω dann

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_S\}.$$

- Liegen unendlich viele Ergebnis vor, dann bezeichnet man ein einzelnes davon häufig mit ω .

- **Ereignis (event):**

- Tritt ein bestimmtes Ergebnis ein, wird dies als Ereignis bezeichnet.
- Enthält das Ereignis genau ein Element der Ergebnismenge, wird es als Elementarereignis bezeichnet.
- Ein Ereignis ist eine Teilmenge der Ergebnismenge Ω , also jede Menge von möglichen Elementarereignissen = jede Teilmenge der Menge Ω einschließlich Ω selbst.

Beispiele:

- Urnen-Beispiel: Mögliche Ereignisse sind beispielsweise $\{\text{gelb, rot}\}$ oder $\{\text{rot, blau, grün}\}$.

- Haushaltseinkommen: Mögliche Ereignisse sind alle möglichen Teilintervalle und Verknüpfungen davon, z.B. $(0, 5000]$, $[1000, 1001)$, $(400, \infty)$, 4000 , etc.

Anmerkungen: Verwendet man die allgemeine Schreibweise mit den ω 's, dann ergibt sich

- im Fall von S Elementarereignissen: $\{\omega_1, \omega_2\}$, $\{\omega_S\}$, $\{\omega_3, \dots, \omega_S\}$, etc.
- im Fall von unendlich vielen Elementarereignissen innerhalb eines Intervalls $\Omega = (-\infty, \infty)$: $(a_1, b_1]$, $[a_2, b_2)$, $(0, \infty)$, etc., wobei immer die untere Grenze kleiner oder gleich der oberen Grenze ist, also $(a_i \leq b_i)$.

- **Zufallsvariable:**

Eine Zufallsvariable ist eine Funktion, die jedem Elementarereignis $\omega \in \Omega$ eine reelle Zahl $X(\omega)$ zuordnet.

Urneneispiel: $X(\omega_1) = 0, X(\omega_2) = 3, X(\omega_3) = 17, X(\omega_4) = 20$.

- **Dichtefunktion**

– Vorbemerkung: Wie wir schon gesehen haben, wird es immer kompliziert, wenn unendlich viele Elementarereignisse vorliegen.

Betrachten wir beispielsweise $\Omega = [0, 4]$. Möchte man nun die Wahrscheinlichkeit berechnen, dass beispielsweise die Zahl π eintritt, dann ist diese Wahrscheinlichkeit 0. Wäre sie nicht 0, dann hätten wir das Problem, dass die Summe der Wahrscheinlichkeiten über alle (überabzählbar viele) Zahlen nicht 1 sein könnte.

Was tun?

- Einen Ausweg bietet folgender Trick: Betrachten wir die Wahrscheinlichkeit, dass die Realisation der Zufallsvariablen X in dem Intervall $[0, x]$ liegt, wobei $x < 4$. Diese Wahrscheinlichkeit lässt sich schreiben als $P(X \leq x)$. Nun kann man fragen, inwieweit sich diese Wahrscheinlichkeit verändert, wenn man das Intervall $[0, x]$ um ein kleines Stück h verlängert. Die Antwort lautet: $P(X \leq x + h) - P(X \leq x)$. Setzt man diese Veränderung in der Wahrscheinlichkeit in Bezug auf die Veränderung der Intervalllänge, erhält man

$$\frac{P(X \leq x + h) - P(X \leq x)}{h}.$$

Lässt man nun die Intervalllänge h gegen Null gehen, bildet also den Grenzwert, erhält man:

$$\lim_{h \rightarrow 0} \frac{P(X \leq x + h) - P(X \leq x)}{h} = f(x).$$

Dieser Grenzwert heißt **Wahrscheinlichkeitsdichtefunktion** oder kurz **Dichtefunktion**, die zu der Wahrscheinlichkeitsfunktion P gehört.

- Wie lässt sich die Dichtefunktion interpretieren?

Schreibt man etwas lässig

$$\frac{P(X \leq x + h) - P(X \leq x)}{h} \approx f(x)$$

und formt dies um zu

$$P(X \leq x + h) - P(X \leq x) \approx f(x)h,$$

dann sieht man, dass die Dichtefunktion $f(x)$ die Rate angibt, mit der sich die Wahrscheinlichkeit verändert, wenn das Intervall $[0, x]$ um h verlängert wird. Die Dichtefunktion gibt also eine **Rate** an.

- Da die Dichtefunktion eine Ableitung ist, gilt umgekehrt in unserem Beispiel, dass

$$\int_0^x f(u)du = P(X \leq x) = F(x).$$

Dabei wird $F(x) = P(X \leq x)$ als **Wahrscheinlichkeitsverteilungsfunktion** bezeichnet. Man erhält natürlich in unserem Beispiel auch, dass

$$\int_0^4 f(u)du = P(X \leq 4) = 1.$$

Allgemein gilt, dass das Integral der Dichtefunktion über den gesamten Bereich der Zufallsvariable 1 ergibt, beispielsweise bei $X(\omega) \in \mathbb{R}$:

$$\int_{-\infty}^{\infty} f(u)du = P(X \leq \infty) = 1.$$

• **Bedingte Wahrscheinlichkeitsfunktion**

Zunächst ein **Beispiel**:

Es bezeichne die Zufallsvariable $X \in [0, \infty)$ den Auszahlungsbetrag in einem Gewinnspiel. Die Wahrscheinlichkeitsfunktion bzw. Verteilungsfunktion $P(X \leq x) = F(x)$ gibt die Wahrscheinlichkeit für einen maximalen Gewinnbetrag von x an. Es ist weiter bekannt, dass zur Ermittlung des Auszahlungsbetrags 2 Maschinen bereitstehen, Maschine A und Maschine B .

Frage: Wie groß ist die Wahrscheinlichkeit für einen maximalen Gewinnbetrag von x falls Maschine A zum Einsatz kommt?

Anders formuliert, wie groß ist die gesuchte Wahrscheinlichkeit, wenn die Bedingung “Maschine A im Einsatz” gilt? Die gesuchte Wahrscheinlichkeit wird als **bedingte Wahrscheinlichkeit** be-

zeichnet und als

$$P(X \leq x|A)$$

notiert. Entsprechend notiert man, falls die Bedingung “Maschine B im Einsatz” gilt, $P(X \leq x|B)$.

Frage: Welcher Zusammenhang besteht nun zwischen der **unbedingten Wahrscheinlichkeit** $P(X \leq x)$ und den beiden **bedingten Wahrscheinlichkeiten** $P(X \leq x|A)$ und $P(X \leq x|B)$?

Zur Beantwortung muss man wissen, mit welcher Wahrscheinlichkeit Maschine A , bzw. Maschine B zum Einsatz kommt. Wenn wir diese Wahrscheinlichkeiten mit $P(A)$ und $P(B)$ bezeichnen, dann können wir die obige Frage beantworten:

$$\begin{aligned} P(X \leq x) &= P(X \leq x|A)P(A) + P(X \leq x|B)P(B), \\ F(x) &= F(x|A)P(A) + F(x|B)P(B). \end{aligned}$$

In unserem Beispiel haben wir genau zwei Elementarereignisse. Der hierfür genannte Zusammenhang lässt sich auf n diskrete Elementarereignisse $\Omega = \{A_1, A_2, \dots, A_n\}$ erweitern:

$$F(x) = F(x|A_1)P(A_1) + F(x|A_2)P(A_2) + \dots + F(x|A_n)P(A_n). \quad (10.1)$$

Bisher haben wir die Bedingung in Form von Ereignissen und nicht in Form von Zufallsvariablen definiert. Ein Beispiel für Letzteres wäre, wenn zur Ermittlung des Auszahlungsbetrags nur eine Maschine zur Verfügung steht, deren Funktionsweise aber von dem vorherigen Auszahlungsbetrag Z abhängt. Dann lautet die bedingte Verteilungsfunktion $F(x|Z = z)$, wobei $Z = z$ bedeutet, dass die Bedingung lautet, dass Zufallsvariable Z genau die Realisation z annimmt. Um wieder den Zusammenhang zwischen der unbedingten und den bedingten Wahrscheinlichkeiten zu erhalten, müssen wir

nun die Summe durch ein Integral ersetzen und die Wahrscheinlichkeit der Bedingung durch die entsprechende Dichtefunktion, da Z ja unendlich viele Werte annehmen kann. Für unser Beispiel ergibt sich dann:

$$F(x) = \int_0^\infty F(x|Z=z)f(z)dz = \int_0^\infty F(x|z)f(z)dz,$$

bzw. allgemein

$$F(x) = \int F(x|Z=z)f(z)dz = \int F(x|z)f(z)dz. \quad (10.2)$$

Noch eine wichtige Eigenschaft:

Wenn die Zufallszahlen X und Z stochastisch unabhängig sind, dann gilt

$$F(x|z) = F(x).$$

• Bedingte Dichtefunktion

Die bedingte Dichtefunktion kann man heuristisch aus der bedingten Verteilungsfunktion in derselben Weise ableiten, wie wir das weiter oben für die unbedingte Dichtefunktion getan haben; es werden lediglich die unbedingten Wahrscheinlichkeiten durch bedingte Wahrscheinlichkeiten ersetzt. Die **bedingte Dichtefunktion** ergibt sich aus

$$\lim_{h \rightarrow 0} \frac{P(X \leq x + h | A) - P(X \leq x | A)}{h} = f(x | A).$$

Falls man endlich viele Bedingungen hat, dann wird (10.1) zu

$$f(x) = f(x | A_1)P(A_1) + f(x | A_2)P(A_2) + \dots + f(x | A_n)P(A_n).$$

Der Zusammenhang (10.2) lautet dann

$$f(x) = \int f(x | Z = z) f(z) dz = \int f(x | z) f(z) dz. \quad (10.3)$$

• **Erwartungswert**

Betrachten wir wieder unser Beispiel der Auszahlungsmaschinen.

Frage: Welchen Auszahlungsbetrag würden Sie “im Mittel” oder “im Durchschnitt” erwarten?

Antwort: $\int_0^\infty xf(x)dx$. Würde die Gewinnauszahlung in n verschiedenen diskreten Beträgen erfolgen, so würde man “im Mittel” erwarten: $\sum_{i=1}^n x_i P(X = x_i)$. Jede mögliche Auszahlung wird mit ihrer Eintrittswahrscheinlichkeit gewichtet aufsummiert. Nicht überraschend bezeichnet man diese Größen auch als Erwartungswert.

Allgemein ist der **Erwartungswert** definiert als

$$E[X] = \int xf(x)dx, \quad X \text{ stetig,}$$

$$E[X] = \sum x_i P(X = x_i), \quad X \text{ diskret.}$$

- **Regeln für den Erwartungswert** z.B. Appendix B in [Wooldridge \(2009\)](#).

1. Für jede Konstante c gilt

$$E[c] = c.$$

2. Für alle Konstanten a und b und Zufallsvariablen X und Y gilt

$$E[aX + bY] = aE[X] + bE[Y].$$

3. Sind die Zufallsvariablen X und Y unabhängig, gilt

$$E[XY] = E[Y]E[X].$$

- **Bedingter Erwartungswert**

Bisher haben wir nicht darauf geachtet, welche Maschine bei der Auszahlungsermittlung zum Einsatz kommt. Interessieren wir uns

hingegen für die erwartete Auszahlung, wenn Maschine A im Einsatz ist, dann müssen wir den **bedingten** Erwartungswert berechnen

$$E[X|A] = \int_0^\infty xf(x|A)dx.$$

Dies geschieht einfach, indem man die unbedingte Dichte $f(x)$ durch die bedingte Dichte $f(x|A)$ ersetzt und die Bedingung in der Notation des Erwartungswertes angibt. Entsprechend lässt sich die erwartete Auszahlung für Maschine B berechnen als

$$E[X|B] = \int_0^\infty xf(x|B)dx.$$

Allgemein erhält man für diskrete Bedingungen

$$E[X|A] = \int xf(x|A)dx, \quad X \text{ stetig,}$$

$$E[X|A] = \sum x_i P(X = x_i|A), \quad X \text{ diskret,}$$

bzw. für stetige Bedingungen

$$E[X|Z = z] = \int x f(x|Z = z) dx, \quad X \text{ stetig},$$

$$E[X|Z = z] = \sum x_i P(X = x_i|Z = z), \quad X \text{ diskret}.$$

Beachte: Häufig verwendet man auch die Kurzformen, so auch in [Wooldridge \(2009\)](#).

$$E[X|z] = \int x f(x|z) dx, \quad X \text{ stetig},$$

$$E[X|z] = \sum x_i P(X = x_i|z), \quad X \text{ diskret}.$$

Entsprechend dem Zusammenhang zwischen unbedingten und bedingten Wahrscheinlichkeiten, existiert ein ähnlicher Zusammenhang auch zwischen dem unbedingten und den bedingten Erwartungswerten. Er lautet

$$E[X] = E [E[X|Z]]$$

und wird als **Law of iterated expectations (LIE)** bezeichnet.

Beweisskizze:

$$\begin{aligned}
 E[X] &= \int x f(x) dx \\
 &= \int x \left[\int f(x|z) f(z) dz \right] dx \quad (\text{Einsetzen von (10.3)}) \\
 &= \int \int x f(x|z) f(z) dz dx \\
 &= \int \underbrace{\int x f(x|z) dx}_{E[X|z]} f(z) dz \quad (\text{Vertauschen von } dx \text{ und } dz) \\
 &= \int E[X|z] f(z) dz \\
 &= E [E[X|Z]].
 \end{aligned}$$

In unserem Beispiel mit den 2 Maschinen ergibt das Gesetz der

iterierten Erwartungen

$$E[X] = E[X|A]P(A) + E[X|B]P(B).$$

Dieses Beispiel macht auch deutlich, dass die bedingten Erwartungswerte $E(X|A)$ und $E(X|B)$ Zufallszahlen sind, die gewichtet mit ihren Eintrittswahrscheinlichkeiten $P(A)$ und $P(B)$ den Erwartungswert $E(X)$ ergeben. Man stelle sich vor, man kennt vor Beginn des Spiels nur die beiden bedingten Erwartungswerte, aber nicht welche Maschine zum Einsatz kommen wird. Dann ist der erwartete Auszahlungsbetrag gerade $E(X)$ und wir müssen die beiden bedingten Erwartungswerte als Zufallsvariablen ansehen. Sobald man weiß, welche Maschine zum Einsatz gekommen ist, ist der dazugehörige bedingte Erwartungswert die Realisation der Zufallsvariablen. Diese Eigenschaft gilt ganz allgemein für bedingte Erwartungswerte.

- **Regeln für bedingte Erwartungen**

z.B. Appendix B in [Wooldridge \(2009\)](#).

1. Für jede Funktion $c(\cdot)$ gilt

$$E[c(X)|X] = c(X).$$

2. Für alle Funktionen $a(\cdot)$ und $b(\cdot)$ gilt

$$E[a(X)Y + b(X)|X] = a(X)E[Y|X] + b(X).$$

3. Sind die Zufallsvariablen X und Y unabhängig, gilt

$$E[Y|X] = E[Y].$$

4. **Law of iterated expectations (LIE)**

$$E[E[Y|X]] = E[Y].$$

5. $E[Y|X] = E[E[Y|X, Z]|X].$

6. Falls $E[Y|X] = E[Y]$, dann $Cov(X, Y) = 0$.

7. Falls $E[Y^2] < \infty$ und $E[g(X)^2] < \infty$ für eine beliebige Funktion $g(\cdot)$, dann gelten

$$E\{[Y - E[Y|X]]^2|X\} \leq E\{[Y - g(X)]^2|X\}$$

$$E\{[Y - E[Y|X]]^2\} \leq E\{[Y - g(X)]^2\}.$$

10.2 Wichtige Regeln der Matrix-Algebra

Matrixaddition

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1K} \\ a_{21} & a_{22} & \dots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{T1} & a_{T2} & \dots & a_{TK} \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1K} \\ c_{21} & c_{22} & \dots & c_{2K} \\ \vdots & \vdots & & \vdots \\ c_{T1} & c_{T2} & \dots & c_{TK} \end{pmatrix}.$$

Haben \mathbf{A} und \mathbf{C} dieselbe Dimension, so gilt

$$\mathbf{A} + \mathbf{C} = \begin{pmatrix} a_{11} + c_{11} & a_{12} + c_{12} & \dots & a_{1K} + c_{1K} \\ a_{21} + c_{21} & a_{22} + c_{22} & \dots & a_{2K} + c_{2K} \\ \vdots & \vdots & & \vdots \\ a_{T1} + c_{T1} & a_{T2} + c_{T2} & \dots & a_{TK} + c_{TK} \end{pmatrix}.$$

Matrixmultiplikation

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{T1} & a_{T2} & \cdots & a_{TK} \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1L} \\ b_{21} & b_{22} & \cdots & b_{2L} \\ \vdots & \vdots & & \vdots \\ b_{K1} & b_{K2} & \cdots & b_{KL} \end{pmatrix}.$$

Das Produkt $\mathbf{C} = \mathbf{AB}$ ist dann definiert, wenn die Zahl der Spalten in \mathbf{A} mit der Anzahl der Zeilen in \mathbf{B} übereinstimmt. Für jedes Element in \mathbf{C} gilt dann die Gleichung

$$c_{ij} = \begin{pmatrix} a_{i1} & \cdots & a_{iK} \end{pmatrix} \begin{pmatrix} b_{1j} \\ \vdots \\ b_{Kj} \end{pmatrix} = a_{i1}b_{1j} + \cdots + a_{iK}b_{Kj} = \sum_{l=1}^K a_{il}b_{lj}.$$

Beachte: Im Allgemeinen gilt, dass $\mathbf{AB} \neq \mathbf{BA}$ ist.

Transponieren einer Matrix

Es sei die (2×3) -Matrix (d.h. 2 Zeilen, 3 Spalten)

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}$$

gegeben. Die transponierte Matrix von \mathbf{A} ist dann die (3×2) -Matrix

$$\mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix}.$$

Es gilt, dass

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}'.$$

Inverse einer Matrix

Sei \mathbf{A} die $(K \times K)$ -Matrix

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1K} \\ a_{21} & a_{22} & \cdots & a_{2K} \\ \vdots & \vdots & & \vdots \\ a_{K1} & a_{K2} & \cdots & a_{KK} \end{pmatrix}.$$

Die Inverse von \mathbf{A} , mit \mathbf{A}^{-1} bezeichnet, ist definiert durch

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}_K \equiv \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Die Einheitsmatrix der Dimension $(K \times K)$ wird mit \mathbf{I}_K bezeichnet.

Die Matrix \mathbf{A} ist dann invertierbar, wenn die Zeilen bzw. Spalten linear unabhängig sind. Mit anderen Worten: Keine Zeile (Spalte) kann als Linearkombination der anderen Zeilen (Spalten) geschrieben werden. Technisch ist dies gleichbedeutend mit der Tatsache, dass die Determinante von \mathbf{A} ungleich Null ist.

Eine nichtinvertierbare Matrix wird häufig auch als singulär bezeichnet.

Die Berechnung von Inversen sollte man lieber dem Computer überlassen. Lediglich für Matrizen mit 2 oder 3 Spalten/Zeilen, ist die Berechnung von maßvoller Komplexität und kann mit geringem Aufwand mit Stift und Papier durchgeführt werden.

Invertieren einer (2×2) -Matrix:Für eine quadratische (2×2) -Matrix

$$\mathbf{B} = \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix}$$

führt die Berechnung der **Determinante** zu

$$\det(\mathbf{B}) = b_{11}b_{22} - b_{21}b_{12}$$

und die **Inverse** ist somit

$$\begin{aligned} \mathbf{B}^{-1} &= \frac{1}{\det(\mathbf{B})} \begin{pmatrix} b_{22} & -b_{12} \\ -b_{21} & b_{11} \end{pmatrix} \\ &= \frac{1}{b_{11}b_{22} - b_{21}b_{12}} \begin{pmatrix} b_{22} & -b_{12} \\ -b_{21} & b_{11} \end{pmatrix}. \end{aligned}$$

Beispiel:

$$\mathbf{C} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \end{pmatrix}, \quad \text{mit} \quad \det(\mathbf{C}) = 0 \cdot (-1) - 1 \cdot 2 = -2$$

$$\mathbf{C}^{-1} = \frac{1}{-2} \begin{pmatrix} -1 & -2 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{pmatrix}$$

Gegenprobe:

$$\mathbf{C}\mathbf{C}^{-1} = \begin{pmatrix} 0 & 2 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & 1 \\ \frac{1}{2} & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Zu Lesen: Weiteres zur Matrixalgebra und ihrer Anwendung im Rahmen des multiplen linearen Regressionsmodells finden Sie in den Appendices D, E.1 in [Wooldridge \(2009\)](#).

10.3 Regeln zur Ableitung von Vektoren und Matrizen

-

$$c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_T \end{pmatrix}, \quad w = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \end{pmatrix},$$

$$z = c'w = \begin{pmatrix} c_1 & c_2 & \cdots & c_T \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \end{pmatrix},$$

$$\frac{\partial z}{\partial w} = c.$$

•

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1T} \\ a_{21} & a_{22} & \cdots & a_{2T} \\ \cdots & \cdots & \cdots & \cdots \\ a_{T1} & a_{T2} & \cdots & a_{TT} \end{pmatrix},$$

$$z = w' A w = \begin{pmatrix} w_1 & w_2 & \cdots & w_T \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1T} \\ a_{21} & a_{22} & \cdots & a_{2T} \\ \cdots & \cdots & \cdots & \cdots \\ a_{T1} & a_{T2} & \cdots & a_{TT} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_T \end{pmatrix}.$$

$$\frac{\partial z}{\partial w} = (A' + A)w.$$

10.4 Daten für die Schätzung der Gravitationsgleichung

Legende für die Daten in importe_ger_2004_ebrd.txt

• Länder und Ländercodes

1	ALB	Albanien	17	GBR	Vereiniges Königreich	33	NLD	Niederlande	Anmerkungen: Tabelle ba- siert auf Table 1 in gravity_data.pdf
2	ARM	Armenien	18	GEO	Georgien	34	NOR	Norwegen	
3	AUT	Österreich	19	GER	Deutschland	35	POL	Polen	
4	AZE	Aserbaidschan	20	GRC	Griechenland	36	PRT	Portugal	
5	BEL	Belgien und Luxemburg	21	HRV	Kroatien	37	ROM	Rumänien	
6	BGR	Bulgarien	22	HUN	Ungarn	38	RUS	Russland	
7	BLR	Weißrussland	23	IRL	Irland	39	SVK	Slowakei	
8	CAN	Kanada	24	ISL	Island	40	SVN	Slowenien	
9	CHE	Schweiz	25	ITA	Italien	41	SWE	Schweden	
10	CYP	Zypern	26	KAZ	Kasachstan	42	TKM	Turkmenistan	
11	CZE	Tschechische Republik	27	KGZ	Kirgisistan	43	TUR	Türkei	
12	DNK	Dänemark	28	LTU	Litauen	44	UKR	Ukraine	
13	ESP	Spanien	29	LVA	Lettland	45	USA	Vereinigte Staaten	
14	EST	Estland	30	MDA	Republik Moldau	46	YUG	Serbien und Montenegro	
15	FIN	Finnland	31	MKD	EYR Mazedonien				
16	FRA	Frankreich	32	MLT	Malta				

Länder, die nur als Herkunftsländer auftauchen:

BIH Bosnien und Herzegowina

TJK Tadschikistan

UZB Usbekistan

CHN China

HKG Hong Kong

JPN Japan

KOR Südkorea

TWN Taiwan

THA Thailand

- **Endogene Variable:**

- TRADE_0_D_0:

- Importe des Landes D aus Land O (d.h. Ausfuhren von Land O nach Land D) in laufenden US-Dollars.

- Produktklassen: Die Handelsströme basieren auf der Aggregation von Handelsströmen, die nach der Standard International Trade Classification, Revision 3 (SITC, Rev.3) auf dem niedrigsten Aggregationsniveau (4- oder 5-stellig) erfasst wurden. Quelle: **UN COMTRADE**

- Nicht enthalten sind Treib- und Schmierstoffe (d.h., insbesondere Kraftstoff- und Erdgasprodukte). Mindestgrenze der zugrunde liegenden aufgespaltenen Handelsströme (auf dem SITC Rev.3 5-stelligen Niveau) beträgt 500 US-Dollar.

- **Erklärende Variablen:**

Herkunftsland (O-Land)

WDI_GDPUSDCR_O	Herkunftsland BIP-Daten; in laufenden US-Dollars	Weltbank - World Development Indicator
WDI_GDPPCUSDCR_O	Herkunftsland BIP-pro Kopf-Daten; in laufenden US-Dollars	Weltbank - World Development Indicator
WEO_GDPCR_O	Ziel- und Herkunftsland BIP-Daten; in laufenden US-Dollars	IWF - World Economic Outlook database
WEO_GDPPCCR_O	Ziel- und Herkunftsland BIP-pro-Kopf-Daten; in laufenden US-Dollars	IWF - World Economic Outlook database
WEO_POP_O	Herkunftsland Bevölkerungsdaten	IWF - World Economic Outlook database
CEPII_AREA_O	Fläche des Herkunftslandes in km ²	CEPII
CEPII_COL45	dummy; d- und o-Land besaßen eine Kolonialbeziehung nach 1945	CEPII
CEPII_COL45_REV	dummy; revidiert durch "Expertenwissen"	CEPII
CEPII_COLONY	dummy; d- und o-Land hatten jemals eine Kolonialbeziehung	CEPII
CEPII_COMCOL	dummy; d- und o-Land teilen einen gemeinsamen Kolonialherren nach 1945	CEPII
CEPII_COMCOL_REV	dummy; revidiert durch "Expertenwissen"	CEPII
CEPII_COMLANG_ETHNO	dummy; d- und o-Land teilen eine gemeinsame Sprache	CEPII
CEPII_COMLANG_ETHNO_REV	wird von mindestens 9% der Bevölkerung gesprochen	CEPII
CEPII_COMLANG_OFF	dummy; d- und o-Land teilen gemeinsame Amtssprache	CEPII
CEPII_CONTIG	dummy; d- und o-Land sind Nachbarstaaten	CEPII
CEPII_DISINT_O	inländische Entfernung in Herkunftsland	CEPII
CEPII_DIST	geodätische Entfernung zwischen d- und o-Land	CEPII
CEPII_DISTCAP	Entfernung zwischen d- und o-Land basierend auf deren Hauptstädten $0.67\sqrt{Fläche/\pi}$	CEPII
CEPII_DISTW	gewichtete Entfernungen, für Details siehe CEPII	CEPII
CEPII_DISTWCES	gewichtete Entfernungen, für Details siehe CEPII	CEPII
CEPII_LAT_O	Breitengrad der Stadt	CEPII
CEPII_LON_O	Längengrad der Stadt	CEPII
CEPII_SMCTRY_REV	dummy; d- und o-Land waren/sind dasselbe Land	CEPII, revidiert
ISO_O	ISO-Code mit drei Buchstaben für Herkunftsland	CEPII
EBRD_TFES_O	EBRD-Maß für den Grad der Liberalisierung der Handels- und Zahlungsströme des o-Landes	EBRD

Zielland (D-Land)		
WDI_GDPUSDCR_D	Zielland BIP-Daten; in laufenden US-Dollars	Weltbank - World Development Indicators
WDI_GDPPCUSDCR_D	Zielland BIP-pro Kopf-Daten; in laufenden US-Dollars	Weltbank - World Development Indicators
WEO_GDPCR_D	Ziel- und Herkunftsland BIP-Daten; in laufenden US-Dollars	IWF - World Economic Outlook database
WEO_GDPPCCR_D	Ziel- und Herkunftsland BIP-pro-Kopf-Daten; in laufenden US-Dollars	IWF - World Economic Outlook database
WEO_POP_D	Zielland Bevölkerungsdaten	IWF - World Economic Outlook database

Anmerkungen: Das EBRD misst Reformbemühungen auf einer Skala von 1 bis 4+ (=4.33); 1 steht für keinen oder nur geringfügigen Fortschritt; 2 zeigt wichtigen Fortschritt an; 3 steht für substantiellen Fortschritt; 4 zeigt umfangreichen Fortschritt an, während 4+ bedeutet, dass das Land die Standard- und die Leistungsnormen fortgeschrittenen Industriestaaten erreicht hat, d.h., von OECD Staaten. Diese Variable ist per Konstruktion qualitativer und nicht kardinaler Art.

- **Dank:** an Richard Frensch, IOS - Leibniz-Institut für Süd- und Südosteuropaforschung, Regensburg und Universität Regensburg, der die Daten zur Verfügung gestellt hat.

- EViews-Befehle, um ausgewählte Daten aus dem Haupt-Workfile zu extrahieren:
 - um Beobachtungen für Exporte nach Deutschland auszuwählen:
im workfile: Proc → Copy/Extract from Current Page
→ By Value to New Page or Workfile:
in Sample - observations to copy: @all if (iso_d="GER"). Objects to copy: auswählen. Page Destination: auswählen.
 - um Beobachtungen für eine Periode auszuwählen, z.B. 2004:
wie oben, nur diesmal: in Sample - observations to copy: 2004 2004
 - um Beobachtungen auszuwählen, die Handelsströme zwischen den USA und Deutschland für alle Perioden anzeigen:
wie oben, nur diesmal: in Sample - observations to copy: @all if (iso_o="USA") and (iso_d="GER")
- **Internetseiten** **CEPII**

10.5 R-Programm für die empirischen Beispiele

```
#####
# EOE_ws19_Emp_Beispiele.R #####
#
#####
# R-Programm zum Reproduzieren der empirischen Beispiele in den
# Folien Einführung in die Ökonometrie, Universität Regensburg
# erstellt von Patrick Kratzer, Roland Weigand und Rolf Tschernig
# Stand: 18.10.2019, 25.08.2020
#
#####
# Hinweise:
# a) Um das Skript ausführen zu können, werden folgende Daten benötigt:
#     - Handelsströme-Beispiele "importe_ger_2004_ebrd.txt",
#     - Löhne-Beispiele "wage1.txt"
#     - Zigaretten-Beispiele "smoke.txt"
# b) Die Daten-Dateien müssen im gleichen Verzeichnis wie das Programm liegen
#     und das working directory muss dem Verzeichnis entsprechen, von dem aus
#     dieses R-Programm aufgerufen wurde. Dazu muss das working directory
#     definiert werden, siehe Hinweise ab Zeile 82.
# c) Zunächst werden die Funktionen stats und SelectCritEviews definiert.
#     Anschließend beginnt das Hauptprogramm in Zeile 75.
# d) Graphiken können als PDF-Datei ausgegeben werden, siehe Hinweis Zeile 81.
#
# Beginn Definition Funktionen
#####
```

```
#####
# Funktion stats #####
# Nützliche Funktion, die bei Eingabe eines Vektors statistische Kennzahlen liefert
# analog zu EViews-Output von "Descriptive Statistics"
#
stats <- function(x) {
  n           <- length(x)
  sigma       <- sd(x) * sqrt((n-1)/n)
  skewness    <- 1/n * sum(((x-mean(x))/sigma)^3)
  kurtosis    <- 1/n * sum(((x-mean(x))/sigma)^4)
  jarquebera <- n/6*((skewness)^2 + 1/4 * ((kurtosis-3))^2)
  pvalue      <- 1- pchisq(jarquebera, df = 2)

  Statistics <- c(mean(x), median(x), max(x), min(x), sd(x),
                  skewness, kurtosis, jarquebera, pvalue)

  names(Statistics) <- c("Mean", "Median", "Maximum", "Minimum", "Std. Dev.",
                         "Skewness", "Kurtosis", "Jarque Bera", "Probability")

  return(data.frame(Statistics))
}
#####
# Ende #####
#####
# Funktion SelectCritEviews #####
# Funktion zur Berechnung von Modellselektionskriterien wie in EViews
# RT, 2011_01_26
SelectCritEviews <- function(model)
```

```

{
  n          <- length(model$residuals)
  k          <- length(model$coefficients)
  fitmeasure <- -2*logLik(model)/n

  aic        <- fitmeasure + k * 2/n
  hq         <- fitmeasure + k * 2*log(log(n))/n
  sc         <- fitmeasure + k * log(n)/n
  sellist    <- list(aic=aic[1],hq=hq[1],sc=sc[1])
  return(t(sellist))
}

#####
# Ende Definition Funktionen
#####

#####
# Beginn Hauptprogramm
#####

#####
# Bestimme Parameter für das R-Programm #####
# 1=Erstelle PDFs von Graphiken, 0=sonst
# Working Directory, in dem die R-Datei und die
# Daten liegen
# MUSS INDIVIDUELL ANGEPASST WERDEN
# In RStudio über "Session" -> "Set Working Directory"
# -> "To Source File Location" zu bestimmen

```

```
# Beispiele: WD = "~/EOE/R-code" oder
# WD = "C:/users/r-code"

##### Ende Parameter Eingabe #####
# Folgende Libraries werden im Verlauf geladen: car,lmtest

# Falls diese nicht installiert sind, werden diese zunächst installiert:
if (!require(car)){
  install.packages("car")
}
if (!require(lmtest)){
  install.packages("lmtest") # benötigt ab Folie 194
}
if (!require(xtable)){
  install.packages("xtable") # benötigt ab Folie 290
}

# Festlegung des Arbeitsverzeichnisses (working directory)
# in welchem sich das R-Programm und die Daten befinden
setwd(WD) # setze es als Working Directory

##### Einlesen der Handelsströme-Daten als data frame
daten_all <-read.table("importe_ger_2004_ebrd.txt", header = TRUE)
# Zuweisung der Variablennamen und
# Eliminieren der Beobachtung Exportland: GER, Importland: GER
attach(daten_all[-20,])

# Zum Ausprobieren, falls importe_ger_2004_ebrd.txt schon eingelesen worden ist
stats(trade_0_d_o)
```

```
##### Einlesen der wage-Daten als data frame
attach(read.table("wage1.txt", header = TRUE))

#####
#####
#####
##### Histogram, Folie 6 #####
#####

# Für Ausgabe im PDF Format Dateiname definieren
if (save.pdf) pdf("r_imports_barplot.pdf", 12, 6)
# Histogramm
barplot(trade_0_d_o*10^-9, names.arg = iso_o, las = 2, col = "lightblue",
         main = "Imports to Germany in 2004 in Billions of US-Dollars")
# Device schließen
if (save.pdf) dev.off()

#####
#####
#####
##### Scatterplot, Folien 8, 11, 60 #####
#####

# Für Ausgabe im PDF Format Dateiname definieren
if (save.pdf) pdf("scatter.pdf", height=6, width=6)
# Scatterplot der beiden Variablen
plot(wdi_gdpusdcr_o, trade_0_d_o, col = "blue", pch = 16)
# Device schließen
if (save.pdf) dev.off()

#####
```

```
#####
# Scatterplot mit (linearer) Regressionsgerade,
# Folien 12, 61

# Für Ausgabe im PDF Format Dateiname definieren
if (save.pdf)  pdf("plot_wdi_vs_trade.pdf", height=4, width=4)
# KQ-Schätzung eines einfachen linearen Regressionsmodells, abgespeichert in ols
ols_trade_wdi  <- lm(trade_0_d_o ~ wdi_gdpusdcr_o)
# Scatterplot der beiden Variablen
plot(wdi_gdpusdcr_o, trade_0_d_o, col = "blue", pch = 16)
# Einzeichnen der linearen Regressionsgeraden mittels abline
abline(ols_trade_wdi, col = "red")
# Hinzufügen einer Legende
legend("bottomright", "Lineare Regression", col = "red", lty = 1, bty = "n")
# Device schließen
if (save.pdf) dev.off()

#####
#####

# Scatterplot mit linearer Regressionsgerade
# und nichtlinearer Regressionsgerade
# in Punkten an Beobachtungen dargestellt, Folie 13

if (save.pdf) pdf("r_imports_scatter_nonlin.pdf", 4, 4)
  # Schätzung Regressionsmodell mit quadratischem Regressor
ols_nonlin <- lm((trade_0_d_o) ~ wdi_gdpusdcr_o + I(wdi_gdpusdcr_o^2)) # quadr.
  # Definiere quadratische Funktion mit geschätzten Parametern
fx <- function(x){ols_nonlin$coefficients[1] +
```

```
ols_nonlin$coefficients[2]*x + ols_nonlin$coefficients[3]*x^2}
# Erstelle Scatterplot
plot(wdi_gdpusdcr_o, trade_0_d_o, col = "blue", pch = 16)
  # Füge lineare Regressionsgerade dazu
abline(ols_trade_wdi, col = "red")
  # Füge Prognosepunkte der quadratischen Regression dazu
lines(wdi_gdpusdcr_o, fx(wdi_gdpusdcr_o),
      col = "green", type="p", pch = 16)
  # Erstelle Legende
legend("bottomright",
       c("linear regression", "nonlinear regression"),
       col = c("red", "green"), lty = c(1,2), bty = "n")
if (save.pdf) dev.off()

#####
#####

# Handelsstrom von USA -> Deutschland, Folie 27
# aus anderer Datei %RT1920

#####
#####

# Regressionsoutput Handelsbeispiel Folie 61
# siehe auch Folie 12

# Anzeige der Ergebnisse der einfachen linearen Regression
summary(ols_trade_wdi)

#####
```

```
#####
# Regressionsoutput Folie 65
summary(lm(wage ~ educ))

#####
# Regressionsoutput Außenhandelsbeispiel Folie 88
summary(lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o)))

#####
# Regressionsoutput Außenhandelsbeispiel Folie 103
summary(lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist)))

#####
# Regressionsoutput Lohnbeispiel Folie 115
summary(lm(log(wage) ~ educ))
```

```
# Regressionsoutput Lohnbeispiel Folie 117

summary(lm(log(wage) ~ educ + exper))

#####
#####

# Bestimmung der Informationskriterien, Folie 177
# Anwendung der Funktion "SelectCritEviews" auf vier
# verschiedene Modelle:

model_1 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o))
coef(model_1)
SelectCritEviews(model_1)

model_2 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist))
coef(model_2)
SelectCritEviews(model_2)

model_3 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
               ebrd_tfes_o)
coef(model_3)
SelectCritEviews(model_3)

model_4 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
               ebrd_tfes_o + log(cepii_area_o))
coef(model_4)
SelectCritEviews(model_4)

#####
```

```
#####
# t-Statistik in Eviews, Folie 195, 205

model_wage_m <- lm(wage ~ 1)
summary(model_wage_m)
# t-Statistik für H_0: mu=6
# mit gerundeten Werten
(5.896 - 6) / 0.161

# mit exakten Werten aus KQ-Schätzung
(coef(summary(model_wage_m))[1] - 6) / coef(summary(model_wage_m))[2]

# mit package car
sqrt(linearHypothesis(model_wage_m, c("(Intercept)=6"))$F[2])

# für Seite 205
# t-Statistik für H_0: mu=5.6
# mit gerundeten Werten
# mit gerundeten Werten
(5.896 - 5.6) / 0.161
# mit exakten Werten
(coef(summary(model_wage_m))[1] - 5.6) / coef(summary(model_wage_m))[2]

#####
# Histogram von "wage", Folie 197, 283

if (save.pdf) pdf("r_wage_hist.pdf", 4, 4)
```

```
hist(wage, breaks = 20, col = "lightblue", prob = T)
curve(dnorm(x, mean = mean(wage), sd = sd(wage)),
      from = -5, to = 25, add = T, col = "red", lty = 2, lwd = 2)
legend("topright", "theoretical\nnormal distribution", col = "red",
      lwd = 2, lty = 2, bty = "n")
box()

if (save.pdf) dev.off()

# Ausgabe der deskriptiven Statistiken und Test auf Normalverteilung
stats(wage)

#####
#####

# Gravitationsgleichung, Folie 227

summary(lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) + ebrd_tfes_o))

#####
#####

# Visualisierung der Residuen, Folie 228

model_3 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
                ebrd_tfes_o)
resid_model_3 <- model_3$resid
trade_0_d_o_fit <- model_3$fitted
```

```

if (save.pdf) pdf("r_resid_model_3.pdf", 5, 3)

par(mfrow = c(1,2))
plot(trade_0_d_o_fit, resid_model_3, col = "blue", pch = 16, main = "Scatterplot")
hist(resid_model_3, breaks = 20, col = "lightblue", prob = T, main = "Histogram")
curve(dnorm(x, mean = mean(resid_model_3), sd = sd(resid_model_3)),
      from = -3, to = 3, add = T, col = "red", lty = 2, lwd = 2)
legend("topleft", "theoretical\nnormal distribution", col = "red", lwd = 2,
       lty = 2, bty = "n")
box()
if (save.pdf) dev.off()

# statistische Auswertung der Residuen
stats(resid_model_3)

#####
#####

# Outputzeile auf Folie 230

summary(lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) + ebrd_tfes_o))
# Befehlszeile für log(wdi_gdpusdcr_o) reinkopieren

#           Estimate Std. Error t value Pr(>|t|)
# log(wdi_gdpusdcr_o)  0.94066   0.06134 15.335 < 2e-16 ***
# t-Statistik für gerundete Werte im Output
(teststat <- (0.94066 - 1)/0.06134)

#####
#####

```

```
#####
# Befehl für Quantil der t-Verteilung auf Folie 230
(crit <- qt(0.975, df = 49 - 3 -1))

#####
# Outputzeilen auf Folie 232, 233
(pval <- 2 * pt(teststat, df = 49 - 3 - 1))      # (Hälfte von 9.262691e-08)
(summary(model_3)$coef[3,])

# t-Statistik (basierend auf Output)
(teststat2 <- (-9.703183e-01 - 0) / 1.526847e-01)

# kritischer Wert
(crit <- qt(0.95, df = 49 - 3 -1))

# p-Wert
(pval <- pt(teststat2, df = 49 - 3 - 1))

#####
# Befehle für Folien 245, 246
(crit <- qt(1-0.05/2, df = 49 - 3 -1))
```

```

summary(lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) + ebrd_tfes_o))

#           Estimate Std. Error t value Pr(>|t|)
# log(wdi_gdpusdcr_o)  0.94066   0.06134 15.335 < 2e-16 ***
# Konfidenzintervall
(0.94066 - 2.014103* 0.06134)
(0.94066 + 2.014103* 0.06134)
#####
#####
# Regressionsoutput auf Folie 250

marketing_102      <-read.table("marketing_102.txt", header = TRUE)

#summary(lm(labsatz ~ log(preis) + log(preis_qualig) +
#           log(preis_qualimo), data=marketing_102))
S      <- marketing_102$absatz
P      <- marketing_102$preis
P_K1   <- marketing_102$preis_qualig
P_K2   <- marketing_102$preis_qualimo
summary(lm(log(S) ~ log(P) + log(P_K1) + log(P_K2)))

#####
#####

# Regressionsoutput auf Folie 252
# verwendet Daten von Folie 250
summary( lm( log(S) ~ log(P) + log(P_K1) + I(log(P_K1)+log(P_K2)) ) )

```

```
#####
#####
#
# Regressionsoutput zu Folie 254, Fortsetzung des Außenhandelsbeispiels
#
model_4 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o + log(cepii_area_o))
summary(model_4)
#
#####
#####
#
# Regressionsoutput auf Folie 257
#
model_2 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist))
summary(model_2)
#
#####
#####
#
# F-Test auf Folie 264
#
model_2_sum <- summary(model_2)
(SSR_model_2 <- (model_2_sum$sigma)^2 * model_2_sum$df[2])
#
model_4_sum <- summary(model_4)
(SSR_model_4 <- (model_4_sum$sigma)^2 * model_4_sum$df[2])
#
# F-Statistik
( (SSR_model_2 - SSR_model_4)/2 ) /
```

```

(SSR_model_4/model_4_sum$df[2])
#####
#####
#####

# F-Test auf Folie 266/267

library(car)

1 - pf(5.24077, df1 = 2, df2 = 44)

model_4 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
               ebrd_tfes_o + log(cepii_area_o))
linearHypothesis(model_4, c("ebrd_tfes_o = 0", "log(cepii_area_o) = 0"))

#####
#####
#####

# Hinweis zu Folie 269, Ermittlung der Kovarianzmatrix

model_4 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
               ebrd_tfes_o + log(cepii_area_o))
vcov(model_4)
coef(summary(model_4))[,2]^2

#####
#####
#####

# Konfidenzellipse auf Folie 272

if (save.pdf) pdf("r_conf_ellipse.pdf", 6, 6)

```

```

model_4 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o + log(cepii_area_o))

confidenceEllipse(model_4, which.coef = c(4, 5), levels = 0.95,
  main = "confidence ellipse", col = "blue")
abline(v = confint(model_4, "ebrd_tfes_o", level = 0.95), lty = 2,
  col = "red", lwd = 2)
abline(h = confint(model_4, "log(cepii_area_o)", level = 0.95), lty = 2,
  col = "red", lwd = 2)

if (save.pdf) dev.off()

#####
#####

# Regressionsoutput auf Folie 277, 278

model_4 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o + log(cepii_area_o))
model_4_h0 <- lm(log(trade_0_d_o)-0.5*ebrd_tfes_o ~ log(wdi_gdpusdcr_o) +
  log(cepii_dist))

summary(model_4_h0)

# F-Statistik auf Basis der Outputs
(SSR_model_4 <- (model_4_sum$sigma)^2 * model_4_sum$df[2])
(SSR_model_4_h0 <- (summary(model_4_h0)$sigma)^2 * summary(model_4_h0)$df[2])
( (SSR_model_4_h0 - SSR_model_4)/2 ) /
  (SSR_model_4/model_4_sum$df[2])

```

```
# F-Statistik mit library(car)
linearHypothesis(model_4, c("ebrd_tfes_o = 0.5", "log(cepii_area_o) = 0"))

#####
#####
#####

# Folie 281: siehe Folie 177

#####
#####
#####

# Folie 284, Dichte der Chi-Quadrat(1)-Verteilung
if (save.pdf) pdf("r_chi_2_1_verteilung.pdf", 6, 3)

curve(dchisq(x, df = 1), from = 0, to = 8, col = 2, ylab = "f(x)", ylim = c(0, 1.5),
       main = expression(paste(chi^2, "(1) - density function")))
abline(v=0)
if (save.pdf) dev.off()

# Folie 284, Dichte und Verteilung verschiedener Chi-Quadrat-Verteilungen
# (nicht in Folien)

if (save.pdf) pdf("r_chi_2_verteilung.pdf", 6, 3)

par(mfrow = c(1, 2))

curve(dchisq(x, df = 1), from = 0, to = 8, col = 1, ylab = "f(x)", ylim = c(0, 0.5),
       main = expression(paste(chi^2, " - density function")))
lines(c(-1, 0), c(0, 0), col = 1)
```

```

grid()
curve(dchisq(x, df = 2), from = 0, to = 8, col = 2, add = T)
curve(dchisq(x, df = 3), from = 0, to = 8, col = 3, add = T)
curve(dchisq(x, df = 5), from = 0, to = 8, col = 4, add = T)
curve(dchisq(x, df = 10), from = 0, to = 8, col = 5, add = T)
legend("topright", c("df = 1", "df = 2", "df = 3", "df = 5", "df = 10"),
       col = 1:5, lty = 1, bty = "n")
curve(pchisq(x, df = 1), from = 0, to = 8, ylab = "F(x)", col = 1, ylim = c(0, 1),
      main = expression(paste(chi^2, " - distribution function")))
lines(c(-1, 0), c(0, 0), col = 1)
grid()
curve(pchisq(x, df = 2), from = 0, to = 8, col = 2, add = T)
curve(pchisq(x, df = 3), from = 0, to = 8, col = 3, add = T)
curve(pchisq(x, df = 5), from = 0, to = 8, col = 4, add = T)
curve(pchisq(x, df = 10), from = 0, to = 8, col = 5, add = T)
# legend

if (save.pdf) dev.off()
#####
#####

# Monte Carlo Simulation auf Folien 290, 291, 292
if (save.pdf) pdf("r_mcarlo.pdf", 6, 4)
par(mfrow = c(2, 3))
set.seed(12345)                      # setze Random seed (für Replizierbarkeit)

reps  <- 1000                         # Anzahl der Replikationen
n      <- c(10, 30, 50, 100, 500, 1000) # Stichprobenumfang für die 6 Auswertungen

means <- matrix(NA, nrow = reps, ncol = 6) # Initialisierung der

```

```

# Matrix mit den simulierten Mittelwerten

for(j in 1:6)
{
  for(i in 1:reps)
  {
    means[i,j] <- mean(3 + (rnorm(n[j])^2-1)*2^-0.5) # Simulation der Mittelwerte
  }
  hist(means[,j], breaks = 30, freq = F, xlab = "",           # graphische Ausgabe
        col = "lightblue", main = paste("n = ",n[j]))        # der Schätzrealisationen
}
if (save.pdf) dev.off()

# Erstelle Tabelle mit Mittelwerten und Standardabweichungen
fx <- function(x) {c(mean(x), sd(x))}
table_output <- apply(means, 2, fx)
# füge Zeile mit wahren Standardabweichungen des Schätzers hinzu
table_output <- rbind(table_output,sqrt(1/n))
# gebe Spalten und Zeilen Namen
rownames(table_output) <- c("Mittelwerte", "Standardabweichungen",
                           "theoret. Standardabw. geg. DGP")
colnames(table_output) <- paste0("n = ",n)
# erstelle Latex-Code für Tabelle
xtable(t(table_output), digits=6)
# Lösche Matrix means aus Simulation
rm(means)

#####
#####
#####
```

```

# Koeffizienten von Modell 3, Folie 313

model_3 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) + ebrd_tfes_o)
coef(model_3)

#####
#####

# Modell 5 auf Folie 318
model_5 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + I(log(wdi_gdpusdcr_o)^2)
               + log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))
summary(model_5)

#####
#####

# Programm zu Folie 320
if (save.pdf) pdf("r_bib_elasticity.pdf", 3, 3)

# Modell 5:
model_5 <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + I(log(wdi_gdpusdcr_o)^2)
               + log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))

# Generiere die Elastizitäten für verschiedene BIPs
elast_gdp <- model_5$coef[2] + 2* model_5$coef[3]*log(wdi_gdpusdcr_o)

# Erstelle Scatterplot
plot(wdi_gdpusdcr_o, elast_gdp, pch = 16, col = "blue", main = "GDP-Elasticity")

if (save.pdf) dev.off()

```

```
#####
#####
#####
# Regressionsoutput Folie 324, wage-Beispiel
ols <- lm(log(wage) ~ female + educ + exper + I(exper^2) + tenure + I(tenure^2))
summary(ols)
#####
#####
#####
# Regressionsoutput Folie 330, wage-Beispiel
femmarr <- female * married
malesing <- (1 - female) * (1 - married)
malemarr <- (1 - female) * married

ols <- lm(log(wage) ~ femmarr + malesing + malemarr + educ + exper + I(exper^2) + tenure + I(tenure^2))
summary(ols)
#####
#####
#####
# Fortsetzung des Lohnbeispiels, Folie 337
ols <- lm(log(wage) ~ female + educ + exper + I(exper^2) + tenure + I(tenure^2) + I(female*educ))
summary(ols)
#####
```

```
#####
# Fortsetzung Außenhandelsbeispiel, Folie 372 ff.

# R-Programm zur FGLS-Schätzung, Kapitel 8 Heteroskedastie
# Florian Brezina, PK, 19.02.2011
# mit Datei importe_ger_2004_ebrd.txt

# Daten einlesen und 20. Beobachtung (Germany) entfernen
# daten <- read.table("importe_ger_2004_ebrd.txt", header = TRUE)[-20,]
# attach(daten)

# definiere Variablen

# definiere log der abhängigen Variable
log_imp <- log(trade_0_d_o)

### Erster Schritt a) KQ-Regression und Berechnung der Residuen

# KQ-Regression
eq_ols_model5 <- lm(log_imp ~ log(wdi_gdpusdcr_o) + I((log(wdi_gdpusdcr_o))^2) +
                      log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))

# Berechne Residuen
res_ols_model5 <- eq_ols_model5$resid

# Berechne gefittete/angepasste Werte
fit_ols_model5 <- fitted.values(eq_ols_model5)

# Plotte die Residuen gegen die gefitteten Werte, um zu untersuchen,
```

```

# ob Heteroskedastie vorliegen könnte
dev.off()
plot(fit_ols_model5, res_ols_model5, pch = 16)

### Erster Schritt b) bis d)

# Quadriere die Residuen und logarithmiere sie anschließend
ln_u_hat_sq <- log(res_ols_model5^2)

# Schätze die Varianzgleichung
eq_h_model5 <- lm(ln_u_hat_sq ~ log(wdi_gdpusdcr_o) + I((log(wdi_gdpusdcr_o))^2) +
                    log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))

# Berechne die gefitteten Werte der logarithmierten Residuenanalyse
ln_u_hat_sq_hat <- fitted.values(eq_h_model5)

# Berechne die h's aus den gefitteten Werten der Varianzregression
h_hat <- exp(ln_u_hat_sq_hat)

### Zweiter Schritt: FGLS-Schätzung

# Schätze FGLS mit den gewichteten weights = 1/h_hat
eq_fgls_model5 <- lm(log_imp ~ log(wdi_gdpusdcr_o) + I((log(wdi_gdpusdcr_o))^2) +
                        log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o),
                        weights = 1/h_hat)
summary(eq_fgls_model5)

# Berechne die gefitteten Werte aus FGLS
fit_fgls_model5 <- fitted.values(eq_fgls_model5)

```

```
# Berechne die Residuen aus FGLS
res_fgls_model5 <- resid(eq_fgls_model5)

# Standardisierung der Residuen mittels der Gewichte
res_fgls_model5_star <- res_fgls_model5*h_hat^(-1/2)

# Plotte die Residuen gegen die gefitteten Werte
plot(fit_fgls_model5, res_fgls_model5_star, pch = 16)

### KQ-Regression mit heteroskedastie-robusten Standardfehlern
library(lmtest)
eq_white_model5 <- coeftest(eq_ols_model5, vcov=hccm(eq_ols_model5, type="hc1"))

# Graphiken/Outputs für Skript
summary(eq_ols_model5)
summary(eq_h_model5)
summary(eq_fgls_model5)
eq_white_model5

if (save.pdf) pdf("r_model_5_fgls.pdf", 6, 3)
par(mfrow = c(1,2))
plot(fit_ols_model5, res_ols_model5, col = "blue", pch = 16, main = "OLS")
plot(fit_fgls_model5, res_fgls_model5_star, col = "blue", pch = 16, main = "FGLS")
if (save.pdf) dev.off()

##### Einschub #####
# ein paar Anmerkungen:
# R^2 und F-Statistik bei R-Output entsprechen den Ergebnissen
# für weighted statistics im Eviews-Output
```

```
# Nachbau des Eviews-Outputs:  
w      <- h_hat^-0.5  
w_scaled <- length(residuals(eq_fgls_model5)) / sum(w) * w  
sum(w_scaled) # Probe  
  
log_imp_star <- log_imp * sqrt(w_scaled) # Wurzel!?  
regressor_star <- model.matrix(eq_fgls_model5) * sqrt(w_scaled)  
  
k <- ncol(model.matrix(eq_fgls_model5))-1  
n <- length(resid(eq_fgls_model5))  
  
# Weighted Statistics  
  
# R-squared  
summary(eq_fgls_model5)$r.squared  
# Adjusted R-squared  
summary(eq_fgls_model5)$adj.r.squared  
# SSR  
(SSR <- sum(w_scaled*(log_imp_star - regressor_star%*%coef(eq_fgls_model5))^2))  
# Mean dependent var  
mean(log_imp * (w_scaled))  
# S.D. dependent var  
sd(log_imp * (w_scaled))  
# S.E. of regression  
sqrt(SSR/(n-k-1))  
  
# Unweighted Statistics  
  
# R-squared
```

```
(r_squared <- 1 - sum(residuals(eq_fgls_model5)^2) /  
  sum((log_imp - mean(log_imp))^2))  
# Adjusted R-squared  
-k/(n-k-1) + (n-1)/(n-k-1)*r_squared  
# Mean dependent var  
mean(log_imp)  
# S.D. dependent var  
sd(log_imp)  
# S.E. of regression  
sqrt(sum(residuals(eq_fgls_model5)^2)/(n-k-1))  
# Sum squared resid  
sum(residuals(eq_fgls_model5)^2)  
  
##### Ende Einschub #####  
  
#####  
#####  
  
# Zigarettenbeispiel ab Folie 385  
  
smoke_all <- read.table("smoke.txt", header = TRUE)  
  
# Erster Schritt  
  
# 1. KQ-Schätzung  
ols_1 <- lm(cigs ~ lincome + lcigpric + educ + age + I(age^2) + restaurn,  
            data=smoke_all)  
summary(ols_1)
```

```
# 2. Speichere die Residuen
u_hat_cig <- resid(ols_1)

# 3. Logarithmiere die quadrierten Residuen
ln_u_sq <- log(u_hat_cig^2)

# 4. Schätzung der Varianzregression mittels KQ führt zu
ols_2 <- lm(ln_u_sq ~ lincome + lcigpric + educ + age + I(age^2) + restaurn,
             data=smoke_all)
summary(ols_2)

# Speichere die Residuen
h_hat_cig <- exp(ln_u_sq - resid(ols_2))
#
# Zweiter Schritt

# Gewichtete KQ-Schätzung mit den Gewichten h_hat_cig^(-1)
ols_3 <- lm(cigs ~ lincome + lcigpric + educ + age + I(age^2) + restaurn,
             weights = h_hat_cig^(-1), data=smoke_all)
summary(ols_3)

# Anmerkung: Im Vergleich zu EViews fehlen einige Statistiken. Siehe Hinweise
#           zu Folie 372 zu deren Berechnung

#####
#####

# Fortsetzung des Zigarettenbeispiels auf Folie 396

ols <- lm(cigs ~ lincome + lcigpric + educ + age + I(age^2) + restaurn,
           data=smoke_all)
```

```

u_hat_sq <- resid(ols)^2
summary(lm(u_hat_sq ~ lincome + lcigpric + educ + age + I(age^2) + restaurn,
           data=smoke_all))

#####
#####
#####

# Fortsetzung Zigarettenbeispiel mit White Test, Folie 401 ff.

# Definition von Funktion für White Test
##### Beginn Funktion whitetest #####
# Function to conduct White test including and without cross terms
# Specification of test equations as in EViews
# Roland Weigand, 2011_01_26, Rolf Tschernig, 2019_10_18, 2020_08_25 (LM test)
# Input:
#       model_est    lm object with estimated model
#       crossterms  1: include cross terms, 0, do not include them
# Output: a list with the following components
#       ftest_result a vector containing the F statistic, the
#                     degrees of freedom and the p value
#       lmtest_result a vector containing the LM statistic,
#                     the degrees of freedom and the p value
#       test_eq      an lm object with the results of the White regression

whitetest <- function(model_est, crossterms=1){

  # Daten aus model extrahieren
  dat <- model_est$model                      # dat is dataframe
  dat$resid_sq <- model_est$resid^2            # resid_sq is added to dataframe
}

```

```

# Formel für die Hilfsregression erstellen
regr <- attr(model_est$terms, "term.labels")
if (crossterms){
  form <- as.formula(paste("resid_sq ~ (", paste(regr, collapse=" + "), ")^2 +"
                        , paste("I(",regr,"^2)", collapse=" + ") ) )
} else {
  form <- as.formula(paste("resid_sq ~ ", paste("I(",regr,"^2)",
                                                collapse=" + ") ) )
}

# Hilfsregression schätzen
test_eq <- lm(form, data=dat)

# Overall F-Test
fstat <- summary(test_eq)$fstatistic
# LM statistic
lmstat <- length(summary(test_eq)$residuals) * summary(test_eq)$r.squared

# Ergebnis berechnen und ausgeben
fstat <- c(fstat[1], fstat[2], fstat[3],
           pf(fstat[1], fstat[2], fstat[3], lower.tail = FALSE))
names(fstat) <- c("F Statistic", "df1", "df2", "p Value")
lmstat <- c(lmstat, summary(test_eq)$df[1] - 1,
            pchisq(lmstat, summary(test_eq)$df[1] - 1, lower.tail = FALSE))
names(lmstat) <- c("LM Statistic", "df", "p Value")
result <- list(lmtest_result = lmtest_result, fstat_result = fstat,
               test_eq = test_eq)
return(result)
}

#####
##### Ende Funktion whitetest #####

```

```
# Anwendung der Funktion

ols <- lm(cigs ~ lincome + lcigpric + educ + age + I(age^2) + restaurn,
           data=smoke_all)
ols_white <- whitetest(ols)
# gebe F-Testergebnis aus
ols_white$ftest_result
# gebe LM-Testergebnis aus
ols_white$lmtest_result
# gebe Testgleichung aus
summary(ols_white$test_eq)

#####
#####

# BP-Test auf Folie 403, Fortsetzung des Außenhandelsbeispiels

bptest(eq_ols_model5)

#####
#####

# White-Test auf Folie 404, 405 (ohne Kreuzprodukte)
# führe White-Test durch, Funktion whitetest() auf Folie 399 definiert
ols_model5_white <- whitetest(eq_ols_model5, crossterms=0)
# gebe F-Testergebnis aus
ols_model5_white$ftest_result
# gebe LM-Testergebnis aus
ols_model5_white$lmtest_result
```

```
# gebe Testgleichung aus
summary(ols_model5_white$test_eq)

#####
#####

# Folie 406

# Breusch Pagan Test für FGLS (funktioniert mit "bptest" leider nicht
# Ergebnisse entsprechen denen von Eviews
log_imp_star  <- log_imp * (w_scaled)
regressor_star <- model.matrix(eq_fgls_model5)[,-1] * (w_scaled)
u_star_sq <- (resid(eq_fgls_model5) * (w_scaled))^2

bpg_eq_fgls <- lm(data.frame(cbind(u_star_sq, regressor_star)))

t_bpg_fgls <- summary(bpg_eq_fgls)$r.squared * n
bp_fgls_res <- c(t_bpg_fgls,
                  1-pchisq(t_bpg_fgls, df = k))
names(bp_fgls_res) <- c("LM-Teststatistik", "p-Wert")
bp_fgls_res
summary(bpg_eq_fgls)

#####
#####

# Folie 407 und 408

# White-Test manuell, erfordert Variablen definiert für Folie 404, 405
```

```
w_scaled_sq <- w_scaled^2
regressor_white <- data.frame(w_scaled_sq, regressor_star^2)

white_eq_fgls <- lm(cbind(u_star_sq , regressor_white))

t_white_fgls <- summary(white_eq_fgls)$r.squared * n
white_fgls_res <- c(t_white_fgls,
                      1-pchisq(t_white_fgls, df = k+1))
names(white_fgls_res) <- c("LM-Teststatistik", "p-Wert")
white_fgls_res
summary(white_eq_fgls)

#####
#####
#####
#           ENDE
#####
#####
#####
```

Listing 10.1:/R_code/EOE_ws19_Emp_Beispiele.R

Literaturverzeichnis

Anderson, J. E., und E. v. Wincoop (2003), “Gravity with Gravitas: A Solution to the Border Puzzle,” *The American Economic Review*, 93, 170–192. [102](#)

Angrist, J. D., und J.-S. Pischke (2015), *Mastering Metrics. The Path from Cause to Effect*, Princeton University Press, Princeton. [24](#)

Bauer, T. K., M. Fertig, und C. M. Schmidt (2009), *Empirische Wirtschaftsforschung*, Springer. [24](#), [73](#)

Casella, G., und R. L. Berger (2002), *Statistical Inference*, 2nd edn., Duxbury - Thomson. **11**

Davidson, R., und J. G. MacKinnon (2004), *Econometric Theory and Methods*, Oxford University Press, Oxford. **259**

Fratianni, M. (2007), “The gravity equation in international trade,” Tech. rep., Dipartimento di Economia, Universita Politecnica delle Marche. **102, 225**

Pindyck, R. S., und D. L. Rubinfeld (1998), *Econometric models and economic forecasts*, Irwin McGraw-Hill. **19**

Stock, J. H., und M. W. Watson (2007), *Introduction to Econometrics*, Pearson, Boston, Mass. **15, 20, 22, 23**

Wooldridge, J. M. (2009), *Introductory Econometrics. A Modern Approach*, 4th edn., Thomson South-Western. **20, 30, 34, 58, 64, 77,**

99, 105, 115, 121, 130, 140, 148, 152, 160, 194, 198, 206, 207, 208,
212, 223, 231, 252, 262, 263, 281, 284, 288, 299, 302, 310, 323, 339,
347, 366, 385, 409, 423, XVI, XVIII, XXI, XXIX