



UNIVERSITÄT REGENSBURG

Methoden der Ökonometrie

Handout

Prof. Dr. Rolf Tschernig

Wintersemester 2021/2022

Version vom 21. Oktober 2021

Hinweis: Dieses Handout ist aus einem umfangreichen Foliensatz hervorgegangen. Ein substantieller Teil der Folien ist in Zusammenarbeit mit Harry Haupt, Universität Passau entstanden. Ich danke Kathrin Kagerer und Stefan Rameseder für ihre hervorragenden Zuarbeiten, wichtigen Korrekturen und substantiellen Verbesserungsvorschläge sehr herzlich. Ebenso danke ich Christoph Rust für seine wichtigen Anmerkungen und Florian Greindl für Korrekturen. Ich danke auch Andreas Kelbel sehr herzlich für die Übertragung des Handouts in das Beamer-Paket. Ebenso danke ich Dominik Ammon für weitere Korrekturen. Ich bitte 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, Methoden der Ökonometrie - Handout, Universität Regensburg, Oktober 2021. Downloaded am [Tag Monat Jahr].

Organisation

Hinweis

Dieses Handout entspricht bis auf Korrekturen von organisatorischen Hinweisen, Links und Tippfehlern und wenigen inhaltlichen Verbesserungen dem Handout vom Wintersemester 2018/2019, Version vom 29. November 2018, das wiederum bis auf Korrekturen und inhaltlichen Verbesserungen, insbesondere in den Abschnitten 5.2 und 5.3 dem Handout vom Wintersemester 2015/2016, Version vom 20. Oktober 2015, entspricht.

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@ur.de

Dominik Ammon, M.Sc. in VWL

Gebäude RW(L), 5. Stock, Raum 516
Universitätsstr. 31, 93040 Regensburg
Tel. (+49) 941/943 2738, Fax (+49) 941/943 4917
Email: dominik.ammon@ur.de

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

Zeitplan und Umfang

Zeitplan für WS 2021/2022

- Woche 1 und 2: **Mathematik-Vorkurs Teil 3**
(Teil des Pflichtmoduls **Methoden der Ökonometrie**)
- Woche 3 bis 15: Pflichtmodul **Methoden der Ökonometrie**
- zu semesterbegleitenden Leistungen siehe Kurshomepage <https://www.uni-regensburg.de/wirtschaftswissenschaften/vwl-tschernig/lehre/master/methoden-der-oekonometrie/index.html>

Format und ECTS

- **4 h Vorlesung** und 2 h Übung
- **10 ECTS**: entspricht ca. 250 h bis 300 h Zeitaufwand für das gesamte Modul

Organisation

Inhalte, Termine und Räume, Downloads, Aktuelles

<https://www.uni-regensburg.de/wirtschaftswissenschaften/vwl-tschernig/lehre/master/methoden-der-oekonometrie/index.html>

Voraussetzung für die Kursteilnahme

- Kenntnis der Inhalte des Mathematik-Vorkurses Teil 1 und 2.
- Hilfreich, aber nicht erforderlich: Kenntnisse eines einführenden Ökonometrikurses, z. B. des Bachelor-Moduls *Einführung in die Ökonometrie*.

Ziele dieses Kurses

(Grund-)Kenntnisse zur Beantwortung folgender Fragen

- **Wie** mache ich eine sorgfältige empirische/ökonometrische Analyse?
- Welche ökonometrischen Methoden gibt es?
- Wie kann ich die Qualität einer empirischen Analyse beurteilen?
- **Warum** und unter welchen Annahmen funktioniert eine ökonometrische Methode?

- Wie kann ich empirische Analysen mit der freien Software R durchführen?

Nutzen

im Studium

- Grundlagen für Masterstudium, insb. für Schwerpunktmodulgruppe *Empirische Wirtschaftsforschung*.
- Basis zum Verstehen weiterführender ökonomischer Lehrbücher.
- Empirische Analysen in anderen Kursen verstehen.
- Empirische Analysen in der Master- oder einer Seminararbeit selber durchführen können (Cassar, Engl, Gürtzgen, Jerger, Kindermann, Knoppik, Lee, Roider, Tschernig, Weber).

im Beruf

- Datenanalysen immer wichtiger (**Big Data, Open Data**)!
- Programmierkenntnisse in vielen beruflichen Tätigkeiten hilfreich.

Notenzusammensetzung und Klausur

Notenzusammensetzung

- Studienbegleitende Leistungen (SBLs) (25%):
Art der SBLs siehe Kurshomepage
- Abschlussklausur (75%)

Abschlussklausur

- Termin: Klausurenperiode
- Dauer: 90 Minuten
- enthält Aufgaben zu Teil 3 des Mathematikvorkurses

Beachte

Um das Modul zu bestehen, reicht die Note 4.0 in der Abschlussklausur **nicht** aus, wenn man in den SBLs eine Gesamtnote schlechter als 4.0 hat.

Software

Im Kurs: Verwendung der Software R.

- Vorteile von R:
 - sehr flexible mathematisch-statistische Programmiersprache.
 - freie Software: <http://www.r-project.org/>.
 - wird in der Wissenschaft und in der Wirtschaft verwendet.
 - schnell wachsende Bibliothek an R-Paketen für verschiedenste Aufgaben.
 - jeder kann selbst Pakete programmieren und der Allgemeinheit zur Verfügung stellen.
 - weite Verbreitung gemäß TIOBE Programming Index: <http://www.tiobe.com/index.php/content/paperinfo/tpci/index.html>.
- Intensive Verwendung von R im Modul. Alle R-Programme im Appendix A des Handouts.
- Verwendung von R:
 - in den Masterkursen **Fortgeschrittene Ökonometrie**, **Applied Financial Econometrics**, **Quantitative Wirtschaftsforschung II**, etc.
 - in Masterarbeiten am Lehrstuhl
- R-Kurs:
 - **Programmieren mit R** (LS Tschernig, zwei Wochen vor Vorlesungsbeginn im WS oder als Screencast)

Alternative Software für ökonometrische Analysen — Übersicht

Graphische Benutzeroberfläche

- **EViews** (EViews-Kurse (Christoph Knoppik), programmierbar, im CIP-Pool vorhanden, Einzellizenz über Lehrstuhl für ca. Euro 80, MA-Veranstaltung: Quantitative Wirtschaftsforschung II)
- **Gretl** (programmierbar, freie Software: <http://gretl.sourceforge.net/>)
- **Stata** (Stata-Kurs (siehe LSF), im CIP-Pool vorhanden)
- **JMulTi** (freie Software: <http://www.jmulti.de/>, MA-Veranstaltung: Quantitative Wirtschaftsforschung II)

(Statistische) Programmiersprachen mit fertigen Programmmodulen

- **R**, siehe oben.
- **Gauss** (einige Lizenzen vorhanden)
- **Ox** (Batch-Version frei)
- **Matlab** (MA-Veranstaltung: Dynamic Macroeconomics)
- **Python** (Data Science und Machine Learning)
- **Fortran, C, C++** (Allgemeine Programmiersprachen mit umfangreichen Numerikbibliotheken)

Computer-Algebra-Sprachen

- **Maple** (UR-Lizenz)
- **Maxima** (freie Software)
- **Mathematica**

Pflichtliteratur

Davidson, R. & MacKinnon, J.G. (2004). *Econometric Theory and Methods*, Oxford University Press (<http://econ.queensu.ca/ETM/>)

Literatur für Mathematikvorkurs für lineare Algebra

- Schmidt, K. & Trenkler, G. (2015). *Einführung in die Moderne Matrix-Algebra. Mit Anwendungen in der Statistik*, 3. Auflage, Springer. [Kompaktes leicht lesbares deutsches Lehrbuch mit sehr vielen Beispielen zum Rechnen mit Matrizen](#) (aus dem Uninetz [Volltext](#) verfügbar)
- Gentle, J.E. (2007) *Matrix Algebra Theory, Computations, and Applications in Statistics*, Springer. [Chapter 2 interessant für Ökonometriker: detaillierte Einführung in Vektorräume](#) (aus dem Uninetz [Volltext](#) verfügbar)
- Fischer, G. (2014) *Lineare Algebra*, 18. Auflage, Vieweg & Teubner. [Abschnitt 1.4 grundlegende Einführung für Mathematiker, Physiker, Ingenieure, usw.](#)(aus dem Uninetz [Volltext](#) verfügbar)
- Lütkepohl, H. (1996) *Handbook of Matrices*, John Wiley & Sons, Chichester. [Hervorragendes Nachschlagewerk zur linearen Algebra und deren verschiedenen Matrizen und dazugehörigen Rechenregeln und Umformungsmöglichkeiten.](#)

Literatur für Mathematikvorkurs zur Wahrscheinlichkeitstheorie

- Casella, G. & Berger, R.L. (2002). *Statistical Inference*, Duxbury - Thomson. [sehr ausführliche, formale Einführung in die Wahrscheinlichkeitstheorie.](#)
- Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I.& Tutz, G. (2016). *Statistik. Der Weg zur Datenanalyse*, 8. Auflage, Springer. [einfache Einführung in die Statistik](#) (aus dem Uninetz [Volltext](#) verfügbar)
- Steland, A. (2016). *Basiswissen Statistik: Kompaktkurs für Anwender aus Wirtschaft, Informatik und Technik*, 4. Auflage, Spinger. [gut geschriebene, kurz gehaltene, technisch präzise Einführung in die Statistik](#) (aus dem Uninetz [Volltext](#) verfügbar)

Literatur zum Wiederholen und Ergänzen

- Kleiber, C. & Zeileis, A. (2008). *Applied Econometrics with R* Springer, Springer. [sehr gute Einführung in R](#) (aus dem Uninetz [Volltext](#) verfügbar)
- Steland, A. (2013). *Basiswissen Statistik: Kompaktkurs für Anwender aus Wirtschaft, Informatik und Technik*, 3. Auflage, Springer. (aus dem Uninetz [hier](#))
- Stock, J.H. & Watson, M.W. (2012). *Introduction to Econometrics*, 3. ed., Person, Addison-Wesley. <https://scholar.harvard.edu/stock/pages/introduction-econometrics>
- Wooldridge, J.M. (2013). *Introductory Econometrics. A Modern Approach*, 5. Ed., Thomson South-Western.

Vertiefende Literatur (in alphabetischer Reihenfolge)

- Angrist, J. & Pischke, J. (2009). *Mostly Harmless Econometrics. An Empiricist's Companion*, Princeton University Press.
(Gut lesbare Einführung in die empirische Evaluationsliteratur)
<http://press.princeton.edu/titles/8769.html>
- Cameron, A.C. and Trivedi, P.K. (2005). *Microeconometrics*, Cambridge University Press.
(Methodik für mikroökonomische Probleme)
<http://cameron.econ.ucdavis.edu/mmabook/mma.html>
- Davidson, R. & MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. Oxford University Press.
Viele Details zur Methodik für nichtlineare Regressionsmodelle,
<http://qed.econ.queensu.ca/dm-book/>
- Greene, W. (2012). *Econometric Analysis*. 7e, Prentice Hall.
Umfassendes Nachschlagewerk mit moderater methodischer Tiefe,
<http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm>
- Hayashi, F. (2000). *Econometrics*, Princeton University Press.
Formal sehr klar aufgebaut.
http://fhayashi.fc2web.com/hayashi_econometrics.htm
- Hansen, B. (2015). *Econometrics* <http://www.ssc.wisc.edu/~bhansen/econometrics/>
- Peracchi, F. (2001). *Econometrics*, John Wiley & Sons.
Der statistische Ansatz zur Regression mit methodischer Tiefe,
<http://eu.wiley.com/WileyCDA/WileyTitle/productCd-0471987646,descCd-tableOfContents.html>
- Ruud, P.A. (2000). *An Introduction to Classical Econometric Theory*. Oxford University Press.
Der geometrische Ansatz mit methodischer Tiefe
- Verbeek, M. (2012). *A Guide to Modern Econometrics*, 4th. ed., Wiley.
- Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*, 2nd. ed., MIT Press.
Viel Intuition und methodische Tiefe
<http://mitpress.mit.edu/books/econometric-analysis-cross-section-and-panel-data>

Noch Organisatorisches

- **PC-Einführung zur Computersprache R** entweder als Screencast oder im Rahmen der ersten Vorlesung nach dem Mathematikvorkurs, siehe [Kurshomepage](#) unter “Aktuelles”
- Infoveranstaltungen zum **Auslandsstudium** immer Anfang des Wintersemesters. Genaue

Daten auf der [Homepage des Akademischen Auslandsamtes](#)

- Nobelpreis-Vorlesung, siehe [Homepage des VWL-Instituts](#)

Inhaltsverzeichnis

I. Mathematischer Vorkurs	1
1. Lineare Algebra	2
1.1. Vektoren	4
1.2. Vektorräume	5
1.3. Euklidischer Raum	9
1.4. Matrizen	13
1.4.1. Addition von Matrizen	15
1.4.2. Skalarmultiplikation	15
1.4.3. Subtraktion von Matrizen	16
1.5. Weitere Operationen mit Matrizen	16
1.5.1. Matrixmultiplikation	16
1.5.2. Rechenregeln für Matrizen	19
1.6. Wichtige spezielle Matrizen	22
1.7. Maßzahlen von Matrizen	24
1.7.1. Spur einer Matrix	24
1.7.2. Rang einer Matrix	25
1.7.3. Determinanten	26
1.8. Matrixinversion	27
1.9. Euklidische Unterräume	29
1.10. Matrizen und lineare Abbildungen	31
1.11. Matrixdarstellung linearer Gleichungssysteme	32
1.12. (Semi-)definite Matrizen	33
1.13. Rechenregeln zum Ableiten von vektorwertigen Funktionen	35
1.14. Partitionierte Matrizen	37
2. Grundlagen der Wahrscheinlichkeitstheorie	38
2.1. Wichtige Grundbegriffe	39
2.2. Wozu brauchen wir Wahrscheinlichkeitstheorie?	40
2.3. Wahrscheinlichkeitsraum	43
2.4. Zufallsvariablen	46
2.5. Verteilungs- und Dichtefunktionen	49
2.5.1. Univariate Verteilungs- und Dichtefunktionen	49
2.5.2. Multivariate Verteilungs- und Dichtefunktionen	54
2.6. Bedingte Wahrscheinlichkeiten	57
2.7. Erwartungswerte und Momente	60
2.8. Bedingte Erwartungswerte und Momente	63

2.9.	Wichtige Wahrscheinlichkeitsverteilungen	67
2.9.1.	Normalverteilung	67
2.9.2.	χ^2 -, t -, F -Verteilung	69
2.9.3.	Ergänzung: Transformationssätze	71
3.	Konvergenz und Grenzwerte	73
3.1.	Konvergenz von Folgen	74
3.2.	Konvergenz von Funktionen	76
3.3.	Fast sichere Konvergenz	78
3.4.	Konvergenz in Wahrscheinlichkeit	79
3.5.	Konvergenz in Verteilung	81
3.6.	Beispiel der gleitenden Hügel: Konvergenz in Wahrscheinlichkeit $\not\Rightarrow$ Fast sichere Konvergenz	83
II.	Ökonometrische Methoden	85
4.	Einführung	86
4.1.	Aussagen über kausale Zusammenhänge	86
4.2.	Was ist Ökonometrie?	96
4.3.	Bestandteile einer empirischen Analyse	96
5.	Grundlagen der Schätz- und Testtheorie	99
5.1.	Stichproben und datengenerierende Prozesse	99
5.2.	Ökonometrische Modelle	101
5.3.	Regressionsmodelle	107
5.4.	Relevante Eigenschaften von Schätzern	113
5.5.	Werkzeuge für die asymptotische Analyse	121
5.5.1.	Gesetz der großen Zahlen — Law of Large Numbers (LLN)	121
5.5.2.	Zentrale Grenzwertsätze	123
5.6.	Grundlagen von Tests	126
6.	Der Kleinst-Quadrate-Schätzer: Ableitung und eine Anwendung	141
6.1.	Vektor- und Matrixdarstellung des multiplen linearen Regressionsmodells	142
6.2.	Der KQ-Schätzer für multiple lineare Regressionsmodelle	143
6.2.1.	Ableitung des KQ-Schätzers als Momentenschätzer	144
6.2.2.	Kleinst-Quadrate-Ableitung des KQ-Schätzers	145
6.3.	Empirische Analyse von Handelsströmen: Teil 1 — ein kurzer Überblick	147
7.	Der Kleinst-Quadrate-Schätzer und dessen geometrische Interpretation	154
7.1.	Die Geometrie des KQ-Schätzers	155
7.1.1.	Orthogonalitätsbedingungen	155
7.1.2.	Orthogonale Projektionen und ihre Eigenschaften	158
7.1.3.	Frisch-Waugh-Lovell Theorem	163
7.2.	Anwendungen des Frisch-Waugh-Lovell Theorems	165

8. Multiple Regression: Interpretation	170
8.1. Parameterinterpretation und funktionale Form	170
8.2. Datenskalierung	175
8.3. Qualitative Daten als Regressoren	176
8.3.1. Dummy-Variable oder Binärvariable	176
8.3.2. Mehrere Untergruppen	180
8.3.3. Interaktionen und Dummyvariablen	182
8.4. Modelle mit quadratischen Regressoren	184
9. Statistische Eigenschaften des KQ-Schätzers: Erwartungswert und Kovarianz	188
9.1. Erwartungstreue des KQ-Schätzers	189
9.1.1. Bedingungen für die Erwartungstreue	190
9.1.2. Vorherbestimmte Regressoren	192
9.2. Konsistenz des KQ-Schätzers	193
9.3. Die Kovarianzmatrix der Parameterschätzer	196
9.4. Die Effizienz unverzerrter KQ-Schätzer	202
9.5. Schätzen der Fehlervarianz	204
9.6. Über- oder fehlspezifizierte lineare Regressionsmodelle	206
10. Modellspezifikation	210
10.1. Modellselektionskriterien	212
10.2. Tests für nichtgeschachtelte Modelle	214
10.3. Empirische Analyse von Handelsströmen: Teil 2	214
11. (Asymptotische) Verteilung des KQ-Schätzers und Tests	217
11.1. Exakte Verteilung des KQ-Schätzers	217
11.2. Asymptotische Verteilung des KQ-Schätzers	218
11.3. Exakte Tests	221
11.3.1. t -Tests: Testen einer einzelnen Restriktion	222
11.3.2. F -Tests: Testen mehrerer Restriktionen	225
11.4. Asymptotische Tests	235
11.4.1. Asymptotischer t -Test	236
11.4.2. Asymptotischer F -Test	237
11.5. Monte-Carlo-Tests und Bootstraptests	238
11.5.1. Monte-Carlo-Tests	238
11.5.2. Bootstraptests	239
11.6. Konfidenzintervalle und -ellipsoide	242
11.6.1. Konfidenzintervalle	242
11.6.2. Konfidenzellipsoide	245
11.7. Empirische Analyse von Handelsströmen: Teil 3	246
12. Univariate Zeitreihenmodelle	256
12.1. Stochastische Prozesse	257
12.2. Lineare stochastische Prozesse und MA-Prozesse	265
12.3. AR-Prozesse	268
12.3.1. AR(1)-Prozesse	269

12.3.2. Komplexe Zahlen	273
12.3.3. AR(p)-Prozesse und mehr	277
12.3.4. KQ-Schätzer für AR(p)-Modelle	282
12.4. Schätzung erster und zweiter Momente im Fall stationärer Prozesse	283
12.4.1. Schätzen des Mittelwertes	283
12.4.2. Schätzen der Autokovarianzfunktion	286
12.4.3. Schätzen der Autokorrelationsfunktion	287
13. Modelle für multivariate Zeitreihen	290
13.1. Multivariate datengenerierende Prozesse	290
13.2. Dynamische ökonometrische Modelle	291
13.3. Bedingungen an exogene Variablen	294
13.4. Dynamische lineare Regressionsmodelle	302
13.5. KQ-Schätzung von dynamischen linearen Regressionsmodellen	305
14. Verallgemeinerter Kleinst-Quadrate-Schätzer und seine Anwendungen	309
14.1. Verallgemeinerter Kleinst-Quadrate-Schätzer	310
14.2. Feasible GLS	313
14.2.1. Modellierung heteroskedastischer Fehler	314
14.2.2. Modelle mit autokorrelierten Fehlern	315
14.3. Heteroskedastie-robuste Standardfehler	315
14.4. Empirische Analyse von Handelsströmen: Teil 4	317
15. Modellüberprüfung	320
15.1. Tests auf Autokorrelation	321
15.2. Tests auf heteroskedastische Fehler	321
15.2.1. Breusch-Pagan-Test	321
15.2.2. White-Test	323
15.3. RESET-Test	324
15.4. Normalitätstest: Lomnicki-Jarque-Bera-Test	325
15.5. Stabilitätstests	325
15.6. Zusammenfassung eines ökonometrischen Modellierungsprozesses	325
15.7. Empirische Analyse von Handelsströmen: Teil 5	325
A. R-Programme	328
A.1. R-Programme für Graphiken in Abschnitt 2.5 zu Verteilungs- und Dichtefunktionen	328
A.2. R-Programm für Monte-Carlo-Simulation im Abschnitt 5.5.1 zum Gesetz der großen Zahlen	333
A.3. R-Programme für Graphiken im Abschnitt 5.6 zu Grundlagen von Tests	334
A.4. R-Programm für empirisches Beispiel zu Handelsströmen, beginnend in Abschnitt 6.3	338
A.5. R-Programm für Graphiken in Abschnitt 7.1 Die Geometrie des KQ-Schätzers	345
A.6. R-Programm für Regressionsergebnisse in Abschnitt 8.3 zu qualitative Daten als Regressoren	349
A.7. R-Programm für Graphiken in Abschnitt 9.1 zu Erwartungstreue des KQ-Schätzers	350

A.8. R-Programm für Monte-Carlo-Simulation im Abschnitt 9.2 zur Konsistenz des KQ-Schätzers	352
A.9. R-Programm zur Darstellung von ifo Geschäftsklimazeitreihen im Abschnitt 12 zu univariaten Zeitreihenmodellen	354
A.10. R-Programm zur Darstellung verschiedener Realisierungen von Zeitreihen im Abschnitt 12.1 zu stochastischen Prozessen	355
A.11. R-Programm für Monte-Carlo-Simulation im Abschnitt 12.2 zu linearen stochastischen Prozessen und MA-Prozessen	356
A.12. R-Programm für Monte-Carlo-Simulation im Abschnitt 12.3.1 zu AR(1)-Prozessen	356
A.13. R-Programm für Monte-Carlo-Simulation im Abschnitt 12.3.3 zu AR(p)-Prozessen und mehr	357
A.14. R-Programm für Schätzung der Autokorrelationsfunktion im Abschnitt 12.4 zur Schätzung erster und zweiter Momente im Fall stationärer Prozesse	358
A.15. R-Programm für die Simulation und Schätzung von AR(1)-Prozessen im Abschnitt 13.5 zur KQ-Schätzung von dynamischen linearen Regressionsmodellen	359
B. R Befehle für die Regressionsanalyse	361
B.1. Übersicht über verfügbare Befehle	361
B.2. Eigene R-Programme	364
C. Daten für die Schätzung der Gravitationsgleichung	366
D. Grundbegriffe Mengen	370

Teil I.

Mathematischer Vorkurs

1. Lineare Algebra

Wozu braucht man lineare Algebra?

- Zur Analyse der Eigenschaften der Lösungen linearer Gleichungssysteme
- Darstellung multivariater Daten
- Darstellung linearer Zusammenhänge
- Zum Lösen linearer Gleichungssysteme
Beispiel: Normalgleichungen zum Berechnen des Kleinst-Quadrate-Schätzers
- Bestandteil von nichtlinearen Optimierungsalgorithmen
- Darstellung geometrischer Zusammenhänge durch Systeme von Zahlen (analytische Geometrie)
Matrizen als geometrische Transformationen
- Alles zusammen: in der Ökonometrie

Übersicht

- Vektoren
- Vektorräume
- Euklidischer Raum und Unterräume
- Matrizen, Rechenregeln, spezielle Matrizen und Maßzahlen
- Matrizen und lineare Abbildungen
- (Semi-)definite Matrizen
- Rechenregeln zum Ableiten von vektorwertigen Funktionen
- Partitionierte Matrizen

Literaturhinweise

- [Schmidt & Trenkler \(2015\)](#) Kompaktes leicht lesbares deutsches Lehrbuch mit sehr vielen Beispielen zum Rechnen mit Matrizen.
- [Gentle \(2007, Chapter 2\)](#) (Volltext-Zugriff im Bereich der UR): detaillierte Einführung in Vektorräume
- [Fischer \(2014, Abschnitt 1.4\)](#) Grundlegende Einführung für Mathematiker, Physiker, Ingenieure, usw.
- [Lütkepohl \(1996\)](#) Hervorragendes Nachschlagewerk zur linearen Algebra und deren verschiedenen Matrizen und dazugehörigen Rechenregeln und Umformungsmöglichkeiten. Oft hilfreich beim Lesen von Fachartikeln.

1.1. Vektoren

Übersicht

- Raum
- Euklidischer Raum
- Vektoren
- Dimension, Länge eines Vektors

Euklidischer Raum und Vektoren

Raum

Ein Raum ist in der Mathematik eine **Menge** mathematischer Objekte **mit** einer zusätzlichen **Struktur**. D.h. es sind Operationen bzgl. der Elemente der Menge möglich. ([http://de.wikipedia.org/wiki/Raum_\(Mathematik\)](http://de.wikipedia.org/wiki/Raum_(Mathematik)))

Beispiele:

- Vektorraum
- Euklidischer Raum (= Vektorraum mit Skalarprodukt)
- Wahrscheinlichkeitsraum (=“Menge mit Mengensystem und Wahrscheinlichkeitsabbildung”)

Euklidischer Raum, n -dimensionaler Raum

- Die einem Euklidischen Raum zugrunde liegende Menge ist die Menge der geordneten n -Tupel \mathbf{x} reeller Zahlen:

$$\mathbb{R}^n = \{\mathbf{x} = (x_1, \dots, x_n) : x_1, \dots, x_n \in \mathbb{R}\}.$$

Die geordneten n -Tupel $\mathbf{x} = (x_1, \dots, x_n)$ werden auch als n -**Vektoren** oder kurz **Vektoren** bezeichnet. Jeder geordnete n -Vektor stellt einen Punkt im n -dimensionalen Euklidischen Raum \mathbb{R}^n dar, kurz: $\mathbf{x} \in \mathbb{R}^n$.

- Die damit verbundene Struktur umfasst als Operationen zwischen den Elementen
 - die Addition,

- die Skalarmultiplikation sowie
- das Skalarprodukt.

Beispiele:

- $n = 1$: $x \in \mathbb{R}^1$ Menge entspricht Zahlengerade, Elemente sind Skalare.
- $n = 2$: $\mathbf{x} \in \mathbb{R}^2$ Menge entspricht der Ebene, Elemente sind zweidimensionale Vektoren.
- $n = 3$: $\mathbf{x} \in \mathbb{R}^3$ Menge entspricht Raum mit Länge, Breite und Höhe.

Weitere Begriffe

- n : **Dimension** von \mathbf{x} . n wird in der linearen Algebra auch als **Länge** bezeichnet (so auch in \mathbb{R} !)

Achtung: Die Länge eines Vektor bezeichnet häufig auch die Euklidische Norm eines Vektors. Siehe Abschnitt 1.2.

- x_i : **Element** oder **Komponente** von \mathbf{x} .

1.2. Vektorräume

Übersicht

- Vektorraum
- Addition und Subtraktion
- Nullvektor
- Inverser Vektor
- Linearkombination
- Geraden
- Skalarprodukt bzw. inneres Produkt

Definition

Eine Menge \mathcal{V} mit den Operationen

- **Addition** $\mathcal{V} \times \mathcal{V} \rightarrow \mathcal{V}$
- **Multiplikation mit einem Skalar (Skalarmultiplikation)** $\mathbb{R} \times \mathcal{V} \rightarrow \mathcal{V}$

wird **linearer Vektorraum** genannt, wenn darüber hinaus

1. für die Addition ein **Nullvektor** und ein **inverses Element** existieren sowie Assoziativität und Kommutativität gelten und
2. für die Multiplikation mit einer Zahl **Distributivität** und **Assoziativität** gelten, sowie **Multiplikation mit Eins** wieder das gleiche Element ergibt, d. h. Multiplikation und Addition sinnvoll verträglich sind.

(adaptiert von [Fischer \(2010, S. 76\)](#))

Bemerkungen

- **Linearkombination:** Verknüpfung der Operationen von Addition und Multiplikation mit Skalaren:

$$\alpha, \beta \text{ skalar, } \mathbf{x}, \mathbf{y} \in \mathcal{V} : \quad \alpha \mathbf{x} + \beta \mathbf{y} \in \mathcal{V}.$$

Jede Linearkombination der Vektoren ist in \mathcal{V} enthalten.

Deshalb ist ein Vektorraum ein **linearer Raum**.

Reellwertige Vektoren und Vektorraum

Die Menge der reellwertigen n -Vektoren $\mathbf{x} \in \mathbb{R}^n$ bildet einen linearen Vektorraum.

Überprüfung – Operationen

Es seien $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

- **Addition** von Vektoren der Länge n

$$\begin{aligned} \mathbb{R}^n \times \mathbb{R}^n &\rightarrow \mathbb{R}^n : \\ (x_1, \dots, x_n) + (y_1, \dots, y_n) &:= (x_1 + y_1, \dots, x_n + y_n) = \mathbf{z} \\ \mathbf{x} + \mathbf{y} &= \mathbf{z} \end{aligned}$$

- **Multiplikation mit einer Zahl** $\lambda \in \mathbb{R}$

$$\mathbb{R} \times \mathbb{R}^n \rightarrow \mathbb{R}^n : \lambda \cdot \mathbf{x} = \lambda \cdot (x_1, \dots, x_n) := (\lambda \cdot x_1, \dots, \lambda \cdot x_n)$$

Überprüfung – Bedingungen für Addition

- **Nullvektor:** Es existiert ein Nullvektor $\mathbf{0} := (0, \dots, 0)$, so dass gilt:

$$\mathbf{0} + \mathbf{x} = \mathbf{x}, \quad \mathbf{0} \cdot \mathbf{x} = \mathbf{0}$$

- **Inverser Vektor:** zu jedem Vektor $\mathbf{x} \in \mathbb{R}^n$ gibt es ein inverses Element $\mathbf{z} \in \mathbb{R}^n$, das ihn mit der Verknüpfung auf den Nullvektor abbildet. Der inverse Vektor ist der **negative Vektor** $-\mathbf{x} = -(x_1, \dots, x_n) := (-x_1, \dots, -x_n)$. Überprüfen!

Überprüfung – Bedingungen für Skalarmultiplikation

- **Distributivität für die Skalarmultiplikation:** Für $\alpha, \beta \in \mathbb{R}$ gilt:

$$\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}, \quad (\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}$$

- **Assoziativität der Skalarmultiplikation**

$$(\alpha\beta) \cdot \mathbf{x} = \alpha \cdot (\beta \cdot \mathbf{x})$$

Weitere Eigenschaften

- **Assoziativität** der Addition:

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z})$$

- **Subtraktion** ergibt sich aus Addition und Multiplikation mit einer Zahl:

$$\mathbf{z} - \mathbf{y} = \mathbf{z} + (-\mathbf{y}) = \mathbf{x}$$

- Zwei Vektoren \mathbf{x} und \mathbf{y} der Länge n sind genau dann gleich, wenn $x_1 = y_1, \dots, x_n = y_n$ gilt.

Beachte: Die **Menge der reellen Zahlen** \mathbb{R} mit den genannten Operationen bildet auch einen Vektorraum (= Spezialfall für $n = 1$).

Geraden im \mathbb{R}^n

Definition

- Zwei verschiedene Punkte $\mathbf{v}, \mathbf{v}' \in \mathbb{R}^n$ bestimmen eine **Gerade**
- Im \mathbb{R}^n : Seien $\mathbf{v}, \mathbf{v}' \in \mathbb{R}^n$ fest. Alle Punkte auf der durch \mathbf{v} und \mathbf{v}' definierten Geraden

sind gegeben durch

$$L = \{\mathbf{x} \in \mathbb{R}^n : \mathbf{x} = \mathbf{v} + \lambda \mathbf{w}, \lambda \in \mathbb{R}\}.$$

wobei $\mathbf{w} = \mathbf{v}' - \mathbf{v}$. Die Menge L ist das Bild der Abbildung $\Phi : \mathbb{R} \rightarrow L \subset \mathbb{R}^n : \lambda \rightarrow \mathbf{v} + \lambda \mathbf{w}$ und wird **Parametrisierung der Geraden** genannt.

Geraden im \mathbb{R}^2 (in der Ebene)

- Spezialfall für $n = 2$.
- Alle Punkte \mathbf{x} einer Gerade im \mathbb{R}^2 lassen sich als eine Gleichung mit zwei Unbekannten $\mathbf{x} = (x_1, x_2)$ und drei festen Koeffizienten a_1, a_2, b darstellen

$$a_1 x_1 + a_2 x_2 = b.$$

Die drei Koeffizienten a_1, a_2, b bestimmen die Lage der Geraden und können aus zwei gegebenen Punkten der Gerade \mathbf{v}, \mathbf{v}' bestimmt werden und vice versa.

- Bestimmung zweier Punkte auf der Gerade für gegebene Koeffizienten: Für gegebenes x_1 lässt sich x_2 eindeutig bestimmen und umgekehrt, sofern $a_1 \neq 0, a_2 \neq 0$. Beispiel:

$$x_1 = 0 : x_2 = \frac{b}{a_2} \quad \text{ein Punkt der Gerade}$$

$$x_2 = 0 : x_1 = \frac{b}{a_1} \quad \text{zweiter Punkt der Gerade}$$

- Zwei Geraden schneiden sich genau in einem Punkt, außer sie sind gleich oder parallel
 \Leftrightarrow zweidimensionales lineares Gleichungssystem hat eine, unendlich viele oder keine Lösung.
- **Schnittpunkt zweier Geraden:** Lösung \mathbf{x} des linearen Gleichungssystems

$$a_1 x_1 + a_2 x_2 = b$$

$$c_1 x_1 + c_2 x_2 = d$$

Auflösen nach x_1, x_2 durch Einsetzen oder mithilfe von **Matrixalgebra**. Siehe Abschnitt [1.4](#).

Vektorraum: Skalarprodukt

Skalarprodukt oder inneres Produkt (scalar product, dot product)

Die Abbildung $\mathcal{V} \times \mathcal{V} \rightarrow \mathbb{R}$ ergibt als Ergebnis einen Skalar.

Die zusätzliche Existenz des Skalarprodukts für einen Vektorraum ermöglicht

1. eine eindeutige Charakterisierung der Beziehung zwischen den Elementen,

2. die Charakterisierung der einzelnen Elemente durch Bestimmung ihrer Länge.

Hinweis: Das Skalarprodukt ist ein spezieller Typ eines inneren Produkts. Innere Produkte können beispielsweise auch für Funktionen definiert sein. Allgemein gilt, dass ein inneres Produkt $\langle \cdot, \cdot \rangle$ als Ergebnis immer eine reelle oder komplexe Größe liefert (Gentle 2007, Sections 2.1.4, 3.2.6).

Generell gilt: Eine

- eindeutige Charakterisierung der Beziehung zwischen den Elementen eines Vektorraums ist gegeben, wenn eine **Metrik** für den Vektorraum existiert,
- eindeutige Charakterisierung der Beziehung einzelner Elemente eines Vektorraums ist gegeben, wenn eine **Norm** für den Vektorraum existiert.

1.3. Euklidischer Raum

Übersicht

- Vektorraum im \mathbb{R}^n
- Euklidischer Raum
- Norm
- Normierter Vektorraum
- Euklidische Norm
- Metrik
- Metrischer Raum
- Orthogonale Vektoren
- Lineare Unabhängigkeit

Skalarprodukt im Vektorraum $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$

$$\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} : \langle \mathbf{x}, \mathbf{y} \rangle := \sum_{i=1}^n x_i y_i \quad (1.1)$$

Definition Euklidischer Raum

Der Vektorraum aller reellwertiger n -Vektoren $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, in dem zusätzlich das Skalarprodukt $\mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R} : \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$ definiert ist, wird als Euklidischer Raum bezeichnet.

Durch die Existenz des Skalarprodukts ist eine anschauliche geometrische Charakterisierung des Euklidischen Raums möglich.

Norm und normierter Vektorraum

Eine **Norm** erlaubt, allgemein formuliert, die quantitative Bewertung einzelner Elemente einer Menge und wie sich zeigen lässt, ihrer Beziehungen zueinander.

Norm für einen Vektorraum

Die Abbildung $\|\cdot\| : \mathcal{V} \rightarrow [0, \infty)$: ordnet jedem Element \mathbf{x} des Vektorraums eine nichtnegative reelle Zahl $\|\mathbf{x}\|$ zu und genügt folgenden Eigenschaften:

1. Wenn $\mathbf{x} \neq 0$, dann gilt $\|\mathbf{x}\| > 0$ und wenn $\|\mathbf{x}\| = 0 \Leftrightarrow \mathbf{x} = 0$.
2. $\|\alpha \mathbf{x}\| = |\alpha| \|\mathbf{x}\|$.
3. $\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ (Dreiecksungleichung).

(Vgl. [Gentle 2007](#), Section 2.1.5)

Normierter Vektorraum

ein Vektorraum, dessen Elemente mit einer Norm bewertet/gemessen werden können.

Verschiedene Vektornormen

- L_2 -Norm oder **Euklidische Norm**:

$$\|\mathbf{x}\|_2 := \sqrt{\sum_{t=1}^n x_t^2}$$

Die Euklidische Norm **misst** die **Länge eines n -dimensionalen Vektors**:

$$\|\mathbf{x}\|_2 := \left(\sum_{t=1}^n x_t^2 \right)^{1/2}$$

Der **Betrag** einer reellen Zahl $|x|, x \in \mathbb{R}$ ist die Euklidische Norm in \mathbb{R} .

- ‡ L_∞ -Norm oder **Tschebyscheff-Norm**: $\|\mathbf{x}\|_\infty := \max_{t \in n} |x_t|$. Z. B. relevant beim Beladen von Fahrzeugen, wenn keine Kante eines zu transportierenden Gegenstandes eine maximale Länge überschreiten darf.
- ‡ L_p -Norm:

$$\|\mathbf{x}\|_p := \left(\sum_{t=1}^n |x_t|^p \right)^{1/p},$$

enthält beide bereits genannten Fälle als Spezialfälle.

Metrik und metrischer Raum

Metrik

Eine Metrik ist eine Abstandsfunktion $d : \mathcal{V} \times \mathcal{V} \rightarrow [0, \infty)$, die folgende Bedingungen erfüllt, wobei \mathcal{V} einen Vektorraum bezeichnet. Für zwei Objekte x und y in \mathcal{V} gilt:

1. $d(y, x) > 0$, wenn $x \neq y$ und $d(y, x) = 0$, falls $x = y$,
2. $d(x, y) = d(y, x)$,
3. $d(x, z) \leq d(x, y) + d(y, z)$.

(Gentle 2007, Section 2.1.7)

Metrischer Raum

Ein normierter Vektorraum ist automatisch ein **metrischer Raum**, da die induzierte Metrik $d(x, y) := \|x - y\|$ allen Anforderungen genügt.

Skalarprodukt, Norm, Metrik

Generell gilt (nicht nur für Euklidischen Raum):

$$\begin{array}{ccccc} \text{Skalarprodukt} & \implies & \text{Norm} & \implies & \text{Metrik} \\ \langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i & \implies & \langle \mathbf{x}, \mathbf{x} \rangle = \sum_{i=1}^n x_i^2 = \|\mathbf{x}\|_2^2 & \implies & d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_2 \end{array}$$

Geometrie von Vektoren im zweidimensionalen Euklidischen Vektorraum

- **Notation**: Im Folgenden schreiben wir: $\|\mathbf{x}\| = \|\mathbf{x}\|_2$.
- Geometrie der **Addition** von Vektoren: Ergebnis ist Diagonale im Parallelogramm.
- Geometrie der **Multiplikation mit einem Skalar** α : $\alpha \mathbf{x}$ ist Vektor parallel zu \mathbf{x} mit unterschiedlicher Länge und möglicherweise mit entgegengesetzter Richtung.

- Geometrie des **Skalarprodukts** bzw. **innere Produkts** zweier Vektoren: Produkt der Längen der beiden Vektoren und dem Kosinus des Winkels θ zwischen beiden (ohne Beweis)

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta. \quad (1.2)$$

für Gültigkeit: Gegeben seien zwei spezielle Vektoren im E^2 :

$$\begin{aligned} \mathbf{w} &= \begin{pmatrix} 1 & 0 \end{pmatrix}, \\ \mathbf{z} &= \begin{pmatrix} \cos \theta & \sin \theta \end{pmatrix}. \end{aligned}$$

Skalarmultiplikation ergibt zwei weitere Vektoren:

$$\begin{aligned} \mathbf{x} &= \alpha \mathbf{w}, \quad \alpha > 0, \\ \mathbf{y} &= \gamma \mathbf{z}, \quad \gamma > 0. \end{aligned}$$

Dann ergeben sich die inneren Produkte bzw. Skalarprodukte

$$\begin{aligned} \|\mathbf{w}\| &= 1, \\ \|\mathbf{z}\| &= (\cos^2 \theta + \sin^2 \theta)^{1/2} = 1, \\ \langle \mathbf{w}, \mathbf{z} \rangle &= w_1 z_1 + w_2 z_2 = \cos \theta \end{aligned}$$

und

$$\begin{aligned} \|\mathbf{x}\| &= |\alpha| \|\mathbf{w}\| = \alpha, \\ \|\mathbf{y}\| &= |\gamma| \|\mathbf{z}\| = \gamma, \\ \langle \mathbf{x}, \mathbf{y} \rangle &= \langle \alpha \mathbf{w}, \gamma \mathbf{z} \rangle = \alpha w_1 \gamma z_1 + \alpha w_2 \gamma z_2 = \alpha \gamma \langle \mathbf{w}, \mathbf{z} \rangle \\ &= \alpha \gamma \cos \theta \\ &= \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta. \end{aligned}$$

Orthogonale Vektoren

- Stehen zwei Vektoren **orthogonal aufeinander** (senkrecht aufeinander), dann und nur dann ist deren inneres Produkt Null, da $\cos 90^\circ = 0 \implies$
Wenn $\langle \mathbf{x}, \mathbf{y} \rangle = 0 \iff$ die Vektoren \mathbf{x} und \mathbf{y} stehen orthogonal zueinander.
- **Cauchy-Schwartz Ungleichung**

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\| \quad \text{bzw.} \quad \langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \langle \mathbf{y}, \mathbf{y} \rangle.$$

Diese folgt aus (1.2) und $-1 \leq \cos \theta \leq 1$.

Lineare Unabhängigkeit

- **Lineare Unabhängigkeit:** k Vektoren \mathbf{x}_i , $i = 1, \dots, k$, (mit positiver Länge) sind linear unabhängig, falls es *keine* $k - 1$ Skalare c_i gibt, so dass gilt:

$$\mathbf{x}_j = \sum_{\substack{i=1 \\ i \neq j}}^k c_i \mathbf{x}_i, \quad 1 \leq j \leq k.$$

Beispiel: Seien die Spalten der $n \times k$ Matrix \mathbf{X} linear unabhängig. Dann existiert nur ein Nullvektor $\boldsymbol{\gamma}$, also kein $\boldsymbol{\gamma}$ mit positiver Länge, so dass

$$\sum_{i=1}^k \mathbf{x}_i \gamma_i = \mathbf{0}, \quad j = 1, \dots, n.$$

1.4. Matrizen

Übersicht

- Definition
- Addition von Matrizen
- Nullmatrix
- Skalarmultiplikation
- Subtraktion von Matrizen

Matrizen

Definition

- Eine **Matrix** \mathbf{A} ist ein rechteckiges Schema von nm Zahlen, $n, m \in \mathbb{N}$,

$$\mathbf{A} := \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ a_{21} & \cdots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = (a_{ij})_{i=1, \dots, m, j=1, \dots, n} = (a_{ij})$$

- **Dimension** einer Matrix: Zahl der Zeilen m und Zahl der Spalten n .
Kurznotation: $(m \times n)$ -Matrix oder $m \times n$ -Matrix.

- Die Einträge a_{ij} , $i = 1, \dots, m$, $j = 1, \dots, n$ werden **Elemente** oder **Koeffizienten** einer Matrix genannt.

Bemerkungen

- **Beachte:** Häufig, so in der Matrixalgebra, wird zusätzlich zur Dimension n ein Vektor als Spalten- oder Zeilenvektor definiert.

$$\mathbf{x} = \underbrace{\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}}_{\text{Spaltenvektor}} \quad \text{oder} \quad \mathbf{x} = \underbrace{(x_1 \ \cdots \ x_n)}_{\text{Zeilenvektor}}$$

In \mathbb{R} wird den Klasse der Vektoren jedoch keine Spalten- oder Zeileneigenschaft zugewiesen. Dies geschieht nur bei der Klasse der Matrizen. Daran denken!

- In \mathbb{R} existiert die Klasse `matrix`. Sie wird benötigt, um Spalten- oder Zeilenvektoren zu definieren.
- Eine $(m \times n)$ -Matrix besteht aus n Spaltenvektoren der Länge m , bzw. m Zeilenvektoren der Länge n . Im Fall reeller Zahlen als Elemente schreibt man

$$\mathbf{A} \in \mathbb{R}^{m \times n}.$$

Denn es liegen n Vektoren der Dimension m vor.

Bemerkungen – Fortsetzung

- Zwei Matrizen der gleichen Dimension sind identisch, wenn alle Koeffizienten übereinstimmen.
- Die Elemente können aus verschiedenen Mengen stammen: z. B. \mathbb{N} , \mathbb{R} , den komplexen Zahlen \mathbb{C} oder auch Polynome.
- Jede Tabelle entspricht einer Matrix.
- Ein Spaltenvektor der Länge m entspricht einer $(m \times 1)$ -Matrix. Ein Zeilenvektor der Länge n entspricht einer $(1 \times n)$ -Matrix.

Grundlegende Operationen mit Matrizen

- Die grundlegenden Operationen **Addition** und **Multiplikation mit einer Zahl** aus Abschnitt 1.1 lassen sich auch auf Matrizen anwenden.
- Alle weiteren Eigenschaften bzgl. dieser Operationen aus Abschnitt 1.1 gelten entsprechend für Matrizen wie im Folgenden zu sehen ist.

1.4.1. Addition von Matrizen

Die Addition von zwei Matrizen \mathbf{A} , \mathbf{B} mit gleicher Dimension m und n ergibt wieder eine $(m \times n)$ -Matrix.

Das (i, j) -te Element ist gerade die Summe der (i, j) -ten Elemente der zu addierenden Matrizen.

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & b_{ij} & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & a_{ij} + b_{ij} & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

Beispiel:

$$\begin{pmatrix} 3 & 4 & 1 \\ 6 & 7 & 0 \\ -1 & 3 & 8 \end{pmatrix} + \begin{pmatrix} -1 & 0 & 7 \\ 6 & 5 & 1 \\ -1 & 7 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 4 & 8 \\ 12 & 12 & 1 \\ -2 & 10 & 8 \end{pmatrix}$$

Beispiel: Vorsicht:

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 2 & -2 \end{pmatrix} + \begin{pmatrix} 3 & 6 \\ 1 & 4 \end{pmatrix} \text{ ist nicht definiert!}$$

Nullmatrix

Definition

Eine $(m \times n)$ -Matrix $\mathbf{0}$ heißt **Nullmatrix**, wenn alle Einträge 0 sind.

Daraus folgt:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} + \begin{pmatrix} 0_{11} & \cdots & 0_{1n} \\ \vdots & 0_{ij} & \vdots \\ 0_{m1} & \cdots & 0_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

bzw.

$$\mathbf{A} + \mathbf{0} = \mathbf{A} \quad \forall \mathbf{A}$$

1.4.2. Skalarmultiplikation

Die Multiplikation einer Matrix \mathbf{A} mit einer Zahl λ ergibt wieder eine Matrix. Dabei wird jedes Element a_{ij} mit λ multipliziert.

$$\lambda \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \cdots & \lambda a_{1n} \\ \vdots & \lambda a_{ij} & \vdots \\ \lambda a_{m1} & \cdots & \lambda a_{mn} \end{pmatrix}.$$

1.4.3. Subtraktion von Matrizen

Definition einer negativen Matrix

Die Matrix $-\mathbf{A}$ ergibt sich aus der Multiplikation einer Matrix mit einem Skalar oder aus der Matrix, die zu \mathbf{A} addiert werden muss, um die Nullmatrix zu erhalten.

$$-\mathbf{A} = - \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} = \begin{pmatrix} -a_{11} & \cdots & -a_{1n} \\ \vdots & -a_{ij} & \vdots \\ -a_{m1} & \cdots & -a_{mn} \end{pmatrix}.$$

Subtraktion

$$\mathbf{A} - \mathbf{B} = \mathbf{A} + (-\mathbf{B}).$$

Daraus folgt:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} - \begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & b_{ij} & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix} = \begin{pmatrix} a_{11} - b_{11} & \cdots & a_{1n} - b_{1n} \\ \vdots & a_{ij} - b_{ij} & \vdots \\ a_{m1} - b_{m1} & \cdots & a_{mn} - b_{mn} \end{pmatrix}$$

1.5. Weitere Operationen mit Matrizen

Übersicht

- Matrixmultiplikation
- Elementweise Multiplikation bzw. Hadamardprodukt
- Transponierte einer Matrix
- Rechenregeln
- Multiplikation von Vektoren: Inneres und äußeres Produkt

1.5.1. Matrixmultiplikation

- **Voraussetzung für die Matrixmultiplikation \mathbf{AB}** von zwei Matrizen: einer $(k \times r)$ -Matrix \mathbf{A} und einer $(m \times n)$ -Matrix \mathbf{B} :
 - Die Zahl der Spalten r von \mathbf{A} entspricht der Zahl der Zeilen m von \mathbf{B} , d.h. \mathbf{A} muss Dimension $(k \times m)$ und \mathbf{B} Dimension $(m \times n)$ aufweisen.

- Diese Voraussetzung ist notwendig und hinreichend.
- Die Reihenfolge der Multiplikation:
 - kann nicht vertauscht werden, wenn $k \neq n$,
 - kann vertauscht werden, wenn $k = n$, jedoch im Allgemeinen mit unterschiedlichem Ergebnis.
- Die Matrixmultiplikation basiert auf dem Skalarprodukt
- **Notation:** $(\mathbf{AB})_{ij}$ bezeichnet das (i, j) -te Element der Matrix \mathbf{AB} .

Berechnung des Matrixprodukts

Der (i, j) -te Eintrag des Matrixprodukts \mathbf{AB} ist definiert als das Skalarprodukt der i -ten Zeile von \mathbf{A} (einem Zeilenvektor) mit der j -ten Spalte von \mathbf{B} (einem Spaltenvektor):

$$\begin{aligned}
 (\mathbf{AB})_{ij} &= (a_{i1} \quad a_{i2} \quad \cdots \quad a_{im}) \cdot \begin{pmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{mj} \end{pmatrix} \\
 &= a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{im}b_{mj} \\
 &= \sum_{h=1}^m a_{ih}b_{hj}
 \end{aligned}$$

Beispiel:

$$\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} aA + bC & aB + bD \\ cA + dC & cB + dD \\ eA + fC & eB + fD \end{pmatrix}.$$

Beachte: Das Produkt in umgekehrter Reihenfolge ist nicht definiert!

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix}$$

Dimension eines Matrixprodukts

Das Matrixprodukt \mathbf{AB} erbt die Zahl der Zeilen r von \mathbf{A} und die Zahl der Spalten n von \mathbf{B} :

$$\begin{array}{rcccl}
 \mathbf{A} & \cdot & \mathbf{B} & = & \mathbf{C} \\
 (k \times m) & \cdot & (m \times n) & = & (k \times n).
 \end{array}$$

- **Gute Praxis:** Vor jeder Matrixmultiplikation die Dimensionen der Matrizen überprüfen! Besonders beim Programmieren!
- **Beachte in R:** Das Matrixprodukt wird mit `A %% B` angegeben. In anderen Sprachen dagegen häufig mit `A*B`.

Kein Kommutativgesetz für die Matrixmultiplikation

- Gegeben sei eine $(n \times m)$ -Matrix **A** und eine $(m \times n)$ -Matrix **B**

$$\begin{array}{rcl} \mathbf{A} & \cdot & \mathbf{B} = \mathbf{C} \\ (n \times m) & \cdot & (m \times n) = (n \times n). \\ \mathbf{B} & \cdot & \mathbf{A} = \mathbf{D} \\ (m \times n) & \cdot & (n \times m) = (m \times m). \end{array}$$

- Selbst wenn **A** und **B** quadratisch sind, d.h. Zeilen- und Spaltenzahl gleich sind, $m = n$, kann $\mathbf{AB} \neq \mathbf{BA}$ auftreten.

Kein Kommutativgesetz für die Matrixmultiplikation

Beispiel:

$$\begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix},$$

während

$$\begin{pmatrix} 1 & -1 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 2 & 2 \end{pmatrix}.$$

Einheitsmatrix

Die $(n \times n)$ -Matrix

$$\mathbf{I} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix},$$

mit $a_{ii} = 1, \forall i$ and $a_{ij} = 0, \forall i \neq j$, wird als Einheitsmatrix bezeichnet.

Eigenschaften der Einheitsmatrix **I**:

- Multiplikative Identität für Matrixmultiplikation:

– für jede $(m \times n)$ -Matrix \mathbf{A} gilt

$$\mathbf{A}\mathbf{I} = \mathbf{A}, \quad (1.3)$$

– für jede $(n \times m)$ -Matrix \mathbf{B} gilt

$$\mathbf{I}\mathbf{B} = \mathbf{B}. \quad (1.4)$$

- \mathbf{I} entspricht der 1 bei den reellen Zahlen.

Elementweise Multiplikation (Hadamardprodukt)

Für zwei $(m \times n)$ -Matrizen \mathbf{A} und \mathbf{B} liefert die elementweise Multiplikation für den (i, j) -ten Eintrag des Hadamardprodukts $\mathbf{A} \odot \mathbf{B}$

$$(\mathbf{A} \odot \mathbf{B})_{ij} = a_{ij}b_{ij}$$

Die resultierende Matrix hat wieder Dimension $(m \times n)$ wie \mathbf{A} und \mathbf{B} .

Beispiel:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \odot \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{pmatrix} = \begin{pmatrix} a_{11}b_{11} & a_{12}b_{12} \\ a_{21}b_{21} & a_{22}b_{22} \end{pmatrix}$$

Bemerkungen:

- In \mathbb{R} wird für das elementweise Produkt $\mathbf{A} * \mathbf{B}$ verwendet! In anderen Sprachen notiert dies das Matrixprodukt!

1.5.2. Rechenregeln für Matrizen

- **Assoziativität für Addition und Matrixmultiplikation:**

$$\begin{aligned} (\mathbf{A} + \mathbf{B}) + \mathbf{C} &= \mathbf{A} + (\mathbf{B} + \mathbf{C}), \\ (\mathbf{A}\mathbf{B})\mathbf{C} &= \mathbf{A}(\mathbf{B}\mathbf{C}) \end{aligned}$$

- **Kommutativgesetz für die Addition**

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$$

- **Distributivität für die Matrixmultiplikation**

$$\begin{aligned} \mathbf{A}(\mathbf{B} + \mathbf{C}) &= \mathbf{A}\mathbf{B} + \mathbf{A}\mathbf{C}, \\ (\mathbf{A} + \mathbf{B})\mathbf{C} &= \mathbf{A}\mathbf{C} + \mathbf{B}\mathbf{C} \end{aligned}$$

Zur Erinnerung: I.A.: $\mathbf{A}\mathbf{B} \neq \mathbf{B}\mathbf{A}$: Matrixmultiplikation nicht kommutativ!

Transponierte einer Matrix

Definition: Transponierte einer Matrix \mathbf{A}

Notation: \mathbf{A}^T oder \mathbf{A}' .

Die Transponierte einer $(k \times n)$ -Matrix \mathbf{A} ist eine $(n \times k)$ -Matrix, die sich durch das Transponieren der einzelnen Zeilen ergibt, so dass (i, j) -te Element von \mathbf{A} zum (j, i) -ten Element von \mathbf{A}^T wird.

Die Berechnung kann auch durch Vertauschen von Zeilen und Spalten erfolgen:

Erste Spalte von \mathbf{A} wird erste Zeile von \mathbf{A}^T ,

Zweite Spalte von \mathbf{A} wird zweite Zeile von \mathbf{A}^T ,

Beispiel:

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix}^T = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \\ a_{13} & a_{23} \end{pmatrix},$$

$$\begin{pmatrix} a_{11} \\ a_{21} \end{pmatrix}^T = (a_{11} \quad a_{21}).$$

Rechenregeln mit transponierten Matrizen

Addition und Multiplikation mit einer Zahl

Gegeben sind zwei $(m \times n)$ -Matrizen \mathbf{A} und \mathbf{B} , sowie ein Skalar α

$$\begin{aligned} (\mathbf{A} + \mathbf{B})^T &= \mathbf{A}^T + \mathbf{B}^T \\ (\mathbf{A} - \mathbf{B})^T &= \mathbf{A}^T - \mathbf{B}^T \\ (\mathbf{A}^T)^T &= \mathbf{A} \\ (\alpha \mathbf{A})^T &= \alpha \mathbf{A}^T \end{aligned}$$

Es ist eine gute Übung, diese Regeln zu beweisen!

Matrixmultiplikation

Gegeben sind eine $(k \times m)$ -Matrix \mathbf{A} und eine $(m \times n)$ -Matrix \mathbf{B} . Es gilt

$$(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T \tag{1.5}$$

- *Beachte das Vertauschen der Reihenfolge!*

- **Notation:** $((\mathbf{AB})^T)_{ij}$ bezeichnet das (i, j) -te Element von $(\mathbf{AB})^T$. Analog zu bisheriger Notation.
- R-Befehl: $\mathfrak{t}(\mathbf{A})$.

Beweis von (1.5):

$$\begin{aligned}
 ((\mathbf{AB})^T)_{ij} &= (\mathbf{AB})_{ji} && \text{(Definition der Transponierten)} \\
 &= \sum_h \mathbf{A}_{jh} \cdot \mathbf{B}_{hi} && \text{(Definition der Matrixmultiplikation)} \\
 &= \sum_h (\mathbf{A}^T)_{hj} \cdot (\mathbf{B}^T)_{ih} && \text{(Definition der Transponierten, zweimal)} \\
 &= \sum_h (\mathbf{B}^T)_{ih} \cdot (\mathbf{A}^T)_{hj} && (a \cdot b = b \cdot a) \text{ f\u00fcr Skalare} \\
 &= (\mathbf{B}^T \mathbf{A}^T)_{ij} && \text{(Definition der Matrixmultiplikation.)}
 \end{aligned}$$

Deshalb gilt $(\mathbf{AB})^T = \mathbf{B}^T \mathbf{A}^T$. QED

Multiplikation von Vektoren

Spezialfall der Transposition einer Matrix

- Ist ein Vektor $\mathbf{x} \in \mathbb{R}^n$ als Spaltenvektor definiert, ergibt die Transposition von \mathbf{x} einen Zeilenvektor mit demselben n -Tupel

$$\mathbf{x}^T = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}^T := (x_1 \quad \cdots \quad x_n)$$

- H\u00e4ufig wird auch die Schreibweise \mathbf{x}' anstelle von \mathbf{x}^T verwendet.

Inneres Produkt

Das Skalarprodukt $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$ impliziert folgende Vektormultiplikationen

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^n x_i y_i$$

und falls \mathbf{x}, \mathbf{y} Spaltenvektoren

$$\langle \mathbf{x}, \mathbf{y} \rangle = (x_1 \quad \cdots \quad x_n) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \mathbf{x}^T \mathbf{y} = \mathbf{y}^T \mathbf{x}$$

- Eine weitere M\u00f6glichkeit ist das \u00e4u\u00dfere Produkt.

Äußeres Produkt (outer product)

Für zwei Spaltenvektoren \mathbf{x}, \mathbf{y} der Länge n erhält man für \mathbf{xy}^T eine $(n \times n)$ -Matrix (siehe Abschnitt 1.5.1)

$$\mathbf{xy}^T = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} (y_1 \ \cdots \ y_n) = \begin{pmatrix} x_1 y_1 & \cdots & x_1 y_n \\ \vdots & \ddots & \vdots \\ x_n y_1 & \cdots & x_n y_n \end{pmatrix}.$$

Vorsicht:

- In beiden Fällen können **nur Vektoren mit der gleichen Länge** miteinander **multipliziert** werden.
- Vektoren unterschiedlicher Länge können nicht miteinander multipliziert werden.

1.6. Wichtige spezielle Matrizen

Übersicht

- Quadratische Matrix
- Diagonalmatrix
- Symmetrische Matrix
- Obere/untere Dreiecksmatrix
- Idempotente Matrix

Quadratische Matrix

Eine $(n \times n)$ -Matrix ist eine quadratische Matrix. Die Zahl der Spalten und Zeilen ist gleich.

Diagonalmatrix

Eine quadratische $(n \times n)$ -Matrix \mathbf{A} ist eine Diagonalmatrix, wenn alle Nichtdiagonalelemente $a_{ij}, i \neq j, i, j = 1, \dots, n$ Null sind.

Symmetrische Matrix

Eine quadratische $(n \times n)$ -Matrix \mathbf{A} ist symmetrisch, wenn für alle $i, j = 1, \dots, n$ $a_{ij} = a_{ji}$, bzw.

$$\mathbf{A} = \mathbf{A}^T$$

gilt.

Obere Dreiecksmatrix

Eine quadratische Matrix \mathbf{A} ist eine obere Dreiecksmatrix, wenn für alle $i > j$, $i, j = 1, \dots, n$ gilt: $a_{ij} = 0$.

Untere Dreiecksmatrix

Eine quadratische Matrix \mathbf{A} ist eine untere Dreiecksmatrix, wenn für alle $i < j$, $i, j = 1, \dots, n$ gilt: $a_{ij} = 0$.

Idempotente Matrix

Eine quadratische Matrix \mathbf{A} heißt **idempotent**, wenn gilt

$$\mathbf{A}\mathbf{A} = \mathbf{A}.$$

Beispiele:

Diagonalmatrizen

$$\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix} \quad \text{oder} \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Symmetrische Matrizen

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix} \quad \text{oder} \quad \begin{pmatrix} 1 & 2 & 3 \\ 2 & 4 & 5 \\ 3 & 5 & 6 \end{pmatrix}$$

Obere Dreiecksmatrizen

$$\begin{pmatrix} a & c \\ 0 & b \end{pmatrix} \quad \text{oder} \quad \begin{pmatrix} 1 & 2 & 9 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$$

Untere Dreiecksmatrizen

$$\begin{pmatrix} a & 0 \\ c & d \end{pmatrix} \quad \text{oder} \quad \begin{pmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{pmatrix}$$

1.7. Maßzahlen von Matrizen

Übersicht

- Spur
- Rang
- Determinante

1.7.1. Spur einer Matrix

Definition

Die **Spur** (**trace**) einer quadratischen Matrix \mathbf{A} ist die Summe der Elemente a_{ii} auf der Diagonale

$$\text{tr}(\mathbf{A}) = \sum_{i=1}^n a_{ii}$$

Die Spur ist eine Abbildung von $\mathbb{R}^{n \times n}$ nach \mathbb{R} .

Beispiel:

$$\text{tr}(\mathbf{I}) = n, \quad \text{tr} \begin{pmatrix} 1 & 3 \\ a & b \end{pmatrix} = 1 + b$$

Rechenregeln

Gegeben sind $(n \times n)$ -Matrizen \mathbf{A} , \mathbf{B} und ein Skalar $\alpha \in \mathbb{R}$:

- $\text{tr}(\mathbf{A}) = \text{tr}(\mathbf{A}^T)$
- $\text{tr}(\alpha \mathbf{A}) = \alpha \text{tr}(\mathbf{A})$
- $\text{tr}(\mathbf{A} + \mathbf{B}) = \text{tr}(\mathbf{A}) + \text{tr}(\mathbf{B})$
- weitere Regeln in [Schmidt & Trenkler \(2006, Abschnitt 3.1\)](#).

Nachprüfen!

1.7.2. Rang einer Matrix

Definition

Der **Rang (rank)** $\text{rk}(\mathbf{A})$ ist eine Abbildung von $\mathbb{R}^{m \times n} \rightarrow \mathbb{N}$, die einer $(m \times n)$ -Matrix \mathbf{A} die maximale Zahl an Vektoren (entweder Zeilen- oder Spaltenvektoren) zuordnet, die linear unabhängig sind.

- Eine $(m \times n)$ -Matrix \mathbf{A} hat **vollen Rang (full rank)**, wenn der Rang der Matrix gleich der kleineren Dimension ist, also

$$\text{rk}(\mathbf{A}) = \begin{cases} m, & \text{falls } m \leq n \text{ und alle } m \text{ Zeilen linear unabhängig sind,} \\ n, & \text{falls } m \geq n \text{ und alle } n \text{ Spalten linear unabhängig sind.} \end{cases}$$

- Eine $(m \times n)$ -Matrix \mathbf{A} hat **vollen Spaltenrang**, wenn $\text{rk}(\mathbf{A}) = n$ ist.
- Eine $(m \times n)$ -Matrix \mathbf{A} hat **vollen Zeilenrang**, wenn $\text{rk}(\mathbf{A}) = m$ ist.

Bemerkungen

- Eine Matrix, die nicht vollen Rang hat, weist ein **Rangdefizit** auf.
- Der Rang ist kleiner als die Spaltenzahl k von \mathbf{X} , falls Spalten von \mathbf{X} linear abhängig sind. Dann
 - lässt sich eine Matrix \mathbf{X}' bilden, die aus k' linear unabhängigen Spalten von \mathbf{X} besteht, so dass $\text{rk}(\mathbf{X}) = k' < k$ und
 - für die Unterräume, siehe Definition in Abschnitt 1.9, $\delta(\mathbf{X}) = \delta(\mathbf{X}')$ gilt,
 - weist auch $\mathbf{X}^T \mathbf{X}$ ein Rangdefizit auf, da $\text{rk}(\mathbf{X}) = \text{rk}(\mathbf{X}^T \mathbf{X}) = k'$, und ist singulär. (Vgl. MLR.3 in Einführung in die Ökonometrie).
- **R-Code** : `rankMatrix()` in R-Paket `Matrix`

Rechenregeln

Gegeben seien $(m \times n)$ -Matrizen \mathbf{A} , \mathbf{B} :

- $0 \leq \text{rk}(\mathbf{A}) \leq \min(m, n)$
- $\text{rk}(\mathbf{A}) = \text{rk}(\mathbf{A}^T) = \text{rk}(\mathbf{A}^T \mathbf{A}) = \text{rk}(\mathbf{A} \mathbf{A}^T)$
- $\text{rk}(\mathbf{A} + \mathbf{B}) \leq \text{rk}(\mathbf{A}) + \text{rk}(\mathbf{B})$
- $\text{rk}(\mathbf{A} \mathbf{C}) \leq \min(\text{rk}(\mathbf{A}), \text{rk}(\mathbf{C}))$
- weitere Regeln in [Schmidt & Trenkler \(2006, Abschnitt 3.2\)](#).

1.7.3. Determinanten

Determinante

- Eine **Determinante** (**determinant**) ist eine Abbildung $\mathbb{R}^{n \times n} \rightarrow \mathbb{R}$, die einer quadratischen Matrix \mathbf{A} eine reelle Zahl zuordnet.
- Die Determinante hat eine wichtige Funktion bei der Bestimmung der Lösungen linearer Gleichungssysteme aber auch in der Geometrie. [Fischer \(2010, Abschnitt 3.1.1\)](#)
- Die Determinante wird mit $|\mathbf{A}|$ oder mit $\det(\mathbf{A})$ notiert.
- Die Berechnung einer Determinante kann rekursiv erfolgen. [Gentle \(2007, Abschnitt 3.1.5\)](#) oder [Schmidt & Trenkler \(2006, Abschnitt 3.3\)](#).
- Für $n \leq 3$ gibt es einfache Berechnungsformeln.

‡ **Geometrische Interpretation:** Der $(n \times 1)$ -Vektor definiert im n -dimensionalen Euklidischen Raum E^n ein n -dimensionales Parallelepipid (= Parallelogramm für $n = 2$), für das sich ein Volumen (für $n = 2$ eine Fläche) berechnen lässt.

Wird ein $(n \times 1)$ -Vektor \mathbf{x} von links mit der Matrix \mathbf{A} multipliziert, entspricht dies einer Abbildung von

$$E^n \longrightarrow E^n : \mathbf{x} \longrightarrow \mathbf{z} = \mathbf{A}\mathbf{x}.$$

Die Determinante $|\mathbf{A}|$ gibt an, um wie viel sich die Volumina, die jeweils durch \mathbf{x} und \mathbf{z} bestimmt werden, unterscheiden (Ein Beispiel für $n = 2$ findet sich in [Davidson & MacKinnon 2004, Section 12.2, pp. 511-512](#)).

Berechnung der Determinante für $n = 2, 3$

- (2×2) -Matrix

$$\mathbf{A} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \det \mathbf{A} = |\mathbf{A}| = ad - bc.$$

- (3×3) -Matrix (Sarrus' Regel)

$$\mathbf{A} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}$$

$$\det \mathbf{A} = |\mathbf{A}| = aei + bfg + cdh - gec - hfa - idb$$

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{matrix} a & b \\ d & e \\ g & h \end{matrix}$$

1.8. Matrixinversion

Übersicht

- Definition einer inversen Matrix
- Berechnung für (2×2) -Matrix
- Existenz
- Rechenregeln

Die Inverse einer Matrix

- ist nur für quadratische Matrizen definiert.
- ergibt sich bei der Lösung eines linearen Gleichungssystems.
- spielt eine zentrale Rolle in der Matrixalgebra.

Inverse einer Matrix

Eine quadratische Matrix \mathbf{A} heißt **invertierbar**, wenn eine quadratische Matrix \mathbf{B} existiert, so dass gilt:

$$\mathbf{AB} = \mathbf{BA} = \mathbf{I}.$$

Die Matrix \mathbf{B} wird als **Inverse** \mathbf{A}^{-1} bezeichnet.

- Die Inverse ist eine Abbildung $\mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$.
- Eine nichtinvertierbare Matrix \mathbf{A} wird als **singulär** bezeichnet.
- Eine invertierbare Matrix \mathbf{A} wird als **regulär** oder als **nicht singulär** bezeichnet.

Berechnung der Inversen für $n = 2, 3$

- (2×2) -Matrix

$$\mathbf{A}^{-1} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{|\mathbf{A}|} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

- (3×3) -Matrix

$$\mathbf{A}^{-1} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}^{-1} = \frac{1}{|\mathbf{A}|} \begin{pmatrix} ei - fh & ch - bi & bf - ce \\ fg - di & ai - cg & cd - af \\ dh - eg & bg - ah & ae - bd \end{pmatrix}$$

- Für $n > 3$ sind komplizierte Verfahren notwendig, die man am Bestem dem Computer überlässt.

Existenz der Inversen

- **Existenz der Inversen:** Die Inverse \mathbf{A}^{-1} existiert dann und nur dann, wenn die Determinante von \mathbf{A} von Null verschieden ist, $|\mathbf{A}| \neq 0$. Dies gilt für alle n !
- **Wichtig:** Ist bei Berechnungen die Determinante nahe Null, kann es zu großen numerischen Ungenauigkeiten kommen. Deshalb wird beim Programmieren die Verwendung der Inversen vermieden, wenn das möglich ist.
- **Existiert die Inverse, ist ein lineares Gleichungssystem**

$$\mathbf{Ax} = \mathbf{b}$$

eindeutig lösbar:

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$$

- Für nichtquadratische und nichtinvertierbare Matrizen gibt es **verallgemeinerte Inverse**.

Rechenregeln für Inverse

\mathbf{A} sei regulär.

- $(\mathbf{A}^{-1})^{-1} = \mathbf{A}$
- $(\mathbf{A}^T)^{-1} = (\mathbf{A}^{-1})^T$
- \mathbf{B} sei regulär. Dann ist \mathbf{AB} regulär und $(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$.
- Ist \mathbf{A} eine Diagonalmatrix, dann gilt: $\mathbf{A}^{-1} = (1/a_{ii})$.

Rechenregeln für Determinanten

Gegeben sind $(n \times n)$ -Matrizen \mathbf{A} , \mathbf{B} und ein Skalar $\lambda \in \mathbb{R}$:

- $|\mathbf{A}| = 0 \iff \text{rk}(\mathbf{A}) < n \iff \mathbf{A}$ ist singulär
- $|\mathbf{A}| \neq 0 \iff \text{rk}(\mathbf{A}) = n \iff \mathbf{A}$ ist regulär
- $|\lambda\mathbf{A}| = \lambda^n |\mathbf{A}|$
- $|\mathbf{AB}| = |\mathbf{A}||\mathbf{B}|$
- $|\mathbf{A}^T| = |\mathbf{A}|$

- $|\mathbf{A}| = \prod_{i=1}^n a_{ii}$, wenn \mathbf{A} Diagonal- oder Dreiecksmatrix ist.
- weitere Regeln z. B. in [Schmidt & Trenkler \(2006, Abschnitt 3.3\)](#).

1.9. Euklidische Unterräume

Übersicht

- Basisvektoren im E^n
- Euklidische Unterräume
- Spaltenraum einer Matrix
- Orthogonales Komplement

Basisvektoren im E^n

Definition

n verschiedene $(n \times 1)$ -Vektoren sind **Basisvektoren**, wenn kein Basisvektor sich als Linearkombination der anderen $(n - 1)$ Basisvektoren darstellen lässt. D. h., die Basisvektoren sind linear unabhängig.

Bemerkungen

- Jedes Element im Euklidischen Raum E^n kann als **Linearkombination** von n **Basisvektoren** dargestellt werden.
- Man sagt dann: **Die n Basisvektoren spannen E^n auf**, d. h. bilden einen Euklidischen Raum E^n . Bezeichnet man die n Basisvektoren mit \mathbf{x}_i , dann ist die Menge aller Vektoren in E^n gegeben durch

$$\left\{ \mathbf{z} \in E^n \mid \mathbf{z} = \sum_{i=1}^n b_i \mathbf{x}_i, b_i \in \mathbb{R}, i = 1, \dots, n \right\}.$$

Euklidische Unterräume

Definition

Reduziert man die Zahl der Basisvektoren auf $k < n$, kann nur noch eine Teilmenge der Vektoren in E^n dargestellt werden. Eine solche Teilmenge bildet einen **Euklidischen Unterraum**.

Notation und Sprechweisen

- Den Unterraum, der von k Basisvektoren $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k\}$ aufgespannt wird, bezeichnen wir mit $\delta(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)$, bzw. $\delta(\mathbf{X})$, falls alle Basisvektoren in der Matrix $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k)$ zusammengefasst werden.

Spaltenraum einer Matrix

- Die Menge der im Unterraum enthaltenen Vektoren \mathbf{z} , d. h. alle Linearkombinationen der Spalten der $(n \times k)$ -Matrix \mathbf{X} , lässt sich beschreiben als

$$\delta(\mathbf{X}) = \delta(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) := \left\{ \mathbf{z} \in E^n \mid \mathbf{z} = \sum_{i=1}^k b_i \mathbf{x}_i, b_i \in \mathbb{R} \right\}. \quad (1.6)$$

- Man sagt, dass der Unterraum $\delta(\mathbf{X})$ dem **Spaltenraum der Matrix \mathbf{X}** entspricht.

Orthogonales Komplement

- Das **orthogonale Komplement** zu dem Unterraum $\delta(\mathbf{X})$ ist ein weiterer Unterraum in E^n , für den gilt:

$$\begin{aligned} \delta^\perp(\mathbf{X}) &= \delta^\perp(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k) \\ &:= \left\{ \mathbf{w} \in E^n \mid \langle \mathbf{w}, \mathbf{z} \rangle = \mathbf{w}^T \mathbf{z} = 0 \text{ für alle } \mathbf{z} \in \delta(\mathbf{X}) \right\}. \end{aligned} \quad (1.7)$$

Frage: Sei $\dim \delta(\mathbf{X}) = k$ die Dimension von $\delta(\mathbf{X})$. Wie groß ist dann $\dim \delta^\perp(\mathbf{X})$?

1.10. Matrizen und lineare Abbildungen

Übersicht

- Abbildung zwischen zwei Vektorräumen
- Lineare Abbildung zwischen zwei Vektorräumen
- Kern und Bild einer linearen Abbildung

Gegeben seien zwei Euklidische Vektorräume, die unterschiedliche Dimensionen n und m aufweisen können. Es sei $\mathbf{x} \in \mathbb{R}^n$ und $\mathbf{y} \in \mathbb{R}^m$. Die **Abbildung**

$$\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad \begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},$$

kurz

$$\mathbf{F}(\mathbf{x}) = \mathbf{y} = \mathbf{A}\mathbf{x}$$

weist jedem Punkt \mathbf{x} im n -dimensionalen Euklidischen Raum \mathbb{R}^n einen Punkt \mathbf{y} im m -dimensionalen Euklidischen Raum \mathbb{R}^m zu.

Lineare Abbildung

Eine Abbildung

$$\mathbf{F} : \mathbb{R}^n \rightarrow \mathbb{R}^m, \quad \mathbf{F}(\mathbf{x}) = \mathbf{y} = \mathbf{A}\mathbf{x}$$

heißt **linear**, wenn folgende Eigenschaften gelten:

1. $\mathbf{F}(\mathbf{x} + \mathbf{z}) = \mathbf{F}(\mathbf{x}) + \mathbf{F}(\mathbf{z})$
2. $\mathbf{F}(\lambda\mathbf{x}) = \lambda\mathbf{F}(\mathbf{x})$

für alle $\mathbf{x}, \mathbf{z} \in \mathbb{R}^n$, $\lambda \in \mathbb{R}$.

Kern und Bild einer linearen Abbildung

Es sei $\mathcal{V} \in \mathbb{R}^n$ und $\mathcal{W} \in \mathbb{R}^m$. Für die Abbildung $\mathbf{F} : \mathcal{V} \rightarrow \mathcal{W}$ bezeichnet

- Im $\mathbf{F} := \mathbf{F}(\mathcal{V})$ das **Bild (image)** dieser Abbildung,

- $\text{Ker } \mathbf{F} := \mathbf{F}^{-1}(\mathbf{0})$ den **Kern** (**kernel**) dieser Abbildung.
- Der Kern kann bei Existenz der Inversen durch $\mathbf{F}^{-1}(\mathbf{y}) = \mathbf{A}^{-1}\mathbf{y}$ mit $\mathbf{y} = \mathbf{0}$ bestimmt werden.
- Der Kern bestimmt die Menge aller $\mathbf{x} \in \mathcal{V}$, deren Bild gerade der Ursprung in \mathcal{W} ist.

1.11. Matrixdarstellung linearer Gleichungssysteme

Übersicht

- Definition geeigneter Matrizen
- Gleichungssystem in Matrixform
- Eindeutige Lösung

Betrachte ein typisches System linearer Gleichungen:

$$\begin{array}{rcccc} a_{11}x_1 + & \cdots & + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + & \cdots & + a_{2n}x_n & = & b_2 \\ \vdots & \vdots & & \vdots & \\ a_{k1}x_1 + & \cdots & + a_{kn}x_n & = & b_k. \end{array}$$

Das lineare Gleichungssystem kann kompakter mit Matrizen dargestellt werden.

Definition geeigneter Matrizen: $(k \times n)$ -Koeffizientenmatrix \mathbf{A} , $(n \times 1)$ -Variablenvektor \mathbf{x} und $(k \times 1)$ -Parametervektor \mathbf{b}

$$\mathbf{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{k1} & \cdots & a_{kn} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \quad \text{und} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}.$$

Gleichungssystem in Matrixform

Das Gleichungssystem lautet nun:

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & a_{ij} & \vdots \\ a_{k1} & \cdots & a_{kn} \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} b_1 \\ \vdots \\ b_k \end{pmatrix}.$$

In kompakter Form

$$\mathbf{Ax} = \mathbf{b}.$$

Das Matrixprodukt \mathbf{Ax} liefert einen $(k \times 1)$ -Vektor, der dem $(k \times 1)$ -Parametervektor \mathbf{b} entspricht, wenn \mathbf{x} eine Lösung des Gleichungssystems darstellt.

Eindeutige Lösung

Ist \mathbf{A} regulär, d. h. invertierbar, dann existiert eine eindeutige Lösung

$$\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}.$$

1.12. (Semi-)definite Matrizen

Übersicht

- Quadratische Form
- Positiv definite und positiv semidefinite Matrizen
- Negativ definite und negativ semidefinite Matrizen
- Indefinite Matrizen

Quadratische Form

$\mathbf{x}^T \mathbf{Ax} = \sum_{i=1}^k \sum_{j=1}^k x_i x_j A_{ij}$ ist eine quadratische Form. Das Ergebnis ist ein Skalar.

Positiv definite und semidefinite Matrizen

- Eine $(k \times k)$ -Matrix \mathbf{A} heißt **positiv definit**, wenn für beliebige $(k \times 1)$ -Vektoren \mathbf{x} mit positiver Norm gilt:

$$\mathbf{x}^T \mathbf{Ax} > 0.$$

- Eine $(k \times k)$ -Matrix \mathbf{A} heißt **positiv semidefinit**, wenn für beliebige $(k \times 1)$ -Vektoren

\mathbf{x} mit positiver Norm gilt:

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0.$$

Negativ definite und semidefinite Matrizen

- Eine $(k \times k)$ -Matrix \mathbf{A} heißt **negativ definit**, wenn für beliebige $(k \times 1)$ -Vektoren \mathbf{x} mit positiver Norm gilt:

$$\mathbf{x}^T \mathbf{A} \mathbf{x} < 0.$$

- Eine $(k \times k)$ -Matrix \mathbf{A} heißt **negativ semidefinit**, wenn für beliebige $(k \times 1)$ -Vektoren \mathbf{x} mit positiver Norm gilt:

$$\mathbf{x}^T \mathbf{A} \mathbf{x} \leq 0.$$

Indefinite Matrizen

Matrizen, die weder positiv, noch negativ (semi-)definit sind, heißen **indefinit**.

- Ist $\mathbf{A} = \mathbf{B}^T \mathbf{B}$, dann ist \mathbf{A} immer positiv semidefinit, da

$$\mathbf{x}^T \mathbf{B}^T \mathbf{B} \mathbf{x} = (\mathbf{B} \mathbf{x})^T (\mathbf{B} \mathbf{x}) = \|\mathbf{B} \mathbf{x}\|^2 \geq 0. \quad (1.8)$$

Wenn \mathbf{B} vollen Rang hat, ist \mathbf{A} positiv definit. Warum?

- Die Diagonalelemente einer positiv definiten Matrix sind positiv. Außerdem existiert für jede positiv definite Matrix \mathbf{A} eine Matrix \mathbf{B} , so dass gilt $\mathbf{A} = \mathbf{B}^T \mathbf{B}$. Dabei ist \mathbf{B} nicht eindeutig.
- Eine Matrix \mathbf{A} heißt **negativ (semi-)definit**, wenn $-\mathbf{A}$ positiv (semi-)definit ist.

Beispiel:

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

ist positiv definit. Denn für jeden Vektor $\mathbf{z} = \begin{pmatrix} z_0 \\ z_1 \end{pmatrix}$ mit $\|\mathbf{z}\| > 0$ gilt

$$\begin{pmatrix} z_0 & z_1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} = z_0^2 + z_1^2 > 0.$$

Beispiel: Die Matrix $\mathbf{M} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ ist indefinit, da sie weder positiv, noch negativ semidefinit ist. Denn für $\mathbf{z} = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}$ erhält man

$$\begin{pmatrix} z_1 & z_2 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = \begin{pmatrix} z_2 & z_1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = 2z_1 z_2.$$

Je nach Wahl von z_1 und z_2 ist das Ergebnis positiv, null oder negativ.

1.13. Rechenregeln zum Ableiten von vektorwertigen Funktionen

Übersicht

- Erste partielle Ableitungen von Skalarprodukten
- Erste partielle Ableitungen von Linearkombinationen
- Erste partielle Ableitungen für quadratische Formen
- ‡ Jacobi-Matrix

Erste partielle Ableitungen von Skalarprodukten

Gegeben sind die $(n \times 1)$ -Spaltenvektoren \mathbf{v} und \mathbf{w} . Für die erste partielle Ableitung des Skalarprodukts $z = \langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^T \mathbf{w} = \mathbf{w}^T \mathbf{v} = \sum_{i=1}^n v_i w_i$ nach w_i gilt dann $\partial z / \partial w_i = v_i$. Sammelt man alle ersten partiellen Ableitungen nach \mathbf{w} in einem Spaltenvektor

$$\frac{\partial z}{\partial \mathbf{w}} = \begin{pmatrix} \frac{\partial z}{\partial w_1} \\ \frac{\partial z}{\partial w_2} \\ \vdots \\ \frac{\partial z}{\partial w_n} \end{pmatrix},$$

ergibt sich

$$\frac{\partial z}{\partial \mathbf{w}} = \mathbf{v}.$$

Erste partielle Ableitungen von Linearkombination

Für $\mathbf{z} = \mathbf{A}\mathbf{w}$ mit

$$\mathbf{z} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

erhält man die ersten partiellen Ableitungen

$$\frac{\partial \mathbf{z}}{\partial \mathbf{w}^T} = \mathbf{A}$$

Erste partielle Ableitungen für quadratische Formen

Für die quadratische Form $v = \mathbf{w}^T \mathbf{A} \mathbf{w}$

$$v = \begin{pmatrix} w_1 & w_2 & \cdots & w_n \end{pmatrix} \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}$$

erhält man die ersten partiellen Ableitungen

$$\frac{\partial v}{\partial \mathbf{w}} = (\mathbf{A} + \mathbf{A}^T) \mathbf{w}.$$

‡ Jacobi-Matrix (Jacobian matrix)

Gegeben sei für $\mathbf{x} \in \mathbb{R}^n$ eine vektorwertige Funktion

$$f : \mathbb{R}^n \longrightarrow \mathbb{R}^m : \mathbf{x} \longrightarrow \mathbf{g}(\mathbf{x}) \equiv \begin{pmatrix} g_1(\mathbf{x}) \\ \dots \\ g_m(\mathbf{x}) \end{pmatrix}.$$

Die $(m \times n)$ -Matrix

$$\mathbf{J}(\mathbf{x}) \equiv \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}^T} \equiv \begin{pmatrix} \frac{\partial g_1(\mathbf{x})}{\partial x_1} & \frac{\partial g_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_1(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial g_m(\mathbf{x})}{\partial x_1} & \frac{\partial g_m(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial g_m(\mathbf{x})}{\partial x_n} \end{pmatrix} \quad (1.9)$$

der partiellen Ableitungen erster Ordnung wird als **Jacobi-Matrix** bezeichnet. Ist die Jacobi-Matrix quadratisch, existiert die Determinante der Jacobi-Matrix (häufig als **Jacobi-Determinante** bezeichnet):

$$|\mathbf{J}(\mathbf{x})| = \left| \frac{\partial \mathbf{g}(\mathbf{x})}{\partial \mathbf{x}^T} \right|. \quad (1.10)$$

1.14. Partitionierte Matrizen

Übersicht

- Addition, Subtraktion und Matrixmultiplikation
- Inversion einer partitionierten Matrix

Partitionierte Matrix

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

wobei die Teilmatrizen \mathbf{A}_{ij} Dimension $(m_i \times n_j)$ haben und $m_1 + m_2 = m$, $n_1 + n_2 = n$ gilt.

Rechenregeln

Auf korrekte Dimensionen der Matrizen und Teilmatrizen achten!

- $\mathbf{A}^T = \begin{pmatrix} \mathbf{A}_{11}^T & \mathbf{A}_{21}^T \\ \mathbf{A}_{12}^T & \mathbf{A}_{22}^T \end{pmatrix}$
- Addition: ersetze in Standardaddition Elemente durch Teilmatrizen.
- Matrixmultiplikation: ersetze in Elemente durch entsprechende Teilmatrizen

$$\mathbf{AB} = \begin{pmatrix} \mathbf{A}_{11}\mathbf{B}_{11} + \mathbf{A}_{12}\mathbf{B}_{21} & \mathbf{A}_{11}\mathbf{B}_{12} + \mathbf{A}_{12}\mathbf{B}_{22} \\ \mathbf{A}_{21}\mathbf{B}_{11} + \mathbf{A}_{22}\mathbf{B}_{21} & \mathbf{A}_{21}\mathbf{B}_{12} + \mathbf{A}_{22}\mathbf{B}_{22} \end{pmatrix},$$

Inversion einer partitionierten Matrix

Die Inverse einer partitionierten Matrix lässt sich folgendermaßen berechnen

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}\mathbf{W}\mathbf{C}\mathbf{A}^{-1} & -\mathbf{A}^{-1}\mathbf{B}\mathbf{W} \\ -\mathbf{W}\mathbf{C}\mathbf{A}^{-1} & \mathbf{W} \end{pmatrix}$$

mit $\mathbf{W} = (\mathbf{D} - \mathbf{C}\mathbf{A}^{-1}\mathbf{B})^{-1}$.

2. Grundlagen der Wahrscheinlichkeitstheorie

Übersicht

- Wichtige Grundbegriffe
- Wozu brauchen wir Wahrscheinlichkeitstheorie
- Zufallsvariablen
- Verteilung- und Dichtefunktionen (uni- und multivariat)
- Bedingte Wahrscheinlichkeiten
- Erwartungswerte und Momente
- Bedingte Erwartungswerte und Momente
- Wichtige Wahrscheinlichkeitsverteilungen

Literaturhinweise

- [Davidson & MacKinnon \(2004, Section 1.2\)](#): knapper Überblick über die Grundlagen der Wahrscheinlichkeitstheorie.
- [Casella & Berger \(2002\)](#): sehr ausführliche, formale Einführung in die Wahrscheinlichkeitstheorie.
- [Fahrmeir et al. \(2016\)](#): einfache Einführung in die Statistik.
- [Steland \(2013\)](#): gut geschriebene, kurz gehaltene, technisch präzise Einführung in die Statistik. (Aus dem Uninetz als Volltext verfügbar [hier](#).)

2.1. Wichtige Grundbegriffe

Übersicht

- Grundgesamtheit bzw. Population
- Stichprobe

Wichtige Grundbegriffe

Definition: Grundgesamtheit bzw. Population (population)

“Menge aller statistischen Einheiten, über die man Aussagen gewinnen will”.

Die Grundgesamtheit ([Fahrmeier et al. 2004](#), Abschnitt 1.3.1, S. 14)

- hängt von der interessierenden Fragestellung ab,
- kann endlich (Anteile einer Produktion mit Qualitätsmängel), unendlich (Menge aller möglichen Verspätungen bei der Bahn) oder hypothetisch (Menge aller potentiellen Käufer) sein.

Die Grundgesamtheit

- kann prinzipiell beobachtbar (alle Studierende der UR, Menge an biologisch produziertem Getreide in einer Region innerhalb eines Jahres) oder
- unbeobachtbar sein (z. B. Auswirkung einer Maßnahme für ein *einzelnes* Individuum)

Definition: Stichprobe

Eine Stichprobe ist typischerweise eine Teilmenge der Grundgesamtheit, die beobachtet werden kann oder bereits wurde und zur Analyse der Grundgesamtheit dienen kann.

Beispiele:

- Teilnehmer einer Vorlesung
- 1 kg Getreide pro 100 zufällig ausgewählten Feldern innerhalb einer Region
- Teilnehmer am sozio-ökonomischen Panel

2.2. Wozu brauchen wir Wahrscheinlichkeitstheorie?

Übersicht

- Illustrierende Aufgabe
- Induktive, deskriptive, explorative Statistik
- Induktive Statistik und Wahrscheinlichkeitstheorie

Wozu brauchen wir Wahrscheinlichkeitstheorie?

Aufgabe zur Geschlechterverteilung

Wie ist die Geschlechterverteilung der Studierenden zu Beginn des VWL/IVWL-Masterstudiums an der Universität Regensburg?

- **Grundgesamtheit:** Alle Studierenden, die dieses Semester einen VWL/IVWL-Master beginnen.
- **Stichprobe:** Alle Studierenden, die dieses Semester den VWL oder IVWL-Master beginnen und in diesem Hörsaal sitzen.

Induktive versus deskriptive versus explorative Statistik

Aussagen über

- Stichprobe/Daten:
 - Beschreibung der wichtigsten Kennzahlen der Daten: **deskriptive Statistik**
 - Suche nachdem, was die Daten über formale Modelle oder Hypothesentests noch preisgeben könnten: **explorative Statistik**
- Grundgesamtheit: **induktive Statistik**

Welche Aussagen sind über Grundgesamtheit möglich?

Inwieweit lassen sich auf Basis der Information in dieser Stichprobe Aussagen über die Geschlechteranteile in der Grundgesamtheit machen?

Mögliche Antworten ohne Wahrscheinlichkeitstheorie

Auf Basis der Stichprobe sind Intervallaussagen über den Anteil von Studentinnen möglich. Diese werden jedoch umso ungenauer, je größer die Grundgesamtheit im Vergleich zur Stichprobe ist.

Mögliche Antworten mit Wahrscheinlichkeitstheorie

- Punktprognosen
- Intervallprognosen mit kürzeren Intervallen und Überdeckungswahrscheinlichkeiten
- erfordern immer zusätzliche Annahmen. **Detailliertere Aussagen** als Aussagen über den möglichen Bereich des Geschlechtsanteils in der Grundgesamtheit **erfordern zusätzliche Annahmen!**

Beispiele:

- Das Geschlechterverhältnis in der Grundgesamtheit entspricht dem in der Stichprobe.
- Es liegt eine Zufallsstichprobe vor.

Fortsetzung der Aufgabe zur Geschlechterverteilung**Antworten ohne Wahrscheinlichkeitstheorie**

Folgende Tabelle ermöglicht Intervallaussagen ohne Wahrscheinlichkeitstheorie, nachdem die aktuellen Stichprobendaten ergänzt worden sind.

Teilgesamtheit	Gesamtzahl	Anzahl weiblich	Anteil weiblich		
			möglicher Bereich	Punktprognose auf Basis der Stichprobe	korrekt
Stichprobe: anwesende Kurs- teilnehmerInnen Methoden der Ökonometrie, die dieses Semester ein VWL/IVWL- Master beginnen					
anwesende + 1 fehlende Kurs- teilnehmerInnen Met. d. Ökonometrie, die dieses Semester ein VWL/IVWL- Master beginnen					
alle Studierende, die dieses Semester ein VWL/IVWL- Master an der WiWi beginnen					

Antworten mit Wahrscheinlichkeitstheorie
erfordern noch etwas Geduld und das Studium der W'Theorie!

2.3. Wahrscheinlichkeitsraum

Übersicht

- Ergebnismenge
- Ereignis
- Elementarereignis
- Sigma-Algebra
- Wahrscheinlichkeitsfunktion
- Wahrscheinlichkeitsraum
- Rechenregel für Wahrscheinlichkeiten

Ergebnismenge (sample space, outcome space)

Definition Ergebnismenge

Die Ergebnismenge Ω enthält alle möglichen Ergebnisse (outcomes) eines Zufallsexperiments.

Die Menge kann **abzählbar** viele oder **überabzählbar** viele Ergebnisse enthalten.

Beispiele:

- Geschlecht eines Studierenden: $\Omega = \{\text{weiblich, männlich}\}$
- 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 (genauer: überabzählbar) viele Ergebnisse vor, dann bezeichnet man ein einzelnes davon häufig mit ω .

Ereignis

Definitionen

- Tritt ein bestimmtes Ergebnis ein, wird dies als **Ereignis (event)** 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.
- Die Ergebnismenge Ω ist ein **sicheres Ereignis**.
- Das **komplementäre Ereignis** A^c zu Ereignis A enthält alle Ereignisse, die in der Ergebnismenge Ω , aber nicht in A sind.

Beispiele:

- Urne: Mögliche Ereignisse sind z. B. {gelb, rot} oder {rot, blau, grün}. Komplementärereignis zu Ereignis $A = \{\text{gelb, rot}\}$ ist $A^c = \{\text{blau, grün}\}$.
- 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 (genauer: überabzählbar) 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 $(a_i \leq b_i)$.

Sigma-Algebra

Vorbemerkungen: Betrachten wir unser Beispiel mit den 4 Kugeln in verschiedenen Farben. Um das Beispiel noch allgemeiner zu machen, bezeichnen wir $\omega_1 = \text{gelb}$, $\omega_2 = \text{rot}$, $\omega_3 = \text{blau}$, $\omega_4 = \text{grün}$: $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4\}$. Nehmen wir nun an, dass wir insbesondere daran interessiert sind, ob bei *einem* Zug folgende Ereignisse eintreten:

$$\mathcal{C} = \{\{\omega_1\}, \{\omega_1, \omega_3, \omega_4\}\},$$

die in der Menge von Teilmengen \mathcal{C} zusammengefasst werden. Wenn wir nun diese Kollektion von Teilmengen \mathcal{C} genauer betrachten, fällt auf, dass zwar das Elementarereignis $\{\omega_1\}$ eintreten kann, aber was machen wir, wenn es *nicht* eintritt. Dann muss ja zwangsläufig das Ereignis $\{\omega_2, \omega_3, \omega_4\}$ eintreten, das aber *nicht* in der Sammlung \mathcal{C} enthalten ist. Das bedeutet, dass wir diesem Ereignis dann auch keine Wahrscheinlichkeit zuordnen können. Da dies keinen Sinn macht, müssen wir die Menge \mathcal{C} mindestens um das Ereignis $\{\omega_2, \omega_3, \omega_4\}$ erweitern. Daraus folgt, dass eine Kollektion von Teilmengen, für die wir jeweils Wahrscheinlichkeiten definieren möchten, bestimmte Eigenschaften aufweisen muss. So muss zumindest immer das Komplement eines Ereignisses in der Kollektion von Teilmengen enthalten sein. Man kann sich auch überlegen, dass beliebige Vereinigungsmengen von Teilmengen ebenfalls in der Kollektion enthalten sein müssen. Erfüllt eine Kollektion von Teilmengen diese Anforderungen, dann wird sie als Sigma-Algebra bezeichnet.

Anmerkung: Eine σ -Algebra ist eine Menge von Ereignissen (Teilmengen), die bezüglich aller enthaltenen Ereignisse die Zuweisung von Wahrscheinlichkeiten zulässt. Für Interessierte die Definition:

‡ Definition einer Sigma-Algebra

Eine Menge von Teilmengen von Ω wird als **Sigma-Algebra** bzw. als σ -Algebra (σ -algebra, σ -field) bezeichnet, wenn für diese Menge von Teilmengen folgende Eigenschaften gelten. Dabei wird eine σ -Algebra häufig mit \mathcal{F} bezeichnet:

1. $\emptyset \in \mathcal{F}$
2. Wenn $A \in \mathcal{F}$, dann $A^c \in \mathcal{F}$
3. Wenn $A_1, A_2, \dots \in \mathcal{F}$, dann $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$

Anmerkung: Im Fall endlich vieler Elementarereignisse ist die σ -Algebra mit der Potenzmenge identisch. Im Fall unendlich vieler Elementarereignisse, beispielsweise im Fall der möglichen Intervalle reeller Zahlen ist die σ -Algebra kleiner als die Potenzmenge. Genau für diesen Fall hat man dieses Konzept entwickelt, da die Potenzmenge "zu groß" sein würde.

Sigma-Algebra und Wahrscheinlichkeitsfunktion

Wahrscheinlichkeitsfunktion

Es sei eine Menge Ω und eine σ -Algebra \mathcal{F} gegeben. Dann ist eine Wahrscheinlichkeitsfunktion P eine Funktion mit Definitionsmenge \mathcal{F} , die folgende Bedingungen erfüllt:

1. $P(A) \geq 0$ für alle $A \in \mathcal{F}$
2. $P(\Omega) = 1$, $P(\emptyset) = 0$.
3. Wenn A_1, A_2, \dots paarweise disjunkt sind, dann $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$

Die Wahrscheinlichkeitsfunktion weist jedem möglichen Ereignis in der σ -Algebra eine Wahrscheinlichkeit zu.

Mehr zur σ -Algebra findet sich in z. B. [Steland \(2010, Abschnitt 2.1.3\)](#) oder [Eine etwas komprimierte Einführung in die Wahrscheinlichkeitstheorie](#).

Man sieht, dass die Definition einer Wahrscheinlichkeitsfunktion nur bezüglich eines Ergebnisraumes Ω und einer dazu passenden σ -Algebra möglich ist. Streng genommen müsste man also zu einer Wahrscheinlichkeitsfunktion P immer dazu sagen, zu welchem Ω und \mathcal{F} sie gehört. Dann erhält man einen

Wahrscheinlichkeitsraum

Das Tripel $(\Omega, \mathcal{F}, \mathcal{P})$ wird **Wahrscheinlichkeitsraum** genannt.

Wenn keine Unklarheiten entstehen, wird häufig auf die Angabe des Wahrscheinlichkeitsraumes verzichtet. Das machen wir hier auch so.

Rechenregel für Wahrscheinlichkeiten

Es seien $A, B \in \mathcal{F}$. Dann gilt

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (2.1)$$

2.4. Zufallsvariablen

Übersicht

- Definition und Beispiele einer Zufallsvariablen
- Realisation einer Zufallsvariablen
- Schreibweisen
- Diskrete und stetige Zufallsvariablen
- Wahrscheinlichkeitsraum von Zufallsvariablen

Zufallsvariablen

Definition

Eine **reelle Zufallsvariable** X ist eine **Funktion** von einer Ergebnismenge Ω nach \mathbb{R} ,

die jedem Elementarereignis $\omega \in \Omega$ eine Zahl $X(\omega)$ zuordnet. Für $X(\omega) \in \mathbb{R}$

$$X : \Omega \mapsto \mathbb{R} : \omega \mapsto X(\omega).$$

Jedes Ereignis $A \in \mathcal{F}$ kann auf eine Menge $\{X(\omega) \in \mathbb{R} | \omega \in A \in \mathcal{F}\}$ abgebildet werden.

Beispiele:

- Studierende: $X(\omega = \text{weiblich}) = 0$, $X(\omega = \text{männlich}) = 1$.
- Urnenbeispiel: $X(\omega_1) = 0$, $X(\omega_2) = 3$, $X(\omega_3) = 17$, $X(\omega_4) = 20$.
- Haushaltseinkommen: $X(\cdot) \geq 0$

Realisation einer Zufallsvariable

Ausprägung x einer Zufallsvariable $X(\omega)$, die in einer Stichprobe beobachtet wurde, so dass $x = X(\omega)$.

Wichtig: Eine Zufallsvariable als solches kann nicht beobachtet werden, da sie eine Funktion aller möglichen Ergebnisse ist.

Schreibweisen von Zufallsvariablen

- In diesem Abschnitt schreiben wir im Folgenden X anstelle von $X(\omega)$. Realisationen oder mögliche Ausprägungen werden mit x bezeichnet.
- In der ökonomischen Literatur wird mangels genügend Symbolen im Allgemeinen nicht zwischen einer Zufallsvariable X und einer möglichen Realisation x unterschieden, sondern beides mit dem gleichen Symbol bezeichnet (Beispiele: abhängige Variable y_t , Fehlerterm u_t im linearen Regressionsmodell)

Arten von Zufallsvariablen

- **Diskrete Zufallsvariablen:** Sie können endlich viele (z. B. binäre Zufallsvariablen) oder unendlich, aber abzählbar viele Werte (z. B. Zählraten $\Omega = \mathbb{N}$) annehmen

$$\sum_{i=1}^{\infty} P(X(\omega) = x_i) = \sum_{i=1}^{\infty} P(X = x_i) = 1$$

- **Stetige Zufallsvariablen:**

- Beispiele: $X \in \mathbb{R}$, $X \in [0, \infty)$.
- Beachte: $P(X = x) = 0$. Wieso?
- Stattdessen betrachtet man Wahrscheinlichkeiten für Intervalle, z. B. $P(X \leq x)$,

$P(a < X \leq b), P(0 < X) \Rightarrow$ kumulative Wahrscheinlichkeitsverteilung.

Wahrscheinlichkeitsraum von Zufallsvariablen

Eine Wahrscheinlichkeitsfunktion für die Zufallsvariable $X(\omega)$ auf Ω kann nur bestimmt werden, wenn

1. eine neue Menge an Elementarereignissen Ω' , die der Bildmenge der Zufallsvariablen für die Elementarereignisse entspricht, vorliegt und
2. eine neue σ -Algebra \mathcal{F}' , die sich aus \mathcal{F} gewinnen lässt.

‡ Details

Die **Wahrscheinlichkeitsfunktion für die Zufallsvariable** hat als Argument $A \in \mathcal{F}'$ im Fall

- diskreter Zufallsvariablen Zahlen,

$$P(X = x) = P(X(\omega) = x) = P(\{\omega \in \Omega | X(\omega) = x\}).$$

- stetiger Zufallsvariablen Intervalle von (reellen) Zahlen mit

$$P(X \in A) = P(X(\omega) \in A) = P(\{\omega \in \Omega | X(\omega) \in A\})$$

Urnenbeispiel:

- $\Omega' = \{X(\omega_1), X(\omega_2), X(\omega_3), X(\omega_4)\} = \{0, 3, 17, 20\}$
- Mögliche σ -Algebra: $\mathcal{F}' = \{\{0\}, \{3\}, \{17\}, \{20\}, \{0, 3\}, \{17, 20\}, \Omega'\}$
- $P(X = 0) = 1/4, P(X = 3) = 1/8, P(X = 17) = 1/8, P(X = 20) = 1/2.$

Dann ergibt sich ein **neuer Wahrscheinlichkeitsraum**. Für $X \in \mathbb{R}$ schreibt man $(\mathbb{R}, \mathcal{B}, \mathcal{P}_X)$. Die σ -Algebra \mathcal{B} ist eine geeignete Menge aller reellen Intervalle, die als Borel-Algebra bezeichnet wird.

Zur Vereinfachung der Notation schreiben wir jedoch weiter häufig $\{\Omega, \mathcal{F}, \mathcal{P}\}$.

2.5. Verteilungs- und Dichtefunktionen

Übersicht

- Univariate Wahrscheinlichkeitsverteilung (cumulative distribution function (CDF))
- Multivariate Verteilungs- und Dichtefunktionen

2.5.1. Univariate Verteilungs- und Dichtefunktionen

Übersicht

- Univariate Wahrscheinlichkeitsverteilung (cumulative distribution function (CDF))
- Eigenschaften von Verteilungsfunktionen
- Wahrscheinlichkeitsdichtefunktionen
- Interpretation Wahrscheinlichkeitsdichtefunktion
- Standardnormalverteilung und Normalverteilung
- CDF einer binären Zufallsvariablen
- Träger
- Quantile und Quantilsfunktionen

Univariate Wahrscheinlichkeitsverteilung

Wahrscheinlichkeitsverteilung (probability distribution, cumulative distribution function (CDF))

Eine Wahrscheinlichkeitsfunktion für eine skalare Zufallsvariable X ist definiert durch

$$F : \mathbb{R} \mapsto [0, 1] : \quad F(x) \equiv P(X \leq x) \\ = P(X(\omega) \in (-\infty, x]). \quad (2.2)$$

Eigenschaften von Verteilungsfunktionen

- $\lim_{x \rightarrow -\infty} F(x) = 0$

- $\lim_{x \rightarrow \infty} F(x) = 1$
- $F(x)$ ist (schwach) monoton steigend
- $P(a < X \leq b) = F(b) - F(a)$
- $F(x) = P(X \leq x) = P(X < x)$, wenn X stetig.

Wahrscheinlichkeitsdichtefunktionen

Motivation von Wahrscheinlichkeitsdichten

Für eine stetige Zufallsvariable Y gilt, dass die Wahrscheinlichkeit 'Y nimmt den Wert y an' gerade Null ist, d.h. $P(Y = y) = 0$. Intuition: Fläche unter einem Integral an einem Punkt ist Null.

Stattdessen muss man für Y ein Intervall betrachten, z.B. $[a, b]$ oder häufig $(-\infty, y]$. Für das letztere erhält man die Wahrscheinlichkeitsverteilung

$$F(y) = P(Y \leq y) \stackrel{Y \text{ stetig}}{=} P(Y < y),$$

die monoton in y wächst. Man kann also auch die Veränderung der Wahrscheinlichkeit betrachten, wenn die Intervalllänge um einen marginalen Betrag $\delta > 0$ zunimmt. Dies ergibt die absolute Veränderung in der Wahrscheinlichkeit

$$P(Y \leq y + \delta) - P(Y \leq y)$$

und die relative Veränderung

$$\frac{P(Y \leq y + \delta) - P(Y \leq y)}{\delta}.$$

Indem man nun die marginale Veränderung δ der Intervalllänge gegen 0 gehen lässt, erhält man die **Wahrscheinlichkeitsdichtefunktion**

$$f(y) = \lim_{\delta \rightarrow 0} \frac{P(Y \leq y + \delta) - P(Y \leq y)}{\delta},$$

die an einigen y positiv sein muss, denn ansonsten würde sich bei einer Änderung der Intervalllänge keine Veränderung der Wahrscheinlichkeit ergeben. Die Wahrscheinlichkeitsdichte gibt also die **Rate** an, mit der sich die Wahrscheinlichkeit ändert, wenn das Intervall marginal verändert wird.

Da

$$P(y < Y \leq y + \delta) = P(Y \leq y + \delta) - P(Y \leq y),$$

erhält man, salopp gesprochen,

$$P(y < Y \leq y + \delta) \approx f(y)\delta.$$

Man kann deshalb die Wahrscheinlichkeit, dass eine Realisation von Y in einem bestimmten Intervall $(y, y + \delta]$ beobachtet wird, mit dem Produkt aus der Dichte und der Intervalllänge approximieren. Diese Approximation ist umso besser, je kleiner δ ist. **Die Dichte ist approximativ proportional zur Wahrscheinlichkeit, dass Y in einem sehr kleinen Intervall um y herum beobachtet wird.**

Wahrscheinlichkeitsdichtefunktion (probability density function) (PDF)

Für eine stetige Zufallsvariable mit differenzierbarer Wahrscheinlichkeitsverteilung $F(x)$ wird die Ableitung erster Ordnung **Wahrscheinlichkeitsdichtefunktion** genannt

$$f(x) \equiv \frac{dF(x)}{dx}, \quad (2.3)$$

$$\int_{-\infty}^x f(z)dz = F(x). \quad (2.4)$$

Wichtige Wahrscheinlichkeitsverteilungen

Standardnormalverteilung

$x \sim N(0, 1)$ für $x \in \mathbb{R}$

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}x^2\right), \quad \Phi(x) = \int_{-\infty}^x \phi(z)dz. \quad (2.5)$$

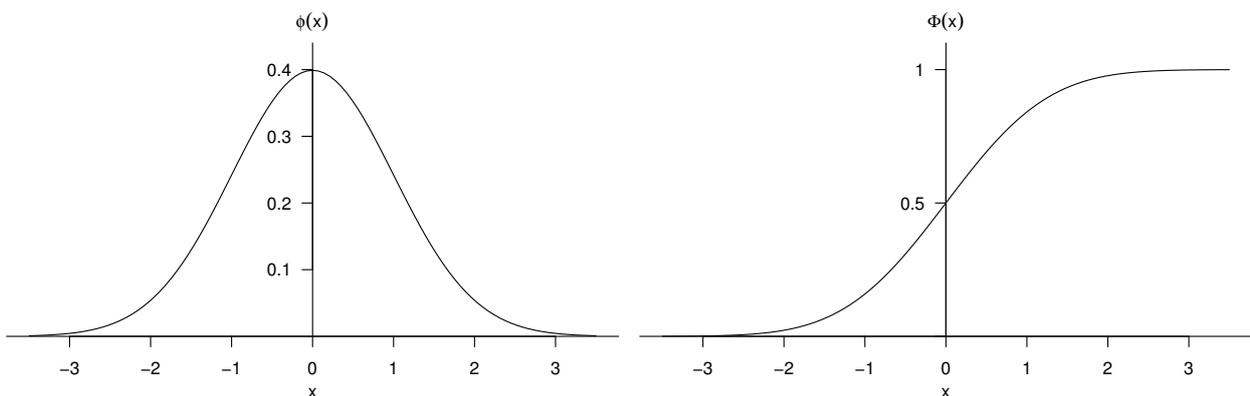


Abbildung 2.1.: PDF und CDF der Standardnormalverteilung (R-Programm siehe Abschnitt A.1, Seite 328)

Normalverteilung

$x \sim N(\mu, \sigma^2)$ mit Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right). \quad (2.6)$$

Beachte: (2.6) kann mit Hilfe des eindimensionalen Transformationsatzes (2.39) abgeleitet werden.

CDF einer binären Zufallsvariable

$$F(x) = \begin{cases} 0 & \text{für } x < 0 \\ p & \text{für } 0 \leq x < 1 \\ 1 & \text{für } x \geq 1. \end{cases} \quad (2.7)$$

Weitere Bemerkungen

- CDFs können Sprünge (=Diskontinuitäten) aufweisen, es können auch CDFs für Zufallsvariablen, die teils stetig, teils diskret sind, definiert werden (z. B. bei zensierten Variablen).
- **Träger (support)**: Gegeben sei eine Zufallsvariable X . Der Bereich, auf dem eine Dichtefunktion $f_X(x)$ positiv ist, wird als **Träger (support)** $\mathcal{X} \subset \mathbb{R}$ einer Dichtefunktion bezeichnet:

$$\mathcal{X} = \{x : f_X(x) > 0\}.$$

- Siehe Abschnitt 2.9 zu Details wichtiger Wahrscheinlichkeitsverteilungen.

Eine tabellarische Übersicht über viele Wahrscheinlichkeitsverteilungen und findet sich auf der Kurshomepage.

Quantile

Quantil

Das α -**Quantil** q_α einer Verteilung für eine Zufallsvariable X ist definiert durch

$$F(q_\alpha) = P(X \leq q_\alpha) = \alpha. \quad (2.8)$$

Die **Quantilsfunktion** lautet:

$$q_\alpha = F^{-1}(\alpha). \quad (2.9)$$

R-Befehle

Berechnen eines Quantils der Standardnormalverteilung: mit `qnorm()`.

Beispiel: Das $P(X \leq q_{0.85}) = 0.85$ -Quantil der Standardnormalverteilung ist $q_{0.85} = 1.036433$. Man erhält es mit dem R-Befehl `qnorm(0.85) = 1.036433`. Es ist in den Graphiken senkrecht und in Rot eingezeichnet. Die blau schraffierte Fläche unter der Dichte ist gerade $\alpha = 0.85$.

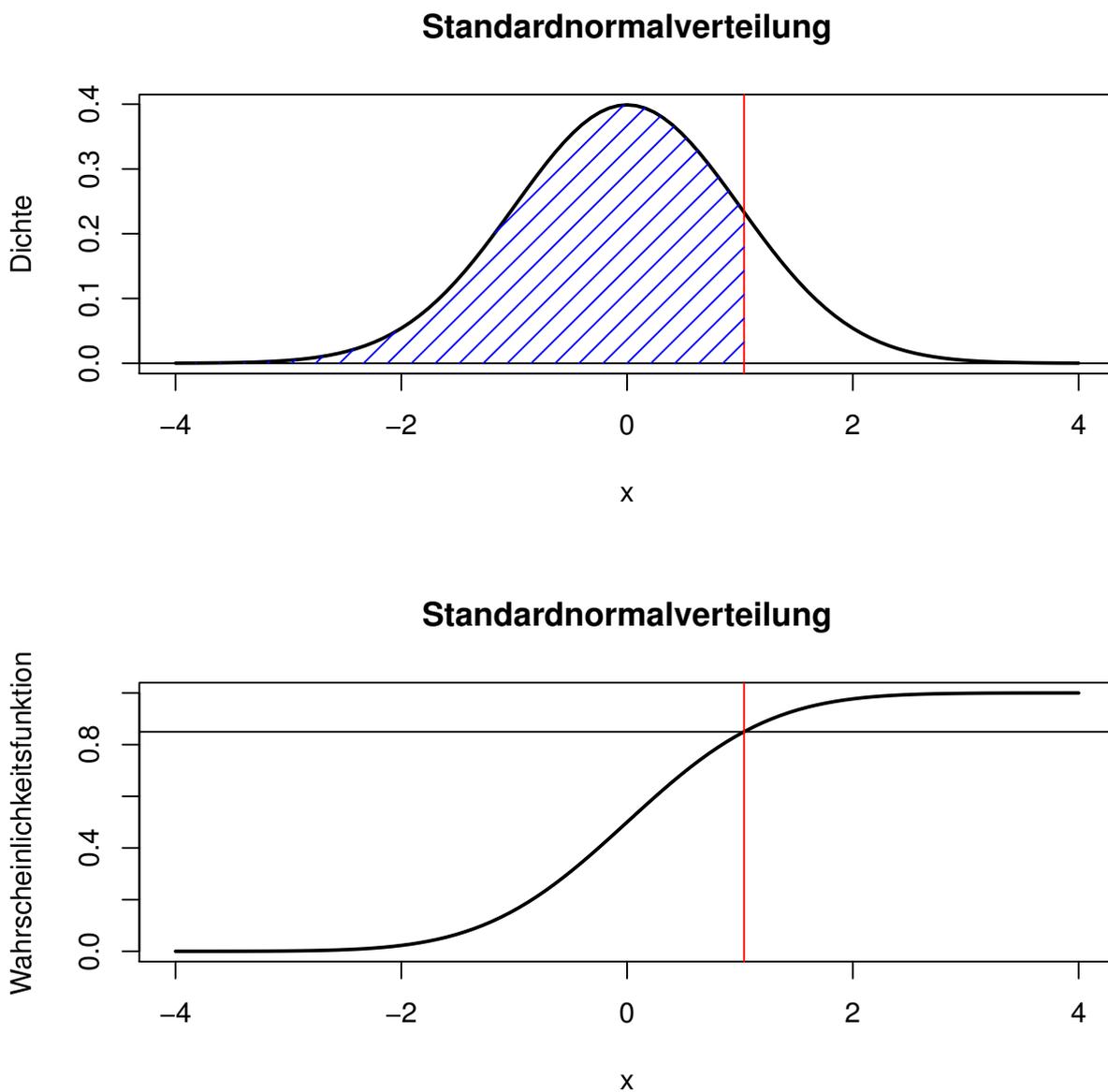


Abbildung 2.2.: 0.85-Quantil der Standardnormalverteilung (R-Programm siehe Abschnitt [A.1](#), Seite [330](#))

Wichtige Quantile

- **Median:** $q_{0.5}$
- **Quartile:** q_α mit $\alpha = 0.25, 0.5, 0.75$
- **Quintile:** q_α mit $\alpha = 0.2, 0.4, 0.6, 0.8$
- **Decile:** q_α mit $\alpha = 0.1, 0.2, \dots, 0.8, 0.9$
- **Percentile:** q_α mit $\alpha = 0.01, 0.02, \dots, 0.98, 0.99$

2.5.2. Multivariate Verteilungs- und Dichtefunktionen

Übersicht

- Multivariate Verteilungs- und Dichtefunktionen
- Gemeinsame Wahrscheinlichkeitsverteilung
- Marginale Wahrscheinlichkeitsverteilung
- Gemeinsame Dichtefunktion

Gemeinsame Wahrscheinlichkeitsverteilung (joint probability distribution function)

für zwei oder mehr Zufallsvariablen X_1, \dots, X_m

$$\begin{aligned} F_{X_1, X_2, \dots, X_m}(x_1, x_2, \dots, x_m) &\equiv P((X_1 \leq x_1) \cap \dots \cap (X_m \leq x_m)) \\ &= P(X_1 \leq x_1, \dots, X_m \leq x_m). \end{aligned} \quad (2.10)$$

Marginale Wahrscheinlichkeitsverteilung (marginal probability distribution)

$$F_{X_i}(x_i) \equiv P(X_i \leq x_i). \quad (2.11)$$

Marginale Wahrscheinlichkeitsdichtefunktion für eine stetige Zufallsvariable X_i

$$f_{X_i}(x_i) \equiv \frac{dF_{X_i}(x_i)}{dx_i}. \quad (2.12)$$

Gemeinsame Wahrscheinlichkeitsdichtefunktion (joint density function)

für zwei oder mehr stetige und Zufallsvariablen $X_1, \dots, X_m \in \mathbb{R}$ mit partiell differenzierbarer CDF:

$$f_{X_1, X_2, \dots, X_m}(x_1, x_2, \dots, x_m) \equiv \frac{\partial^m F_{X_1, X_2, \dots, X_m}(x_1, x_2, \dots, x_m)}{\partial x_1 \partial x_2 \cdots \partial x_m}, \quad (2.13)$$

$$= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_m} f_{X_1, X_2, \dots, X_m}(z_1, z_2, \dots, z_m) dz_1 dz_2 \cdots dz_m,$$

$$F_{X_1}(x_1) = F_{X_1, X_2, \dots, X_m}(x_1, \infty, \dots, \infty).$$

Zusammenhang zwischen marginalen und gemeinsamen Dichten

Es gilt, z. B. im Fall von drei Zufallsvariablen

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X_1, X_2, X_3}(x_1, z_2, z_3) dz_2 dz_3. \quad (2.14)$$

Notation: Davidson & MacKinnon (2004) verzichten auf die Indexierung von F und f . Mit Ausnahme dieses Abschnittes erfolgt dies zur Vereinfachung der Notation auch in diesen Unterlagen, falls die Indexierung aus dem Zusammenhang leicht erschlossen werden kann.

Bivariate Normalverteilung

$$f_{X_1, X_2}(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \frac{x_1 - \mu_1}{\sigma_1} \frac{x_2 - \mu_2}{\sigma_2} + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\} \quad (2.15)$$

Multivariate Normalverteilung: siehe (2.31) in Abschnitt 2.9.1.

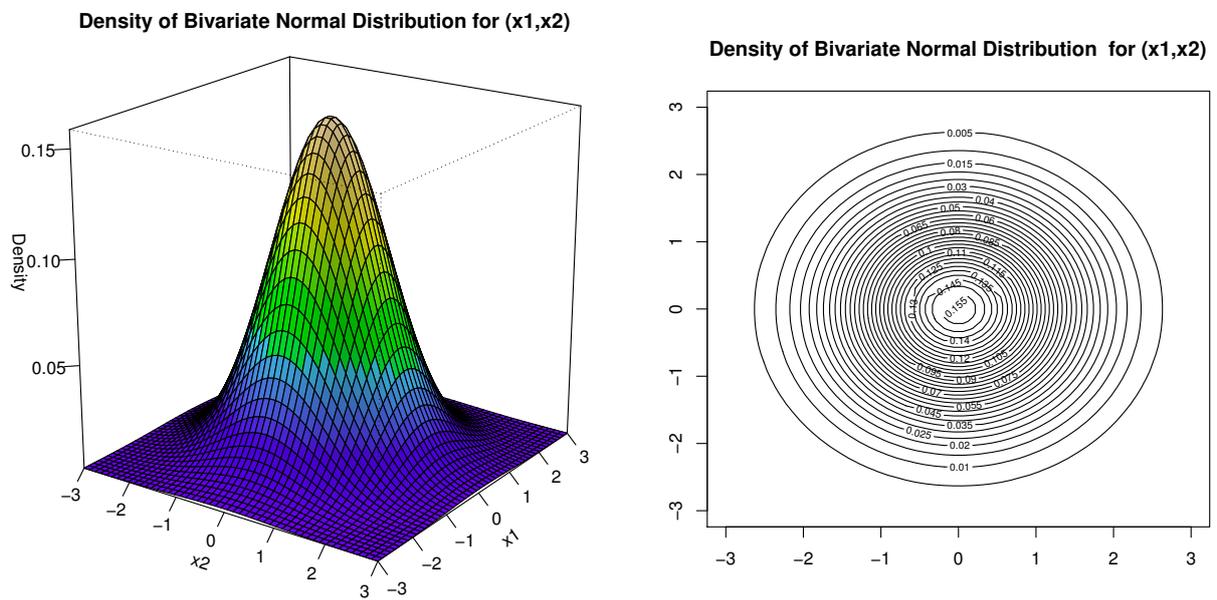


Abbildung 2.3.: PDF der bivariaten Normalverteilung (R-Programm siehe Abschnitt A.1, Seite 331)

2.6. Bedingte Wahrscheinlichkeiten

Übersicht

- Motivation
- Zusammenhang mit gemeinsamer Wahrscheinlichkeit
- Grundregel für bedingte Wahrscheinlichkeiten
- Zusammenhang mit unbedingten Wahrscheinlichkeiten
- Bedingung auf Zufallsvariablen
- Bedingte Wahrscheinlichkeitsdichte
- Bedingte Normalverteilung
- Zusammenhang zwischen marginaler und bedingter Dichte
- Stochastische Unabhängigkeit und bedingte Dichte / Verteilung

- **Beispiel zur Motivation:** Es bezeichne die Zufallsvariable $X \in [0, \infty)$ den Auszahlungsbetrag in einem Gewinnspiel. Die Wahrscheinlichkeitsfunktion bzw. Verteilungsfunktion $P(X \leq x) = F_X(x)$ 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 , wenn Maschine A zum Einsatz kommt?

Anders formuliert, wie groß ist die gesuchte Wahrscheinlichkeit, wenn die Bedingung "Maschine A im Einsatz" gilt? Man nennt deshalb die gesuchte Wahrscheinlichkeit auch **bedingte Wahrscheinlichkeit** und man schreibt

$$P(X \leq x|A).$$

Entsprechend notiert man, falls die Bedingung "Maschine B im Einsatz" gilt, $P(X \leq x|B)$.

- **Zusammenhang mit gemeinsamer Wahrscheinlichkeit** Sei E das Ereignis $(X \leq x)$. Wenn jemand nur glücklich ist, wenn er eine Auszahlung in Höhe von maximal x von Maschine B bekommt, dann möchte er die Wahrscheinlichkeit $P(E \cap B)$ bestimmen. Diese Wahrscheinlichkeit ergibt sich durch den

Multiplikationssatz

Für zwei Ereignisse E, B aus der Kollektion aller möglichen Ereignisse \mathcal{F} gilt:

$$P(E \cap B) = P(B)P(E|B), \quad P(B) > 0.$$

Die Kenntnis über Realisation von B kann helfen, genauere Aussagen über die mögliche Realisation von E zu machen.

Der Multiplikationssatz ergibt sich aus der

Definition einer bedingten Wahrscheinlichkeit

Für zwei Ereignisse E, B aus der Kollektion aller möglichen Ereignisse \mathcal{F} gilt:

$$P(E|B) = \frac{P(E \cap B)}{P(B)}, \quad P(B) > 0.$$

Beispiele:

- $B \in E$: $P(E|B) = 1$; Z. B. Maschine B zahlt immer einen Minimumbetrag größer Null, aber ein Maximum kleiner x .
- E und B disjunkt: $P(E|B) = 0$.

Satz von Bayes

Für zwei Ereignisse E, B aus der Kollektion aller möglichen Ereignisse \mathcal{F} gilt:

$$P(E|B) = \frac{P(B|E)P(E)}{P(B)}, \quad P(B), P(E) > 0.$$

- Zusammenhang 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:

Zusammenhang zwischen unbedingten und bedingten Wahrscheinlichkeiten

$$\begin{aligned} P(E) &= P(E \cap A) + P(E \cap B) \\ P(X \leq x) &= P(X \leq x|A)P(A) + P(X \leq x|B)P(B) \\ F_X(x) &= F_{X|Maschine}(x|A)P(A) + F_{X|Maschine}(x|B)P(B) \end{aligned}$$

(Die Ergebnismenge mit den Elementarereignissen für die Maschinenwahl ist $\Omega = \{A, B\}$.)

Ersetzt man das Ereignis B durch das zu A komplementäre Ereignis A^c , erhält man das allgemein gültige

Gesetz der totalen Wahrscheinlichkeit

$$P(E) = P(E \cap A) + P(E \cap A^c)$$

$$P(E) = P(E|A)P(A) + P(E|A^c)P(A^c)$$

- **Bedingung auf Zufallsvariablen:** 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 $F_{X|Z}(x|Z = z)$, wobei $Z = z$ bedeutet, dass die Bedingung lautet, dass Zufallsvariable Z genau die Realisation z annimmt. Ist Z stetig und $Z \in [0, \infty)$, müssen wir, um wieder den Zusammenhang zwischen der unbedingten und den bedingten Wahrscheinlichkeiten zu erhalten, 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:

$$F_X(x) = \int_0^\infty F_{X|Z}(x|Z = z)f_Z(z)dz = \int_0^\infty F_{X|Z}(x|z)f_Z(z)dz$$

bzw. allgemein:

Zusammenhang zwischen unbedingter und bedingten Verteilungsfunktionen

$$F_X(x) = \int F_{X|Z}(x|Z = z)f_Z(z)dz = \int F_{X|Z}(x|z)f_Z(z)dz \quad (2.16)$$

Bedingte Wahrscheinlichkeitsdichte (conditional probability distribution function)

für Zufallsvariable X_1 gegeben eine Zufallsvariable X_2 oder mehrere Zufallsvariablen X_2, \dots, X_m :

$$f_{X_1|X_2}(x_1|x_2) \equiv \frac{f_{X_1, X_2}(x_1, x_2)}{f_{X_2}(x_2)}, \quad (2.17)$$

vorausgesetzt, dass $f_{X_2}(x_2) > 0$,

$$f_{X_1|X_2, \dots, X_m}(x_1|x_2, \dots, x_m) \equiv \frac{f_{X_1, \dots, X_m}(x_1, x_2, \dots, x_m)}{f_{X_2, \dots, X_m}(x_2, \dots, x_m)}, \quad (2.18)$$

vorausgesetzt, dass

$$f_{X_2, \dots, X_m}(x_2, \dots, x_m) > 0.$$

Bedingte Normalverteilung:

Es sei $\mu(X) = E[Y|X]$ und $\sigma^2(X) = Var(Y|X)$. Dann sind folgende Schreibweisen

äquivalent:

$$Y|X \sim N(\mu(x), \sigma^2(x))$$

$$f_{Y|X}(y|x) = \frac{1}{\sqrt{2\pi\sigma^2(x)}} \exp\left(-\frac{1}{2} \frac{(y - \mu(x))^2}{\sigma^2(x)}\right) \quad (2.19)$$

Wichtige Eigenschaften:

Berechnen der marginalen Dichte aus der bedingten Dichte

$$f_X(x) = \int f_{X|Z}(x|Z=z)f_Z(z)dz = \int f_{X|Z}(x|z)f_Z(z)dz. \quad (2.20)$$

Stochastische Unabhängigkeit

Gilt

$$F_{X_1, X_2}(x_1, x_2) = F_{X_1, X_2}(x_1, \infty)F_{X_1, X_2}(\infty, x_2) = P(X_1 \leq x_1) P(X_2 \leq x_2), \quad (2.21)$$

werden die Zufallsvariablen X_1 und X_2 als **stochastisch unabhängig** oder **unabhängig** bezeichnet und es gilt

$$f_{X_1, X_2}(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2). \quad (2.22)$$

Entsprechende Faktorisierungen gelten für mehr als zwei ZV. Wenn die Zufallszahlen X_1 und X_2 stochastisch unabhängig sind, dann gilt:

$$F_{X_1|X_2}(x_1|x_2) = F_{X_1}(x_1), \quad (2.23a)$$

$$f_{X_1|X_2}(x_1|x_2) = f_{X_1}(x_1). \quad (2.23b)$$

2.7. Erwartungswerte und Momente

Übersicht

- Definitionen und Regeln
- Ungleichungen für Erwartungswerte
- Momente zweiter Ordnung: Varianz, Kovarianz, Korrelation
- Regeln
- Höhere Momente: unzentrierte und zentrierte Momente

- Schiefe, Wölbung

Erwartungswerte, bzw. erste Momente

- **Erwartungswert einer diskrete Zufallsvariable X mit endlich** vielen möglichen Ausprägungen x_i , $m < \infty$,

$$E[X] = \sum_{i=1}^m x_i P(X = x_i)$$

- **Erwartungswert einer diskrete Zufallsvariable X mit unendlich** vielen Ausprägungen x_i

$$E[X] = \sum_{i=1}^{\infty} x_i P(X = x_i)$$

Beachte: Dieser Erwartungswert existiert nur, wenn

$$\sum_{i=1}^{\infty} |x_i| P(X = x_i) < \infty.$$

- **Erwartungswert einer stetigen Zufallsvariable $X \in \mathbb{R}$**

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

Beachte: Dieser Erwartungswert existiert nur, wenn

$$\int_{-\infty}^{\infty} |x| f(x) dx < \infty.$$

- **Erwartungswert einer stetigen Zufallsvariable X auf Träger $\mathcal{X} = (a, b) \subset \mathbb{R}$**

$$E[X] = \int_a^b x f(x) dx$$

Beachte: Dieser Erwartungswert existiert immer, sofern $f(x) < \infty$ für $x \in \mathcal{X}$.

Regeln für den Erwartungswert

z. B. [Wooldridge \(2009, Appendix B\)](#)

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 stochastisch unabhängig, gilt

$$E[XY] = E[X]E[Y].$$

- bzw. allgemeiner: Sind die Zufallsvariablen X und Y stochastisch unabhängig und gilt für alle Funktionen $f(x)$ und $g(y)$, dass $E[|f(X)|] < \infty$ und $E[|g(Y)|] < \infty$, dann gilt

$$E[f(X)g(Y)] = E[f(X)]E[g(Y)].$$

Ungleichungen für Erwartungswerte

1. $E[|X + Y|] \leq E[|X|] + E[|Y|]$
2. **Jensen-Ungleichung:** Ist $g(x)$ konvex, dann gilt: $E[g(X)] \geq g(E[X])$. Das Ungleichheitszeichen gilt strikt, wenn $g(x)$ strikt konvex ist. Ist $g(x)$ konkav, kehrt sich das Ungleichheitszeichen um.

Momente zweiter Ordnung

Varianz, Kovarianz und Korrelation

$$\begin{aligned}
 \text{Var}(X) &= E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f(x) dx, \\
 \text{Cov}(X, Y) &= E[(X - E[X])(Y - E[Y])] \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[X])(y - E[Y]) f_{X,Y}(x, y) dx dy, \\
 \text{Corr}(X, Y) &= \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}. \tag{2.24}
 \end{aligned}$$

Regeln

- $\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - E[X]^2$ (Verschiebungssatz),
- $\text{Var}(a + bX) = \text{Var}(bX) = b^2 \text{Var}(X)$,
- $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])] = E[XY] - E[X]E[Y]$
- $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$,

Höhere Momente

- **zweites (unzentriertes) Moment:** $m_2(X) = \int_{-\infty}^{\infty} x^2 f(x) dx$

- Es sei $\mu = E[X] = m_1(X)$ und $\sigma = \sqrt{\text{Var}(X)} = \sqrt{\tilde{m}_2(X)}$.
- **k -tes (unzentriertes) Moment:**

$$m_k(X) = E[X^k] = \int_{-\infty}^{\infty} x^k f(x) dx$$

- **k -tes zentriertes Moment:**

$$\tilde{m}_k(X) = E[(X - E(X))^k] = \int_{-\infty}^{\infty} (x - m_1(X))^k f(x) dx$$

- **Schiefe (Skewness)** (drittes zentriertes Moment)

$$\frac{E[(X - E[X])^3]}{\sigma^3} = \frac{\int_{-\infty}^{\infty} (x - \mu)^3 f(x) dx}{\sigma^3}.$$

- **Wölbung (Kurtosis)**

$$\frac{E[(X - E[X])^4]}{\sigma^4} = \frac{\int_{-\infty}^{\infty} (x - \mu)^4 f(x) dx}{\sigma^4}.$$

Beispiele:

- Die Schiefe von symmetrischen Dichten ist 0.
- Die Wölbung einer standardnormalverteilten Zufallsvariablen ist 3.

2.8. Bedingte Erwartungswerte und Momente

Übersicht

- Definitionen und Regeln
- Gesetz der iterierten Erwartungen
- Regeln für bedingte Erwartungen
- Regeln
- Regeln für bedingte Varianzen und Kovarianzen

Bedingter Erwartungswert

- **Definition:** 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

$$E[X|A] = \int_0^{\infty} x f(x|A) dx.$$

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

$$E[X|B] = \int_0^{\infty} x f(x|B) dx.$$

Ist noch nicht “realisiert”, welche $M = A, B$ im Einsatz ist, ist der **bedingte Erwartungswert**

$$E[X|M] = \int_0^{\infty} x f(x|M) dx = g(M)$$

eine **Funktion** mit Argument M . Damit ist der bedingte Erwartungswert eine **Zufallsvariable**. Dies gilt allgemein.

Abhängig davon, ob die Bedingung bzw. X stetig oder diskret sind, unterscheidet sich die **Berechnung des bedingten Erwartungswertes** etwas

X	=	stetig	diskret	Bedingung
$E[X A]$	=	$\int x f(x A) dx$	$\sum x_i P(X = x_i A)$	diskret
$E[X Z = z]$	=	$\int x f(x Z = z) dx$	$\sum x_i P(X = x_i z)$	stetig

Beachte: Häufig verwendet man auch die Kurzformen, so auch in [Wooldridge \(2009\)](#), z. B.

$$E[X|z] = \int x f(x|z) dx.$$

- **Gesetz der iterierten Erwartungen (Law of iterated expectations (LIE)):** Entsprechend dem Zusammenhang zwischen unbedingten und bedingten Wahrscheinlichkeiten, existiert ein ähnlicher Zusammenhang auch zwischen dem unbedingten und den bedingten Erwartungswerten. Er lautet

$$E[X] = E[E(X|Z)] = E[g(Z)], \quad g(Z) = E(X|Z)$$

und wird als **Gesetz der iterierten Erwartungen** bezeichnet.

Beweisskizze:

$$\begin{aligned}
 E[X] &= \int x f(x) dx \\
 &= \int x \left[\int f(x|z) f(z) dz \right] dx \quad (\text{Einsetzen von (2.20)}) \\
 &= \int \int x f(x|z) f(z) dz dx \\
 &= \int \underbrace{\int x f(x|z) dx}_{E[X|z]} f(z) dz \quad (\text{Vertauschen von } dx \text{ und } dz) \\
 &= \int E[X|z] f(z) dz \\
 &= E[E(X|Z)].
 \end{aligned}$$

In unserem **Beispiel** mit den 2 Maschinen ergibt das Gesetz der iterierten Erwartungen

$$\begin{aligned}
 E[X] &= E[X|A]P(A) + E[X|B]P(B), \\
 E[X] &= g(A)P(A) + g(B)P(B)
 \end{aligned} \tag{2.25}$$

Dieses Beispiel macht nochmals 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.

Regeln für bedingte Erwartungen

(z. B. [Wooldridge \(2009, Appendix B\)](#))

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. **Gesetz der iterierten Erwartungen (LIE)**

$$\begin{aligned}
 E[Y] &= E[E(Y|X)] \\
 E[Y|X] &= E[E(Y|X, Z)|X]
 \end{aligned}$$

5. Falls $E[Y^2] < \infty$, $E[g(X)^2] < \infty$ für eine beliebige Funktion $g(\cdot)$, dann:

$$\begin{aligned} E\{[Y - E(Y|X)]^2 | X\} &\leq E\{[Y - g(X)]^2 | X\} \\ E\{[Y - E(Y|X)]^2\} &\leq E\{[Y - g(X)]^2\}. \end{aligned}$$

Regeln für bedingte Varianzen und Kovarianzen

Verschiebungssatz für (Ko-)Varianzen, etc.

$$\text{Var}(Y|X) = E[(Y - E[Y|X])^2 | X] = E[Y^2 | X] - E[Y|X]^2, \quad (2.26)$$

$$\begin{aligned} \text{Cov}(Y, X|Z) &= E[(Y - E[Y|Z])(X - E[X|Z]) | Z] \\ &= E[YX | Z] - E[Y|Z]E[X|Z], \end{aligned} \quad (2.27)$$

$$\text{Var}(Y) = E[\text{Var}(Y|X)] + \text{Var}(E[Y|X]). \quad (2.28)$$

Ein Beweis für (2.28) findet sich für die multivariate Variante (9.6).

Zusammenhänge zwischen bedingten Erwartungen und Kovarianzen

Es gilt für zwei Zufallsvariablen Y und X :

$$E[Y|X] = E[Y] \implies \text{Cov}(Y, X) = 0, \quad (2.29a)$$

$$E[Y|X] = 0 \implies E[Y] = 0 \text{ und } \text{Cov}(Y, X) = 0, \quad (2.29b)$$

$$\text{Cov}(Y, X) \neq 0 \implies E[Y|X] \neq 0, \quad (2.29c)$$

$$\text{Cov}(Y, X) = 0 \text{ \& } E[Y] = 0 \implies E[YX] = E[XE(Y|X)] = 0, \quad (2.29d)$$

$$E[Y] = 0 \not\Rightarrow E[Y|X] = 0, \quad (2.29e)$$

$$\text{Cov}(Y, X) = 0 \not\Rightarrow E[Y|X] = 0. \quad (2.29f)$$

Beispiel: Für $Y = X^2$ und $E(X) = E(X^3) = 0$ gilt $\text{Cov}(Y, X) = 0$, da $\text{Cov}(X^2, X) = E[X^3] - E[X^2]E[X] = 0$, aber $E[Y|X] = X^2 \neq 0$.

Nachweise via

$$\begin{aligned} \text{Cov}(Y, X) &= E[YX] - E[Y]E[X] = E[E[YX|X]] - E[E[Y|X]]E[X] \\ &= E[XE[Y|X]] - E[E[Y|X]]E[X] \end{aligned}$$

- (2.29b): Ist $E[Y|X] = 0$, muss $\text{Cov}(Y, X) = 0$ folgen.
- (2.29c): Wäre diese Aussage falsch und $E[Y|X] = 0$ folgen, folgt $\text{Cov}(Y, X) = 0$ aus (2.29b), was zu einem Widerspruch führt.
- (2.29f): Aus $\text{Cov}(Y, X) = 0$ folgt lediglich $E[XE[Y|X]] = E[E[Y|X]]E[X]$, nicht jedoch $E[Y|X] = 0$.

2.9. Wichtige Wahrscheinlichkeitsverteilungen

Übersicht

- Normalverteilung
- χ^2 , t -, F -Verteilung
- ‡ Transformationssätze

2.9.1. Normalverteilung

Übersicht

- Standardnormalverteilung
- Normalverteilung
- Multivariate Standardnormalverteilung
- Multivariate Normalverteilung
- Bivariate Normalverteilung
- Linearkombinationen multivariate normalverteilter Zufallsvektoren
- IID und NID

- **Standardnormalverteilung:** $x \sim N(0, 1)$ mit Dichtefunktion (2.5)

$$\phi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}z^2\right). \quad (2.5)$$

- **Normalverteilung:** $x \sim N(\mu, \sigma^2)$ mit Dichte

$$f(x) = \frac{1}{\sqrt{\sigma^2 2\pi}} \exp\left(-\frac{1}{2} \frac{(x - \mu)^2}{\sigma^2}\right) = \frac{1}{\sigma} \phi\left(\frac{x - \mu}{\sigma}\right). \quad (2.6)$$

Beachte: (2.6) kann mit Hilfe des eindimensionalen Transformationssatzes (2.39) abgeleitet werden.

- **Multivariate Standardnormalverteilung:** $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_n)$ mit Dichte

$$\phi(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\mathbf{z}^T \mathbf{z}\right). \quad (2.30)$$

Man beachte, dass diese Darstellung äquivalent ist zu (vgl. hierzu (2.22))

$$\phi(\mathbf{z}) = \phi(z_1)\phi(z_2)\cdots\phi(z_n).$$

Ein multivariat standardnormalverteilter Zufallsvektor \mathbf{z} setzt sich also aus unabhängig und identisch verteilten (genauer standardnormalverteilten) Zufallsvariablen z_1, \dots, z_n zusammen. Umgekehrt: n i.i.d. standardnormalverteilte Zufallszahlen lassen sich als multivariat standardnormalverteilter Zufallsvektor schreiben. Beachte: Ohne die i.i.d. Voraussetzung geht das nicht!

- **Multivariate Normalverteilung:**

$$\mathbf{x} = \mathbf{A}\mathbf{z} + \boldsymbol{\mu} \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega}), \quad \text{wobei } \boldsymbol{\Omega} = \mathbf{A}\mathbf{A}^T \quad (2.31)$$

und für die $(r \times n)$ -Matrix \mathbf{A} , $r \leq n$, $rk(\mathbf{A}) = r$ gilt. Dichtefunktion:

$$f(x_1, x_2, \dots, x_r) = f(\mathbf{x}) = \frac{1}{(2\pi)^{r/2}} (\det(\boldsymbol{\Omega}))^{-1/2} \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Omega}^{-1}(\mathbf{x} - \boldsymbol{\mu})\right). \quad (2.32)$$

- **Bivariate Normalverteilung (2.15).** Siehe Abschnitt 2.5 für Plot.

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x_1-\mu_1}{\sigma_1}\right)^2 - 2\rho\frac{x_1-\mu_1}{\sigma_1}\frac{x_2-\mu_2}{\sigma_2} + \left(\frac{x_2-\mu_2}{\sigma_2}\right)^2\right]\right\}$$

- **Linearkombinationen multivariat normalverteilter Zufallsvektoren**

$$\begin{aligned} \text{Es gilt für} \quad \mathbf{w} &= \mathbf{b} + \mathbf{B}\mathbf{x} \quad \text{mit } \mathbf{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) : \\ \mathbf{w} &\sim N(\mathbf{b} + \mathbf{B}\boldsymbol{\mu}, \mathbf{B}\boldsymbol{\Sigma}\mathbf{B}^T). \end{aligned} \quad (2.33)$$

- **Notation**

– Die Zufallsvariablen v_t , $t = 1, \dots, n$ sind **unabhängig und identisch verteilt** bzw. **independently and identically distributed (IID)**:

$$v_t \sim IID(E[v_t], Var(v_t)).$$

– Die Zufallsvariablen v_t , $t = 1, \dots, n$ sind **unabhängig und identisch normalverteilt** bzw. **independently and identically normally distributed (NID)**:

$$v_t \sim NID(E(v_t), Var(v_t)).$$

In Matrixnotation entspricht dies mit $\mu_v = E[v_t]$, $\sigma_v^2 = Var(v_t)$

$$\begin{aligned} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} &\sim N\left(\begin{pmatrix} \mu_v \\ \mu_v \\ \vdots \\ \mu_v \end{pmatrix}, \begin{pmatrix} \sigma_v^2 & 0 & \cdots & 0 \\ 0 & \sigma_v^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_v^2 \end{pmatrix}\right), \\ \mathbf{v} &\sim N(\mu_v \boldsymbol{\iota}, \sigma_v^2 \mathbf{I}). \end{aligned}$$

Vgl. zur Definition von $\boldsymbol{\iota}$ (7.8).

2.9.2. χ^2 -, t -, F -Verteilung

Übersicht

- χ^2 -Verteilung
- Student t -Verteilung
- F -Verteilung

 χ^2 -Verteilung

- Sind z_1, \dots, z_m i.i.d. standardnormalverteilt, $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I}_m)$, so ist die Summe der quadrierten Zufallsvariablen

$$y = \sum_{i=1}^m z_i^2 = \mathbf{z}^T \mathbf{z} = \|\mathbf{z}\|^2$$

χ^2 -verteilt mit m Freiheitsgraden. In Kurzschreibweise:

$$y \sim \chi^2(m).$$

- Erwartungswert: $E(y) = m$,

$$\text{da } E\left(\sum_{i=1}^m z_i^2\right) = \sum_{i=1}^m E(z_i^2) = m.$$

- Varianz: $Var(y) = 2m$, da

$$\begin{aligned} E[(y - m)^2] &\stackrel{\text{Unabhängigkeit}}{=} mVar(z_i^2) \\ &= mE\left[(z_i^2 - 1)^2\right] \\ &= m\left(E[z_i^4] - 2 + 1\right) \\ &= 2m. \end{aligned}$$

- Wenn $y_1 = \sum_{i=1}^{m_1} z_i^2 \sim \chi^2(m_1)$ und $y_2 = \sum_{i=m_1+1}^m z_i^2 \sim \chi^2(m_2)$, $m = m_1 + m_2$, unabhängig sind, dann gilt

$$y = y_1 + y_2 \sim \chi^2(m).$$

- Ist \mathbf{x} ein multivariat normalverteilter $(m \times 1)$ -Vektor mit nichtsingulärer Kovarianzmatrix $\mathbf{\Omega}$, $\mathbf{x} \sim N(\mathbf{0}, \mathbf{\Omega})$, dann ist

$$y = \mathbf{x}^T \mathbf{\Omega}^{-1} \mathbf{x} \sim \chi^2(m). \quad (2.34)$$

Beweis: Da Ω regulär ist, existiert eine Zerlegung $\Omega = \mathbf{A}\mathbf{A}^T$, so dass $\mathbf{z} = \mathbf{A}^{-1}\mathbf{x}$ die Kovarianzmatrix \mathbf{I} aufweist. Dann gilt $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$ und

$$E \left[\mathbf{A}^{-1} \mathbf{x} \mathbf{x}^T (\mathbf{A}^{-1})^T \right] = \mathbf{A}^{-1} \Omega (\mathbf{A}^{-1})^T = \mathbf{A}^{-1} \mathbf{A} \mathbf{A}^T (\mathbf{A}^T)^{-1} = \mathbf{I}.$$

- Ist \mathbf{P} eine Projektionsmatrix mit $\text{rk } \mathbf{P} = r < m$ und $\mathbf{z} \sim N(\mathbf{0}, \mathbf{I})$, gilt

$$\mathbf{z}^T \mathbf{P} \mathbf{z} \sim \chi^2(r). \quad (2.35)$$

Beweis: Man nehme an, dass \mathbf{P} auf die r linear unabhängigen Spalten der $(m \times r)$ -Matrix \mathbf{Z} projiziert. Dann ist $\mathbf{P} = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T$ und man erhält

$$\mathbf{z}^T \mathbf{P} \mathbf{z} = \underbrace{\mathbf{z}^T \mathbf{Z}}_{\mathbf{w}^T} \underbrace{(\mathbf{Z}^T \mathbf{Z})^{-1}}_{\text{inverse Kovarianzmatrix mit Rang } r} \underbrace{\mathbf{Z}^T \mathbf{z}}_{\mathbf{w}}.$$

Da für den $(r \times 1)$ -Vektor $\mathbf{w} \sim N(\mathbf{0}, \mathbf{Z}^T \mathbf{Z})$ gilt, gilt wegen (2.34)

$$\mathbf{w}^T (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{w} \sim \chi^2(r).$$

- Für $m \rightarrow \infty$ gilt, dass eine $\chi^2(m)$ -verteilte Zufallsgröße in Verteilung gegen eine normalverteilte Zufallsgröße $N(m, 2m)$ konvergiert.

Student t -Verteilung

- Gegeben sei eine standardnormalverteilte Zufallsvariable $z \sim N(0, 1)$ und eine davon stochastisch unabhängige χ^2 -verteilte Zufallsgröße $y \sim \chi^2(m)$ mit m Freiheitsgraden. Dann ist die Zufallsvariable

$$t = \frac{z}{(y/m)^{1/2}} \sim t(m) \quad (2.36)$$

t -verteilt mit m Freiheitsgraden.

- Die Dichte der t -Verteilung ist symmetrisch und glockenförmig.
- Es existieren alle Momente der t -Verteilung bis zum $m - 1$ Moment. Die t -Verteilung mit $m = 1$ heißt auch **Cauchy-Verteilung**. Man beachte, dass weder Erwartungswert noch Varianz existieren, da die Verteilung zu viel Masse in den Flanken aufweist.
- **Erwartungswert:** Für $m > 1$: $E(t) = 0$, **Varianz:** Für $m > 2$: $\text{Var}(t) = m/(m - 2)$.
- Die t -Verteilung nähert sich mit zunehmender Zahl an Freiheitsgraden der Standardnormalverteilung an. Man kann hier asymptotisch argumentieren: Mit $m \rightarrow \infty$ gilt $\text{plim } m \rightarrow \infty y/m = 1$, da y eine Summe von m quadrierten unabhängigen standardnormalverteilten Zufallsvariablen ist. Mit Slutsky's Theorem gilt damit auch $\text{plim}_{m \rightarrow \infty} (y/m)^{1/2} = 1$ und somit

$$\text{plim}_{m \rightarrow \infty} \frac{z}{(y/m)^{1/2}} = z \sim N(0, 1).$$

F-Verteilung

- Gegeben seien zwei stochastisch unabhängige χ^2 -verteilte Zufallsvariablen $y_1 \sim \chi^2(m_1)$ und $y_2 \sim \chi^2(m_2)$. Dann folgt die Zufallsvariable

$$F = \frac{y_1/m_1}{y_2/m_2} \sim F(m_1, m_2) \quad (2.37)$$

einer **F-Verteilung mit m_1 und m_2 Freiheitsgraden.**

- Für $m_2 \rightarrow \infty$ nähert sich die Zufallsvariable $m_1 F$ einer $\chi^2(m_1)$ -Verteilung an, da $\text{plim}_{m_2 \rightarrow \infty} y_2/m_2 = 1$. Falls $t \sim t(m_2)$, dann gilt $t^2 \sim F(1, m_2)$.

2.9.3. Ergänzung: Transformationssätze**Transformationssätze**

- ‡ **Eindimensionaler Transformationssatz (change of variable):** Gegeben sei eine stetige Zufallsvariable $X \in \mathbb{R}$ mit Dichtefunktion $f_X(x) > 0$.

Gegeben sei weiter eine Zufallsvariable $Y = g(X)$, wobei die Funktion $g(\cdot)$ stetig und umkehrbar sei, so dass

$$x = g^{-1}(y). \quad (2.38)$$

Außerdem seien $g(\cdot)$ und $g^{-1}(\cdot)$ einmal differenzierbar.

Dann lässt sich für die Zufallsvariable Y die Dichtefunktion $f_Y(y)$ durch

$$f_Y(y) = \left| \frac{d}{dy} g^{-1}(y) \right| f_X(g^{-1}(y)) \quad (2.39)$$

berechnen (Casella & Berger 2002, Theorem 2.1.5).

- ‡ **Mehrdimensionaler Transformationssatz:** Gegeben sei ein stetiger $(m \times 1)$ -Zufallsvektor $\mathbf{x} \in \mathcal{X} \subset \mathbb{R}^m$ mit Dichtefunktion $f_{\mathbf{x}}(\mathbf{x}) > 0$. Weiter sei ein $(m \times 1)$ -Zufallsvektor

$$\mathbf{y} = \mathbf{g}(\mathbf{x}) = \mathbf{a} + \mathbf{A}\mathbf{x} \quad (2.40)$$

gegeben.

Ist \mathbf{A} invertierbar (siehe Casella & Berger (2002, Section 4.6, p. 185) für Bedingungen für den Fall, dass $\mathbf{g}(\mathbf{x})$ in (2.40) nichtlinear ist), gilt

$$\mathbf{x} = \mathbf{h}(\mathbf{y}) = \mathbf{A}^{-1}(\mathbf{y} - \mathbf{a})$$

und (siehe Abschnitt 6.2.2)

$$\frac{\partial \mathbf{x}}{\partial \mathbf{y}^T} = \frac{\partial \mathbf{h}(\mathbf{y})}{\partial \mathbf{y}^T} = \mathbf{A}^{-1}.$$

Dann lässt sich für den Zufallsvektor \mathbf{y} die Dichtefunktion $f_{\mathbf{y}}(\mathbf{y})$ durch

$$f_{\mathbf{y}}(\mathbf{y}) = \left| \frac{\partial \mathbf{h}(\mathbf{y})}{\partial \mathbf{y}^T} \right| f_{\mathbf{x}}(\mathbf{h}(\mathbf{y})) = |\mathbf{A}^{-1}| f_{\mathbf{x}}(\mathbf{A}^{-1}(\mathbf{y} - \mathbf{a})) \quad (2.41)$$

berechnen, wobei $\left| \frac{\partial \mathbf{h}(\mathbf{y})}{\partial \mathbf{y}^T} \right|$ die Determinante der Jacobi-Matrix $\frac{\partial \mathbf{h}(\mathbf{y})}{\partial \mathbf{y}^T}$ bezeichnet, siehe (1.10) für weitere Details. (Davidson 2000, Theorem B.9.2)

3. Konvergenz und Grenzwerte

In vielen Bereichen der Mathematik treten Konvergenzen auf, ohne dass bei der angewandten Benutzung klar wird, dass es sich bei den Konstrukten um Grenzwertprozesse handelt. Man nehme sich die Funktion $f(x) = x^2$. Die Ableitungsfunktion ist $f'(x) = 2x$, einfach durch algebraische Formeln „abgeleitet“. Der tatsächliche Ableitungsvorgang würde aber folgendermaßen aussehen:

Beispiel: Ableitung einer einfachen Funktion $f(x) = x^2$:

$$\begin{aligned} f'(x) &:= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\ &= \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} \\ &= \lim_{h \rightarrow 0} \frac{2xh + h^2}{h} = \lim_{h \rightarrow 0} \frac{2xh}{h} + \lim_{h \rightarrow 0} \frac{h^2}{h} = 2x + 0 = 2x \end{aligned}$$

Egal, ob nun Ableitung, Integral, Stetigkeit oder Funktionenfolgen, Grenzwerte kommen in ganz unterschiedlicher Form vor. Um nun die „neuen“, für die Ökonometrie wichtigen Formen von Konvergenz zu verstehen, werden die Standardfälle kurz wiederholt.

In der Mathematik und Wahrscheinlichkeitstheorie gibt es verschiedene Arten von Konvergenz, von denen wir die folgenden behandeln:

Übersicht: Arten von Konvergenz

1. Konvergenz von Zahlenfolgen - Grundgerüst aller Konvergenztheorien
2. Konvergenz von Funktionenfolgen
3. Konvergenz von Zufallsvariablenfolgen
 - Fast sichere Konvergenz
 - Konvergenz in Wahrscheinlichkeit
 - Konvergenz in Verteilung

Für die Ökonometrie ist insbesondere 3) mit deren Formen relevant. Zu deren Verständnis muss man 1) und 2) verstehen.

3.1. Konvergenz von Folgen

Sei (a_n) eine Folge reeller Zahlen, d. h., eine Abbildung $f : \mathbb{N} \rightarrow \mathbb{R}$ mit $n \mapsto a_n \in \mathbb{R}$ (man stelle sich $a_n := \frac{1}{n}$ vor). Man schreibt statt $f(n) = a_n$ auch oft $(a_n) = (a_n)_{n \in \mathbb{N}} = \{a_1, a_2, a_3, \dots\}$ für die Menge der Folgenglieder.

Beispiele:

$$f(n) = (a_n)_{n \in \mathbb{N}} = \left(\frac{1}{n}\right)_{n \in \mathbb{N}} = \left\{1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots\right\}.$$

$$f(n) = (a_n)_{n \in \mathbb{N}} = (n^2)_{n \in \mathbb{N}} = \{1, 4, 9, 16, \dots\}.$$

$$f(n) = a_n = \frac{f(x+1/n) - f(x)}{1/n} \stackrel{h:=1/n}{=} \frac{f(x+h) - f(x)}{h} \text{ für festes } x.$$

Konvergente Folge

Eine Folge (a_n) heißt konvergent in \mathbb{R} , falls es eine Zahl $a \in \mathbb{R}$, mit folgender Eigenschaft gibt:

- Für alle $\epsilon > 0$ gibt es ein Folgenglied indiziert mit $N \in \mathbb{N}$ (genauer $N(\epsilon)$, da es von ϵ abhängt), so dass

$$|a_n - a| < \epsilon \text{ für alle späteren Folgenglieder } n > N \text{ gilt.}$$

Die Zahl a heißt **Grenzwert** oder **Limes** der Folge und man schreibt

$$\lim_{n \rightarrow \infty} a_n = a \text{ oder } a_n \rightarrow a \text{ für } n \rightarrow \infty$$

Beachte: Im Konvergenzfall ist der Grenzwert der Folge eindeutig bestimmt!

Beispiel: $f(n) := (a_n) := \frac{50}{n} \cdot |\sin(0.1 \cdot n)|$. Die Folgenglieder sind in Abbildung 3.1 als Kreise eingezeichnet. Man vermutet nun, dass der Grenzwert $a = a_\infty = 0$ ist. Um dies zu zeigen, muss man für alle $\epsilon > 0$ beweisen, dass es ein Folgenglied $N(\epsilon)$ gibt, ab dem alle weiteren Folgenglieder im ϵ -Abstand von a liegen. Gibt man sich nun ein $\epsilon_1 = 1$ vor, so sieht man, dass die Folgenglieder 26 bis 43 bereits Abstand kleiner 1 von a_∞ haben, jedoch die Folgenglieder 44 bis 50 außerhalb dieses Bereichs liegen. Man wählt sich nun ein $N(\epsilon_1) = 51$ (oder ein höheres Folgenglied) und hofft, dass alle weiteren in diesem Abstand zum Grenzwert liegen und könnte dies mit weiteren, kleineren ϵ_i fortsetzen. Dies zeigt jedoch nicht die Konvergenz der Folge! Die Definition sagt deutlich „für alle“ $\epsilon > 0$! Ein Beweis würde folgendermaßen aussehen: Da man zeigen muss, dass es für alle ϵ gilt, gibt man sich ein kleines $\epsilon > 0$ vor, das beliebig klein werden kann, aber für einen

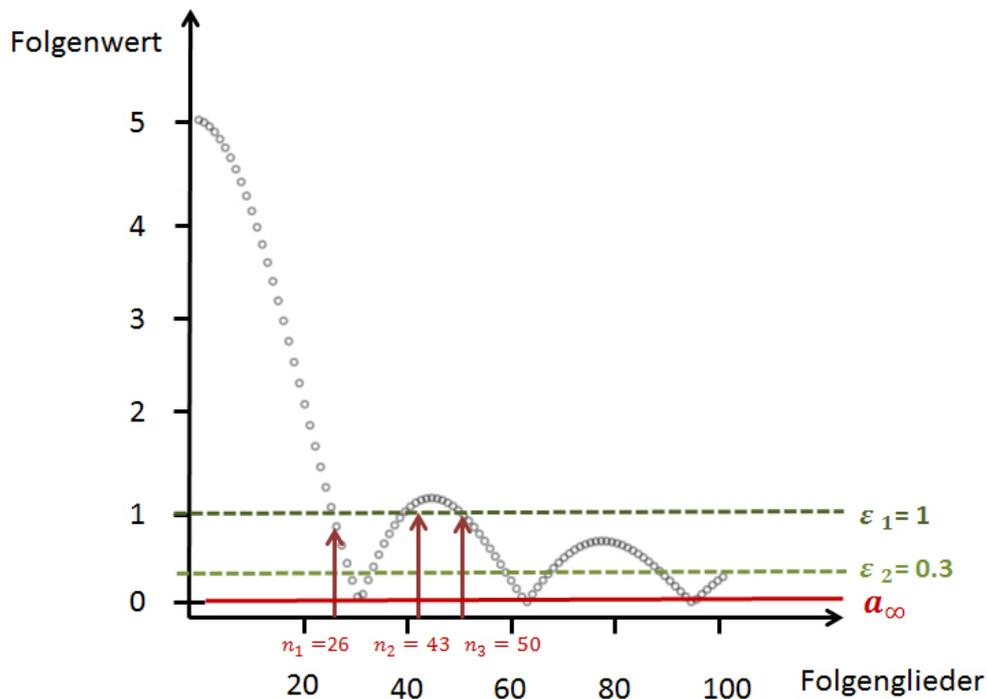


Abbildung 3.1.: Die ersten 100 Folgenglieder von $\frac{50}{n} \cdot |\sin(0.1 \cdot n)|$

Moment fest vorgegeben ist. Nun muss man zeigen, dass es ein Folgenglied $N(\epsilon)$ in Abhängigkeit des gerade vorgegebenen ϵ gibt, ab dem alle weiteren Folgenglieder a_n mit $n > N(\epsilon)$ im ϵ -Abstand von $a_\infty = 0$ liegen. Die Aufgabe besteht also im Finden dieses $N(\epsilon)$, so dass gilt:

$$|a_{N(\epsilon)} - 0| < \epsilon$$

Umstellen der Bedingung liefert das gewünschte Ergebnis:

$$|a_N| < \epsilon \iff \frac{50}{N} \cdot \underbrace{|\sin(0.1 \cdot N)|}_{\leq 1} < \epsilon \iff \frac{50}{N} < \epsilon \iff N > \frac{50}{\epsilon}$$

Gibt man sich nun ein ϵ vor, so weiß man, dass alle Folgenglieder mit Index größer als $\frac{50}{\epsilon}$ im ϵ -Bereich um den Grenzwert liegen. Ein Beweis würde nun folgendermaßen aussehen: Sei $\epsilon > 0$ beliebig klein, aber fest vorgegeben. Wähle dann den Folgenindex $N(\epsilon)$, so dass einerseits $N(\epsilon) > \frac{50}{\epsilon}$ und $N(\epsilon) \in \mathbb{N}$. Damit gilt für die nachfolgenden Folgenglieder mit Index $n \geq N(\epsilon)$:

$$\begin{aligned} |a_n - a| &\stackrel{a=0}{=} |a_n - 0| \stackrel{\text{Def.}}{=} \left| \frac{50}{n} \cdot \sin(0.1 \cdot n) \right| = \left| \frac{50}{n} \right| \cdot |\sin(0.1 \cdot n)| \\ &\stackrel{|\sin(0.1 \cdot n)| \leq 1}{\leq} \left| \frac{50}{n} \right| \stackrel{n \geq N(\epsilon)}{\leq} \left| \frac{50}{N(\epsilon)} \right| \stackrel{N(\epsilon) > \frac{50}{\epsilon}}{<} \left| \frac{50}{\frac{50}{\epsilon}} \right| = |\epsilon| \stackrel{\epsilon > 0}{=} \epsilon \end{aligned}$$

Insgesamt hat man damit gezeigt: Für beliebiges $\epsilon > 0$ gibt es ein $N(\epsilon) \in \mathbb{N}$, so dass für alle späteren Folgenglieder mit Index $n \geq N(\epsilon)$ gilt:

$$|a_n - a| < \epsilon$$

Rechenregeln von konvergenten Zahlenfolgen x_n, y_n :

- Falls $\lim_{n \rightarrow \infty} x_n = x$ und $\lim_{n \rightarrow \infty} y_n = y$, dann folgt auch:
 $\lim_{n \rightarrow \infty} x_n + \lim_{n \rightarrow \infty} y_n = \lim_{n \rightarrow \infty} (x_n + y_n)$.
- Falls $\lim_{n \rightarrow \infty} x_n = x$ und $\lim_{n \rightarrow \infty} y_n = y$, dann folgt auch:
 $\lim_{n \rightarrow \infty} x_n \cdot \lim_{n \rightarrow \infty} y_n = \lim_{n \rightarrow \infty} (x_n \cdot y_n)$.
- Falls $\lim_{n \rightarrow \infty} x_n = x$ und $\lim_{n \rightarrow \infty} y_n = y$ und $y \neq 0$, dann:
 $\lim_{n \rightarrow \infty} x_n / \lim_{n \rightarrow \infty} y_n = \lim_{n \rightarrow \infty} (x_n / y_n)$.

Bemerkung: Die Umkehrungen gelten im Allgemeinen nicht : Da $0 = (-1)^n + (-1)^{n+1}$ folgt aus $\lim_{n \rightarrow \infty} 0 = 0$ nicht, dass die Grenzwerte $\lim_{n \rightarrow \infty} (-1)^n$ und $\lim_{n \rightarrow \infty} (-1)^{n+1}$ existieren, was klarerweise nicht der Fall ist

3.2. Konvergenz von Funktionen

Während vorher eine skalare Folge definiert wurde, geht man nun einen Schritt weiter und betrachtet Funktionenfolgen bzw. -scharen $f_n(x)$ (man stelle sich $f_n(x) = x + \frac{1}{n}$ vor).

Für jedes feste $n \in \mathbb{N}$ ist der Ausdruck $f_n(x)$ eine Funktion $f_n(x) : \mathbb{D} \rightarrow \mathbb{R}$, wobei \mathbb{D} der Definitionsbereich der Funktion ist, und für jedes feste $x_0 \in \mathbb{D}$ ist der Ausdruck $f_n(x_0)$ eine Folge $\mathbb{N} \rightarrow \mathbb{R}$. Man definiert nun die (**punktweise** - im Gegensatz zur **gleichmäßigen** -) Konvergenz einer Funktionenfolge als den Grenzwert der Folge $f_n(x_0)$ für festes x_0 :

Punktweise Konvergenz

Eine Funktionenfolge $(f_n(x))$ heißt punktweise konvergent, falls für jeden festen Punkt $x_0 \in \mathbb{D}$ die Zahlenfolge $(f_n(x_0))_{n \in \mathbb{N}}$ konvergiert.

Durch

$$f(x) := \lim_{n \rightarrow \infty} f_n(x)$$

ist dann eine Funktion $f : \mathbb{D} \rightarrow \mathbb{R}$ definiert. Eigenschaften der f_n wie Stetigkeit, Integrierbarkeit, Differenzierbarkeit, also Eigenschaften die auch Grenzwerte verlangen, übertragen sich **im Allgemeinen nicht!** (deswegen der Begriff der gleichmäßigen Konvergenz).

Beispiel: $f_n(x) := x^n$ mit Definitionsbereich $\mathbb{D} = [0, 1]$. Einige Elemente der Folgenchar sind in Abbildung 3.2 zu sehen. Da für alle $n \in \mathbb{N}$ und eine Zahl $x \in (0, 1)$ gerade $x^n < x$ gilt, aber $1^n = 1$ ist, ist klar, dass der rot eingezeichnete Grenzwert der Funktionenfolge $f_n(x)$ folgendermaßen aussieht

$$f(x) = \begin{cases} 0 & \text{falls } x < 1 \\ 1 & \text{falls } x = 1 \end{cases}$$

Bewiesen wird dies wieder mit der $\epsilon - N$ -Definition für Zahlenfolgen von oben, wobei man dies für „jedes“ $x \in \mathbb{D} = [0, 1]$ einzeln macht. Hier reduziert es sich aber auf eine Fallunterscheidung von $x \in [0, 1)$ und $x = 1$. Auffallend hierbei ist, dass jedes Folgenglied $f_n(x)$ eine stetige Funktion darstellt, während der Grenzwert eine Unstetigkeitsstelle an der Stelle 1 besitzt.

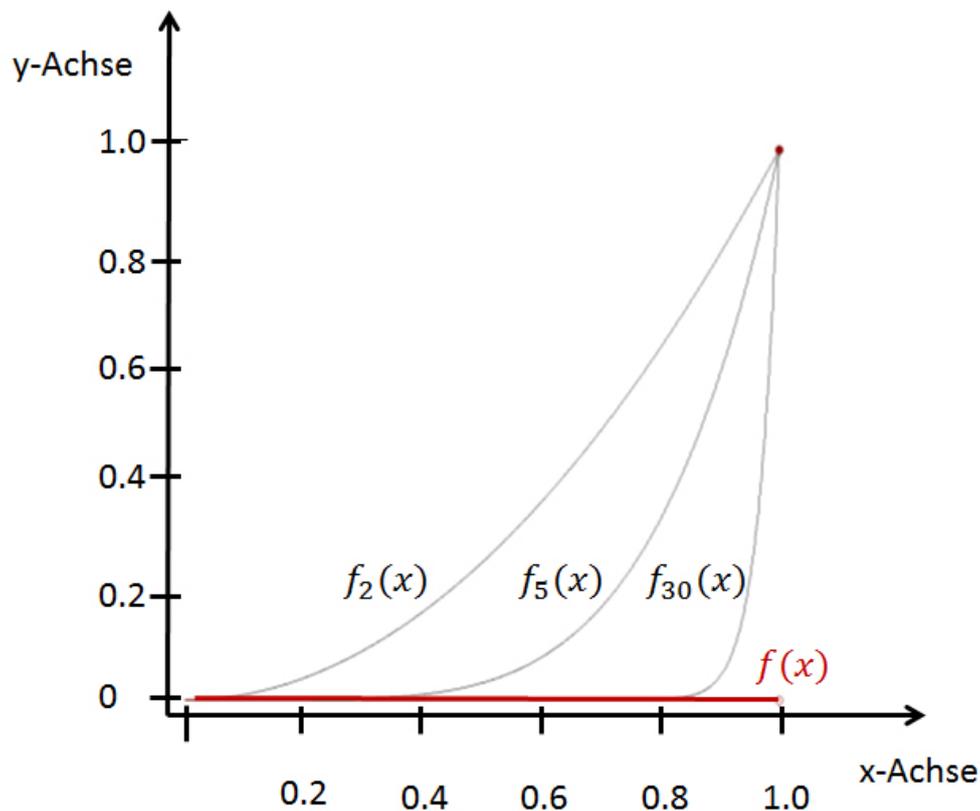


Abbildung 3.2.: Drei Folgenglieder der Funktionenchar $f_n(x) = x^n$

3.3. Fast sichere Konvergenz

Betrachtet man Zufallsvariablen als Funktionen nach \mathbb{R} , für die es eine Wahrscheinlichkeitsfunktion zur Bewertung gibt, so kann man eine spezielle punktweise Konvergenz für Zufallsvariablen bilden, die fast sichere Konvergenz. Sei $X_n(\omega)$ eine Zufallsvariablenfolge. $X_n(\omega)$ konvergiert fast sicher gegen $X(\omega)$, in Zeichen $X_n \xrightarrow{a.s.} X$ (almost surely), falls die Wahrscheinlichkeit der Menge der Ergebnisse ω , an denen sich $\lim_{n \rightarrow \infty} X_n$ und X unterscheiden, 0 ist. In formaler Schreibweise:

$$\begin{aligned} X_n \xrightarrow{a.s.} X &\iff P\left(\lim_{n \rightarrow \infty} X_n = X\right) = P\left(\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}\right) = 1 \\ &\iff P\left(\{\omega \in \Omega \mid \lim_{n \rightarrow \infty} X_n(\omega) \neq X(\omega)\}\right) = 0 \end{aligned}$$

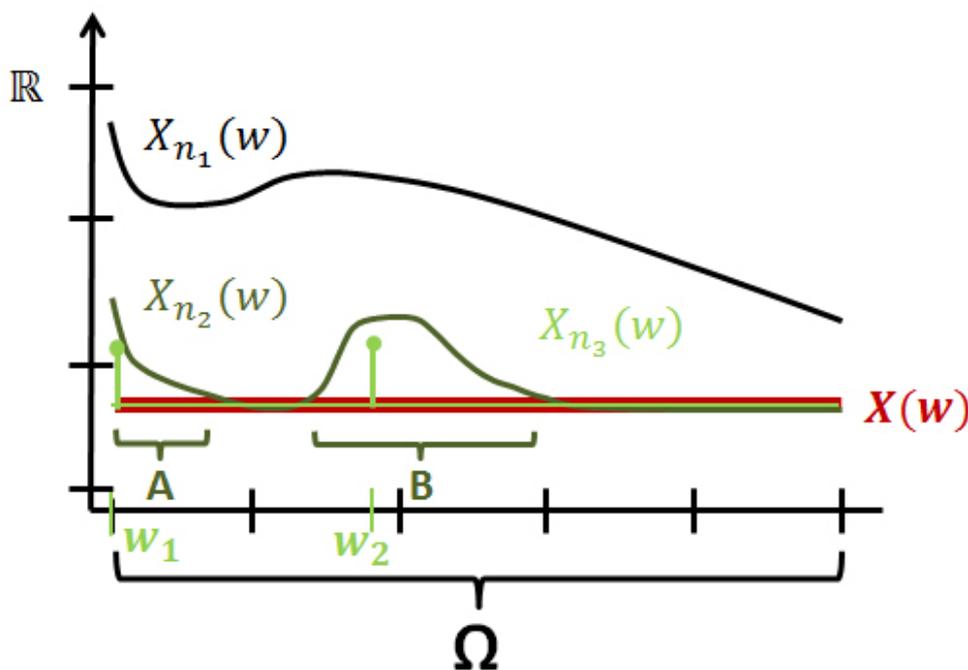


Abbildung 3.3.: Drei Zufallsvariablenfolglglieder mit $n_1 < n_2 < n_3$

Während bei der punktweisen Konvergenz gefordert wurde, dass für jedes $x \in \mathbb{R}$ $\lim_{n \rightarrow \infty} f_n(x) = f(x)$ gelten muss, hat man mit der Wahrscheinlichkeit die Möglichkeit zu fordern, dass sich der Limes der Folge und der Grenzwert „um Wahrscheinlichkeit 0“ unterscheiden. Man lässt also zuerst die Folge gegen unendlich laufen und betrachtet danach die Wahrscheinlichkeit, an denen sich beide unterscheiden. Betrachtet man im Bild rechts das Folglglied $X_{n_1}(\omega)$, so unterscheidet es sich in jedem ω von $X(\omega)$, und da $P(\Omega) = 1$ ist dies noch keine Annäherung.

Nimmt man ein späteres Folgenglied $X_{n_2}(\omega)$, so unterscheiden sich $X_{n_2}(\omega)$ und $X(\omega)$ nur noch in $A \cup B$ mit Wahrscheinlichkeit $P(A \cup B) = p < 1$ und $X_{n_3}(\omega)$ und $X(\omega)$ nur noch in $\{\omega_1, \omega_2\}$. Geht man davon aus, dass X eine stetige Zufallsvariable ist, Ω also überabzählbar viele Elemente hat, so ist die Wahrscheinlichkeit $P(\{\omega_1, \omega_2\}) = P(\{\omega_1\}) + P(\{\omega_2\}) = 0$ und (der Grenzwert) $X_{n_3}(\omega)$ und $X(\omega)$ unterscheiden sich in einer Menge $\{\omega \in \Omega \mid X_{n_3}(\omega) \neq X(\omega)\}$ mit Wahrscheinlichkeit 0. Man sagt, sie sind fast überall gleich bzw. die so skizzierte Folge $X_n(\omega)$ konvergiert fast sicher gegen $X(\omega)$.

Rechenregeln von fast sicher konvergenten Zufallsvariablenfolgen X_n, Y_n :

- Falls $X_n \xrightarrow{a.s.} X$ und $Y_n \xrightarrow{a.s.} Y$, dann folgt auch: $X_n + Y_n \xrightarrow{a.s.} X + Y$.
- Falls $X_n \xrightarrow{a.s.} X$ und $Y_n \xrightarrow{a.s.} Y$, dann folgt auch: $X_n \cdot Y_n \xrightarrow{a.s.} X \cdot Y$.

Bemerkung: Addition und Multiplikation erhalten die fast sichere Konvergenz.

3.4. Konvergenz in Wahrscheinlichkeit

2. Idee: Betrachte nicht mehr die Wahrscheinlichkeit des Limes, sondern den Limes der Wahrscheinlichkeit der Differenzmengen.

Konvergenz in Wahrscheinlichkeit (convergence in probability)

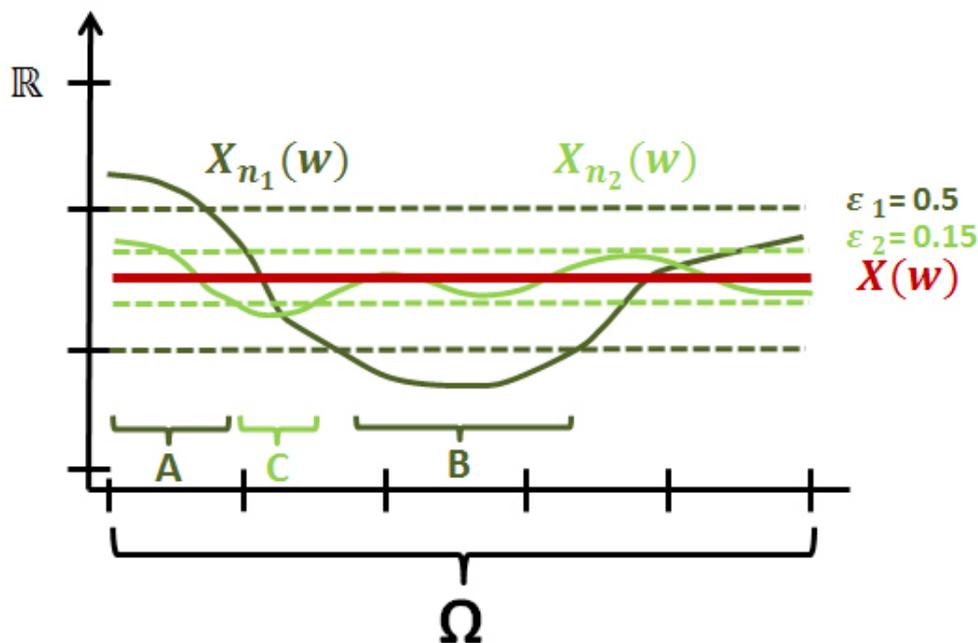
Sei $X_n(\omega)$ eine Folge von Zufallsvariablen. X_n konvergiert in Wahrscheinlichkeit gegen X , in Zeichen $X_n \xrightarrow{P} X$ (in probability), falls für alle $\epsilon > 0$ gilt, dass

$$\begin{aligned} \lim_{n \rightarrow \infty} p_n &:= \lim_{n \rightarrow \infty} P(|X_n - X|_{\mathbb{R}} > \epsilon) = 0 \\ \iff \lim_{n \rightarrow \infty} P(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)|_{\mathbb{R}} > \epsilon\}) &= 0 \end{aligned}$$

Während man vorher zuerst den Limes der Zufallsvariablenfolge bildete und dann die Unterschiede zum Grenzwert mit der Wahrscheinlichkeit bewertete, geht man nun anders vor. Für jedes Folgenglied $X_n(\omega)$ und für jedes ϵ kann man die Wahrscheinlichkeit $p_n := P(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)| > \epsilon\})$ berechnen und bekommt somit eine Zahlenfolge p_n auf $[0, 1]$. Geht diese Zahlenfolge gegen 0, so liegt X_n mit Wahrscheinlichkeit 1 in einem ϵ -Band um X und man sagt, dass X_n in Wahrscheinlichkeit gegen X konvergiert, in Zeichen $X_n \xrightarrow{P} X$ oder $\text{plim}_{n \rightarrow \infty} X_n = X$. Formal ausgedrückt: Eine Zufallsvariablenfolge X_n konvergiert in Wahrscheinlichkeit gegen X , falls für alle $\epsilon > 0$ gilt, dass

$$\lim_{n \rightarrow \infty} p_n := \lim_{n \rightarrow \infty} P(|X_n - X|_{\mathbb{R}} > \epsilon) = \lim_{n \rightarrow \infty} P(\{\omega \in \Omega \mid |X_n(\omega) - X(\omega)|_{\mathbb{R}} > \epsilon\}) = 0$$

Betrachtet man Abbildung 3.4 mit zwei Folgengliedern und zwei verschiedenen ϵ , so macht dies keine Aussage mehr über die Identität zweier Zufallsvariablen. Es macht nur noch eine

Abbildung 3.4.: Zwei Zufallsvariablenfolgglieder mit $n_1 < n_2$

Aussage darüber, dass wenn man es genauer wissen will (ϵ kleiner macht), man ein Folgglied finden wird, so dass die Wahrscheinlichkeit der Ergebnisse ω , bei denen der Abstand größer als ϵ ist, 0 ist. Im Beispiel $X_{n_1}(\omega)$ und ϵ_1 ist es gerade die Menge $A \cup B$, in denen sich $X_{n_1}(\omega)$ von $X(\omega)$ um mehr als ϵ_1 unterscheiden. Man könnte nun $X_{n_2}(\omega)$ mit dem gleichen ϵ_1 benutzen; da $X_{n_2}(\omega)$ im ϵ_1 -Band liegt, ist die Wahrscheinlichkeit für Ergebnisse außerhalb 0. Verengt man nun das Band auf ϵ_2 , so unterscheiden sich $X_{n_2}(\omega)$ und $X(\omega)$ nur noch in C . Hat die Menge C Wahrscheinlichkeit 0, so müsste man keinen neuen Index finden, ist die Wahrscheinlichkeit ungleich 0, so findet man wieder ein späteres Folgglied, ab dem die Menge der „Ausreißer“ Wahrscheinlichkeit 0 hat.

Rechenregeln von in Wahrscheinlichkeit konvergenten Zufallsvariablenfolgen X_n, Y_n :

- Falls $X_n \xrightarrow{P} X$ und $Y_n \xrightarrow{P} Y$, dann folgt auch: $X_n + Y_n \xrightarrow{P} X + Y$.
- Falls $X_n \xrightarrow{P} X$ und $Y_n \xrightarrow{P} Y$, dann folgt auch: $X_n \cdot Y_n \xrightarrow{P} X \cdot Y$.

Bemerkung: Addition und Multiplikation erhalten die Konvergenz in Wahrscheinlichkeit.

Konvergenz in Wahrscheinlichkeit für Zufallsvektoren

Es bezeichne \mathbf{y}^n einen $(n \times 1)$ -Zufallsvektor, dessen Dimension mit n variiert.

Eine Vektorfunktion $\mathbf{a}_n : \mathbb{R}^n \rightarrow \mathbb{R}^m : \mathbf{a}_n := \mathbf{a}(\mathbf{y}^n)$ konvergiert in Wahrscheinlichkeit gegen $\mathbf{a}_0 \in \mathbb{R}^m$, falls

$$\lim_{n \rightarrow \infty} P(\|\mathbf{a}(\mathbf{y}^n) - \mathbf{a}_0\|_{\mathbb{R}^m} < \epsilon) = 1.$$

Bemerkung: $\|\cdot\|_{\mathbb{R}^m} : \mathbb{R}^m \rightarrow \mathbb{R}$ (vgl. Definition von oben).

Beispiel: $\hat{\mu} : \mathbb{R}^n \rightarrow \mathbb{R} : \hat{\mu}(\mathbf{y}^n) = \frac{1}{n} \sum_{t=1}^n y_t$. Hier ist also $m = 1$. Bei $\hat{\beta}$, dem KQ-Schätzer mit k Regressoren ist $m = k$.

Rechenregeln für in Wahrscheinlichkeit konvergierende Zufallsvektoren

Seien $\{\mathbf{x}_n\}$ und $\{\mathbf{y}_n\}$ Folgen von Zufallsvektoren. Falls $\text{plim}_{n \rightarrow \infty} \mathbf{x}_n$, und $\text{plim}_{n \rightarrow \infty} \mathbf{y}_n$, dann gilt:

$$\text{plim}_{n \rightarrow \infty} (\mathbf{x}_n \pm \mathbf{y}_n) = \text{plim}_{n \rightarrow \infty} \mathbf{x}_n \pm \text{plim}_{n \rightarrow \infty} \mathbf{y}_n, \quad (3.1a)$$

$$\text{plim}_{n \rightarrow \infty} (\mathbf{x}_n^T \mathbf{y}_n) = (\text{plim}_{n \rightarrow \infty} \mathbf{x}_n)^T (\text{plim}_{n \rightarrow \infty} \mathbf{y}_n), \quad (3.1b)$$

Diese Regeln gelten auch, wenn die Zufallsvektoren durch Matrizen mit Zufallsvariablen mit entsprechenden Eigenschaften ersetzt werden.

3.5. Konvergenz in Verteilung

3. Idee: Betrachte nicht mehr die Zufallsvariablen als Funktionen, sondern betrachte nur noch die Verteilungsfunktionen der Zufallsvariablen.

Konvergenz in Verteilung (convergence in distribution)

Sei $X_n(\omega)$ eine Folge von Zufallsvariablen. X_n konvergiert in Verteilung gegen X , in Zeichen $X_n \xrightarrow{d} X$ (in distribution), falls für die Funktionenfolge der Verteilungen F_n von X_n und der Verteilung F von X gilt, dass

$$\lim_{n \rightarrow \infty} F_n = F \text{ punktweise}$$

Beispiel: Es sei $\{X_n\}$ die weiter oben definierte Folge von Zufallsvariablen. Man erinnere sich: $X_n \xrightarrow{P} X$, wobei $X \sim N(\mu, \sigma^2)$ ist. Sei nun Z eine normalverteilte Zufallsvariable mit Erwartungswert μ und Varianz σ^2 . Dann gilt $X_n \xrightarrow{d} Z$. Damit haben X und Z die gleiche Verteilung, sind aber verschiedene Zufallsvariablen!

Zusammenhang zwischen den Konvergenzkonzepten

$$X_n \xrightarrow{ptw.} X \implies X_n \xrightarrow{a.s.} X \implies X_n \xrightarrow{P} X \implies X_n \xrightarrow{d} X. \quad (3.2)$$

(Ein Beispiel, weshalb die Umkehrung des dritten Folgepfeils nicht gilt, findet sich im

BA-Modul Zeitreihenökonomie, Abschnitt 5.1.4., ein Beispiel für die Umkehrung des zweiten Folgepfeils im Beispiel der gleitenden Hügel.)

Theorem über stetige Abbildungen (Continuous Mapping Theorem) (CMT)

- Sei $h(\cdot)$ eine stetige Funktion.

Falls $X_n \xrightarrow{a.s.} X$, dann gilt $h(X_n) \xrightarrow{a.s.} h(X)$.

Falls $X_n \xrightarrow{p} X$, dann gilt $h(X_n) \xrightarrow{p} h(X)$.

Falls $X_n \xrightarrow{d} X$, dann gilt $h(X_n) \xrightarrow{d} h(X)$.

Bemerkung: Stetige Transformationen erhalten die Konvergenzkonzepte.

- Für Folgen von $(k \times 1)$ -Zufallsvektoren \mathbf{x}_n gilt entsprechend:
Gegeben sei eine stetige vektorwertige Funktion $\mathbf{h} : \mathbb{R}^k \rightarrow \mathbb{R}^m$, dann gilt für $\star \in \{a.s., p, d\}$

$$\text{Falls } \mathbf{x}_n \xrightarrow{\star} \mathbf{x}, \text{ dann gilt } \mathbf{h}(\mathbf{x}_n) \xrightarrow{\star} \mathbf{h}(\mathbf{x}). \quad (3.3)$$

(Vgl. z. B. [Vaart \(1998, Theorem 3.2\)](#).)

Betrachten Sie zur Verdeutlichung die Funktion

$$f(x) = \begin{cases} 0 & x < 2 \\ 1 & x \geq 2 \end{cases}$$

und die Zahlenfolge $a_n = 2 - \frac{1}{n}$. Dann ist offensichtlich $\lim_{n \rightarrow \infty} a_n = 2$ monoton von links. Man sieht nun schnell, dass der Grenzwert der Funktionswerte ungleich dem Funktionswert des Grenzwerts ist:

$$\lim_{n \rightarrow \infty} f(a_n) \stackrel{a_n < 2 \forall n}{=} \lim_{n \rightarrow \infty} 0 = 0 \neq 1 = f(2) = f(\lim_{n \rightarrow \infty} a_n)$$

Bisher wurde noch nicht untersucht, auf welche Art und Weise Konvergenz in Verteilung beibehalten wird, sehr berühmt ist dabei folgendes Theorem:

Slutzky's Theorem

Sei $\mathbf{x}_n \xrightarrow{d} \mathbf{x}$ und $\mathbf{y}_n \xrightarrow{p} \mathbf{c}$ mit $\mathbf{c} \in \mathbb{R}^p$ konstant. Dann gilt

$$\mathbf{y}_n^T \mathbf{x}_n \xrightarrow{d} \mathbf{c}^T \mathbf{x}. \quad (3.4)$$

Vgl. [Vaart \(1998, Theorem 3.6\)](#).

Das Theorem gilt auch, falls \mathbf{y} und \mathbf{c} durch entsprechend dimensionierte Matrizen ersetzt werden.

Hinweis: Der Term Slutsky's Theorem wird in der statistischen und ökonometrischen Literatur nicht einheitlich verwendet. So bezeichnet Davidson (2000) in seinem Theorem 3.1.3 als Slutsky's Theorem die Aussage (3.3) für Skalare und Konvergenz in Wahrscheinlichkeit.

Fazit:

- Konvergenzbegriffe, die unterschiedliche „Abstände“ Eigenschaften von (Zufalls-)Variablen messen (und deren Zusammenhänge in und unter ihnen).
- Für ökonometrische Aussagen sind im Allgemeinen nur die beiden „letzten“, Konvergenz in Wahrscheinlichkeit und in Verteilung, interessant.
- Für ein besseres Verständnis der ökonometrischen Theorie benötigt man auch die Kenntnis der anderen Konvergenzbegriffe.

3.6. Beispiel der gleitenden Hügel: Konvergenz in Wahrscheinlichkeit $\not\Rightarrow$ Fast sichere Konvergenz

Betrachtet man anfangs der Einfachheit halber die doppelt indizierte Zufallsvariablenfolge X_{nk} auf $\Omega = [0, 1]$, wobei $n, k \in \mathbb{N}$, $k \leq n$ und $I_{[a,b]}(\omega) := \begin{cases} 1 & \text{falls } \omega \in [a, b] \\ 0 & \text{falls } \omega \notin [a, b] \end{cases}$

$$\begin{aligned} X_{11} &= I_{[0,1]} \\ X_{21} &= I_{[0,1/2]} & X_{22} &= I_{[1/2,1]} \\ X_{31} &= I_{[0,1/3]} & X_{32} &= I_{[1/3,2/3]} & X_{33} &= I_{[2/3,1]} \\ &\dots \end{aligned}$$

so erkennt man schnell deren Struktur. Die Graphen dieser Zufallsvariablen sind „gleitende Hügel“, die mit wachsendem n immer schmaler werden (n ist die Anzahl der Intervalle, in denen man das Einheitsintervall unterteilt, k das Intervall, in dem man sich befindet). Gemäß der lexikographischen Ordnung kann man aus X_{nk} eine Zufallsvariable Y_n erzeugen:

$$Y_1 = X_{11}, Y_2 = X_{21}, Y_3 = X_{22}, Y_4 = X_{31}, Y_5 = X_{32}, \dots$$

Damit erkennt man:

- Y_n konvergiert in Wahrscheinlichkeit gegen 0, denn für beliebiges $\epsilon \in (0, 1)$ gilt $P(|X_{nk} - 0| > \epsilon) = P(X_{nk} > \epsilon) \stackrel{\substack{n \text{ gleichgro\ss e} \\ \text{Intervalle}}}{\cong} \frac{1}{n} \xrightarrow{n \rightarrow \infty} 0$ (in Worten: die Wahrscheinlichkeit sich auf einem Hügel zu befinden geht gegen 0)
- Y_n konvergiert nicht fast sicher gegen 0, denn die Zufallsvariablenfolge Y_n konvergiert an keinem Punkt punktweise (konvergiert sie nirgends punktweise, so trivialerweise auch nicht fast sicher, da dabei die Menge aller Ausreißer Wahrscheinlichkeit 0 haben): Zu zeigen: für

alle $\omega \in \Omega$ existiert $\epsilon > 0$, so dass für alle $N \in \mathbb{N}$ ein $n > N$ existiert mit $|Y_n - 0| = Y_n > \epsilon$. Sei $\omega \in \Omega$ fest, aber beliebig, $\epsilon = 0.5$, $N \in \mathbb{N}$ fest, aber beliebig und ohne Einschränkung gilt $Y_N(\omega) = 0$. Sei $Y_N = X_{n'k'}$ und ohne Einschränkung $\omega = \frac{k^*}{n^*} \notin [k'/n', (k'+1)/n']$ (falls irrational, nutze eine rationale Annäherung; falls $k^* = 1$ oder $k^* = n^*$, trivial). Dann liegt ω in jedem Intervall $[\omega - 1/r, \omega + 1/r] = [\frac{rk^* - n^*}{rn^*}, \frac{rk^* + n^*}{rn^*}]$ für beliebiges $r \in \mathbb{N}$. Wähle nun r mit dem Archimedischen Axiom, so dass einerseits $rn^* > n'$ und andererseits $[\omega - 1/r, \omega + 1/r] \subset [0, 1]$. Definiere $\tilde{k} = rk^* - n^* \in \mathbb{N}$ (wegen „andererseits“), dann ist $Y_{\hat{n}}(\omega) := X_{rn^*, \tilde{k}}(\omega) = 1 > \epsilon = 0.5$ für $\hat{n} > N$.

Siehe z. B. [Casella & Berger \(2002, Example 5.5.8, p. 234-5\)](#).

Teil II.

Ökonometrische Methoden

4. Einführung

Ziel dieses Kurses:

Erlernen der ökonometrischen Grundlagen in Theorie und Praxis, die für eine sorgfältige empirische Analyse ökonomischer Fragestellungen relevant sind.

Beispiele ökonomischer Fragen, die eine empirische Analyse erfordern:

- Führt eine Reduktion der Klassengröße (Schule, Universität) zu besseren Lernergebnissen? (Vgl. [Stock & Watson 2007](#), Section 1.1.)
- Wirkung von Fortbildungsmaßnahmen der Bundesanstalt für Arbeit (BA): Erhöht sich hierdurch das verbleibende Lebensinkommen, reduziert sich die Arbeitslosigkeitsdauer?
- Kommt es in den kommenden Jahren zu einer höheren Inflation?
- Welche Faktoren beeinflussen die länder-spezifischen Importe nach Deutschland?
- Verstehen dynamischer Prozesse
- Welche Faktoren beeinflussen die Länge und Intensität von Konjunkturzyklen?
- Welche Faktoren beeinflussen das Wirtschaftswachstum, die Einkommens- und Vermögensverteilung?

In den genannten Beispielen geht es häufig um das Bestimmen von kausalen Variablen.

4.1. Aussagen über kausale Zusammenhänge

Die Kenntnis kausaler Zusammenhänge ist **Voraussetzung für die Evaluation** von geplanten oder durchgeführten (wirtschaftspolitischen, betrieblichen, etc.) Maßnahmen.

Kausalität

- Gängiges Verständnis: “causality means that a specific action leads to a specific, measurable consequence ([Stock & Watson 2007](#), p. 8)

- Genau betrachtet: Die Wirkung einer Aktion ist im Einzelfall **immer** unbekannt, denn
 - ist Individuum/Einheit/Variable i von einer Aktion / Maßnahme betroffen, kann man für i nur das Ergebnis mit der Aktion beobachten, nicht jedoch das Ergebnis, wenn die Aktion nicht durchgeführt worden wäre.
 - Ist alternativ Individuum/Einheit/Variable i von dieser Aktion nicht betroffen, kennt man nur das Ergebnis ohne Aktion, aber nicht das Ergebnis für dieses i , wenn es von der Aktion betroffen gewesen wäre.

Der jeweils nicht eingetretene Fall ist der **kontrafaktische Zustand** und würde die Antwort auf eine “**Was wäre wenn?**“-Frage liefern.

In der Sprache der Ökonometrie: **Der individuelle Erfolg** einer Maßnahme ist **immer unbeobachtbar**, da es er immer eine kontrafaktische Größe enthält.

- Messbar ist jedoch unter bestimmten Voraussetzungen die **durchschnittliche Wirkung einer Aktion** auf eine Gruppe von Individuen.
- **Definition von Kausalität:** Im Folgenden bezeichnen wir **eine Aktion oder Maßnahme als kausal**, wenn eine **durchschnittliche Wirkung** einer Aktion messbar ist.

A “**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 & Watson 2007, p. 9)”.

- **Ceteris paribus:** Werden alle anderen kausalen Variablen außer die interessierende Variable konstant gehalten und nur die interessierende Aktion durchgeführt, betrachtet man das Ergebnis der Aktion **ceteris paribus (c. p.)**.
- Beachte: Für die Existenz eines kausalen Effekts ist Voraussetzung, dass eine Aktion auch auf einzelne Individuen einen Effekt hat.

Messbarkeit von kausalen Zusammenhänge

Eine Quantifizierung der durchschnittlichen Wirkung einer Aktion, also die Quantifizierung eines **kausalen Zusammenhangs** auf Basis eines ökonometrischen Modells ist nur möglich, wenn

1. ein **ideales kontrolliertes Zufallsexperiment** durchgeführt werden kann oder wenn
2. eine Stichprobe von einem **Quasi-Experiment** vorliegt und spezifische **Identifikationsannahmen** getroffen werden oder wenn
3. das ökonometrische Modell aus einem kausal interpretierbaren (**ökonomischen**) **Modell** abgeleitet wurde, das eine für die Fragestellung hilfreiche Approximation der Realität darstellt.

In diesem Kurs betrachten wir nur Fall 3.

Die Fälle 1. und 2. sind genauer in den Abschnitten 2.2 und 2.3 in **Weiterführende Fragen der Ökonometrie** beschrieben und werden in der MA-Veranstaltung **Methoden der Politikevaluation** detailliert behandelt. Sie spielen in der **Evaluationsforschung** eine große Rolle.

Kontrolliertes Zufallsexperiment

Eine Gesamtgruppe von Individuen wird in eine **Maßnahmengruppe / Teilnehmergruppe (treatment group)** und eine **Kontrollgruppe** unterteilt. Letztere enthält alle Individuen, die an einer Maßnahme nicht teilnehmen. Das zentrale Merkmal eines kontrollierten Zufallsexperiments ist, dass die Teilnehmer der **Maßnahmengruppe zufällig** ausgewählt werden.

Beispiel: Klassengröße Zu Beginn eines Schuljahres werden Schüler (und Lehrer, etc.) einer Schule zufällig auf kleine und große Klassen aufgeteilt. Auf diese Weise wird vermieden, dass SchülerInnen mit bestimmten Eigenschaften vornehmlich in einer Klassengröße zu finden sind.

Quasi-Experiment / Natürliches Experiment

Häufig ist es aus rechtlichen, ethischen, ökonomischen oder anderen Gründen nicht möglich, ein kontrolliertes Zufallsexperiment durchzuführen.

Unter bestimmten **zusätzlichen** Annahmen an die Grundgesamtheit können daraus gewonnene Stichprobenbeobachtungen behandelt werden, als ob ein kontrolliertes Zufallsexperiment vorliegen würde.

Beachte: Für viele ökonomische Fragestellungen, z. B. makroökonomischer Art, sind weder kontrollierte Zufallsexperimente durchführbar, noch können natürliche Experimente beobachtet werden. Dann kann die kausale Interpretation eines ökonometrischen Modells nur auf Grundlage eines ökonomischen Modells, das dem ökonometrischen Modell zugrundeliegt, erfolgen.

Simultanität und Kausalität

Beobachtet man zwei Variablen y und x in der gleichen Zeitperiode, bspw. für das Jahr 2014, dann ist es möglich, dass

1. beide Variable eine **simultane Beziehung** ausweisen, sich also gegenseitig und **gleichzeitig**, beeinflussen

$$x \longleftrightarrow y$$

oder

2. eine Variable für die andere **kausal** ist, bspw. x für y ,

$$x \longrightarrow y$$

oder

3. beide Variablen durch eine dritte Variable beeinflusst werden, sich aber nicht gegenseitig beeinflussen

$$z \longrightarrow \begin{cases} x \\ y \end{cases}, x \not\rightarrow y, y \not\rightarrow x$$

oder

4. kein Einfluss besteht

$$x \not\rightarrow y.$$

Beachte:

- Prinzipiell muss man bei einer empirischen Analyse mit mehreren potentiellen Variablen in einer Zeitperiode davon ausgehen, dass Simultanität, also Fall 1, vorliegt. Erst ein theoretisch oder statistisch begründeter Ausschluss einer Wirkungsrichtung ermöglicht Simultanität auszuschließen und Fall 2, genau eine kausale Beziehung, zu erhalten. Oder auch Fall 4 zu erhalten, dass keine kausale Beziehung vorliegt.
- Ob Fall 2. vorliegt, kann unter bestimmten Voraussetzungen statistisch getestet werden (siehe Veranstaltung **Methoden der Politikevaluation**).
- Eine statistische Überprüfung von Fall 4. kann mit Methoden der Modellselektion erfolgen, siehe Kapitel 10.

Beispiel: Einflussfaktoren von Importen

Ziel/Wissenschaftliche Fragestellung:

Ermittle die Faktoren, die die Importe nach Deutschland beeinflussen, und quantifiziere ihren Einfluss.

Erste Überlegungen: Welche Variablen einer Zeitperiode könnten relevant sein und welche Wirkungsrichtungen könnten zwischen diesen existieren?

Der $(m \times 1)$ -Vektor $\mathbf{s}_t \in \mathbb{R}^m$ enthält alle Variablen, die für die Analyse relevant sein könnten. Im Folgenden sind beispielhaft und unvollständig (!) aufgeführt:

- Humankapital des Exportlandes (s_1)
- Koloniale Vergangenheit beider Länder (s_2)
- Bruttoinlandsprodukt des Importlandes (s_3)
- BIP des Exportlandes (s_4)
- Entfernung zu Exportland (s_5)
- Fläche des Exportlandes (s_6)
- Offenheit in einem Land (s_7)
- Importe (s_8)
- unspezifiziert (s_9)

Abbildung 4.1 zeigt diese Variablen und unterstellt für alle Variablenpaare zunächst **Simultanität**.

Um die Zahl der simultanen Beziehungen zu reduzieren, erscheint folgende Annahme auf jeden Fall gerechtfertigt:

Annahme: Fläche und Entfernung werden nicht durch andere Variablen beeinflusst.

Man erhält Abbildung 4.2.

Eine weitere Reduktion simultaner Beziehungen erfolgt am Besten durch ein ökonomisches Modell, das allerdings weitere Annahmen erfordert. In Abschnitt 6.3 wird ein ökonomisches Modell dargestellt, das es ermöglicht, alle verbleibenden simultanen Beziehungen zwischen Importen (s_8) und den anderen Variablen entweder ganz zu eliminieren oder in eine kausale Wirkungsrichtung umzuwandeln.

Man erhält Abbildung 4.3, in der im Vergleich zur Abbildung 4.2 sich die Zahl und Art der Beziehungen zwischen den Importen und den anderen Variablen verändert hat, nicht jedoch die Zahl und Art der Beziehungen zwischen den potenziellen Einflussvariablen.

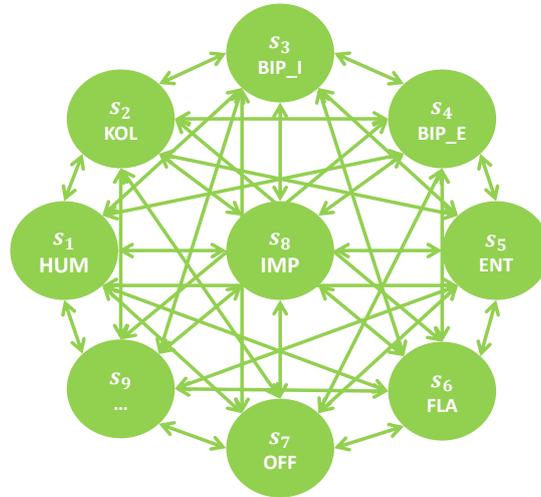


Abbildung 4.1.: Einflussfaktoren auf Handelsströme: mögliche simultane Beziehungen

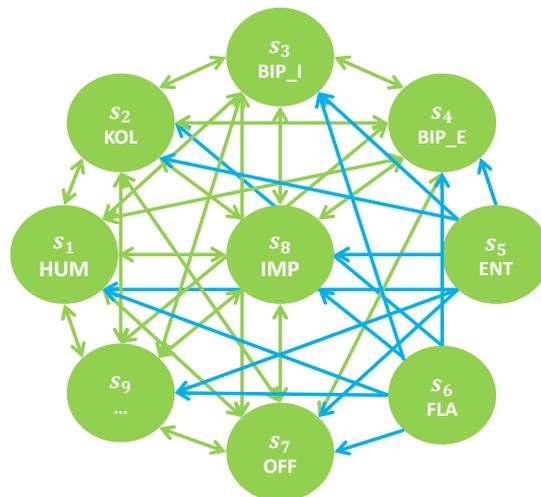


Abbildung 4.2.: Einflussfaktoren auf Handelsströme: erste Reduktion an simultanen Beziehungen: blaue Pfeile zeigen kausale Beziehungen an

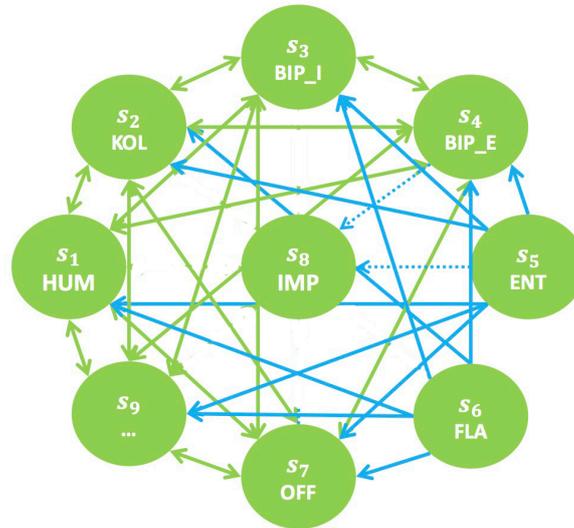


Abbildung 4.3.: Einflussfaktoren auf Handelsströme: kausale und simultane Beziehungen auf Basis eines ökonomischen Modells plus weitere relevante Einflussfaktoren; gestrichelte Pfeile stellen die relevanten kausalen Beziehungen dar (später Modell 2).

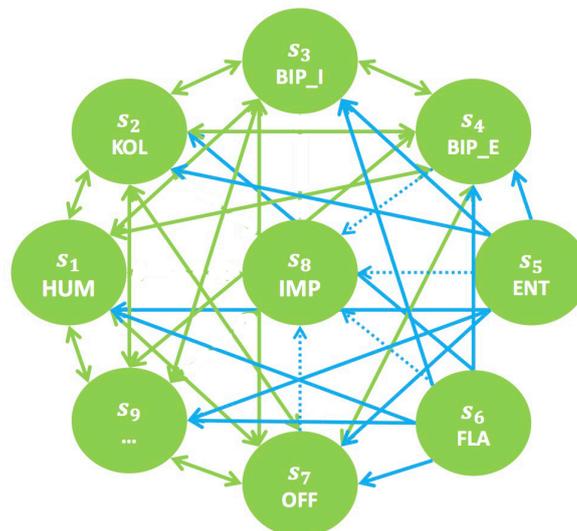


Abbildung 4.4.: Einflussfaktoren auf Handelsströme: kausale und simultane Beziehungen auf Basis eines ökonomischen Modells plus weitere relevante Einflussfaktoren; gestrichelte Pfeile stellen die relevanten kausalen Beziehungen dar (später Modell 4).

Um die Ausgangsfrage zu beantworten, müssen die verbleibenden kausalen Beziehungen in Abbildung 4.3 quantifiziert werden. Dies erfolgt im einfachsten Fall mit einem multiplen linearen Regressionsmodell.

In der Realität stellt das unterstellte ökonomische Modell vermutlich eine zu starke Vereinfachung dar, so dass weitere Einflussfaktoren geeignet berücksichtigt werden müssen. Ein Beispiel zeigt Abbildung 4.4. Wie dies effizient geschieht, ist ebenfalls Gegenstand dieses Kurses.

Multiplere lineares Regressionsmodell

- Um die (durchschnittliche) Wirkung einer Aktion, in obigem Beispiel eine Änderung der Wirtschaftskraft im Exportland auf die Importe quantifizieren zu können, ist es i. Allg. nicht ausreichend, nur diese beiden Variablen in einem Modell zu betrachten. Stattdessen müssen i. Allg. alle relevanten kausalen Variablen im Modell berücksichtigt werden.
- Hat man gut begründet, dass die Variablen z_1, \dots, z_{k-1} die Variable y kausal beeinflussen und ist man an der Quantifizierung des kausalen Effekts der Variable z_1 auf die Variable y interessiert und unterstellt einen linearen Zusammenhang (in den Parametern β_1, \dots, β_k), dann ergibt sich ein Beispiel eines **multiplen linearen Regressionsmodells**

$$y = \beta_1 + \beta_2 z_1 + \dots + \beta_k z_{k-1} + u. \quad (4.1)$$

- Die Variable u wird als **Fehlerterm** bezeichnet, der alle nicht modellierbaren / nicht modellierten Einflüsse enthält. Eine Möglichkeit, u genauer zu interpretieren, ist mit (5.29) gegeben.
- Die Variable y wird als **endogene Variable** bezeichnet, da sie durch das Modell erklärt wird / werden soll.
- Die für y als kausal postulierten Variablen z_j werden als **exogene Variablen** bezeichnet.

In den folgenden Kapiteln werden die Eigenschaften und Annahmen bzgl. dieser Variablen genauer spezifiziert werden.

- Eine Quantifizierung des kausalen Effekts von z_1 auf y erfolgt, indem β_2 bestimmt wird. Hierfür ist eine
 1. Stichprobe mit geeigneten Daten und
 2. ein geeignetes ökonometrisches Schätzverfahren notwendig.

Beispiel: Einflussfaktoren von Importen Aus Abbildung 4.3 folgt das

Modell mit $y = s_8, z_1 = s_4, z_2 = s_5$

$$\text{Importe} = \beta_1 + \beta_2 \text{BIP} + \beta_3 \text{Entfernung} + u \quad (4.2)$$

- In Kapitel 12 und 13 werden auch dynamische Modelle betrachtet. Hierbei wird immer unterstellt, dass verzögerte endogene Variablen, also Variablen, die in Vorperioden der betrachteten Periode liegen, kausal sind. Um eine allgemeine Notation zu ermöglichen, werden die Variablen auf der rechten Seite typischerweise mit x_j bezeichnet.

$$y = \beta_1 + \beta_2 x_2 + \dots + \beta_k x_k + u \quad (4.3)$$

Das **multiple lineare Regressionsmodell** (4.3) ist ein zentrales Werkzeug für alle im Folgenden genannten Ziele empirischer Analysen und steht deshalb im Mittelpunkt dieses Kurses.

- Die Variablen y und x haben in der Literatur verschiedene Bezeichnungen.

Namen für Variablen in Regressionsmodellen

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

Simultane Gleichungsmodelle

- In manchen Analysen besteht zwischen dem interessierenden y und anderen Variablen s_i eine simultane Beziehung. Dann liegen mehrere endogene Variablen vor, die explizit oder indirekt gemeinsam modelliert werden müssen.
- Ein einfaches Beispiel für ein simultanes Gleichungsmodell findet sich in Abschnitt 5.2, Gleichungen (5.8) und (5.9).
- Die Schätzung von simultanen Gleichgewichtsmodellen wird im MA-Modul **Fortgeschrittene Ökonometrie** behandelt.

Ziele empirischer Analysen

1. Identifikation und Quantifizierung von kausalen Zusammenhängen
2. Falsifizierung postulierter ökonomischer Zusammenhänge
3. Punkt- und Intervallprognosen

4. Analyse und Evaluation von erfolgten/geplanten Maßnahmen (wirtschaftspolitisch, betrieblich, etc.)
5. Bewertung von Unsicherheit und Risiko

4.2. Was ist Ökonometrie?

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 & Watson 2007, p. 7)”,
- bietet 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, (siehe u. a. Kapitel 11)
- erlaubt die **Quantifizierung** der **Risiken** von Vorhersagen, Entscheidungen und sogar ihrer eigenen Analyse, (siehe u. a. Abschnitt 9.3 und folgende)
- erlaubt die **Quantifizierung** von **kausalen Zusammenhängen**, die sich aus einem ökonomischem Modell ergeben.

Grundsätzlich:

- Quantitative Antworten beinhalten immer **Unsicherheit**. Unsicherheit besteht bezüglich:
 - des “wahren” (datengenerierenden) Mechanismus,
 - der Auswahl der Variablen in der empirischen Analyse,
 - der Messung der Variablen,
 - der Wahl des ökonometrischen Modells,
 - der statistischen Qualität des Schätz- oder Prognoseverfahrens.
- Zur Quantifizierung von **Unsicherheit** ist der Werkzeugkasten der **Wahrscheinlichkeitstheorie** sehr nützlich, aber nicht nur dafür

4.3. Bestandteile einer empirischen Analyse

Eine empirische Analyse sollte einer strukturierten Vorgehensweise folgen, die im Verlauf des Kurses begründet werden wird. Sie ist wie folgt gegliedert:

I. Ökonomischer Analyseteil

1. Wissenschaftliche Fragestellung

- Sorgfältige Formulierung der interessierenden Fragestellung bzw. des Problems.

2. Ökonomisches Modell

- Spezifizierung eines ökonomischen Modells.
- Identifizieren von kausalen und simultanen Beziehungen.
- Gewinnen von Hypothesen, die empirisch überprüft werden sollen.
- Interpretation von Modellparametern.

3. Datenverfügbarkeit

- Welche Daten sind hinsichtlich des ökonomischen Modells erforderlich und liegen bereits vor, bzw. lassen sich beschaffen?

II. Ökonometrischer Modellierungsprozess

1. Auswahl einer Klasse ökonometrischer Modelle

- Berücksichtigen der Variablen aus dem ökonomischen Modell und deren Verfügbarkeit.
- Berücksichtigen der funktionalen Zusammenhänge aus dem ökonomischen Modell bzw. Approximation derselben.
- Überlegen, ob datengenerierender Mechanismus (DGP) in Modellklasse enthalten sein könnte.
- Ggf. formulieren statistisch überprüfbarer Hypothesen bezüglich des DGPs.
- Wahl von Schätzmethoden mit günstigen Schätzeigenschaften: Welches Schätzverfahren ist geeignet und möglichst effizient, d. h. nutzt die Stichprobeninformation möglichst gut aus? Welche Eigenschaften hat das gewählte Schätzverfahren?

2. Beschaffen von Daten: Erheben einer Stichprobe

- Charakterisierung der Stichprobenerhebung.

3. Spezifizieren, Schätzen und Auswählen von einem oder mehreren ökonometrischen Modellen:

- Verwenden geeigneter Schätzmethoden.
- Verwenden geeigneter Modellwahlmethoden.

4. Überprüfen der gewählten Modelle

- Ist das gewählte Modell korrekt spezifiziert? Falls ja, so fehlen keine relevanten erklärenden Variablen, die funktionale Form ist korrekt gewählt und die Annahmen bezüglich der Fehler sind erfüllt.
- Sind die Annahmen für das gewählte Schätzverfahren erfüllt, so dass die statistischen Eigenschaften des Schätzverfahrens gelten und die Inferenz zulässig ist?
- Falls Annahmen verletzt, Spezifikation und Schätzen alternativer Modelle mit ggf. anderen Variablen und/oder Wahl alternativer Schätzverfahren → Gehe wieder zu Schritt 1 oder Schritt 3.

5. Verwenden der geprüften Modelle:

- Testen der statistischen Hypothesen: Werden die postulierten (ökonomischen) Hypothesen durch die Daten statistisch widerlegt?
- Prognosen
- Interpretation interessierender Parameter

Die für die einzelnen Schritte relevanten ökonometrischen Verfahren werden in den folgenden Kapiteln besprochen.

5. Grundlagen der Schätz- und Testtheorie

5.1. Stichproben und datengenerierende Prozesse

Es bezeichne \mathbf{s}_t einen $m \times 1$ -Vektor von Zufallsvariablen.

Stichproben

- Eine **Stichprobe** ist eine Teilmenge der Grundgesamtheit, die erhoben werden kann (=Zufallsvektor) oder bereits erhoben wurde (Realisation eines Zufallsvektors). Eine Stichprobe der **Stichprobengröße** n ist gegeben durch

$$\{\mathbf{s}_t, t = 1, \dots, n\}.$$

- Die stochastischen Eigenschaften einer Stichprobe sind vollständig durch die **gemeinsame Dichte aller n Stichprobenbeobachtungen** beschrieben:

$$f_{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n}(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n).$$

Bzgl. dieser gemeinsamen Dichte ist eine Stichprobe eine **mögliche zukünftige** oder **bereits erfolgte Realisation** mit n Beobachtungen \mathbf{s}_t .

- **Arten von Stichproben:**
 - **Zufallsstichprobe:** Die n Stichprobenbeobachtungen $\mathbf{s}_t, t = 1, \dots, n$ werden **zufällig** gezogen, d. h. sie sind **unabhängig und identisch verteilt (independently and identically distributed) (IID)**, d. h. zusätzlich zu (2.22) sind die marginalen Dichten identisch,

$$f_{\mathbf{s}_1, \dots, \mathbf{s}_n}(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n) = f_{\mathbf{s}}(\mathbf{s}_1) f_{\mathbf{s}}(\mathbf{s}_2) \cdots f_{\mathbf{s}}(\mathbf{s}_n) = \prod_{t=1}^n f_{\mathbf{s}}(\mathbf{s}_t). \quad (5.1)$$

Der große Vorteil von Zufallsstichproben ist, dass ausschließlich die gemeinsame / marginale Dichte $f_{\mathbf{s}}(\mathbf{s}_t)$ bestimmt werden muss und nicht die gemeinsame Dichte aller Stichprobenbeobachtungen. Alle Stichprobenbeobachtungen sind Ziehungen aus der gleichen Dichte.

- Es bestehen **stochastische Abhängigkeiten** zwischen den einzelnen Stichprobenbeob-

achtungen \mathbf{s}_t , d. h. die Zerlegung (5.1) gilt nicht. Dann gilt

$$\begin{aligned}
 f_{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n}(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n) &= f_{\mathbf{s}_n | \mathbf{s}_{n-1}, \dots, \mathbf{s}_1}(\mathbf{s}_n | \mathbf{s}_{n-1}, \dots, \mathbf{s}_1) \\
 &\quad f_{\mathbf{s}_{n-1} | \mathbf{s}_{n-2}, \dots, \mathbf{s}_1}(\mathbf{s}_{n-1} | \mathbf{s}_{n-2}, \dots, \mathbf{s}_1) \\
 &\quad \cdots f_{\mathbf{s}_2 | \mathbf{s}_1}(\mathbf{s}_2 | \mathbf{s}_1) f_{\mathbf{s}_1}(1) \\
 &= \prod_{t=1}^n f_{\mathbf{s}_t | \mathbf{s}_{t-1}, \dots, \mathbf{s}_1}(\mathbf{s}_t | \mathbf{s}_{t-1}, \dots, \mathbf{s}_1)
 \end{aligned} \tag{5.2}$$

indem das Theorem von Bayes mehrfach angewendet wird. Die gemeinsame Dichte lässt sich bei abhängigen Beobachtungen als ein Produkt von bedingten Dichten darstellen.

- Notiert der Index t die Beobachtungen die Zeit, sind die Beobachtungen eindeutig sortiert. Dann wird die zeitlich geordnete Kollektion von Zufallsvariablen $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ als **stochastischer Prozess** bezeichnet und eine beobachtete Stichprobe als **Zeitreihe**. Zu deren Modellierung werden **Zeitreihenmodelle** verwendet. Ist $m = 1$ und s_t ein Skalar, so liegt eine **univariate Zeitreihe** vor. Ist $m > 1$ und \mathbf{s}_t ein Vektor, untersucht man eine **multivariate Zeitreihe**. Eine Einführung in univariate Zeitreihenmodelle findet sich in Abschnitt 12.3.1.

Es ist möglich, dass die bedingten Dichten $f_{\mathbf{s}_t | \mathbf{s}_{t-1}, \dots, \mathbf{s}_1}(\mathbf{s}_t | \mathbf{s}_{t-1}, \dots, \mathbf{s}_1)$ von der Zeit t abhängen. Sie können beispielsweise von Saisonkomponenten oder einem Zeittrend abhängen. Dies wird entweder durch geeignete Indizes an den bedingten Dichten oder entsprechende Variablen in der Bedingung der Dichten kenntlich gemacht. Mehr dazu in Kapitel 13.

- Liegen für alle Einheiten im Querschnitt Zeitreihendaten vor, spricht man von **Paneldaten**, siehe MA-Veranstaltung **Angewandte Mikroökonomie**.

Datengenerierender Mechanismus, datengenerierender Prozess (data generating process (DGP)):

- In der Ökonometrie/Statistik wird anstelle des Begriffs **Grundgesamtheit** häufig das Konzept **datengenerierender Mechanismus** oder **datengenerierender Prozess (data generating process (DGP))** verwendet. Damit wird der **stochastische Mechanismus** bezeichnet, der die beobachteten Stichprobendaten $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$ in der realen Welt erzeugt haben kann (Davidson & MacKinnon 2004, Sections 1.5, 3.1).
- Der DGP, der einer **Zufallsstichprobe** von n Beobachtungen $\{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n\}$ zugrundeliegt, wird vollständig durch die gemeinsame / marginale ($m = 1$) Dichte $f_{\mathbf{S}}(\mathbf{s})$ bestimmt.
- Im Fall **abhängiger Stichprobenbeobachtungen**, typischerweise bei **Zeitreihen**, die durch einen **stochastischen Prozess** erzeugt werden, müssen aufgrund (5.2) die bedingten Dichten $f_{\mathbf{s}_{t-1} | \mathbf{s}_{t-2}, \dots, \mathbf{s}_1}(\mathbf{s}_{t-1} | \mathbf{s}_{t-2}, \dots, \mathbf{s}_1)$ betrachtet werden.

Beispiel: DGP für tägliche Renditen des DAX:

Annahme bzgl. DGP: Die täglichen DAX-Renditen sind unabhängig und identisch

normalverteilt

$$y_t \sim NID(\mu_0, \sigma_0^2). \quad (5.3)$$

Der Erwartungswert μ_0 und die Varianz σ_0^2 sind fest, aber unbekannt. Alternative Schreibweise, vgl. (2.6):

$$f(y_t; \mu_0, \sigma_0^2) = \frac{1}{\sigma_0^2} \phi\left(\frac{y_t - \mu_0}{\sigma_0}\right) \quad (5.4)$$

- Wie im vorherigen DAX-Beispiel werden in diesem Text Parameter eines DGP immer mit Index 0 notiert.

Ist man nur am Teil des DGPs für die endogenen Variablen gegeben die kausalen Variablen interessiert, so zerlegt man die Dichte $f(\mathbf{s}_t)$ geeignet in bedingte Dichten.

Es bezeichne

$$\mathbf{s}_t = \begin{pmatrix} \mathbf{w}_t \\ \mathbf{y}_t \\ \mathbf{z}_t \end{pmatrix} = \begin{cases} \text{Variablen ohne direkten Einfluss auf } \mathbf{y}_t \\ \text{zu erklärende/endogene Variablen} \\ \text{erklärende/exogene Variablen} \end{cases} \quad (5.5)$$

Einflussfaktoren von Importen $\mathbf{w}_t = (s_1 \ s_2 \ s_3 \ s_9)^T$, $y_t = s_8$, $\mathbf{z}_t = (s_4 \ s_5 \ s_6 \ s_6)$, s_3 ist irrelevant, wenn nur ein Importland und eine Periode betrachtet wird.

Dann ist (allgemein) folgende Faktorisierung

$$f_{\mathbf{S}}(\mathbf{s}_t) = f_{\mathbf{W}|\mathbf{Y},\mathbf{Z}}(\mathbf{w}_t|\mathbf{y}_t, \mathbf{z}_t) f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}_t|\mathbf{z}_t) f_{\mathbf{Z}}(\mathbf{z}_t) \quad (5.6)$$

sinnvoll.

Zur Erklärung von \mathbf{y}_t **gegeben die erklärenden Variablen \mathbf{z}_t** ist ausschließlich die bedingte Dichte

$$f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}_t|\mathbf{z}_t)$$

notwendig. Weder die bedingte Dichte $f_{\mathbf{W}|\mathbf{Y},\mathbf{Z}}(\mathbf{w}|\mathbf{y}, \mathbf{z})$, noch die gemeinsame Dichte $f_{\mathbf{Z}}(\mathbf{z}_t)$ müssen betrachtet werden, was den Modellierungsprozess wesentlich vereinfacht!

5.2. Ökonometrische Modelle

Zur Vereinfachung der Notation betrachten wir in diesem Abschnitt ausschließlich Modelle für Zufallsstichproben.

Modelle für Stichproben mit stochastisch abhängigen Beobachtungen werden in Abschnitt 13.2 behandelt und stellen eine Erweiterung der Modelle aus diesem Abschnitt dar.

- Ein **ökonomisches Modell** \mathbb{M} ist eine Familie von Funktionen $M(\cdot)$ in Abhängigkeit von den Daten und einem $p^* \times 1$ Parametervektor $\boldsymbol{\psi}$. Die Funktionen können (ökonomische) Zusammenhänge beschreiben und enthalten implizit oder explizit eine vollständige oder teilweise Beschreibung des DGP oder zumindest eine Approximation des DGP (Davidson 2000, Section 4.1.1). Die Menge an möglichen und erlaubten Parametern wird als **Parameterraum** Ψ bezeichnet

$$\mathbb{M} \equiv \{M(\mathbf{s}_t; \boldsymbol{\psi}), \boldsymbol{\psi} \in \Psi\}, \quad \Psi \subseteq \mathbb{R}^{p^*}. \quad (5.7)$$

- **Strukturelle Form** eines Modells: Wesentliche Parameter des Modells können (ökonomisch) interpretiert werden.

Beispiel: Angebots- und Nachfragefunktionen Die Variable q_A bezeichnet die Angebotsmenge für ein Gut gegeben dessen Preis p , die Variable q_N die Nachfragemenge desselben Gutes gegeben dessen Preis p , die Variablen z_1 und z_2 sind exogen. Die Fehlerterme sind mit \tilde{u}_1 und \tilde{u}_2 notiert. Ein einfaches Beispiel für eine Angebots- und eine Nachfragefunktion ist dann

$$\text{Angebotsgleichung: } q_A = \alpha_1 p + \beta_1 z_1 + \tilde{u}_1, \quad (5.8a)$$

$$\text{Nachfragegleichung: } q_N = \alpha_2 p + \beta_2 z_2 + \tilde{u}_2, \quad (5.8b)$$

$$\text{Gleichgewichtsbedingung: } q_A = q_N = q. \quad (5.8c)$$

Liegt Marktträumung vor, sind q_N und q_A prinzipiell unbeobachtbar, da nur die Marktträumungsmenge q beobachtbar ist. Die Parameter α_1 und α_2 sind ökonomisch interpretierbar und zwar als die Preiselastizität des Angebots beziehungsweise die Preiselastizität der Nachfrage.

Beispiel: simultanes Gleichungsmodell Das Gleichungssystem (5.8) lässt sich als simultanes Gleichungsmodell schreiben, indem man $y_1 = q_A$ und $y_2 = q_N$ setzt und die Nachfragegleichung durch $-\alpha_2$ dividiert und umstellt. Man erhält dann ein simultanes Gleichungsmodell mit zwei endogenen Variablen

$$y_1 = \alpha_{12} y_2 + \beta_{11} + \beta_{12} z_1 + u_1, \quad (5.9a)$$

$$y_2 = \alpha_{21} y_1 + \beta_{21} + \beta_{23} z_2 + u_2. \quad (5.9b)$$

In (5.9) (und in (5.8)) stehen auf der rechten Seite nicht nur kausale Variablen, so wie in (4.1), sondern auch endogene Variablen. Der Parameter α_{12} misst den kausalen Effekt $y_2 \rightarrow y_1$ und der Parameter α_{21} den kausalen Effekt $y_1 \rightarrow y_2$. Unterstellt man zusätzlich

$$E[u_1 | z_1, z_2] = 0, \quad E[u_2 | z_1, z_2] = 0, \quad (5.9c)$$

$$\sigma_1^2 = \text{Var}(u_1 | z_1, z_2), \quad \sigma_2^2 = \text{Var}(u_2 | z_1, z_2), \quad (5.9d)$$

ergibt sich für das simultane Gleichungsmodell (5.9) mit $\mathbf{s}_t = (y_{1t} \ y_{2t} \ z_{1t} \ z_{2t})^T$ der Parametervektor

$$\boldsymbol{\psi} = (\alpha_{12} \ \beta_{11} \ \beta_{12} \ \sigma_1^2 \ \alpha_{21} \ \beta_{21} \ \beta_{23} \ \sigma_2^2). \quad (5.10)$$

- Es ist möglich, wie im Fall eines simultanen Gleichungsmodellen (5.9), dass die Elemente von strukturellen Modellen keine Menge von bedingten Dichten oder Teile davon (wie z. B. bedingte Erwartungswerte) enthalten. In solchen Fällen muss das strukturelle Modell umgeformt werden, damit die Elemente des abgeleiteten Modells bedingte Dichten oder Teile davon sind.

Notation: Bei dieser Umformung entsteht auch ein neuer Parametervektor $\boldsymbol{\theta}$ der Länge p , der sich aus dem Parametervektor $\boldsymbol{\psi}$ der strukturellen Form ergibt und typischerweise weniger Parameter, $p \leq p^*$ hat. Wir schreiben dann

$$\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\psi}) \in \boldsymbol{\Theta} \quad \text{für alle } \boldsymbol{\psi} \in \boldsymbol{\Phi}. \quad (5.11)$$

Der dazugehörige Parameterraum ist also $\boldsymbol{\Theta}$.

- Werden gemeinsame Dichten für \mathbf{s}_t betrachtet, die von einem Parametervektor $\boldsymbol{\psi}$ abhängen, schreibt man

$$f_{\mathbf{S}}(\mathbf{s}_t; \boldsymbol{\theta}) = f_{\mathbf{S}}(\mathbf{s}_t; \boldsymbol{\theta}(\boldsymbol{\psi})).$$

Damit lässt sich die Menge aller Dichten, die durch ein strukturelles Modell \mathbb{M} impliziert werden, schreiben als

$$\mathbb{M}_D \equiv \{f_{\mathbf{S}}(\mathbf{s}_t; \boldsymbol{\theta}(\boldsymbol{\psi})), \boldsymbol{\psi} \in \boldsymbol{\Psi}\}. \quad (5.12)$$

In vielen Standardfällen entspricht das strukturelle Modell bereits der Modelldefinition (5.12). Deshalb definieren Davidson & MacKinnon (2004, Section 3.1) ein ökonometrisches Modell als eine Menge \mathbb{M}_D an möglichen DGPs. Die hier verwendete Definition (5.7) ist jedoch allgemeiner.

- Ein Parametervektor $\boldsymbol{\psi}$ eines Modells, für den die durch das Modell implizierte Dichte $f_{\mathbf{S}}(\mathbf{s}_t; \boldsymbol{\theta})$ der Dichte des DGP $f_{\mathbf{S}}(\mathbf{s}_t)$ entspricht, also

$$f_{\mathbf{S}}(\mathbf{s}_t; \boldsymbol{\theta}(\boldsymbol{\psi}_0)) = \underbrace{f_{\mathbf{S}}(\mathbf{s}_t)}_{\text{DGP}} \quad (5.13)$$

gilt, wird mit $\boldsymbol{\psi}_0$ notiert und häufig als **wahrer Parametervektor** oder **korrekter Parametervektor** bezeichnet. Entsprechendes gilt für $\boldsymbol{\theta}_0 = \boldsymbol{\theta}(\boldsymbol{\psi}_0)$.

- Modelle \mathbb{M}_D werden auch als Modelle in **reduzierter Form** bezeichnet. Die interessierenden Parameter $\boldsymbol{\theta}$ eines Modells in reduzierter Form sind nur interpretierbar, wenn die strukturelle Form und reduzierte Form eines Modells identisch sind. Mehr dazu in Abschnitt 13.3.

Beispiel: DAX-Renditen – Fortsetzung: Das Modell mit $s_t = y_t$ umfasst alle möglichen DGPs der Art

$$y_t \sim NID(\mu, \sigma^2), \quad \boldsymbol{\theta} = \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \in \boldsymbol{\Theta} \in (\mathbb{R} \times \mathbb{R}^+). \quad (5.14)$$

Bzw.:

$$\mathbb{M} = \mathbb{M}_D = \left\{ f(y_t; \mu, \sigma^2) := \frac{1}{\sigma} \phi\left(\frac{y_t - \mu}{\sigma}\right), \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+ \right\}.$$

Das strukturelle Modell ist ein Menge von Dichten \mathbb{M}_D .

Beispiel: simultanes Gleichungsmodell — reduzierte Form

Um die folgenden Darstellungen zu vereinfachen, wird das simultane Gleichungssystem (5.9) in Matrixschreibweise geschrieben

$$\underbrace{\begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{21} & 1 \end{pmatrix}}_{\mathbf{A}} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \underbrace{\begin{pmatrix} \beta_{11} \\ \beta_{21} \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \beta_{12} & 0 \\ 0 & \beta_{23} \end{pmatrix}}_{\mathbf{B}} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix},$$

so dass die **strukturelle Form** in Matrixschreibweise lautet

$$\mathbf{A}\mathbf{y} = \boldsymbol{\beta} + \mathbf{B}\mathbf{z} + \mathbf{u}. \quad (5.15)$$

Falls \mathbf{A} regulär ist, ergibt sich die **reduzierte Form** durch Inversion

$$\mathbf{y} = \underbrace{\mathbf{A}^{-1}\boldsymbol{\beta}}_{\mathbf{b}} + \underbrace{\mathbf{A}^{-1}\mathbf{B}}_{\mathbf{D}}\mathbf{z} + \underbrace{\mathbf{A}^{-1}\mathbf{u}}_{\boldsymbol{\varepsilon}},$$

$$\mathbf{y} = \mathbf{b} + \mathbf{D}\mathbf{z} + \boldsymbol{\varepsilon}.$$

Um für (5.9) ein Modell der Art \mathbb{M}_D zu erhalten, muss zusätzlich

- eine Annahme bzgl. der gemeinsamen Verteilung der Fehler u_{1t} und u_{2t} und
- eine Annahme bzgl. der gemeinsamen Verteilung der exogenen Variablen z_{1t} und z_{2t} und deren stochastisches Verhältnis zu den Fehlern getroffen werden, z. B.

$$\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} | z_1, z_2 \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \boldsymbol{\Sigma}_u \right).$$

Dann gilt

$$\boldsymbol{\varepsilon} | \mathbf{z} \sim N(0, \boldsymbol{\Sigma}), \quad \boldsymbol{\Sigma} = \mathbf{A}^{-1}\boldsymbol{\Sigma}_u(\mathbf{A}^{-1})^T, \quad (5.16)$$

$$\mathbf{y} | \mathbf{z} \sim N(\mathbf{b} + \mathbf{D}\mathbf{z}, \boldsymbol{\Sigma}) \quad (5.17)$$

bzw.

$$f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z}; \mathbf{b}, \mathbf{D}, \boldsymbol{\Sigma}) = \frac{1}{2\pi} \det(\boldsymbol{\Sigma})^{-1/2} \exp \left(-\frac{1}{2}(\mathbf{y} - \mathbf{b} - \mathbf{D}\mathbf{z})^T \boldsymbol{\Sigma}^{-1}(\mathbf{y} - \mathbf{b} - \mathbf{D}\mathbf{z}) \right)$$

$$\mathbb{M}_D = \left\{ f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z}; \mathbf{b}, \mathbf{D}, \boldsymbol{\Sigma}), \mathbf{b}, \mathbf{D}, \boldsymbol{\Sigma} \text{ Elemente des Parameterraums } \boldsymbol{\Theta} \right\}.$$

Noch ausführlicher wird dies in Abschnitt 13.2 im allgemeineren Rahmen von multivariaten Zeitreihenmodellen dargestellt.

- **Korrekt spezifiziertes ökonometrisches Modell:** Ein Modell ist

- **korrekt spezifiziert**, falls $\text{DGP} \in \mathbb{M}_D$,
- **fehlspezifiziert**, falls $\text{DGP} \notin \mathbb{M}_D$

gilt.

Beispiel: DAX-Renditen – Fortsetzung: Ist $\theta_0 = (\mu_0, \sigma_0)^T \in \Theta$, dann enthält das Modell (5.14) auch den tatsächlichen DGP (5.3) mit μ_0 und σ_0^2 und das Modell ist korrekt spezifiziert.

Ist der DGP der DAX-Renditen jedoch durch eine t -Verteilung

$$y_t/\sigma_0 \text{ IID } t(m_0), \quad m_0 = 5$$

gegeben, ist das Modell (5.14) fehlspezifiziert.

Beispiel: DAX-Renditen – Fortsetzung: Das Modell (5.14) ist vollständig spezifiziert. Ein Modell $y_t \sim \text{IID}(\mu, \sigma^2)$ ist unvollständig spezifiziert, da eine Verteilungsannahme fehlt.

- Davidson & MacKinnon (2004, Section 1.3) nennen ein parametrisches Modell **vollständig spezifiziert**, wenn es möglich ist, nach Zuweisung von Zahlenwerten zu allen im Modell vorhandenen Parametern Realisationen der abhängigen Variable y_t zu generieren. Ansonsten ist es **partiell spezifiziert**.
- Lässt sich aus dem Modell M in struktureller Form ein Modell in der reduzierten Form M_D ableiten, so sagen wir, dass das Modell M **vollständig spezifiziert** ist.
- Ist ein strukturelles Modell M **vollständig und darüber hinaus korrekt** spezifiziert, existiert ein Parametervektor $\theta_0 = \theta(\psi_0)$, für den die Dichte in M_D dem DGP entspricht:

$$M_D \supseteq M_D(\mathbf{s}_t; \theta_0) \equiv f(\mathbf{s}_t; \theta_0) = \underbrace{f_{\mathbf{s}}(\mathbf{s}_t)}_{\text{DGP}}. \quad (5.18)$$

Beispiel: Simultanes Gleichungsmodell

Ist für das simultane Gleichungsmodell (5.9) mit Parametervektor (5.10) keine gemeinsame Verteilung der kausalen Variablen z_1 und z_2 gegeben, sondern ausschließlich eine gemeinsame Verteilung (5.16) der Fehler, dann kann “nur” ein Modell für bedingte Dichten abgeleitet werden.

- **Modellklassen:**
 - **Univariate Modelle:** $s_t = y_t$, ist ein Skalar, $m = 1$.
 - **Multivariate Modelle:** \mathbf{s}_t ist ein Vektor, $m > 1$.
- Ökonometrische Modelle, in denen die implizierten DGPs durch Funktionen in Abhängigkeit von den möglichen Variablen und einem Parametervektor ψ mit fester Länge $p^* < \infty$ unterschieden werden, werden als **parametrische ökonometrische Modelle** bezeichnet.
- In der ökonometrischen Theorie und Praxis spielen jedoch auch **semiparametrische Modelle** und **nichtparametrische Modelle** eine Rolle. Eine kurze Einführung bietet Davidson & MacKinnon (2004, Section 15.5). Eine ausführliche Darstellung liefert die Monographie von Li & Racine (2007).

Bedingte Modelle (conditional models)

- Ist man ausschließlich an der Erklärung der endogenen Variablen \mathbf{y} gegeben die kausalen Variablen \mathbf{z} interessiert, ist es ausreichend, bedingte Modelle zu betrachten. Auf Basis der Faktorisierung des DGPs in (5.6) ergibt sich ein **bedingtes ökonometrisches Modell** (für bedingte Dichten)

$$\mathbb{M}_D \equiv \{f_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z}; \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\} \quad (5.19)$$

Die Variablen \mathbf{z} werden außerhalb des Modells bestimmt. Die Unterscheidung in endogene und exogene Variablen wurde bereits bei der ersten Darstellung des multiplen linearen Regressionsmodells (4.1) in Abschnitt 4.1 durchgeführt:

- **Endogene(n) Variable(n)**: Variable(n) wird/werden durch den im Modell beschriebenen Mechanismus generiert.
- **Exogene Variablen**: Variablen, die außerhalb des Modells bestimmt werden können (da sie keine simultane Beziehung zu den endogenen Variablen aufweisen).
- **Wichtig**: Häufig ist nicht klar, ob eine Variable s_j mit y eine simultane Beziehung aufweist oder für y kausal ist. Dann muss dies im Rahmen des Modellierungsprozesses bestimmt werden (was allerdings nur unter bestimmten Voraussetzungen möglich ist). Das erfordert, dass im Modell zunächst s_j als simultan zugelassen wird und gleichzeitig Parameterwerte im Modell existieren, für die s_j kausal wird. Ein Beispiel hierfür findet sich in den Abschnitten 13.2 und 13.3.
- **Im Folgenden bis zu Kapitel 13 unterstellen wir, dass nur eine endogene Variable zu erklären ist und alle erklärenden Variablen kausal sind.** Unter dieser Voraussetzung sind die strukturelle und reduzierte Form identisch und es gilt $\boldsymbol{\theta} = \boldsymbol{\psi}$.

Beispiel: Einflussfaktoren von Importen

Um für *Importe* auf Basis der Regressiongleichung (4.2) eine auf *BIP* und *Entfernung* bedingte Dichte zu erhalten, muss noch eine Annahme für die Verteilung der Fehler und deren Verhältnis zu den bedingenden Zufallsvariablen *BIP* und *Entfernung* gemacht werden. Häufig wird eine bedingte Normalverteilung

$$u|BIP, Entfernung \sim NID(0, \sigma^2)$$

unterstellt. Unter dieser Annahme ergibt sich unmittelbar, dass die *Importe* bedingt normalverteilt sind:

$$Importe|BIP, Entfernung \sim NID(\beta_1 + \beta_2 BIP + \beta_3 Entfernung, \sigma^2). \quad (5.20)$$

Bzw. mit $\boldsymbol{\theta} = (\beta_1 \ \beta_2 \ \beta_3 \ \sigma^2)^T$

$$\begin{aligned} \mathbb{M} &= \mathbb{M}_D = \{f(Importe|BIP, Entfernung; \boldsymbol{\psi}) \\ &\equiv \frac{1}{\sigma^2} \phi\left(\frac{Importe - \beta_1 - \beta_2 BIP - \beta_3 Entfernung}{\sigma}\right), \\ &\boldsymbol{\theta} \in \boldsymbol{\Theta} \in (\mathbb{R}^3 \times \mathbb{R}^+)\}. \end{aligned}$$

- Für eine empirische Analyse werden häufig verschiedene Modelle \mathbb{M}_i , $i = 1, 2, \dots, I$ betrachtet. Mit Hilfe von **Methoden zur Modellselektion** wird dann versucht, ein korrektes Modell zu wählen. Dazu später mehr.
- In der Praxis sind ökonometrische Modelle (fast) immer fehlspezifiziert. Für bedingte vollständig spezifizierete Modelle heißt das

$$f(\mathbf{y}|\mathbf{z}) \notin \mathbb{M}_{D,i} \quad , i = 1, 2, \dots, I.$$

Man versucht dann, aus den verschiedenen Modellen, $i = 1, 2, \dots, I$, ein Modell zu wählen, das für den Untersuchungszweck eine möglichst gute Approximation an den DGP liefert. Die sich hieraus ergebenden Konsequenzen ignorieren wir aber in diesem Kurs.

Informationsmengen

- Die Menge aller potentiellen Variablen, die für eine vorliegende Fragestellung und ein dafür zu verwendendes Modell zur Erklärung der endogenen Variablen \mathbf{y}_t als **kausale** Variablen in Frage kommen können, wird häufig als **Informationsmenge** bezeichnet und mit Ω_t abgekürzt. Die Informationsmenge hängt bei Zeitreihen typischerweise vom Zeitpunkt t ab, deshalb der Index t , siehe Abschnitt 13.1.
- Die Menge aller Variablen, die in einem vorliegenden Modell zur Erklärung der endogenen Variablen \mathbf{y}_t als **kausale** Variablen verwendet werden, ist ebenfalls eine **Informationsmenge** und wird im Folgenden $\mathcal{I}_t \subset \Omega_t$ abgekürzt.

5.3. Regressionsmodelle

Für viele (ökonomischen) Fragestellungen ist es nicht notwendig, den DGP oder die bedingte Dichte vollständig zu modellieren.

Notation: Alle erklärenden Variablen und eine Konstante, falls notwendig, werden in dem $(1 \times k)$ -Zeilenvektor

$$\mathbf{X}_t \equiv (X_{t1} \quad \cdots \quad X_{tk})$$

zusammengefasst. Liegt eine Konstante vor, gilt

$$\mathbf{X}_t = (1 \quad z_{1t} \quad \cdots \quad z_{k-1,t}).$$

Sehr häufig ist es ausreichend, einzelne Charakteristika der (bedingten) Dichten zu modellieren, insbesondere

- den bedingten Erwartungswert $E[y_t|\mathbf{X}_t]$ und/oder
- die bedingte Varianz $Var(y_t|\mathbf{X}_t)$ oder auch
- bedingte Quantile.

Regressionsmodelle:

- Ein bedingtes Modell zur Modellierung des bedingten Erwartungswertes $E[y_t|\mathbf{X}_t]$ wird als **Regressionsmodell** bezeichnet.
- Aus der Identität

$$y_t = \underbrace{E[y_t|\mathbf{X}_t]}_{\text{systematischer Teil}} + \underbrace{y_t - E[y_t|\mathbf{X}_t]}_{\text{unsystematischer Teil}}$$

wird durch Spezifikation der Funktion des bedingten Erwartungswertes $E[y_t|\mathbf{X}_t]$ ein Regressionsmodell.

Der bedingte Erwartungswert wird als systematischer Teil bezeichnet und im Folgenden mit $m(\mathbf{X}_t) = E[y_t|\mathbf{X}_t]$ notiert, wobei der bedingte Erwartungswert bzgl. der Dichte des DGP berechnet wird. Der unsystematische Teil wird **Fehlerterm** oder als **Störterm** bezeichnet. Im Zusammenhang mit dem korrekten Funktion $m(\mathbf{X}_t)$ wird der Fehlerterm mit ε_t angegeben

$$y_t = m(\mathbf{X}_t) + \varepsilon_t. \quad (5.21)$$

- Die Funktion des bedingten Erwartungswertes $m(\mathbf{X}_t)$ ist im Allgemeinen nicht bekannt. Nimmt man an, dass die Funktion $m(\mathbf{X}_t)$ linear in den Parametern β_1, \dots, β_k ist, erhält man

$$m(\mathbf{X}_t, \boldsymbol{\beta}) = x_{t1}\beta_1 + x_{t2}\beta_2 + \dots + x_{tk}\beta_k = \mathbf{X}_t\boldsymbol{\beta}, \quad \boldsymbol{\beta} := \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} \quad (5.22)$$

Ist diese Annahme korrekt und existiert ein $\boldsymbol{\beta} = \boldsymbol{\beta}_0 \in \Psi$, so dass

$$m(\mathbf{X}_t, \boldsymbol{\beta}_0) = m(\mathbf{X}_t) = E[y_t|\mathbf{X}_t], \quad (5.23)$$

dann ist die **Funktion des bedingten Erwartungswertes korrekt spezifiziert** und $\boldsymbol{\beta}_0$ wird als **korrekter Parametervektor** bezeichnet.

- Für alle Parametervektoren im Parameterraum, $\boldsymbol{\beta} \in \Psi$ erhält man das **multiple lineare Regressionsmodell**

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + u_t \quad (5.24)$$

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t. \quad (5.25)$$

- **Notation:** Für eine vorgegebene Stichprobe $\{(y_t, \mathbf{X}_t), t = 1, \dots, n\}$ müsste man strenggenommen $u_t(\boldsymbol{\beta})$ statt u_t schreiben, also

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t(\boldsymbol{\beta}), \quad (5.26)$$

da sich ja der Fehlerterm ändern muss, wenn sich $\boldsymbol{\beta}$ ändert, solange alle y_t und \mathbf{X}_t gleich bleiben. Wir werden diese Notation später verwenden, wenn es um die Interpretation des KQ-Schätzers geht. Ansonsten verwenden wir, wie in der Literatur üblich, einfach u_t .

- **Notation:** Im Unterschied zu Wooldridge (2009) beginnen Davidson & MacKinnon (2004) den Index der Parameter bei 1 und zählen bis k . Der Kurs folgt i. Allg. Davidson & MacKinnon (2004), auch in anderen Notationsfragen.
- **Ist die Funktion des bedingten Erwartungswertes korrekt spezifiziert,** dann gilt (5.22) für $\beta = \beta_0$. Einsetzen von β_0 in das multiple lineare Regressionsmodell (5.25) und Bestimmen des bedingten Erwartungswertes zeigt, dass dann

$$E[u_t | \mathbf{X}_t] = 0 \quad (5.27)$$

gilt. Deshalb liegt es nahe, dass man zur Schätzung von β_0 die Anforderung (5.27) verwendet. Dies führt direkt zum Kleinst-Quadrate-Schätzer (6.5), der in den folgenden Kapiteln behandelt wird.

- **Mögliche Interpretation des Fehlerterms** Die Bedingung (5.27) kann folgendermaßen interpretiert werden. Es gibt für y_t weitere kausale Faktoren \mathbf{v}_t , die von den explizit berücksichtigten kausalen Faktoren $z_{1t}, \dots, z_{k-1,t}$ stochastisch unabhängig sind und auch nicht im Vektor \mathbf{s}_t in (5.5) berücksichtigt wurden. Beeinflussen diese als Linearkombination y_t , enthält die Gleichung alle kausalen Variablen

$$y_t = \beta_{20}z_{1t} + \beta_{30}z_{2t} + \dots + \beta_{k0}z_{k-1,t} + \mathbf{v}_t^T \boldsymbol{\gamma}_0. \quad (5.28)$$

Falls es möglich wäre, neben den $z_{1t}, \dots, z_{k-1,t}$ auch \mathbf{v}_t zu beobachten, dann könnte man y_t bei Kenntnis der wahren Parameter exakt vorhersagen, da $E[y_t | z_{1t}, \dots, z_{k-1,t}, \mathbf{v}_t]$ gerade der rechten Seite von (5.28) und damit y_t entspricht.

Sind die \mathbf{v}_t nicht bekannt, kann nur der bedingte Erwartungswert

$$\begin{aligned} E[y_t | z_{1t}, \dots, z_{k-1,t}] &= \beta_{20}z_{1t} + \beta_{30}z_{2t} + \dots + \beta_{k0}z_{k-1,t} + E[\mathbf{v}_t | z_{1t}, \dots, z_{k-1,t}]^T \boldsymbol{\gamma}_0 \\ &= \beta_{20}z_{1t} + \beta_{30}z_{2t} + \dots + \beta_{k0}z_{k-1,t} + \underbrace{E[\mathbf{v}_t]^T}_{\equiv \beta_1} \boldsymbol{\gamma}_0 \end{aligned}$$

bestimmt werden. Das zweite Gleichheitszeichen folgt wegen der stochastischen Unabhängigkeit von \mathbf{v}_t und den $z_{jt}, j = 1, \dots, k-1$. Damit entspricht die Konstante β_1 gerade der Linearkombination aus den unbedingten Erwartungswerten aller nicht berücksichtigten Faktoren \mathbf{v}_t und dem Parametervektor $\boldsymbol{\gamma}_0$.

Alternativ ist es möglich, dass man nicht an den einzelnen Einflüssen \mathbf{v}_t interessiert ist, so dass es ausreichend ist, den bedingten Erwartungswert zu betrachten.

Der Fehlerterm u_t ergibt sich dann aus

$$u_t = \mathbf{v}_t^T \boldsymbol{\gamma}_0 - E[\mathbf{v}_t]^T \boldsymbol{\gamma}_0 \quad (5.29)$$

aus den individuellen Abweichungen von diesem Mittelwert. Außerdem gilt (5.27).

Deshalb wird dieser Teil des Modells als unsystematischer Teil bezeichnet.

Zur Erinnerung: In der empirischen Analyse kann Kausalität nur in der durchschnittlichen Wirkung einer Aktion bestimmt werden.

Wichtig: Würde man einige Elemente von \mathbf{v}_t in die Regression aufnehmen, würde dies zu einer Reduktion der Varianz des Fehlerterms führen. Könnte man \mathbf{v}_t komplett aufnehmen, verschwindet die Varianz, da ja y_t exakt prognostiziert werden kann.

- Regressionsmodelle gehören zu den **bedingten Modellen**, da die Regressoren nicht im Modell erklärt werden.
- **Einfaches lineares Regressionsmodell:**

$$y_t = \beta_1 + \beta_2 x_t + u_t. \quad (5.30)$$

- Regressionsmodelle sind entweder
 - **korrekt spezifiziert** (DGP im Modell enthalten). oder
 - **fehlspezifiziert** (DGP im Modell nicht enthalten).

Beispiel für fehlspezifiziertes Modell: DGP

$$y_t = \beta_{10} + \beta_{20}x_t + \beta_{30}x_t^2 + v_t, \quad E[v_t|x_t] = 0, \quad \beta_{30} \neq 0 \quad (5.31)$$

Modell: das einfache lineare Regressionsmodell (5.30).

Der bedingte Erwartungswert gegeben den DGP lautet:

$$y_t = \beta_{10} + \beta_{20}x_t + \underbrace{\beta_{30}x_t^2 + v_t}_{u_t}$$

$$E[y_t|x_t] = \beta_{10} + \beta_{20}x_t + \underbrace{E[u_t|x_t]}_{=\beta_{30}x_t^2 \neq 0},$$

so dass Bedingung $E[u_t|x_t] = 0$ in (5.30) verletzt ist und der DGP nicht im Modell (5.30) enthalten ist.

Hinweis: Zur Analyse einer spezifischen Fragestellung ist es unter bestimmten Bedingungen möglich, fehlspezifizierte Modelle zu verwenden. Dazu gehört die Auswahl eines adäquaten Schätzverfahrens, z. B. die Verwendung des **Instrumentvariablen-schätzers (IV-Schätzers)** oder des **GMM-Schätzers** jeweils mit geeigneten Instrumenten, siehe MA-Modul **Fortgeschrittene Ökonometrie** oder [Davidson & MacKinnon \(2004, Chapter 8 and 9\)](#).

- Regressionsmodelle sind **vollständig spezifiziert**, wenn alle Parameter der bedingten Dichte in der Modellierung berücksichtigt werden. D. h. insbesondere, dass die Verteilung des Störterms modelliert wird.

Beispiele zu vollständig und partiell spezifizierten Regressionsmodellen:

- Das Regressionsmodell

$$\ln(\text{Importe}_t) = \beta_1 + \beta_2 \ln(\text{BIP}_t) + u_t, \quad (5.32)$$

$$u_t | \ln(\text{BIP}_t) \sim NID(0, \sigma^2)$$

ist vollständig spezifiziert.

- Wird dagegen in (5.32) nur $u_t | \ln(BIP_t) \sim IID(0, \sigma^2)$ im Modell spezifiziert, bleibt die Verteilung des Störterms offen, und das Modell ist partiell spezifiziert. Ist der DGP im Modell enthalten, ist das Modell partiell, aber korrekt spezifiziert.

- **Eigenschaft des wahren Parametervektors β_0** im multiplen linearen Regressionsmodell der **Grundgesamtheit**:

- Im korrekt spezifizierten Modell gilt $E[y_t | \mathbf{X}_t] = \mathbf{X}_t \beta_0$ und damit (5.27) $E[u_t | \mathbf{X}_t] = 0$, wobei gemäß (5.26) $u_t = u_t(\beta_0)$. Daraus ergibt sich

$$E[u_t x_{tj}] = 0 \quad j = 1, \dots, k, \quad E[u_t \mathbf{X}_t] = \mathbf{0}. \quad (5.33)$$

Nach Multiplikation von (5.25) mit \mathbf{X}_t^T erhält man

$$\begin{aligned} \mathbf{X}_t^T y_t &= \mathbf{X}_t^T \mathbf{X}_t \beta_0 + \mathbf{X}_t^T u_t \\ E[\mathbf{X}_t^T y_t] &= E[\mathbf{X}_t^T \mathbf{X}_t] \beta_0 + \underbrace{E[\mathbf{X}_t^T u_t]}_{=0} \\ \beta_0 &= E[\mathbf{X}_t^T \mathbf{X}_t]^{-1} E[\mathbf{X}_t^T y_t], \end{aligned} \quad (5.34)$$

sofern $E[\mathbf{X}_t^T \mathbf{X}_t]$ invertierbar ist. Da (5.33) bei korrekter Spezifikation nur für den wahren Parametervektor gilt, kann diese Bedingung zur Ableitung eines Schätzers verwendet werden. Da die Bedingungen (5.33) zweite Momente des DGPs enthält, werden sie als **Momentenbedingungen** bezeichnet. Ändern sich die Momente, ändert sich auch der Parametervektor.

- Mit Hilfe der **Momentenbedingungen** lassen sich in vielen Fällen Schätzer ableiten. In Abschnitt 6.2.1 wird gezeigt, dass die Momentenbedingungen (5.33) den KQ-Schätzer implizieren.
- Jetzt stellt sich die Frage, was passiert, wenn die **Momentenbedingung bei einem fehlspezifizierten Modell** angewendet wird. Wir betrachten den Fall, dass (5.21) den DGP beschreibt, wobei die Regressionsfunktion $m(\cdot)$ auch nichtlinear in den Parameter sein kann, so dass

$$y_t = \mathbf{X}_t \beta + \underbrace{m(\mathbf{X}_t) - \mathbf{X}_t \beta}_{u_t(\beta)} + \varepsilon_t. \quad (5.35)$$

Im Folgenden wird der Erwartungswert bzgl. der Dichten des unbekanntes DGPs berechnet. Der bedingte Erwartungswert des Fehlerterms $u_t(\beta)$ ist

$$\begin{aligned} E[u_t(\beta) | \mathbf{X}_t] &= E[m(\mathbf{X}_t) - \mathbf{X}_t \beta + \varepsilon_t | \mathbf{X}_t] \\ &= m(\mathbf{X}_t) - \mathbf{X}_t \beta + \underbrace{E[\varepsilon_t | \mathbf{X}_t]}_{=0} \\ &= m(\mathbf{X}_t) - \mathbf{X}_t \beta = \begin{cases} = 0 & \text{falls korrekt spezifiziert und } \beta = \beta_0, \\ \neq 0 & \text{möglich, falls Modell fehlspezifiziert.} \end{cases} \end{aligned} \quad (5.36)$$

Im zweiten Fall kann für einzelne Beobachtungen ein **Approximationsfehler** auftreten.

Ganz analog zum Vorgehen bei korrekter Spezifikation erhält man

$$\mathbf{X}_t^T y_t = \mathbf{X}_t^T \mathbf{X}_t \boldsymbol{\beta} + \mathbf{X}_t^T u_t(\boldsymbol{\beta}), \quad (5.37)$$

$$E[\mathbf{X}_t^T y_t] = E[\mathbf{X}_t^T \mathbf{X}_t] \boldsymbol{\beta} + E[\mathbf{X}_t^T u_t(\boldsymbol{\beta})]. \quad (5.38)$$

Wendet man die Momentenbedingung (5.33) entsprechend auf $E[\mathbf{X}_t^T u_t(\boldsymbol{\beta})]$ an und existiert ein $\boldsymbol{\beta}_{00}$, so dass

$$E[\mathbf{X}_t^T u_t(\boldsymbol{\beta}_{00})] = 0 \quad (5.39)$$

gilt, ergibt sich

$$\boldsymbol{\beta}_{00} = E[\mathbf{X}_t^T \mathbf{X}_t]^{-1} E[\mathbf{X}_t^T y_t]. \quad (5.40)$$

Der Parametervektor $\boldsymbol{\beta}_{00}$ wird häufig als **pseudo-wahrer Parametervektor** bezeichnet.

Interpretation des pseudo-wahren Parametervektors:

1. **Wenn eine Konstante im Modell ist**, also $x_{t1} = 1$ ist, dann folgt aus der Momentenbedingung (5.39), dass der unbedingte Erwartungswert der Fehler Null ist, da $E[x_{t1} u_t(\boldsymbol{\beta}_{00})] = E[u_t(\boldsymbol{\beta}_{00})] = 0$ gilt.

Dann ergibt sich

$$\begin{aligned} \text{Cov}(\mathbf{X}_t, u_t(\boldsymbol{\beta}_{00})) &= E[\mathbf{X}_t^T u_t(\boldsymbol{\beta}_{00})] - E[\mathbf{X}_t^T] E[u_t(\boldsymbol{\beta}_{00})] \\ &= E[\mathbf{X}_t^T u_t(\boldsymbol{\beta}_{00})] = E[\mathbf{X}_t^T (m(\mathbf{X}_t) - \mathbf{X}_t \boldsymbol{\beta}_{00})] + E[\mathbf{X}_t^T \varepsilon_t] \\ &= E[\mathbf{X}_t^T u_t(\boldsymbol{\beta}_{00})] = E[\mathbf{X}_t^T (m(\mathbf{X}_t) - \mathbf{X}_t \boldsymbol{\beta}_{00})]. \end{aligned} \quad (5.41)$$

Die Momentenbedingung (5.39) garantiert deshalb (bei einer Konstante im Modell), dass für den **pseudo-wahren Parametervektor** $\boldsymbol{\beta}_{00}$ die **Kovarianz zwischen den Regressoren \mathbf{X}_t und den Approximationsfehlern $m(\mathbf{X}_t) - \mathbf{X}_t \boldsymbol{\beta}_{00}$ Null ist**. Mit anderen Worten, die Approximationsfehler und die Regressoren \mathbf{X}_t sind unkorreliert. In diesem Fall lässt sich (5.40) auch schreiben als

$$\boldsymbol{\beta}_{00} = \text{Var}(\mathbf{X}_t)^{-1} \text{Cov}(\mathbf{X}_t^T, y_t).$$

‡ Beweis: $\text{Cov}(\mathbf{X}_t^T, y_t) = E[\mathbf{X}_t^T y_t] - E[\mathbf{X}_t^T] E[y_t] = E[\mathbf{X}_t^T y_t] - E[\mathbf{X}_t^T] E[\mathbf{X}_t] \boldsymbol{\beta}_{00} - E[\mathbf{X}_t^T \varepsilon_t] = E[\mathbf{X}_t^T y_t] + (\text{Var}(\mathbf{X}_t) - E[\mathbf{X}_t^T \mathbf{X}_t]) \boldsymbol{\beta}_{00}$, so dass

$$\text{Cov}(\mathbf{X}_t^T, y_t) - \text{Var}(\mathbf{X}_t) \boldsymbol{\beta}_{00} = E[\mathbf{X}_t^T y_t] - E[\mathbf{X}_t^T \mathbf{X}_t] \boldsymbol{\beta}_{00}.$$

2. Eine zweite Interpretation des pseudo-wahren Parametervektors ergibt sich aus folgender Argumentation.

Man betrachtet den Erwartungswert der quadratischen Abweichungen der endogenen Variable y_t von deren **linearen Prognosen** $\mathbf{X}_t \boldsymbol{\beta}_{00}$ (linear in den Parametern $\boldsymbol{\beta}$). Dieser

Erwartungswert wird im folgenden Abschnitt als mittlerer quadratischer Fehler (mean squared error) (MSE) bezeichnet werden.

Zunächst betrachtet man für einen *gegebenen DGP* den MSE für ein multiples lineares Regressionsmodell für einen beliebigen Parametervektor β

$$\begin{aligned} MSE(y_t, \mathbf{X}_t; \beta) &\equiv E[(y_t - \mathbf{X}_t\beta)^2] = E[y_t^2 - 2y_t\mathbf{X}_t\beta + \beta^T\mathbf{X}_t^T\mathbf{X}_t\beta] \\ &= E[y_t^2] - 2E[y_t\mathbf{X}_t]\beta + \beta^T E[\mathbf{X}_t^T\mathbf{X}_t]\beta \end{aligned} \quad (5.42)$$

und sucht dann den Parametervektor, der diesen mittleren quadratischen Fehler minimiert. Dies erfolgt durch Ableiten des $MSE(y_t, \mathbf{X}_t; \beta)$ bzgl. β und Nullsetzen

$$\frac{\partial MSE(y_t, \mathbf{X}_t; \beta)}{\partial \beta} = -2E[\mathbf{X}_t^T y_t] + 2E[\mathbf{X}_t^T \mathbf{X}_t]\beta \stackrel{!}{=} \mathbf{0}. \quad (5.43)$$

Hieraus folgt wiederum

$$\beta_{00} = E[\mathbf{X}_t^T \mathbf{X}_t]^{-1} E[\mathbf{X}_t^T y_t]. \quad (5.40)$$

Damit liefert der **pseudo-wahre Parametervektor** β_{00} auch die **beste lineare Prognose** von y_t im Sinne eines minimalen mittleren quadratischen Fehlers (MSE) des Regressionsmodells.

Ausblick: Die Gleichung (5.43) motiviert darüber hinaus eine mögliche Ableitung des Kleinst-Quadrate-Schätzers im Abschnitt 6.2.2.

Offensichtlich gilt $\beta_{00} = \beta_0$, wenn das Regressionsmodell korrekt spezifiziert ist.

5.4. Relevante Eigenschaften von Schätzern

Notation für Erwartungswerte von Matrizen:

$$E[\mathbf{X}] = \begin{pmatrix} E[x_{11}] & E[x_{12}] & \cdots & E[x_{1k}] \\ E[x_{21}] & E[x_{22}] & \cdots & E[x_{2k}] \\ \vdots & \vdots & \ddots & \vdots \\ E[x_{n1}] & E[x_{n2}] & \cdots & E[x_{nk}] \end{pmatrix} \quad (5.44)$$

Schätzer und Schätzung

- Das Modell enthalte p Parameter, die in dem $(p \times 1)$ -Parametervektor θ zusammengefasst sind. Ein **Schätzer** $\tilde{\theta}(y_1, \dots, y_n)$ für den Parametervektor θ ist eine (**vektorwertige**) **Funktion**, die als Argument ausschließlich Stichprobenbeobachtungen (y_1, \dots, y_n) enthält und dazu dient, Schätzwerte von θ zu bestimmen, die in einem noch näher zu spezifizierenden Sinne möglichst nahe an θ liegen. Ein Schätzer $\tilde{\theta}(y_1, \dots, y_n)$ ist eine **Funktion von Zufallsvariablen**, da die Stichprobenbeobachtungen vor ihrer Erhebung Zufallsvariablen sind.

- Wird ein Schätzer auf Basis einer erhobenen Stichprobe berechnet, erhält man eine **Schätzung** von θ_0 .
- Im Allgemeinen weicht die Schätzung von den Parameterwerten θ_0 des tatsächlichen vorliegenden DGP ab. Diese Abweichungen werden als **Schätzfehler** $\tilde{\theta}(\mathbf{y}) - \theta_0$ bezeichnet. Der Parametervektor des tatsächlich vorliegenden DGPs wird, wie bereits ausgeführt, häufig als **wahrer Parametervektor** bezeichnet.

Auswahlkriterien für Schätzer

- Die Wahl der Schätzmethode hängt von der gewählten Bewertung der Schätzfehler ab, die wiederum von der Fragestellung abhängt. Eine Bewertung des Schätzfehlers für einen Parameter i ist mit Hilfe der **Verlustfunktion (loss function)** $L(\tilde{\theta}_i(\mathbf{y}), \theta_i)$ möglich. Typische Verlustfunktionen für skalare Parameter θ sind:

– **Quadratische Verlustfunktion:**

$$L_{sq}(\tilde{\theta}(\mathbf{y}), \theta_0) := (\tilde{\theta}(\mathbf{y}) - \theta_0)^2 \quad (5.45)$$

Die quadratische Verlustfunktion misst das Quadrat des Euklidischen Abstands (Länge) zwischen dem geschätzten $\tilde{\theta}$ und dem wahren Parameter θ_0 .

– **Absolutbetrag des Schätzfehlers:**

$$L_{abs}(\tilde{\theta}(\mathbf{y}), \theta_0) := |\tilde{\theta}(\mathbf{y}) - \theta_0|. \quad (5.46)$$

– **Asymmetrische Verlustfunktion:** Beispiel:

$$\begin{aligned} L_{abs}(\tilde{\theta}(\mathbf{y}), \theta_0) := & a |\tilde{\theta}(\mathbf{y}) - \theta_0| 1(\tilde{\theta}(\mathbf{y}) - \theta_0 > 0) \\ & + b |\tilde{\theta}(\mathbf{y}) - \theta_0| 1(\tilde{\theta}(\mathbf{y}) - \theta_0 < 0), \quad a, b > 0, \end{aligned} \quad (5.47)$$

wobei $1(\cdot)$ die Indikatorfunktion bezeichnet.

- Der Wert der Verlustfunktion hängt von der Stichprobe ab. Um einen stichprobenunabhängigen Wert zu bekommen, betrachtet man den Erwartungswert der Verlustfunktion

$$E[L(\tilde{\theta}(\mathbf{y}), \theta_0)], \quad (5.48)$$

wobei der Erwartungswert bzgl. der Stichprobenbeobachtungen \mathbf{y} bestimmt wird, die durch den DGP generiert werden können. Dieser Erwartungswert misst den erwarteten Verlust eines Schätzers und wird in der Statistik als **Risiko** eines Schätzers für Parameter θ bezeichnet.

Interpretation: Wird die Verlustfunktion für eine große Anzahl verschiedener Stichproben von demselben DGP berechnet, liegt der Durchschnitt nahe dem Risiko.

- Das Risiko bzgl. der quadratischen Verlustfunktion für einen skalaren Parameter wird auch als **mittlerer quadratischer Fehler (mean squared error (MSE))**

$$MSE(\tilde{\theta}(\mathbf{y})) := E[(\tilde{\theta}(\mathbf{y}) - \theta_0)^2] \quad (5.49)$$

bezeichnet.

Werden alle p Parameter zusammen betrachtet, erhält man die **Matrix der mittleren quadratischen Fehler**:

$$MSE(\tilde{\boldsymbol{\theta}}(\mathbf{y})) := E \left[(\tilde{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}_0) (\tilde{\boldsymbol{\theta}}(\mathbf{y}) - \boldsymbol{\theta}_0)^T \right]. \quad (5.50)$$

Die MSE-Matrix lässt sich in zwei wichtige Bestandteile zerlegen, die im Folgenden definiert werden.

- **Erwartungswert minimiert MSE**

- **Verzerrung (bias)** eines Schätzers $\tilde{\boldsymbol{\theta}}(\mathbf{y})$:

$$B(\tilde{\boldsymbol{\theta}}(\mathbf{y})) := E[\tilde{\boldsymbol{\theta}}(\mathbf{y})] - \boldsymbol{\theta}_0 \quad (5.51)$$

- **Varianz-Kovarianzmatrix / Kovarianzmatrix / Varianzmatrix** eines Schätzers $\tilde{\boldsymbol{\theta}}(\mathbf{y})$:

$$Var(\tilde{\boldsymbol{\theta}}(\mathbf{y})) := E \left[(\tilde{\boldsymbol{\theta}}(\mathbf{y}) - E[\tilde{\boldsymbol{\theta}}(\mathbf{y})]) (\tilde{\boldsymbol{\theta}}(\mathbf{y}) - E[\tilde{\boldsymbol{\theta}}(\mathbf{y})])^T \right] \quad (5.52)$$

Die Varianz-Kovarianzmatrix lautet im Detail (zur besseren Lesbarkeit wird wie i. Allg. üblich die Abhängigkeit von der Stichprobe nicht angegeben):

$$\begin{aligned} Var(\tilde{\boldsymbol{\theta}}) &:= E \left[(\tilde{\boldsymbol{\theta}} - E[\tilde{\boldsymbol{\theta}}]) (\tilde{\boldsymbol{\theta}} - E[\tilde{\boldsymbol{\theta}}])^T \right] \\ &= \begin{pmatrix} Var(\tilde{\theta}_1) & Cov(\tilde{\theta}_1, \tilde{\theta}_2) & \cdots & Cov(\tilde{\theta}_1, \tilde{\theta}_p) \\ Cov(\tilde{\theta}_2, \tilde{\theta}_1) & Var(\tilde{\theta}_2) & \cdots & Cov(\tilde{\theta}_2, \tilde{\theta}_p) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(\tilde{\theta}_p, \tilde{\theta}_1) & Cov(\tilde{\theta}_p, \tilde{\theta}_2) & \cdots & Var(\tilde{\theta}_p) \end{pmatrix}. \end{aligned} \quad (5.53)$$

- **Zerlegung der MSE-Matrix:** Generell lässt sich die MSE-Matrix in die Varianz-Kovarianzmatrix des Schätzers und das äußere Produkt der Verzerrungen zerlegen:

$$MSE(\tilde{\boldsymbol{\theta}}(\mathbf{y})) = Var(\tilde{\boldsymbol{\theta}}(\mathbf{y})) + B(\tilde{\boldsymbol{\theta}}(\mathbf{y})) B(\tilde{\boldsymbol{\theta}}(\mathbf{y}))^T, \quad (5.54)$$

- Ein **unverzerrter Schätzer** wird auch als **erwartungstreu** bezeichnet. Im Erwartungswert über alle möglichen Stichproben entspricht der Schätzer dem Parametervektor des tatsächlichen DGP.

$$E[\tilde{\boldsymbol{\theta}}(\mathbf{y})] = \boldsymbol{\theta}_0. \quad (5.55)$$

Interpretation: Erwartungstreue impliziert, dass bei einer großen Anzahl an Stichproben der Durchschnittswert aller Schätzungen sehr nahe am wahren Wert liegt.

- Ist ein Schätzer erwartungstreu, d. h. $E[\tilde{\boldsymbol{\theta}}(\mathbf{y})] = \boldsymbol{\theta}_0$, entspricht der MSE gerade der Varianz des Schätzers.

- **Eigenschaften von Varianz-Kovarianzmatrizen**

- Varianz-Kovarianzmatrizen sind **symmetrisch** und **immer positiv semidefinit**, meist jedoch positiv definit, da aufgrund ihrer Definition (1.8) gilt.

- Die Inverse einer Varianz-Kovarianzmatrix

$$Var(\tilde{\boldsymbol{\theta}})^{-1}$$

wird als **Präzisionsmatrix (precision matrix)** bezeichnet.

- **Vergleich von Varianz-Kovarianzmatrizen zwei Schätzer $\hat{\boldsymbol{\theta}}$ und $\tilde{\boldsymbol{\theta}}$**

Sind zwei Schätzer erwartungstreu und verwendet man als Auswahlkriterium den MSE, wählt man den Schätzer, der von beiden die kleinere Varianz aufweist.

Im skalaren Fall ($p = 1$) ist dies einfach, da sich beide Varianzen leicht vergleichen lassen. Ist $p > 1$, muss man zwei Varianz-Kovarianzmatrizen vergleichen. Der Schätzer $\hat{\boldsymbol{\theta}}$ weist eine “kleinere” Varianz-Kovarianz-Matrix als der Schätzer $\tilde{\boldsymbol{\theta}}$ auf, wenn die folgende Differenz der Präzisionsmatrizen

$$Var(\hat{\boldsymbol{\theta}})^{-1} - Var(\tilde{\boldsymbol{\theta}})^{-1}$$

positiv semidefinit und nicht Null ist. Sind beide Varianz-Kovarianzmatrizen positiv definit, gilt äquivalent, dass die Differenz

$$Var(\tilde{\boldsymbol{\theta}}) - Var(\hat{\boldsymbol{\theta}})$$

positiv semidefinit und nicht Null ist (Davidson & MacKinnon 2004, Section 3.5, Seite 105 und Exercise 3.8).

Interpretation: Die Eigenschaft einer **positiv semidefiniten** Differenz der Varianz-Kovarianzmatrizen bedeutet, dass jede Linearkombination der Differenz nicht negativ ist. Insbesondere gilt

$$Var(\tilde{\theta}_j) \geq Var(\hat{\theta}_j), \quad j = 1, \dots, p. \quad (5.56)$$

- **Korrelationsmatrix** eines Schätzers $\tilde{\boldsymbol{\theta}}$:

Vgl. zur Definition einer Korrelation (2.24)

$$Corr(\tilde{\boldsymbol{\theta}}) := \left(\frac{Cov(\tilde{\theta}_i, \tilde{\theta}_j)}{(Var(\tilde{\theta}_i)Var(\tilde{\theta}_j))^{1/2}} \right)_{i=1, \dots, p, j=1, \dots, p} \quad (5.57)$$

Die Korrelationsmatrix lässt sich ebenfalls in Matrixschreibweise darstellen als

$$Corr(\tilde{\boldsymbol{\theta}}) = \left(diag(Var(\tilde{\boldsymbol{\theta}})) \right)^{-1/2} Var(\tilde{\boldsymbol{\theta}}) \left(diag(Var(\tilde{\boldsymbol{\theta}})) \right)^{-1/2}, \quad (5.58)$$

wobei $diag(\mathbf{A})$ eine Diagonalmatrix bezeichnet, die auf der Diagonale die Diagonalelemente der Matrix \mathbf{A} enthält.

Wesentlich bei der Korrelationsmatrix ist, dass alle Elemente auf der Diagonale 1 sind und alle Nichtdiagonalelemente im Intervall $[-1, 1]$ liegen.

R-Befehle

Berechnen der Korrelationsmatrix aus einer Kovarianzmatrix mit `cov2cor()`.

- **Wünschenswerte Anforderungen** an einen Schätzer:
 1. minimales Risiko oder
 2. minimales Risiko bei Erwartungstreue, d. h. minimale Varianz
- **Effizienz** eines Schätzers: Wird der MSE als Auswahlkriterium für das Risiko gewählt und betrachtet man Schätzer aus einer Klasse, die ausschließlich unverzerrte Schätzer enthält, wird ein Schätzer der betrachteten Klasse als effizient bezeichnet, wenn er in dem oben bestimmten Sinne die kleinstmögliche Varianz aufweist.

Konkret: Ein Schätzer $\hat{\beta}$ ist der **effiziente Schätzer in einer Klasse von unverzerrten Schätzern** $\tilde{\beta}$, wenn gilt, dass die Matrix der Differenz der Varianz-Kovarianzmatrizen $Var(\tilde{\beta}) - Var(\hat{\beta})$ **positiv semidefinit** ist.

- **Kenntnis** der Wahrscheinlichkeitsverteilung des Schätzers für jede Stichprobengröße n . Man bezeichnet diese Verteilung als **exakte Wahrscheinlichkeitsverteilung** eines Schätzers.

Wichtige Eigenschaften eines Schätzers für endliche Stichproben

- Erwartungstreue
- Varianz-Kovarianzmatrix und Korrelationsmatrix
- Effizienz bzw. allgemeiner Risiko
- Exakte Wahrscheinlichkeitsverteilung

Beispiel: der Schätzer des Erwartungswertes μ :

- Ein möglicher Schätzer des Erwartungswertes ist gegeben durch das **arithmetische Mittel** aller Stichprobenbeobachtungen

$$\hat{\mu}(\mathbf{y}) := \frac{1}{n} \sum_{t=1}^n y_t. \quad (5.59)$$

$\hat{\mu}(\mathbf{y})$ ist ein Spezialfall des Kleinst-Quadrate-Schätzers (6.5).

– Berechnen der **Verzerrung**:

$$\begin{aligned} E[\hat{\mu}(\mathbf{y})] - \mu_0 &= E\left[\frac{1}{n} \sum_{t=1}^n y_t\right] - \mu_0 = \frac{1}{n} \sum_{t=1}^n E[y_t] - \mu_0 \\ &\stackrel{IID}{=} \frac{1}{n} \sum_{t=1}^n \mu_0 - \mu_0 = \mu_0 - \mu_0 = 0. \end{aligned}$$

Der Schätzer des Erwartungswertes ist im Fall einer Zufallsstichprobe erwartungstreu.

- Berechnen der **Varianz** des Schätzers:

$$\text{Var}(\hat{\mu}(\mathbf{y})) = \text{Var}\left(\frac{1}{n} \sum_{t=1}^n y_t\right) \stackrel{\text{IID}}{=} \frac{1}{n^2} \sum_{t=1}^n \text{Var}(y_t) = \frac{\sigma_0^2}{n} \quad (5.60)$$

- **MSE** des Schätzers: entspricht der Varianz, da der Schätzer erwartungstreu ist. Der MSE entspricht hier auch dem Risiko bzgl. der quadratischen Verlustfunktion.

Beachte: Das Risiko des Erwartungswertschätzers nimmt mit zunehmender Stichprobengröße n mit der Rate n ab.

- **Verteilung** der Schätzers: Aufgrund der Modellannahme (5.14) ist der Schätzer $\hat{\mu}(\mathbf{y}) = \frac{1}{n} \sum_{t=1}^n y_t$ ist eine Linearkombination von unabhängig und identisch normalverteilter y_t . Deswegen gilt

$$\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}) \quad \text{mit } \boldsymbol{\mu} = \mu \boldsymbol{\iota}, \boldsymbol{\Sigma} = \sigma^2 \mathbf{I},$$

wobei $\boldsymbol{\iota}$ ein $(n \times 1)$ -Vektor mit Einsen ist. Die Summe $\boldsymbol{\iota}^T \mathbf{y} = \sum_{t=1}^n y_t$ ist wegen (2.33) ebenso normalverteilt mit

$$\boldsymbol{\iota}^T \mathbf{y} \sim N(\boldsymbol{\iota}^T \boldsymbol{\mu}, \boldsymbol{\iota}^T \boldsymbol{\Sigma} \boldsymbol{\iota}).$$

Wegen $\boldsymbol{\iota}^T \boldsymbol{\mu} = n\mu$ und $\boldsymbol{\iota}^T \boldsymbol{\Sigma} \boldsymbol{\iota} = n\sigma^2$ erhält man

$$\begin{aligned} \boldsymbol{\iota}^T \mathbf{y} &\sim N(n\mu, n\sigma^2) \quad \text{und} \\ \hat{\mu}(\mathbf{y}) &\sim N\left(\mu, \frac{\sigma^2}{n}\right). \end{aligned} \quad (5.61)$$

Der Schätzer des Erwartungswertes $\hat{\mu}(\mathbf{y})$ ist also ebenso normalverteilt.

- Ein anderer möglicher Schätzer ist

$$\tilde{\mu}(\mathbf{y}) = \frac{1}{2}(y_1 + y_n). \quad (5.62)$$

Bestimmen Sie wieder alle Eigenschaften und vergleichen Sie diese. Zeigen Sie, dass im Vergleich des arithmetischen Mittels (5.59) und (5.62) der erstgenannte effizient ist.

Asymptotische Eigenschaften

Prinzipiell hängen die bisher betrachteten Kenngrößen Verzerrung, Varianz, Risiko, MSE und Verteilung von der Stichprobengröße und dem DGP ab. Dabei kann die Abhängigkeit von Parametern des DGP sehr unpraktisch sein, da diese ja gerade unbekannt sind und so eine Auswahl eines geeigneten Schätzers nicht gut möglich ist. Deshalb werden auch Kenngrößen betrachtet, die in solchen Fällen in einem geeigneten Sinne unabhängig vom DGP sind und zumindest garantieren, dass sich die Eigenschaften eines betrachteten Schätzers mit wachsendem Stichprobenumfang „wünschenswerten“ Eigenschaften, z. B. Erwartungstreue, nähern. Man „betreibt“ dann **Asymptotik** oder **asymptotische Theorie**: man indiziert die Schätzfunktion mit der Stichprobengröße n und untersucht die Eigenschaften von $\tilde{\boldsymbol{\theta}}_n$ für $n \rightarrow \infty$. Man untersucht also die Konvergenzeigenschaften einer Funktionenfolge, siehe Mathematikvorkurs Kapitel 3.

Wichtige asymptotische Eigenschaften eines Schätzers

- Konsistenz
- Asymptotische Varianz
- Asymptotische Effizienz
- Asymptotische Verteilung

Die Eigenschaften im Einzelnen:

- **Konsistenz:** ist ein Schätzer verzerrt, kann man fragen, ob das Ausmaß der Verzerrung mit zunehmender Stichprobengröße geringer wird und der Schätzer gegen den wahren Parametervektor θ_0 konvergiert, wenn der Stichprobenumfang gegen unendlich strebt. "Konvergenz" bedeutet hier **Konvergenz des Schätzers in Wahrscheinlichkeit**

$$\text{plim}_{n \rightarrow \infty} \tilde{\theta}_n = \theta_0 \quad (5.63)$$

oder **fast sichere Konvergenz**

$$\tilde{\theta}_n \xrightarrow{a.s.} \theta_0. \quad (5.64)$$

Konsistenz impliziert, dass

1. der Schätzer asymptotisch erwartungstreu (unverzerrt) ist

$$\lim_{n \rightarrow \infty} E[\tilde{\theta}_n] = \theta_0.$$

2. die Varianz des Schätzers für $n \rightarrow \infty$ gegen Null geht.

- Ist ein Schätzer nicht konsistent, wird er als **inkonsistent** bezeichnet.

Beispiel: Das arithmetische Mittel als Schätzer des Erwartungswertes: (5.59) ist konsistent, da es für beliebiges n erwartungstreu ist und die Varianz mit $n \rightarrow \infty$ gegen Null geht, siehe (5.60).

- **Asymptotische Varianz-Kovarianzmatrix:** Im Allg. hängt die Varianz-Kovarianzmatrix von der Stichprobengröße n ab. Wenn, wie für Konsistenz notwendig, die Varianzen und Kovarianzen des Schätzers für $n \rightarrow \infty$ gegen Null gehen, lassen sich für den Grenzfall $n \rightarrow \infty$ die Varianz-Kovarianzmatrizen von verschiedenen Schätzern nicht vergleichen. Auch kann deshalb für den Grenzfall keine nicht-degenerierte Wahrscheinlichkeitsverteilung existieren. Beides setzt voraus, dass die Abhängigkeit der Varianz-Kovarianzmatrix $\text{Var}(\hat{\theta}_n)$ von der Stichprobengröße für $n \rightarrow \infty$ eliminiert werden kann.

Um diese Abhängigkeit der Varianz von $\hat{\theta}$ zu verhindern, muss man $\hat{\theta}_n$ mit einem vom Stichprobenumfang abhängigen **Faktor** $r(n)$ multiplizieren, **der verhindert, dass $\text{Var}(r(n) \cdot \hat{\theta}_n)$ gegen Null konvergiert** oder **gegen Unendlich divergiert**. Es kann auch erforderlich sein, dass für jeden Parameterschätzer $\hat{\theta}_{in}$ ein spezifischer Faktor $r_i(n)$ notwendig ist. Diese Faktoren werden als **Konvergenzraten** bezeichnet. Als Ergebnis erhält man die **asymptotische Varianz-Kovarianzmatrix**, die häufig mit $\text{asyVar}(\tilde{\theta}_n)$ notiert wird.

Beispiel: das arithmetische Mittel als Schätzer des Erwartungswertes:

Das Kollabieren oder Divergieren der Varianz von $\hat{\mu}_n - \mu_0$ wird verhindert, indem man $\hat{\mu}_n - \mu_0$ mit dem vom Stichprobenumfang abhängigen Faktor $r(n) = \sqrt{n}$ multipliziert. Denn aus $Var(\hat{\mu}_n) = n^{-1}\sigma_0^2$ folgt

$$Var(\sqrt{n}(\hat{\mu}_n - \mu_0)) = nVar(\hat{\mu}_n - \mu_0) = n\frac{\sigma_0^2}{n} = \sigma_0^2 = \text{asyVar}(\hat{\mu}_n). \quad (5.65)$$

Damit ist σ_0^2 die asymptotische Varianz des arithmetischen Mittels und die Konvergenzrate beträgt \sqrt{n} .

Beispiel: der ineffiziente Erwartungswertschätzer (5.62)

Zeigen Sie, dass für diesen Erwartungswertschätzer $r(n) = 1$ gilt. Damit ist dessen Konvergenzrate kleiner als die Rate des arithmetischen Mittels, weshalb letzterer vorzuziehen ist.

• **Asymptotische Verteilung:**

- Die asymptotische Verteilung ist die Grenzverteilung, die sich für $n \rightarrow \infty$ ergibt. Später wird dies noch genauer definiert werden.

Beispiel: Schätzer des Erwartungswertes: Die Verteilung bzw. Dichte $f(\hat{\mu}; \mu_0, \sigma_0^2/n)$ des Schätzers des Erwartungswertes $\hat{\mu}$ hängt von der Stichprobengröße ab, da dessen Varianz von der Stichprobengröße abhängt, siehe (5.61).

Die Normalverteilung wird von der Stichprobengröße n unabhängig, wenn die asymptotische Varianz eingesetzt werden kann. Das erreicht man, indem man anstelle die mit der Konvergenzrate multiplizierte Folge von Zufallsvariablen $\sqrt{n}(\hat{\mu}_n - \mu_0)$ betrachtet:

$$\sqrt{n}(\hat{\mu}_n - \mu_0) \sim N(0, \sigma_0^2). \quad (5.66)$$

Da die Verteilung $N(0, \sigma_0^2)$ von der Stichprobengröße unabhängig ist, ist diese auch für $n \rightarrow \infty$ gültig und damit die Grenzverteilung

$$\sqrt{n}(\hat{\mu}_n - \mu_0) \xrightarrow{d} N(0, \sigma_0^2). \quad (5.67)$$

- **Wann ist die Kenntnis der asymptotischen Verteilung nützlich?**

Falls die Normalverteilungsannahme nicht getroffen werden kann, funktioniert die Ableitung in (5.66) nicht mehr. Kann man also nur

$$y_t \sim IID(\mu_0, \sigma_0^2), \quad t = 1, 2, \dots, n, \quad (5.68)$$

voraussetzen, ist es nicht möglich, die **exakte Wahrscheinlichkeitsverteilung** des Schätzers

$$F_n(z) := P(\hat{\mu}_n \leq z)$$

zu bestimmen. Ist jedoch in so einem Fall die asymptotische Verteilung bekannt, kann diese approximativ statt der unbekannt exakten Verteilung verwendet werden. Für den vorliegenden Fall existiert die asymptotische Verteilung, siehe Abschnitt 5.5.2.

- **Asymptotische Effizienz eines Schätzers**

Für zwei asymptotisch normalverteilte Schätzer $\hat{\theta}_n$ und $\tilde{\theta}_n$, beide mit Konvergenzrate $r(n) = \sqrt{(n)}$, ist $\hat{\theta}_n$ asymptotisch relativ effizienter als $\tilde{\theta}_n$, wenn die Differenz von deren asymptotischen Varianz-Kovarianzmatrizen $asyVar(\tilde{\theta}_n) - asyVar(\hat{\theta}_n)$ positiv semidefinit ist (Wooldridge 2010, Definition 3.11). Asymptotische Effizienz spielt in Kapitel 14 eine Rolle.

5.5. Werkzeuge für die asymptotische Analyse

5.5.1. Gesetz der großen Zahlen — Law of Large Numbers (LLN)

Ein Gesetz der großen Zahlen nennt Voraussetzungen, unter denen das arithmetische Mittel in Wahrscheinlichkeit oder sogar fast sicher gegen den wahren Mittelwert konvergiert.

- **Schwaches Gesetz der großen Zahlen von Chintschin (Khinchine's Weak Law of Large Numbers (WLLN))** Sei z_t , $t = 1, 2, \dots, n$, eine IID-Folge von Zufallsvariablen mit endlichem Erwartungswert μ . Dann gilt für das arithmetische Mittel $\hat{\mu}_n = n^{-1} \sum_{t=1}^n z_t$

$$\hat{\mu} \xrightarrow{P} \mu, \quad (5.69a)$$

$$\text{bzw. } \text{plim}(\hat{\mu}) = \mu. \quad (5.69b)$$

(Siehe z. B. Davidson (1994, Theorem 23.5) — Beweis zu schwierig.)

- **Zwei Versionen des LLN**

– Schwaches LLN (WLLN):

$$\hat{\mu} \xrightarrow{P} \mu.$$

– Starkes LLN (SLLN):

$$\hat{\mu} \xrightarrow{a.s.} \mu.$$

- Es gibt auch LLN für verschiedene nicht-IID-Fälle, siehe z. B. Davidson (2000, Section 3.2).
- **Beachte**, dass z_t auch eine Funktion einer anderen Zufallsvariable sein kann, beispielsweise eine Potenz einer Zufallsvariablen oder das Produkt von zwei verschiedenen Zufallsvariablen.

Beispiel: Schätzer des Erwartungswertes: Sind die Voraussetzungen eines der Gesetze der großen Zahlen erfüllt, ist das arithmetische Mittel ein **konsistenter Schätzer** des Erwartungswertes.

Liegt eine Zufallsstichprobe vor und weist der DGP einen endlichen Erwartungswert auf, dann gilt beispielsweise das schwache Gesetz der großen Zahlen von Chintschin

$$\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^n y_t \xrightarrow{P} \mu_0 \quad \text{bzw.} \quad \text{plim}_{n \rightarrow \infty} \hat{\mu}_n = \mu_0.$$

Beispiel: Vergleich von zwei Schätzern des Erwartungswertes mittels Monte-Carlo-Simulation

In einer Monte-Carlo-Simulation werden die Schätzeigenschaften des arithmetischen Mittels (5.59) und des ineffizienten Mittelwertschätzers (5.62) verglichen.

Aufbau:

- **DGP**

$$y_t = \mu_0 + \sigma_0 u_t, \quad u_t = (\varepsilon_t - m) / \sqrt{2m}, \quad \varepsilon_t \sim i.i.d. \chi^2(m) \quad (5.70)$$

$$\mu_0 = 1, \quad \sigma_0 = 2, \quad m = 1 \quad (5.71)$$

Die Dichte der Fehler u_t ist asymmetrisch, da die ε_t aus einer χ^2 -Verteilung mit m Freiheitsgraden gezogen werden und anschließend standardisiert werden.

- $R = 10000$ Realisationen von Stichproben mit jeweils Stichprobengrößen $n = 10, 50, 100, 500$. Beachte: Die Voraussetzungen des schwachen Gesetzes der großen Zahlen (LLN) sind erfüllt.
- Berechnung des arithmetischen Mittels und der Standardabweichung für jede Stichprobengröße und jeden Schätzer

R-Code, siehe Abschnitt A.2, Seite 333.

R-Output

	N	mu_hat_mean	mu_hat_sd	mu_tilde_mean	mu_tilde_sd
[1,]	10	1.0014164	0.31831308	1.0020431	0.7099616
[2,]	50	0.9991498	0.14162425	0.9960143	0.7054847
[3,]	100	0.9990515	0.09997354	0.9900695	0.6896356
[4,]	500	1.0003874	0.04474699	1.0058432	0.7074540

Man sieht, dass beide Schätzer erwartungstreu sind. Die Standardabweichung des alternativen Schätzers `mu_tilde` (5.62) ist bei jeder Stichprobengröße größer als die Standardabweichung des arithmetischen Mittels `mu_hat`. Darüber hinaus wird die Standardabweichung des arithmetischen Mittels mit zunehmender Stichprobengröße kleiner. Das erste Ergebnis illustriert die Effizienz des arithmetischen Mittels, das zweite Ergebnis, dass das arithmetische Mittel ein konsistenter Schätzer ist.

Die Verteilungen der Schätzer sind in den Abbildungen 5.1 und 5.2 zu sehen.

Es fällt auf, dass

- der ineffiziente Schätzer unabhängig von der Stichprobengröße eine schiefe Verteilung (wie die Fehler) hat, die Dichte des arithmetischen Mittels jedoch mit zunehmender Stichprobengröße symmetrischer wird — und wie sich im nächsten Abschnitt zeigen wird, gegen die Dichte der Normalverteilung konvergiert.

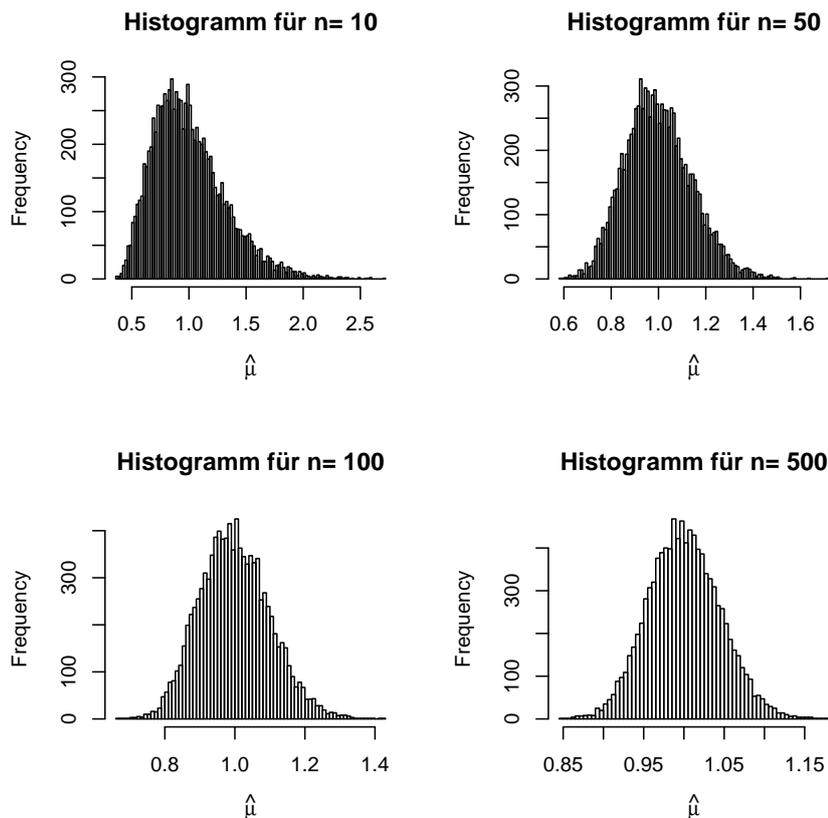


Abbildung 5.1.: Histogramme des arithmetischen Mittels (R-Programm siehe Abschnitt A.2, Seite 333) DGP siehe Gleichung (5.70)

5.5.2. Zentrale Grenzwertsätze

Vorbemerkung: Prinzipiell ist ein zentraler Grenzwertsatz von zentraler Bedeutung um in sehr allgemeinen Fällen eine Grenzverteilung für einen Schätzer bestimmen zu können. Dabei gibt es verschiedene Versionen von zentralen Grenzwertsätzen, die sich in ihren Voraussetzungen unterscheiden.

Beispiel: Schätzer des Erwartungswertes:

- Liegt eine Zufallsstichprobe vor, aber es nicht bekannt, welche Verteilung der DGP hat, also welche Verteilung z. B. die Rendite des DAX hat, dann funktioniert die Ableitung der Grenzverteilung via (5.66) nicht.
- Da die Existenz asymptotische Varianz die Konvergenzrate $r(n) = \sqrt{(n)}$ erfordert, ist zu fragen gegen welche asymptotische Verteilung die Folge der Zufallsvariablen $\sqrt{n}(\hat{\mu}_n - \mu_0)$ konvergiert, wenn y_t beispielsweise IID, aber nicht normalverteilt ist?

Die Antwort liefert für diesen Fall der **Zentrale Grenzwertsatz (central**

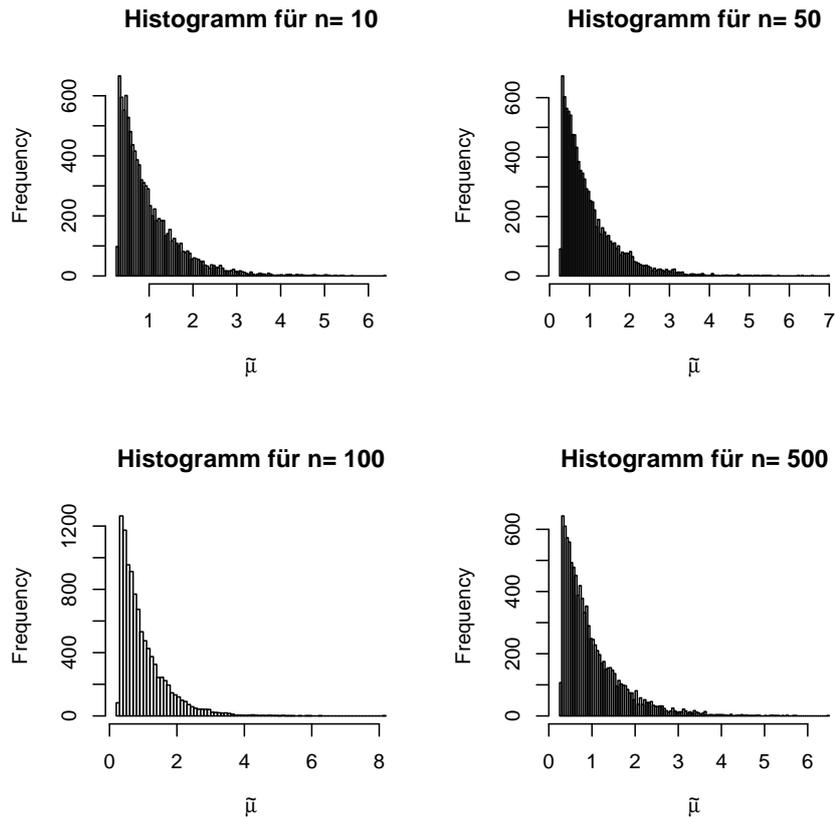


Abbildung 5.2.: Histogramme des ineffizienten Erwartungswertschätzers (5.62) (R-Programm siehe Abschnitt A.2, Seite 333) DGP siehe Gleichung (5.70)

limit theorem (CLT)) von Lindeberg und Lévy.

- **Zentraler Grenzwertsatz für IID-Zufallsvariablen** (Lindeberg-Lévy Theorem) Es sei $y_t \sim IID(\mu_0, \sigma_0^2)$, $t = 1, 2, \dots$, $|\mu_0| < \infty$, $0 < \sigma_0^2 < \infty$. Dann gilt für den Schätzer des Erwartungswertes $\hat{\mu}_n = \frac{1}{n} \sum_{t=1}^n y_t$:

$$\sqrt{n}(\hat{\mu}_n - \mu_0) \xrightarrow{d} N(0, \sigma_0^2).$$

Beweis: (Für eine Beweisidee siehe z.B. [Hendry \(1995, Section A.5\)](#)) □

Bemerkungen:

– Man kann alternativ auch

$$\sqrt{n}(\hat{\mu}_n - \mu_0) \xrightarrow{d} z, \quad z \sim N(0, \sigma_0^2)$$

schreiben, aber nicht (wie irrtümlich in [Davidson & MacKinnon \(2004, Section 4.5, p. 149\)](#))

$$\text{plim}_{n \rightarrow \infty} \sqrt{n}(\hat{\mu}_n - \mu_0) = z \sim N(0, \sigma_0^2),$$

weil dieser Wahrscheinlichkeitslimes nicht existiert; siehe für einen Beweis hierfür z. B. [Davidson \(1994, Section 23.1\)](#).

- Unabhängig von der Art der marginalen Verteilung von y_t **konvergiert der mit \sqrt{n} skalierte Schätzer des Erwartungswertes in Verteilung gegen eine Normalverteilung**, solange y_t eine endliche Varianz aufweist. Man sagt, dass der **Schätzer des Erwartungswertes asymptotisch normalverteilt** ist.
- **Alternative Schreibweise des zentralen Grenzwertsatzes:** Es bezeichne die Zufallsvariable $X_n = \hat{\mu}_n$. Dann ist die **exakte Wahrscheinlichkeitsverteilung** des arithmetischen Mittels für die Stichprobengröße n gegeben durch

$$F_n(z) := P(\hat{\mu}_n \leq z).$$

Der zentrale Grenzwertsatz sagt, dass die Folge der Verteilungsfunktionen $F_n(z)$ gegen die Verteilungsfunktion $F(z) = \Phi(z)$ punktweise konvergiert

$$\lim_{n \rightarrow \infty} F_n(z) = \Phi(z).$$

- Der zentrale Grenzwertsatz sagt nichts darüber aus, wie gut die asymptotische Verteilung die exakte Verteilung $F_n(z)$ für eine gegebene Stichprobengröße n approximiert. Um hierüber Aussagen zu gewinnen, sind i. Allg. Computersimulationen notwendig.

Beispiel: Vergleich von zwei Schätzern des Erwartungswertes mittels Monte-Carlo-Simulationen (Fortsetzung aus Abschnitt 5.5.1) Die Histogramme des arithmetischen Mittels in Abbildung 5.1 illustrieren gut den zentralen Grenzwertsatz. Die Histogramme für den ineffizienten Schätzer in Abbildung 5.2 deuten darauf hin, dass kein zentraler Grenzwertsatz gilt. Der Grund hierfür ist, dass unabhängig von der Stichprobengröße immer genau zwei Beobachtungen in die Schätzung eingehen und damit kein CLT gelten kann.

- **Zentraler Grenzwertsatz für heterogene, aber stochastisch unabhängige Zufallsvariablen** Häufig sind die y_t nicht IID, sondern sind nur unabhängig, aber nicht identisch verteilt, zum Beispiel, wenn sie eine unterschiedliche Varianz aufweisen, $y_t \sim (\mu_0, \sigma_t^2)$, $t = 1, 2, \dots$. Dann gilt für die Varianz von $\sqrt{n}\hat{\mu}_n$

$$\text{Var}(\sqrt{n}\hat{\mu}_n) = \text{Var}\left(\frac{1}{\sqrt{n}} \sum_{t=1}^n y_t\right) = \frac{1}{n} \sum_{t=1}^n \text{Var}(y_t) = \frac{1}{n} \sum_{t=1}^n \sigma_t^2.$$

Sofern die $\text{Var}(y_t)$ einige Bedingungen erfüllen, z. B. $0 < \text{Var}(y_t) < c < \infty$, für alle $t = 1, 2, \dots$, gilt ein zentraler Grenzwertsatz

$$\sqrt{n}(\hat{\mu}_n - \mu_0) \xrightarrow{d} N\left(0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \text{Var}(y_t)\right). \quad (5.72)$$

Bedingungen an die Folge der Varianzen sind notwendig, um folgende Fälle auszuschließen:

- Würde z. B. für ein festes $0 < a < 1$ gelten, dass $\text{Var}(y_t) = \sigma_0^2 a^t \rightarrow 0$ mit $t \rightarrow \infty$, dann ist $\sum_{t=1}^{\infty} \text{Var}(y_t) = \sigma_0^2 \frac{1}{1-a}$ und somit ergibt sich für

$$\text{Var}(\sqrt{n}\hat{\mu}_n) = \frac{1}{n} \sigma_0^2 \frac{1}{1-a} \rightarrow 0 \text{ für } n \rightarrow \infty,$$

die Varianz von $\sqrt{n}\hat{\mu}_n$ verschwindet also asymptotisch. Damit ist natürlich keine (sinnvolle) Grenzverteilung möglich.

- Würde entsprechend gelten $Var(y_t) = \sigma_0^2 t \rightarrow \infty$, dann erhält man

$$Var(\sqrt{n}\hat{\mu}_n) = \frac{1}{n}\sigma_0^2 \frac{n(n+1)}{2} \rightarrow \infty \text{ mit } n \rightarrow \infty.$$

Bedingungen, die sicherstellen, dass eine Grenzverteilung existiert, werden häufig als **Regularitätsbedingungen** bezeichnet.

Beispiel: Schätzer des Erwartungswertes: Dieser zentrale Grenzwertsatz ist hilfreich, wenn die unbedingte Varianz der DAX-Renditen von der Zeit abhängt, also beispielsweise vom Wochentag.

- **Zentrale Grenzwertsätze für Vektoren**

- **‡ Cramér-Wold Device:** Für eine Folge von Zufallsvektoren \mathbf{x}_n gilt

$$\mathbf{x}_n \xrightarrow{d} \mathbf{x}$$

dann und nur dann, wenn für alle zulässigen Vektoren $\boldsymbol{\lambda}$ gilt:

$$\boldsymbol{\lambda}^T \mathbf{x}_n \xrightarrow{d} \boldsymbol{\lambda}^T \mathbf{x}.$$

- **Multivariater Grenzwertsatz:** Gegeben seien die unabhängig verteilten $(r \times 1)$ -Zufallsvektoren \mathbf{v}_t mit Erwartungswert $\boldsymbol{\mu}_0$ und Varianz-Kovarianzmatrix $Var(\mathbf{v}_t)$. Dann gilt unter geeigneten Regularitätsbedingungen für den Schätzer des multivariaten Erwartungswertes $\hat{\boldsymbol{\mu}}_n = \frac{1}{n} \sum_{t=1}^n \mathbf{v}_t$

$$\sqrt{n}(\hat{\boldsymbol{\mu}}_n - \boldsymbol{\mu}_0) \xrightarrow{d} N\left(\mathbf{0}, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n Var(\mathbf{v}_t)\right). \quad (5.73)$$

5.6. Grundlagen von Tests

Statistische Tests werden angewendet, um

- ökonomische Hypothesen zu überprüfen,
- bei der Modellspezifikation und Modellüberprüfung ökonometrischer Modelle (relevante Regressorvariablen, funktionale Form der Regressionsfunktion, Verletzung von Annahmen,...).

Statistischer Test:

- Stichprobenbasiertes Entscheidungsverfahren um zu entscheiden, ob eine Hypothese abgelehnt werden muss.

- Die Hypothese muss sich auf Eigenschaften von Wahrscheinlichkeitsverteilungen beziehen, die in dem betrachteten Modell enthalten sind.
- Zur Entscheidung stehen genau zwei Alternativen: die Hypothese nicht abzulehnen oder abzulehnen.

Bestandteile eines statistischen Tests

:

1. Hypothesenpaar
2. Teststatistik
3. Entscheidungsregel
4. Entscheidung und Interpretation

Zu 1.: Zwei **disjunkte Hypothesen** über ein oder mehrere Elemente des Parametervektors $\boldsymbol{\theta} \in \Theta$, wobei $\boldsymbol{\theta}$ die Parameter der in Betracht gezogenen Wahrscheinlichkeitsverteilungen bezeichnet.

- **Nullhypothese** $H_0 : \boldsymbol{\theta} \in \Theta_{H_0}$.
- **Alternativhypothese** $H_1 : \boldsymbol{\theta} \in \Theta_{H_1}$.

Dabei deckt die Vereinigung der beiden Hypothesen $\Theta_{H_0} \cup \Theta_{H_1} = \Theta$ den gesamten Parameterraum Θ ab. (Vgl. zu Parameterraum Abschnitt 5.1.)

Beispiel: Test bzgl. des Erwartungswertes der DAX-Renditen:

- Ökonomische Frage: Ist die durchschnittliche tägliche Rendite des DAX Null?
- Statistischer Test soll im Rahmen des bisher betrachteten Modells

$$y_t \sim NID(\mu, \sigma^2), \quad \mu \in \mathbb{R}, \sigma \in \mathbb{R}^+. \quad (5.14)$$

durchgeführt werden. Diese Annahme legt die Menge der möglichen Wahrscheinlichkeitsverteilungen fest: Normalverteilungen mit Varianz $\sigma^2 > 0$ und Erwartungswert μ .

- Allgemeine Formulierung des Hypothesenpaars bezüglich des Erwartungswertes μ :

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

Im vorliegenden Fall ist $\mu_{H_0} = 0$ und damit

$$H_0 : \mu = 0 \quad \text{versus} \quad H_1 : \mu \neq 0.$$

- Wir haben noch nichts über den anderen Modellparameter, die Varianz σ^2 gesagt. Die vollständige Formulierung des Hypothesenpaares umfasst den

gesamten Parametervektor $\boldsymbol{\theta} = \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \in \Theta = (\mathbb{R} \times \mathbb{R}^+)$:

$$H_0 : \boldsymbol{\theta} = \begin{pmatrix} \mu \\ \sigma \end{pmatrix} \in \Theta_{H_0} = (\{\mu_{H_0}\} \times \mathbb{R}^+) \quad \text{versus}$$

$$H_1 : \boldsymbol{\theta} \in \Theta_{H_1} = ((\mathbb{R} \setminus \{\mu_{H_0}\}) \times \mathbb{R}^+)$$

- Wäre der Erwartungswert μ bekannt, könnte sofort entschieden werden, ob die Nullhypothese korrekt ist.
- In der Praxis kann eine Entscheidung nur auf Basis einer Stichprobe und einer *Schätzung* $\hat{\mu}(y_1, y_2, \dots, y_n)$ des Erwartungswertes getroffen werden. Ein statistischer Test liefert diese Entscheidung. Dabei sollte er gewisse Optimalitätskriterien erfüllen. Dazu später mehr.

Zu 2.: Eine **Teststatistik** λ ist eine Funktion, die aus den Stichprobenwerten \mathbf{y} berechnet wird: $\lambda = \lambda(\mathbf{y})$. Hinweis: Vor Beobachten einer Stichprobe ist eine Teststatistik eine Zufallsvariable, nach Beobachten einer Stichprobe eine Realisation einer Zufallsvariable, also eine Zahl.

Zu 3.: Eine **Entscheidungsregel**, die festlegt, für welche Werte von λ die **Nullhypothese** H_0 **abgelehnt** und für welche Werte die **Nullhypothese nicht abgelehnt** wird. Genauer: Der Wertebereich von λ wird in zwei disjunkte Teilbereiche unterteilt:

- **Ablehnungsbereich (rejection region), kritischer Bereich \mathcal{C}** Liegt die Teststatistik λ innerhalb des kritischen Bereichs, wird H_0 abgelehnt:

Lehne H_0 ab, falls $\lambda \in \mathcal{C}$.

- **Nicht-Ablehnungsbereich** Liegt die Teststatistik λ innerhalb des Nicht-Ablehnungsbereichs, wird H_0 *nicht abgelehnt*:

Lehne H_0 nicht ab, falls $\lambda \notin \mathcal{C}$.

- **Kritische Werte:** Eine oder mehrere Grenze(n) c zwischen Ablehnungs- und Nicht-Ablehnungsbereich.

Hinweis: Statt dem Symbol λ wird bei t -Tests typischerweise das Symbol t oder bei F -Tests häufig das Symbol F verwendet.

Beispiel: Test bzgl. des Erwartungswertes (mean) der DAX-Renditen — Fortsetzung:

Die Nullhypothese kann mit einem t -test wie folgt durchgeführt werden. Die einzelnen Elemente werden anschließend abgeleitet und begründet:

Zu 2.: Teststatistik des t -Tests:

$$t(\mathbf{y}) = \frac{\hat{\mu} - \mu_{H_0}}{\hat{\sigma}_{\hat{\mu}}} = \frac{\left(\frac{1}{n} \sum_{t=1}^n y_t\right) - \mu_{H_0}}{\sqrt{\frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2} \frac{1}{n}} \quad (5.74)$$

Zu 3.:

- Ablehnungsbereich: $\mathcal{C} = (-\infty, -1.96) \cup (1.96, \infty)$
- Nicht-Ablehnungsbereich: $(-1.96, 1.96)$
- Kritische Werte: $c_l = -1.96, c_r = 1.96$.
- Entscheidungsregel: Lehne H_0 ab, falls $t(\mathbf{y}) \in \mathcal{C}$.

Durchführen des Tests:

- Stichprobe: Tägliche Renditen des DAX vom 25.03.1993 bis 30.09.2015, insgesamt 5652 Beobachtungen (R-Programm siehe Abschnitt A.3, Seite 334, Daten `dax19930325_20150930.xlsx`):
- Für $H_0 : \mu_{H_0} = 0$ und $\hat{\mu} = 0.00004130056$ und $\hat{\sigma}_{\hat{\mu}} = 0.00002342752$ ergibt sich die Teststatistik

$$t(\mathbf{y}) = \frac{0.00004130056 - 0}{0.00002342752} = 1.762908$$

- $t(\mathbf{y}) \in \mathcal{C} \Rightarrow$ Lehne H_0 nicht ab.

Wie groß ist die Wahrscheinlichkeit einer Fehlentscheidung?

Eigenschaften eines Tests:

- **Fehler 1. Art (Type I error):** Der Fehler erster Art eines Tests gibt an, mit welcher Wahrscheinlichkeit die Nullhypothese H_0 für eine noch nicht erhobene Stichprobe verworfen wird, obwohl H_0 in der Grundgesamtheit korrekt ist:

$$\begin{aligned} \text{Intuitive (schlampige) Schreibweise:} & \quad P(\text{Lehne } H_0 \text{ ab} | H_0 \text{ ist wahr}) \\ \text{Exakte Schreibweise:} & \quad \alpha(\boldsymbol{\theta}) = P(\lambda \in \mathcal{C}; \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta_{H_0}. \end{aligned} \quad (5.75)$$

Beachte: Der Fehler 1. Art hängt möglicherweise von $\boldsymbol{\theta}$ ab!

Beispiel: Einseitiger t -Test, siehe später.

- **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:

$$\begin{aligned} \text{Intuitive (...) Schreibweise:} & \quad P(\text{Lehne } H_0 \text{ nicht ab} | H_1 \text{ ist wahr}) \\ \text{Exakte Schreibweise:} & \quad \beta(\boldsymbol{\theta}) = P(\lambda \notin \mathcal{C}; \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta_{H_1}. \end{aligned} \quad (5.76)$$

- **Gütefunktion (power function)** eines Tests: Die Gütefunktion eines Tests gibt die Ablehnungswahrscheinlichkeit für einen bestimmten Parametervektor $\boldsymbol{\theta} \in \Theta$ an

$$\begin{aligned} \pi(\boldsymbol{\theta}) &= P(\text{Lehne } H_0 \text{ ab}; \boldsymbol{\theta}) \\ &= 1 - P(\lambda \notin \mathcal{C}; \boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta. \end{aligned} \quad (5.77)$$

Beachte: Die Gütefunktion ist für den gesamten Parameterraum Θ definiert.

- **Güte, Macht (power) π** eines Tests: Die Güte eines Tests gibt an, mit welcher Wahrscheinlichkeit $\pi(\boldsymbol{\theta})$ die Nullhypothese für ein bestimmtes $\boldsymbol{\theta}$ abgelehnt wird, **wenn** $\boldsymbol{\theta} \in \Theta_{H_1}$ ist:

$$\pi(\boldsymbol{\theta}) = 1 - \beta(\boldsymbol{\theta}), \quad \boldsymbol{\theta} \in \Theta_{H_1}.$$

- **Größe (size)** eines Tests: In vielen Fällen hängt der Fehler 1. Art von $\boldsymbol{\theta} \in \Theta_{H_0}$ ab. Das **Supremum** der Fehler 1. Art über alle möglichen $\boldsymbol{\theta} \in \Theta_{H_0}$ wird als **Größe (size)** eines Tests bezeichnet:

$$\sup_{\boldsymbol{\theta} \in \Theta_{H_0}} P(\lambda \in \mathcal{C}; \boldsymbol{\theta}) \quad (5.78)$$

Bestimmen des kritischen Bereichs \mathcal{C}

- **Testverteilung:** $P(\lambda \leq x; \boldsymbol{\theta})$ — notwendig für die Bestimmung der Gütefunktion $\pi(\boldsymbol{\theta})$ (5.77) eines Tests.
 - unter $H_0 : P(\lambda \leq x; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_{H_0}$ — notwendig für die Bestimmung des kritischen Bereichs \mathcal{C} .
 - unter $H_1 : P(\lambda \leq x; \boldsymbol{\theta}), \boldsymbol{\theta} \in \Theta_{H_1}$ — notwendig für die Bestimmung der Güte eines Tests.

- **Optimal wäre**, den Fehler 1. Art so klein wie möglich zu machen und gleichzeitig die Gütefunktion eines Tests so groß wie möglich zu machen. Das geht leider nicht. Deshalb begrenzt man den Fehler 1. Art und möchte dann die Güte $\pi(\boldsymbol{\theta})$ für alle $\boldsymbol{\theta} \in \Theta_{H_1}$ maximieren.
- **Signifikanzniveau (level of significance, level)**: Deshalb wird ein **Signifikanzniveau** α vorgegeben, das den Fehler 1. Art beschränkt:

$$P(\text{Lehne } H_0 \text{ ab}; \boldsymbol{\theta}) = P(\lambda \in \mathcal{C}; \boldsymbol{\theta}) \leq \alpha \quad \text{für alle } \boldsymbol{\theta} \in \Theta_{H_0}. \quad (5.79)$$

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

- **Falls mehrere Tests zur Auswahl stehen**, die das Signifikanzniveau α einhalten, dann wählt man den Test, der die Powerfunktion $\pi(\boldsymbol{\theta})$ für $\boldsymbol{\theta} \in \Theta_{H_1}$ maximiert.
- Siehe folgendes Beispiel zur Ableitung der t -Statistik und des relevanten kritischen Bereichs. Tests zum Überprüfen von Hypothesen mit mehreren Parametern werden im Abschnitt 11.3.2 abgeleitet.

Beispiel: Test bzgl. des Erwartungswertes — Fortsetzung: Ableiten der Teststatistik zum Testen einer Nullhypothese bzgl. des Erwartungswertes bei bekannter Standardabweichung und Bestimmen des kritischen Bereichs

1. Unter den getroffenen Annahmen ist der Schätzer des Erwartungswertes normalverteilt, siehe (5.61)

$$\hat{\mu}(\mathbf{y}) \sim N\left(\mu, \frac{\sigma^2}{n}\right). \quad (5.61)$$

2. Allerdings hängt die Verteilung von unbekanntem Parametern ab. Dies wird durch Standardisieren vermieden:

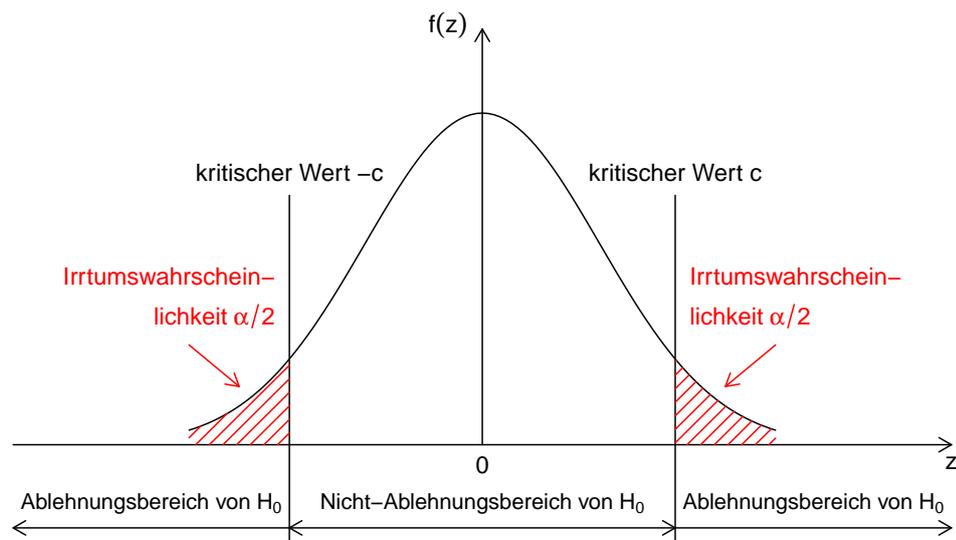
$$\frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1) \quad (5.80)$$

Fall A: $\sigma = \sigma_0$ bekannt: Unter $H_0 : \mu = \mu_{H_0}$ kann $\frac{\hat{\mu} - \mu}{\sigma_0/\sqrt{n}}$ berechnet werden und man erhält eine standardnormalverteilte Teststatistik

$$z(\mathbf{y}) = \frac{\hat{\mu} - \mu_{H_0}}{\sigma_0/\sqrt{n}} \sim N(0, 1). \quad (5.81)$$

Fall B: σ unbekannt: siehe allgemeine Ableitung von (5.74) im Rahmen des KQ-Regressionsmodells in Abschnitt 11.3.1.

3. Bestimmen des kritischen Bereichs \mathcal{C} für Fall A (Fall B geht analog):
 - a) Festlegen des Signifikanzniveaus α .
 - b) Graphik der Wahrscheinlichkeitsdichte von $z(\mathbf{y})$ unter H_0 :



Wann sollte man H_0 ablehnen?

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

- i. der geschätzte Erwartungswert $\hat{\mu}$ weit weg von μ_{H_0} (unter H_0). Das könnte Evidenz für $H_1 : \mu \neq \mu_{H_0}$ sein. Man sollte dann H_0 ablehnen.
- ii. Oder die Standardabweichung $\sigma_{\hat{\mu}} = \sigma_0/\sqrt{n}$ der geschätzten Abweichung ist klein im Vergleich zur Differenz $\hat{\mu} - \mu_{H_0}$.

D. h., man sollte H_0 ablehnen, wenn z sehr groß oder sehr klein ist.

Der kritische Bereich ist demnach

$$\mathcal{C} = (-\infty, c_l) \cup (c_r, \infty).$$

Bestimmen der kritischen Werte c_l, c_r mit Hilfe des vorgegebenen Signifikanzniveaus (5.79). Üblicherweise teilt man das vorgegebene Signifikanzniveau α auf beide Flanken symmetrisch auf. Der Fehler 1. Art ist dann kleiner oder gleich dem Signifikanzniveau α , wenn gilt

$$P(z < c_l; \mu_{H_0}, \sigma_0) \leq \alpha/2 \quad \text{und} \quad P(z > c_r; \mu_{H_0}, \sigma_0) \leq \alpha/2, \quad (5.82)$$

$$F(z; \mu_{H_0}, \sigma_0) \leq \alpha/2 \quad \text{und} \quad 1 - F(z; \mu_{H_0}, \sigma_0) \leq \alpha/2. \quad (5.83)$$

Unter H_0 ist z standardnormalverteilt (2.5), so dass gilt

$$\Phi(c_l) \leq \alpha/2 \quad \text{und} \quad (1 - \Phi(c_r)) \leq \alpha/2, \quad (5.84)$$

Idealerweise soll das Gleichheitszeichen gelten, weil dann das Signifikanzniveau den Fehler 1. Art genau kontrolliert. Der kritische Wert c_l entspricht gerade dem $\alpha/2$ -Quantil (2.9) der Standardnormalverteilung

$$c_l = q_{\alpha/2} = \Phi^{-1}(\alpha/2), \quad c_r = q_{1-\alpha/2} = \Phi^{-1}(1 - \alpha/2).$$

Aufgrund der Symmetrie der Normalverteilungsdichte ergibt sich $c_l = -c_r = -c$. Somit erhält man für $\alpha = 0.05$ die kritischen Werte $-c = -1.96$ bzw. $c = 1.96$. Siehe z. B. Table G.1 in [Wooldridge \(2009\)](#) oder berechne c mit dem R-Befehl `c <- qnorm(p=1-alpha/2)`, wobei `alpha` das Signifikanzniveau angibt.

Berechnen der Macht (Güte)

- Allgemeines Vorgehen: Bestimme zunächst Gütefunktion, d. h. Dichtefunktion der Teststatistik für ein beliebiges $\theta \in \Theta$. Da die Güte des Tests im Allg. von θ abhängt, berechnet man die Güte des Tests für alle oder ausgewählte $\theta \in \Theta_{H_1}$. Die Güte wird berechnet, indem die Fläche unter der Dichtefunktion im kritischen Bereich bestimmt wird.

Beispiel: Test bzgl. des Erwartungswertes — Fortsetzung: Im Folgenden wird weiterhin σ_0 als bekannt vorausgesetzt:

- Sowohl unter H_0 als auch H_1 gilt gegeben den Erwartungswert μ_0 und die Standardabweichung σ_0 des DGP gemäß (5.80), also

$$\frac{\hat{\mu} - \mu_0}{\sigma_0/\sqrt{n}} \sim N(0, 1).$$

Durch Erweitern erhält man

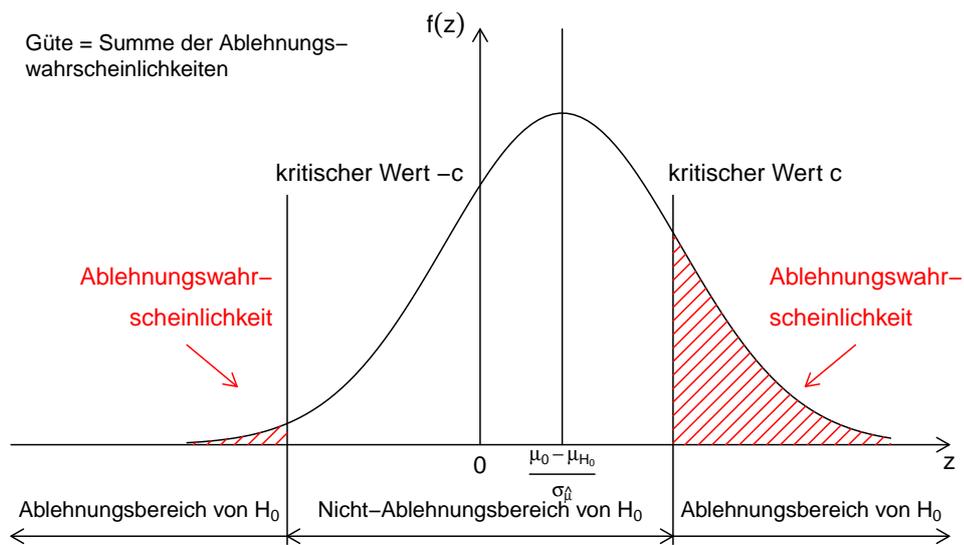
$$\frac{\hat{\mu} + \mu_{H_0} - \mu_{H_0} - \mu_0}{\sigma_0/\sqrt{n}} = \frac{\hat{\mu} - \mu_{H_0}}{\sigma_0/\sqrt{n}} + \frac{\mu_{H_0} - \mu_0}{\sigma_0/\sqrt{n}} = \underbrace{\frac{\hat{\mu} - \mu_{H_0}}{\sigma_0/\sqrt{n}}}_{z(\mathbf{y})} - \underbrace{\frac{\mu_0 - \mu_{H_0}}{\sigma_0/\sqrt{n}}}_m$$

und somit erhält man

$$z(\mathbf{y}) = \frac{\hat{\mu} - \mu_{H_0}}{\sigma_0/\sqrt{n}} \sim N\left(\frac{\mu_0 - \mu_{H_0}}{\sigma_0/\sqrt{n}}, 1\right),$$

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

- **Fazit:** Gilt H_1 , so ist die Dichte und auch die Verteilung der Teststatistik $z(\mathbf{y})$ um $(\mu_0 - \mu_{H_0})/(\sigma_0/\sqrt{n})$ verschoben.
- In der Abbildung der Dichte unter H_1 (für ein konkretes $\mu_0 \neq \mu_{H_0}$) ergibt sich die Güte bzw. power aus der Summe der beiden schraffierten Flächen: $\pi(\theta) = P(z < -c; \theta) + P(z > c; \theta)$, $\theta \in \Theta_{H_1}$.



- Für ein gegebenes $\sigma_{\hat{\mu}}$ steigt die Güte des Tests mit steigender Differenz zwischen dem Nullhypothese-Wert μ_{H_0} und dem wahren Wert μ_0 . Es ist dann “einfacher”, eine falsche Nullhypothese abzulehnen.
- Für gegebene Parameterwerte $\theta \in \Theta$ lässt die Powerfunktion berechnen und graphisch darstellen.

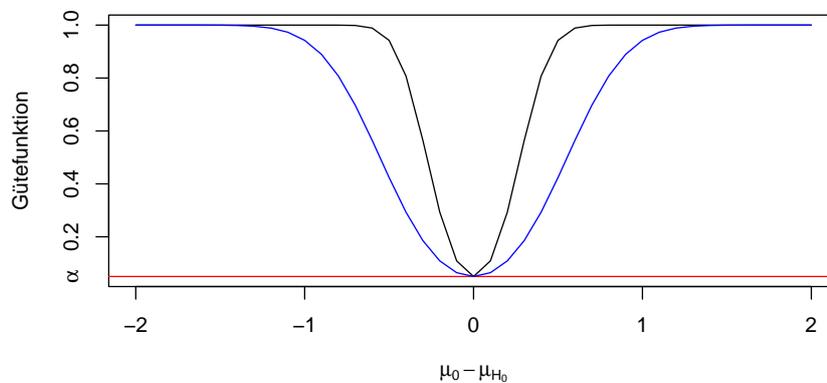


Abbildung 5.3.: Gütefunktion für Test bzgl. Erwartungswert (R-Programm siehe Abschnitt A.3, Seite 336)
 Verwendete Parameter: Signifikanzniveau $\alpha = 0.05$ (rote horizontale Linie), $\mu_0 - \mu_{H_0} \in [-2, 2]$,
 $n = 50$, $\sigma = 1$ (schwarze Linie), $\sigma = 2$ (blaue Linie).

- **Eigenschaften der Güte:** Güte eines Tests steigt mit
 - größerem Abstand zwischen korrektem Wert und Nullhypothese und/oder
 - sinkender Standardabweichung σ und/oder
 - mit Stichprobengröße n .

Fazit: Statistische Tests erfordern mindestens die Kenntnis der Wahrscheinlichkeitsverteilung

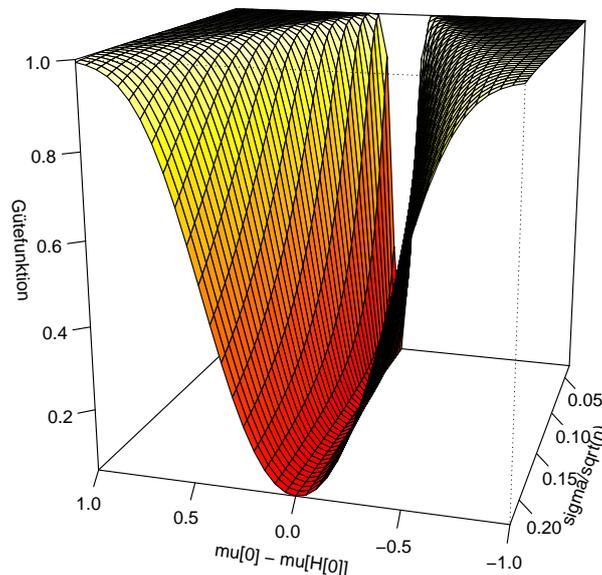


Abbildung 5.4.: Darstellung der Gütefunktion für z gegeben $\alpha = 0.05$ auf einem Gitter (R-Programm siehe Abschnitt A.3, Seite 337) Parameterbereich $\mu_0 - \mu_{H_0} \in [-1, 1]$, $\sigma_{\hat{\mu}} = \sigma_0/\sqrt{n} \in [1/\sqrt{20}, 1/\sqrt{1000}]$

der Teststatistik unter H_0 , aber um die Power zu bestimmen auch unter H_1 .

Test bzgl. Erwartungswert bei unbekannter Varianz: t -Statistik (5.74) (Fall B, Seite 131)

In der Praxis ist die Varianz σ^2 unbekannt. Das generelle Vorgehen wird anhand des Tests bzgl. des Erwartungswertes illustriert.

Beispiel: Test bzgl. des Erwartungswertes — Fortsetzung:

Lösung: Die Varianz σ^2 wird mittels

$$s^2 = \frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2$$

geschätzt (vgl. (9.25)). Einsetzen von s in (5.80) ergibt

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

und die bereits für den Test des Erwartungswertes verwendete sogenannte **t -Statistik (5.74)**

$$t(\mathbf{y}) = \frac{\hat{\mu} - \mu_{H_0}}{\hat{\sigma}_{\hat{\mu}}} = \frac{\left(\frac{1}{n} \sum_{t=1}^n y_t\right) - \mu_{H_0}}{\sqrt{\frac{1}{n-1} \sum_{t=1}^n (y_t - \bar{y})^2 \frac{1}{n}}}$$

Diese t -Statistik ist nicht mehr normalverteilt. In Abschnitt 11.3.1 wird gezeigt, dass diese t -Statistik einer t -Verteilung mit $n - 1$ Freiheitsgraden (kurz t_{n-1}) folgt.

Damit gilt

$$t(\mathbf{y}) = \frac{\hat{\mu} - \mu_{H_0}}{\hat{\sigma}_{\hat{\mu}}} \sim t_{n-1}.$$

Zu den Eigenschaften der (symmetrischen) t -Verteilung (2.36) siehe Teil I. Mathematischer Vorkurs. 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, kann man z. B. in Table G.2 in [Wooldridge \(2009\)](#)) nachschauen oder diese mit dem R-Befehl `c <- qt(p=1-alpha/2, df=n-k)` berechnen, wobei `alpha` das Signifikanzniveau und `k` die Zahl der geschätzten Parameter, hier $k = 1$, angeben.

Ein- und zweiseitige Hypothesentests mit dem t -Test

Die Möglichkeit von einseitigen Tests besteht, wenn **ein** Element θ_j des Parametervektors $\boldsymbol{\theta}$, wie z. B. der Erwartungswert μ , getestet werden soll. Sowohl für die ein- als auch den zweiseitigen Test lautet die t -Statistik allgemein

$$t(\mathbf{y}) = \frac{\hat{\theta}_j - \theta_{j,H_0}}{\hat{\sigma}_{\hat{\theta}_j}}.$$

- **Zweiseitige Tests**

$$H_0 : \theta_j = \theta_{j,H_0} \quad \text{versus} \quad H_{H_1} : \theta_j \neq \theta_{j,H_0}.$$

- **Einseitige Tests**

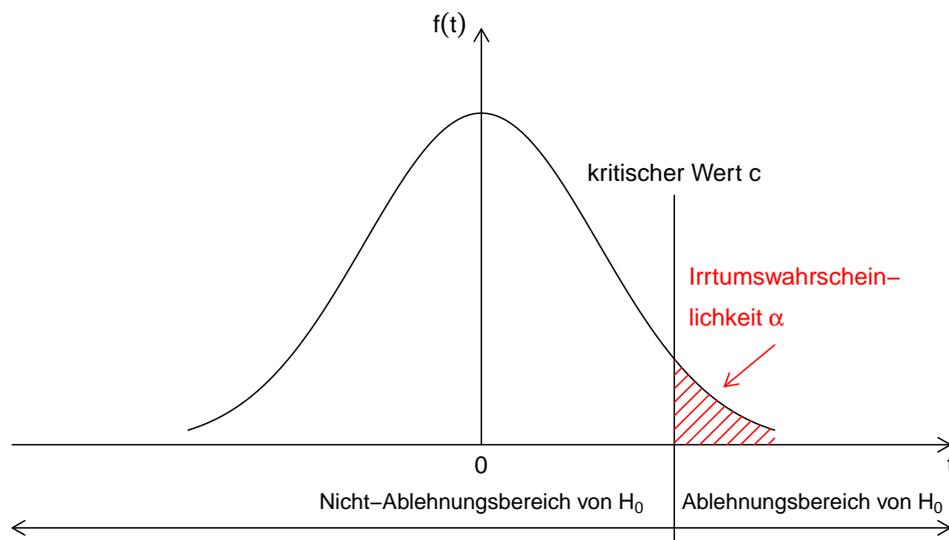
- **Test mit rechtsseitiger Alternative**

$$H_0 : \theta_j \leq \theta_{j,H_0} \quad \text{versus} \quad H_1 : \theta_j > \theta_{j,H_0}$$

Beachte: Häufig, so auch in [Wooldridge \(2009\)](#), liest man $H_0 : \theta_j = \theta_{j,H_0}$ versus $H_1 : \theta_j > \theta_{j,H_0}$. 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.

- * **Kritischer Wert:**

Dichte der t -Teststatistik für $\theta_{j,0} = \theta_{j,H_0}$, wobei $\theta_{j,0}$ den Parameterwert des DGP bezeichnet:



Es wird kein Ablehnungsbereich auf der linken Seite benötigt, da alle $\theta_j < \theta_{j,H_0}$ Elemente von H_0 sind und somit in den Nicht-Ablehnungsbereich fallen.

Fehler 1. Art und Größe eines einseitigen Tests: Angenommen, für $\theta_{j,0}$ des DGP gilt $\theta_{j,0} < \theta_{j,H_0}$, so dass H_0 vorliegt. Da die Lage der Dichte der Teststatistik $t(\mathbf{y})$ von $\frac{\theta_{j,0} - \theta_{j,H_0}}{\sigma_{\hat{\theta}_j}}$ abhängt (vgl. Graphik auf Seite 133 für $\theta_j = \mu$) befindet sich die Dichte für $\theta_{j,0} - \theta_{j,H_0} < 0$ links von der Dichte für $\theta_j = \theta_{j,H_0}$. Entsprechend ist der schraffierte Bereich, also der Fehler 1. Art, im ersten Fall kleiner als im zweiten Fall. Damit entspricht der Fehler 1. Art für $\theta_j = \theta_{j,H_0}$ gerade der Größe (5.78) des Tests. Da das gewählte Signifikanzniveau α die Größe eines Tests vorgibt, wird also der kritische Wert für $\theta_j = \theta_{j,H_0}$ bestimmt.

* **Entscheidungsregel:**

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

Beispiel: Test bzgl. des Erwartungswertes (mean) der DAX-Renditen
— **Fortsetzung: Sind die DAX-Renditen positiv?**

* Hypothesenpaar:

$$H_0 : \mu \leq 0 \quad \text{versus} \quad H_1 : \mu > 0$$

* Bestimmung des kritischen Wertes: Für $\alpha = 0.05$ erhält man aus der t -Verteilung mit 1151 Freiheitsgraden $c \leftarrow \text{qt}(p=0.95, df=1151)$ den kritischen Wert 1.646179.

* Berechnung der Teststatistik: Wie im Fall des zweiseitigen Tests (5.74):

$$t(\mathbf{y}) = \frac{0.00004130056 - 0}{0.00002342752} = 1.762908$$

* **Testentscheidung:** Da

$$t = 1.763 > c = 1.645,$$

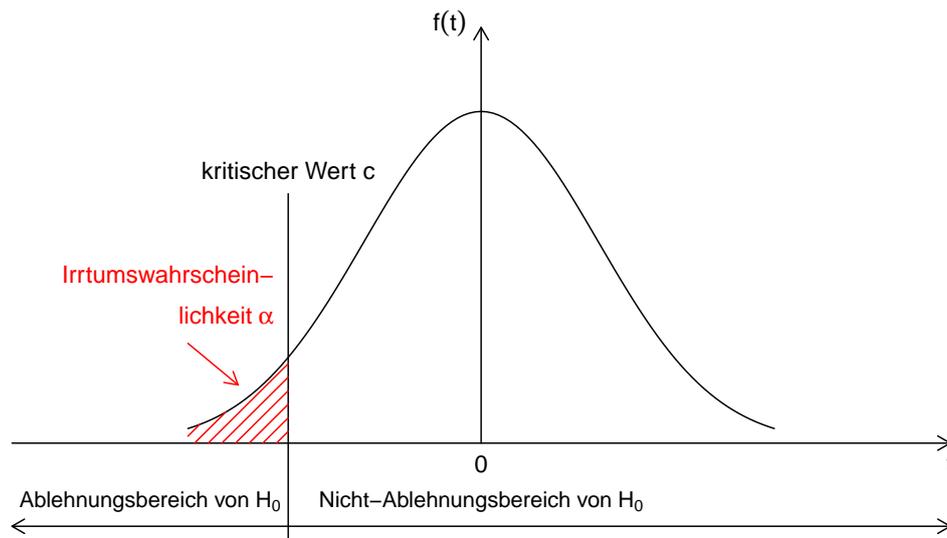
wird die Nullhypothese abgelehnt. Es liegt statistische Evidenz für einen positiven Erwartungswert der täglichen DAX-Renditen vor.

* Welches Testergebnis erhalten Sie für ein Signifikanzniveau von 1%?

– **Test mit linksseitiger Alternativhypothese**

$$H_0 : \theta_j \geq \theta_{j,H_0} \quad \text{versus} \quad H_1 : \theta_j < \theta_{j,H_0}.$$

Dichte der t -Teststatistik für $\theta_{j,0} = \theta_{j,H_0}$:



Vorgehen wie bei rechtsseitiger Alternativhypothese, nur spiegelverkehrt.

- **Fazit:** Unterschied einseitiger und zweiseitiger Tests: Fläche des gegebenen Signifikanzniveaus ist einseitig konzentriert oder zweiseitig halbiert.
- **Vorteil einseitiger Tests**

– **Da man mittels statistischer Tests Hypothesen nicht bestätigen, sondern nur ablehnen kann**, wählt man üblicherweise die Alternativhypothese so, dass sie die Vermutung widerspiegelt, die statistisch "gestützt" werden soll.

Betrifft die Vermutung also nur eine Seite, da die andere Seite nicht interessiert oder aus ökonomischen Gründen ausgeschlossen werden kann, ist ein einseitiger Test möglich.

– Beim einseitigen Test kann das vorgegebene Signifikanzniveau auf einer Seite konzentriert werden, so dass der kritische Wert im Vergleich zum zweiseitigen Test im Absolutbetrag kleiner und eine Ablehnung der Nullhypothese wahrscheinlicher wird und damit die Power steigt, wenn die Nullhypothese in der Grundgesamtheit falsch ist.

Beispiel: Test bzgl. des Erwartungswertes (mean) der DAX-Renditen

— **Fortsetzung:** Ist man ausschließlich daran interessiert, ob die DAX-Renditen positiv sind, ist ein einseitiger Test wie oben möglich. Während $H_0 : \mu = 0$ beim zweiseitigen Test gegeben ein Signifikanzniveau von 0.05 nicht abgelehnt werden kann, ist dies beim rechtsseitigen Test möglich.

– **Wichtig:** Ein einseitiger Test ist jedoch nur gerechtfertigt, wenn die Seite, die in der Null-

hypothese enthalten ist, nicht interessiert oder aus ökonomischen Gründen ausgeschlossen werden kann.

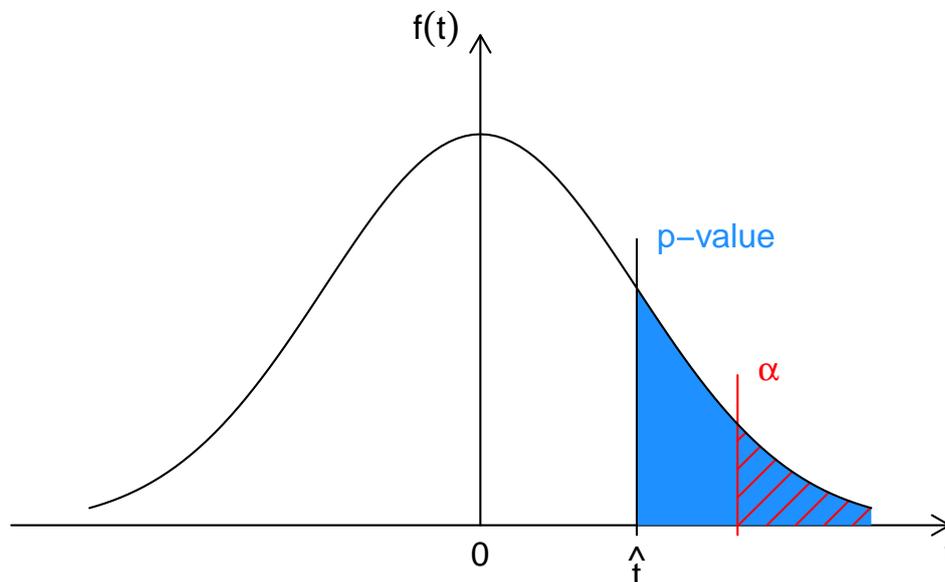
p -Werte (p -values)

- Für eine gegebene Stichprobe lässt sich für jede Teststatistik das *größte* Signifikanzniveau berechnen, bei dem die berechnete Teststatistik *gerade noch nicht* zu einer Ablehnung der Nullhypothese geführt hätte. Würde man das Signifikanzniveau weiter erhöhen, würde die Nullhypothese abgelehnt werden. Diese Wahrscheinlichkeit nennt man **p -value (probability value)**.
- Man sagt auch, dass der p -value das *kleinste* Signifikanzniveau angibt, bei dem die Nullhypothese *gerade noch* abgelehnt werden kann. Siehe Davidson & MacKinnon (2004, Section 4.2, pages 126-127) oder Wooldridge (2009, Abschnitt 4.2, p. 133).
- Im Falle eines einseitigen t -Tests mit rechtsseitiger Alternative erhält man

$$P(X > t(\mathbf{y})|\mathbf{y}, \theta_{j,H_0}) := p, \quad (5.85a)$$

$$\text{bzw.} \quad P(X \leq t(\mathbf{y})|\mathbf{y}, \theta_{j,H_0}) = 1 - p, \quad (5.85b)$$

da $P(X > t(\mathbf{y})|\mathbf{y}, \theta_{j,H_0}) = 1 - P(X \leq t(\mathbf{y})|\mathbf{y}, \theta_{j,H_0})$.



- **Entscheidungsregel mit p -Werten:** Anstelle zu prüfen, ob die Teststatistik im kritischen Bereich liegt, kann man den p -Wert mit dem Signifikanzniveau vergleichen:
Lehne H_0 ab, falls der p -value kleiner als das Signifikanzniveau α ist.

Linksseitiger Test: $p = P(t < t(\mathbf{y})|\mathbf{y}, \theta_{j,H_0}),$

Rechtsseitiger Test: $p = P(t > t(\mathbf{y})|\mathbf{y}, \theta_{j,H_0}),$

Zweiseitiger Test: $p = P(t < -|t(\mathbf{y})||\mathbf{y}, \theta_{j,H_0}) + P(t > |t(\mathbf{y})||\mathbf{y}, \theta_{j,H_0})$

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

$$H_0 : \theta_j = 0 \quad \text{versus} \quad H_1 \theta_j \neq 0.$$

Literatur: Davidson & MacKinnon (2004, Section 4.2) oder als Einstieg Wooldridge (2009, Appendix C.6).

6. Der Kleinst-Quadrate-Schätzer: Ableitung und eine Anwendung

Je nach DGP und Eigenschaften der Stichprobendaten kommen zur Schätzung der Parameter des **multiplen linearen Regressionsmodells**

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t. \quad (5.24)$$

unterschiedliche Schätzer zum Einsatz.

Wiederholung von Abschnitt 5.4: Ein **Schätzer** für den Parametervektor β ist eine vektorwertige Funktion $\tilde{\beta}(\mathbf{X}_1, y_1, \dots, \mathbf{X}_n, y_n)$ der Stichprobe $\{(\mathbf{X}_t, y_t), t = 1, \dots, n\}$.

Wichtige Schätzer für das multiple lineare Regressionsmodell

- **Kleinst-Quadrate-Schätzer** (KQ-Schätzer) — ordinary least squares estimator (OLS estimator)
⇒ alle Kapitel bis einschließlich Kapitel 11
- **verallgemeinerter Kleinst-Quadrate-Schätzer** — generalized least squares estimator (GLS estimator)
⇒ Kapitel 14
- **Instrumentvariablenschätzer** (IV-Schätzer) — (IV estimator)
⇒ MA-Modul **Fortgeschrittene Ökonometrie**
- **verallgemeinerter Momentenschätzer** — generalized method of moment estimator (GMM estimator)
⇒ MA-Modul **Fortgeschrittene Ökonometrie**
- **Maximum-Likelihood-Schätzer** (ML-Schätzer) — (ML estimator)
⇒ MA-Modul **Fortgeschrittene Ökonometrie**

Die Schätzer unterscheiden sich in ihren Voraussetzungen, Eigenschaften und Anwendungsmöglichkeiten.

6.1. Vektor- und Matrixdarstellung des multiplen linearen Regressionsmodells

- Notation:

$$\mathbf{X}_t = (x_{t1} \quad \cdots \quad x_{tk}), \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}.$$

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

- Matrixdarstellung

– für eine **Stichprobenbeobachtung** t :

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t. \quad (5.25)$$

– für die **gesamte Stichprobe**

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}, \quad (6.1)$$

- Vektordarstellung:

Das Regressionsmodell für die gesamte Stichprobe (6.1) lässt sich auch als **Vektoraddition** darstellen:

$$\mathbf{y} = \mathbf{x}_1 \beta_1 + \mathbf{x}_2 \beta_2 + \cdots + \mathbf{x}_k \beta_k + \mathbf{u}, \quad (6.2)$$

wobei die folgenden Variablenvektoren aus jeweils allen n Stichprobenbeobachtungen

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad \mathbf{x}_i = \begin{pmatrix} x_{1i} \\ x_{2i} \\ \vdots \\ x_{ni} \end{pmatrix}, \quad i = 1, \dots, k, \quad (6.3)$$

bestehen.

Weitere Matrixnotation für später

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} = (\mathbf{x}_1 \quad \mathbf{x}_2 \quad \cdots \quad \mathbf{x}_k) = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix}$$

$$\mathbf{X}^T = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^T \\ \mathbf{x}_2^T \\ \vdots \\ \mathbf{x}_k^T \end{pmatrix} = (\mathbf{X}_1^T \quad \mathbf{X}_2^T \quad \cdots \quad \mathbf{X}_n^T)$$

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \quad (6.4a)$$

$$= \begin{pmatrix} \sum_{t=1}^n x_{t1}^2 & \sum_{t=1}^n x_{t1}x_{t2} & \cdots & \sum_{t=1}^n x_{t1}x_{tk} \\ \sum_{t=1}^n x_{t2}x_{t1} & \sum_{t=1}^n x_{t2}^2 & \cdots & \sum_{t=1}^n x_{t2}x_{tk} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{t=1}^n x_{tk}x_{t1} & \sum_{t=1}^n x_{tk}x_{t2} & \cdots & \sum_{t=1}^n x_{tk}^2 \end{pmatrix} \quad (6.4b)$$

$$= \begin{pmatrix} \mathbf{x}_1^T \mathbf{x}_1 & \mathbf{x}_1^T \mathbf{x}_2 & \cdots & \mathbf{x}_1^T \mathbf{x}_k \\ \mathbf{x}_2^T \mathbf{x}_1 & \mathbf{x}_2^T \mathbf{x}_2 & \cdots & \mathbf{x}_2^T \mathbf{x}_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{x}_k^T \mathbf{x}_1 & \mathbf{x}_k^T \mathbf{x}_2 & \cdots & \mathbf{x}_k^T \mathbf{x}_k \end{pmatrix} \quad (6.4c)$$

$$= \sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t \quad (6.4d)$$

6.2. Der KQ-Schätzer für multiple lineare Regressionsmodelle

- Kleinst-Quadrate-Schätzer (KQ-Schätzer) (ordinary least squares estimator (KQ-Schätzer) von β im multiplen linearen Regressionsmodell (6.1):

$$\hat{\beta} = \left(\sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t \right)^{-1} \sum_{t=1}^n \mathbf{X}_t^T y_t \quad (6.5a)$$

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (6.5b)$$

Ableitung in Matrixnotation in Abschnitt 6.2.2.

- Regressionsmodell der Stichprobe:
 - Stichprobenregressionsfunktion (sample regression function)

$$\hat{y} = \mathbf{X} \hat{\beta} \quad (6.6)$$

- KQ-Schätzwerte/angepasste Werte/Prognosewerte (fitted values): \hat{y}
- Residuen: $u(\tilde{\beta}) = y - X\tilde{\beta}$
- KQ-Residuen (OLS residuals): $\hat{u} = y - X\hat{\beta}$

Im Weiteren werden die KQ-Residuen \hat{u} häufig einfach als Residuen bezeichnet.

- **Eigenschaften des KQ-Schätzers für das einfache multiple Regressionsmodell**

- Die **statistischen** Schätzeigenschaften sind abhängig von der Art der Datengenerierung, bzw. von den Eigenschaften der Grundgesamtheit. Sie können niemals verifiziert werden, da die Datengenerierung unbeobachtbar ist. Ihre Analyse erfordert die Methoden der **Wahrscheinlichkeitstheorie** \implies Kapitel 9 und folgende.
- Die **numerischen** Eigenschaften gelten immer und sind unabhängig von der Datengenerierung. Sie können mit algebraischen oder geometrischen Methoden untersucht werden \implies Kapitel 7.

6.2.1. Ableitung des KQ-Schätzers als Momentenschätzer

- **Basis des Momentenschätzers:** Gesetz der Großen Zahlen (law of large numbers (LLN)), vgl. Abschnitt 5.5.1.
- **Einfachster Fall von (5.34):**

$$y_t = \beta_1 + u_t, \quad E[u_t] = 0, \quad (6.7a)$$

so dass

$$\beta_1 = E[y_t] \quad (6.7b)$$

gerade dem Erwartungswert (dem ersten Moment) von y_t entspricht.

- Unter bestimmten Voraussetzungen (z. B. Vorliegen einer Zufallsstichprobe, vgl. (5.69)) rechtfertigt das **Gesetz der Großen Zahlen**, einen Erwartungswert $E[y_t]$ mit dem arithmetischen Mittel $\frac{1}{n} \sum_{t=1}^n y_t$ einer Stichprobe y_1, \dots, y_n zu schätzen,

$$\hat{E}[y_t] = \frac{1}{n} \sum_{t=1}^n y_t,$$

so dass die Genauigkeit des Schätzers mit der Stichprobengröße zunimmt. Mehr dazu in Abschnitt 5.5.1.

- β_1 kann also geschätzt werden, indem der Erwartungswert $E[y_t]$ mit dem arithmetischen Mittel geschätzt wird

$$\hat{\beta}_1 = \frac{1}{n} \sum_{t=1}^n y_t.$$

- Dieses Prinzip funktioniert wegen (5.34) auch für den KQ-Schätzer des multiplen linearen Regressionsmodell, da die Erwartungswerte $E[\mathbf{X}_t^T \mathbf{X}_t]$ und $E[\mathbf{X}_t^T y_t]$ in (5.34) bzw. (5.40) wieder durch Berechnen der Mittelwerte der Matrizen $\mathbf{X}_t^T \mathbf{X}_t$ bzw. der Vektoren $\mathbf{X}_t^T y_t$ geschätzt werden können. Man erhält:

$$\hat{E}[\mathbf{X}_t^T \mathbf{X}_t] = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t$$

$$\hat{E}[\mathbf{X}_t^T y_t] = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T y_t$$

$$\hat{\beta} = \left(\sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t \right)^{-1} \sum_{t=1}^n \mathbf{X}_t^T y_t \quad (6.5a)$$

$$= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (6.5b)$$

Dabei wurde in der vorletzten Zeile $1/n$ gekürzt. Die Matrixdarstellung (6.5) des KQ-Schätzers in der letzten Zeile folgt aus der Anwendung der Matrixregeln.

In der Ableitung von (6.5) verwendete Annahmen:

1. Es liegt eine Zufallsstichprobe vor.
2. Die Matrix $\mathbf{X}^T \mathbf{X}$ ist invertierbar — dies erfordert $rk(\mathbf{X}) = k$.

Damit wird auf jeden Fall β_{00} , wie in (5.40) definiert, geschätzt.

Soll $\hat{\beta}$ den korrekten Parametervektor β_0 des DGP schätzen, muss (5.34) gelten.

4. Damit (5.34) gilt, ist zusätzlich erforderlich (siehe Abschnitt 5.3), dass
 - a) das multiple lineare Regressionsmodell korrekt spezifiziert ist, d. h. der DGP in (5.25) enthalten ist, und
 - b) die Fehler gegeben die Regressoren im Erwartungswert Null sind, d. h. $E[u_t | \mathbf{X}_t] = 0$, so dass (5.33) gilt.

- Literatur: Davidson & MacKinnon (Vgl. 2004, Section 1.5).

6.2.2. Kleinst-Quadrate-Ableitung des KQ-Schätzers

- Gegeben ist das multiple lineare Regressionsmodell (6.1)

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

- Idee des Kleinst-Quadrate-Schätzers: Minimiere die **Summe der Quadrate der Residuen (Sum of Squared Residuals (SSR))**, also die Zielfunktion

$$S(\beta) = \sum_{t=1}^n u_t(\beta)^2 = \sum_{t=1}^n (y_t - \mathbf{X}_t \beta)^2. \quad (6.8)$$

Diese Zielfunktion ergibt sich, indem das statistische Risiko (5.48) auf Basis der quadratischen Verlustfunktion (5.46), also (5.42), mit dem arithmetischen Mittel geschätzt wird.

- Eine mögliche **Alternative** zur KQ-Zielfunktion (6.8): Minimierung der Summe der Absolutbeträge

$$S_M(\boldsymbol{\beta}) = \sum_{t=1}^n |u_t(\boldsymbol{\beta})| = \sum_{t=1}^n |y_t - \mathbf{X}_t \boldsymbol{\beta}| \quad (6.9)$$

liefert Schätzung des bedingten **Medians**, also des bedingten 50%-Quantils. Diese Zielfunktion ergibt sich, indem das statistische Risiko (5.48) auf Basis des Absolutbetrages des Schätzfehlers (5.46) mit dem arithmetischen Mittel geschätzt wird.

- Residuenquadratsumme in Matrixschreibweise:

$$\begin{aligned} S(\boldsymbol{\beta}) &= \sum_{t=1}^n u_t(\boldsymbol{\beta})^2 \\ &= \mathbf{u}(\boldsymbol{\beta})^T \mathbf{u}(\boldsymbol{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \mathbf{y}^T \mathbf{y} - 2\mathbf{y}^T \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}^T \mathbf{X}^T \mathbf{X}\boldsymbol{\beta}. \end{aligned}$$

Minimieren: Ableiten nach $\boldsymbol{\beta}$, Nullsetzen, ...

- Ableitung des KQ-Schätzers in Matrixalgebra, siehe Abschnitt 1.13 zu den Rechenregeln:
 - Aus dem Vektor der partiellen Ableitungen erster Ordnung

$$\frac{\partial S(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = -2\mathbf{X}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{X}\boldsymbol{\beta},$$

erhält man durch Nullsetzen die **Bedingungen erster Ordnung (first-order conditions (foc))**

$$\mathbf{X}^T \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{y}. \quad (6.10)$$

Diese werden auch als die **Normalgleichungen** bezeichnet.

- Ist $\mathbf{X}^T \mathbf{X}$ invertierbar — dies erfordert $rk(\mathbf{X}) = k$ —, ergibt sich wieder der KQ-Schätzer (6.5)

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

- $\hat{\boldsymbol{\beta}}$ ist ein **eindeutiges Minimum** der Zielfunktion $S(\boldsymbol{\beta})$, wenn für den Rang $rk(\mathbf{X})$ der Matrix \mathbf{X} gilt: $rk(\mathbf{X}) = k$.
- Zur Interpretation des KQ-Schätzers $\hat{\boldsymbol{\beta}}$ siehe das Ende von Abschnitt 5.3 und das spätere Kapitel 8.

6.3. Empirische Analyse von Handelsströmen: Teil 1 — ein kurzer Überblick

(Eine in manchen Teilen vereinfachte Version findet sich im Kursmaterial für die BA-Vorlesung [Einführung in die Ökonometrie](#) in Kapitel 1 und folgende.)

Die folgenden Schritte entsprechen dem Abschnitt [4.3 Bestandteile einer empirischen Analyse](#)

I. Ökonomischer Analyseteil

I.1 Ziel/Wissenschaftliche Fragestellung:

- Ermittle die Faktoren, die die Importe nach Deutschland beeinflussen, und quantifiziere ihren Einfluss.
- **Ein erster, grober (empirischer) Versuch:**

Daten: Importe nach Deutschland aus 54 Herkunftsländern im Jahr 2004 (in laufenden US-Dollars)

Datenbeschreibung	Einheit	Abkürzung	Quelle
Importe von Deutschland	laufende US-Dollars	trade_0_d_o	UN COMTRADE
Herkunftsland BIP-Daten	laufende US-Dollars	wdi_gdpusdcr_o	Weltbank - World Development Indicators

(Siehe Appendix [C](#) für genaue Datenbeschreibungen.)

R-Code zur Erzeugung des Streudiagramms in [Abbildung 6.1](#):

Der folgende R-Code sind ein Teil des R-Programms in [Abschnitt A.4](#), Seite [336](#). Bemerkung: Die eingerückten Befehle sind nur nötig, wenn PDF-Datei erzeugt werden soll.

```
#####
#                               Beginn Hauptprogramm
#####
save.pdf      <- 1                # 1=Erstelle PDFs von Graphiken, 0=sonst

# 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")
}

# Festlegung des Arbeitsverzeichnisses (working directory)
# in welchem sich das R-Program und die Daten befinden
```

```

WD          <- getwd() # Bestimme Verzeichnis der R-Datei und
setwd(WD)   # setze es als Working Directory

# Einlesen der 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)

#####
# Abschnitt 6.3
#####

##### Scatterplot mit (linearer) Regressionsgerade #####
# I.1 Ziel/Wissenschaftliche Fragestellung: erster empirischer Versuch

# Für Ausgabe im PDF Format Dateiname definieren
if (save.pdf) pdf("plot_wdi_vs_trade.pdf", height=6, width=6)

# KQ-Schätzung eines einfachen linearen Regressionsmodells, abgespeichert in ols
ols        <- 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, col = "red")
# Hinzufügen einer Legende
legend("bottomright", "Lineare Regression", col = "red", lty = 1, bty = "n")

# Device schließen
if (save.pdf) dev.off()

```

Listing 6.1: ./R_code/4_ff_Beispiel_Handelsstroeme.R

- 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?
- Wie legen wir die Gerade durch die Punktwolke?
- 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?
- Ist der mögliche Zusammenhang tatsächlich linear?
- Und: wie sehr dürfen die Ergebnisse für eine andere Stichprobe abweichen, z. B. für 2003?

I.2 Ökonomisches Modell: Einfachste Form einer Gravitationsgleichung:

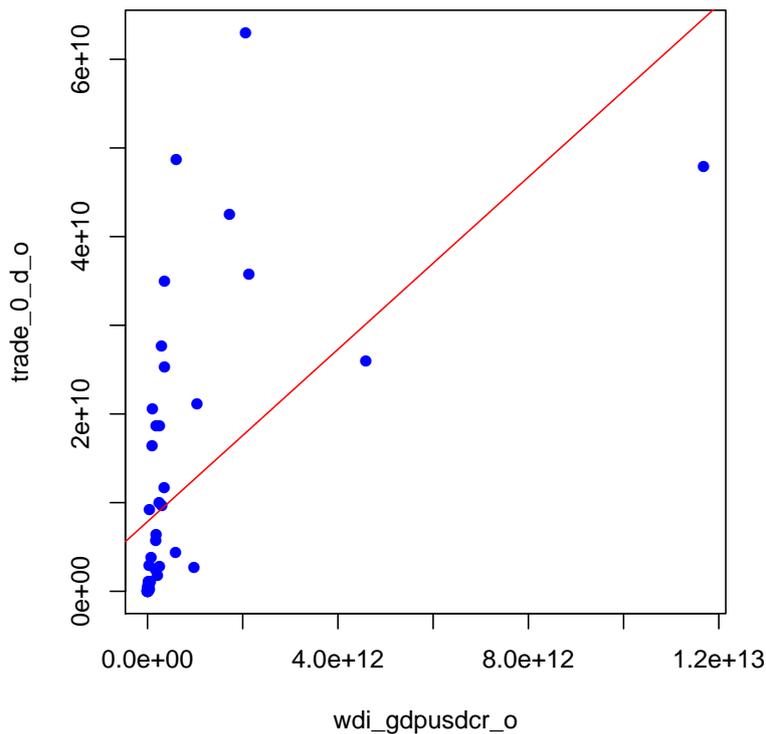


Abbildung 6.1.: Scatterplot (Streudiagramm) zu Handelsstromdaten versus GDP

- Kurze Einführung zu Gravitationsgleichungen: z. B. in [Frtianni \(2007\)](#). Eine theoretische Fundierung der Gravitationsgleichung findet sich in [Anderson & Wincoop \(2003\)](#).
- Unter idealisierten Annahmen wie vollständige Spezialisierung der Produktion, identischen Konsumpräferenzen in den Ländern, keinen Transport- und Handelskosten, werden Handelsströme zwischen Länderpaaren in Abhängigkeit vom jeweiligen Einkommen der gepaarten Länder und ihrer Entfernung zueinander erklärt:

$$M_{ijt} = A_0 Y_{it}^{\alpha_1} Y_{jt}^{\alpha_2} d_{ij}^{\alpha_3} \quad (6.11)$$

M_{ijt} :Export von Land i nach Land j in Periode t

Y_{it} :Realeinkommen in Land i in Periode t

d_{ij} :Entfernung zwischen Land i und Land j (verschiedene Maße möglich)

- Aus der ökonomischen Theorie der Gravitationsgleichungen, siehe [Frtianni \(2007\)](#), ergeben sich die **Hypothesen**,
 - dass $\alpha_1, \alpha_2 > 0, \alpha_3 < 0$ und
 - unter bestimmten Voraussetzungen die **Hypothese, dass die BIP-Elastizitäten gleich 1** sind

$$\alpha_1 = \alpha_2 = 1.$$

Diese **Hypothesen** lassen sich bei Vorliegen geeigneter Daten **statistisch testen**.

- Doppelindex ij kann in *einen* Index l umgewandelt werden.
- **Vereinfachung: Betrachtung nur einer Zeitperiode und einer Richtung**, nämlich der Importe von Deutschland im Jahr 2004. Eine so vereinfachte Gravitationsgleichung lautet

$$Importe_i = e^{\beta_1} Y_i^{\beta_2} d_i^{\beta_3}. \quad (6.12)$$

Durch Logarithmieren erhält man

$$\ln(Importe_i) = \beta_1 + \beta_2 \ln(Y_i) + \beta_3 \ln(d_i). \quad (6.13)$$

Interpretation der Parameter, vgl. (8.2):

- β_2 : BIP-Elastizität der Importe.
- β_3 : Entfernungselastizität der Importe

Eine **ökonomische Hypothese**:

Die BIP-Elastizität der Importe ist 1: $\beta_2 = 1$.

I.3 Datenverfügbarkeit

Für unser Beispiel sind alle verfügbaren Daten in Appendix C inklusive genauer Datenbeschreibungen aufgelistet.

II. Ökonometrisches Modell:

1. Auswahl einer Klasse ökonometrischer Modelle

- Wahl der Klasse **multipler linearer Regressionsmodelle**: Es wird angenommen, dass das logarithmierte theoretische Modell (6.13) nach Erweiterung um länderspezifische Merkmale und einen stochastischen Fehlerterm den systematischen Teil, vgl. Abschnitt 5.3, korrekt spezifiziert. Zusammen mit dem unsystematischen Teil (Störterm) erhält man ein **multiple lineares Regressionsmodell**

$$\ln(M_{ijt}) = \beta_1 + \beta_2 \ln Y_{it} + \beta_3 \ln Y_{jt} + \beta_4 \ln d_{ij} + \mathbf{F}_{ijt} \boldsymbol{\beta}_5 + u_{ij}, \quad (6.14)$$

\mathbf{F}_{ij} : spezifische Merkmale für Exporte von i nach j .

- Die Berücksichtigung verschiedener Perioden erfordert Paneldatenmodelle, siehe z. B. Davidson & MacKinnon (2004, Chapter 7.8).
- Die Beschränkung auf Importe (6.12) nach Deutschland und Querschnittsdaten ergibt

$$\ln(Importe_i) = \beta_1 + \beta_2 \ln(BIP_i) + \beta_3 \ln(Entfernung_i) + \mathbf{F}_i \boldsymbol{\beta}_5 + u_i. \quad (6.15)$$

- Beachte: Da die Variablen in \mathbf{F}_i noch nicht gewählt sind, sind viele Regressionsmodelle mit jeweils unterschiedlichen Variablen denkbar, die (6.15) erfüllen. Da zumindest alle Modelle multiple lineare Regressionsmodelle sind, spricht man von der Wahl einer Modellklasse.

2. Beschaffen von Daten: Erheben einer Stichprobe

- Welche Güter sollen in Importen enthalten sein?
- Wie Messung der Entfernung zwischen Ländern?
- Welche Variablen sollen in \mathbf{F}_i enthalten sein? Mögliche (und verfügbare) Variablen: Offenheit, Bevölkerung, Fläche, Koloniale Vergangenheit.
- Wie Messung von Offenheit?, etc.

Für die folgenden Schätzungen ist eine Stichprobe mit einer großen Zahl alternativer Variablen verfügbar. Siehe Appendix C für genaue Datenbeschreibungen.

Wichtig: Auswahl und Messung der Variablen kann empirische Ergebnisse substantiell beeinflussen.

3. Spezifizieren, Schätzen und Auswählen eines ökonometrischen Modells

- Zunächst vernachlässigen wir alle Variablen in \mathbf{F}_i und lineares Regressionsmodell mit BIP und Entfernung als erklärenden Variablen:

Modell 2:

$$\ln(\text{Importe}_i) = \beta_1 + \beta_2 \ln(\text{BIP}_i) + \beta_3 \ln(\text{Entfernung}_i) + u_i. \quad (6.16)$$

(Das Modell 1 enthält nur das BIP als Regressor und wird im Abschnitt 10.3 betrachtet.)

- Schätzen von Modell 2 mit dem Kleinst-Quadrate-Schätzer (KQ-Schätzer):

R-Code (Ausschnitt aus dem R-Programm in Abschnitt A.4)

```
# Die Nummerierung der Regressionsmodelle orientiert sich an
# den Modellen im Skript, Abschnitt 10.3

# Ausführen einer linearen Regression und Speichern der Ergebnisse als Objekt
mod_2_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist))

# Anzeige der Regressionsergebnisse
summary(mod_2_kq)
```

Listing 6.2: ./R_code/4_ff_Beispiel_Handelsstroeme.R

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
(1 observation deleted due to missingness)
Multiple R-squared:  0.8838,    Adjusted R-squared:  0.8787
F-statistic: 174.9 on 2 and 46 DF,  p-value: < 2.2e-16

```

4. Überprüfen des geschätzten Modells

Ist das Modell korrekt spezifiziert?

- Fehlen Variablen?
- Ist der Zusammenhang tatsächlich linear in den Logarithmen?
- Sind die Voraussetzungen zur Anwendung des KQ-Schätzers erfüllt? Ist der KQ-Schätzer zum Schätzen von (6.16) überhaupt geeignet?

Zu a): Allererster Check: Verändern sich die Parameterschätzungen, wenn weitere Variablen im Regressionsmodell berücksichtigt werden, z. B. Offenheit

Modell 3a:

$$\ln(\text{Importe}) = \beta_1 + \beta_2 \ln(\text{BIP}) + \beta_3 \ln(\text{Entfernung}) \quad (6.17)$$

$$+ \beta_4 \text{Offenheit} + \beta_6 \ln \text{Flaeche} + u. \quad (6.18)$$

R-Code (Ausschnitt aus dem R-Programm in Abschnitt A.4)

```

# Verwenden des formula-Befehls
mod_3a_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o
mod_3a_kq      <- lm(mod_3a_formula)
# Anzeige der Regressionsergebnisse des zweiten linearen Regressionsmodells
summary(mod_3a_kq)

```

Listing 6.3: ./R_code/4_ff_Beispiel_Handelsstroeme.R

R-Output

```

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
(1 observation deleted due to missingness)
Multiple R-squared:  0.8995,    Adjusted R-squared:  0.8928
F-statistic: 134.2 on 3 and 45 DF,  p-value: < 2.2e-16

```

Ist der Unterschied in den Schätzungen zwischen beiden Modellspezifikationen relevant? Zum Überprüfen kann ein t -Test verwendet werden, siehe Kapitel 11.

Statt der Variablen Offenheit oder auch als zusätzliche Variable könnte die Variable Fläche verwendet werden. Die Auswahl eines Modells kann mit Modellselektionsverfahren erfolgen, siehe Abschnitt 10.1.

Zu b) und c): Testverfahren zur Modelldiagnose werden in Kapitel 15 besprochen.

5. Verwenden des überprüften Modells

Ergibt die Modellüberprüfung keine Probleme mehr, dann können wir das Modell verwenden:

- Interpretation der Parameter des Modells. Siehe Abschnitte 8.1 und 8.4 zur Interpretation der Parameter in unterschiedlich spezifizierten Modellen.
- Durchführen von Hypothesentests:
 - Gibt es einen kausalen Zusammenhang zwischen Importen und Wirtschaftsleistung des Exportlandes? Voraussetzung ist, dass $\beta_2 \neq 0$.
 - Überprüfen der bereits gestellten Hypothese: Ist die BIP-Elastizität der Importe Eins?
 - Entsprechende Tests werden im Kapitel zur Asymptotik und zum Testen in Teil 3 in Abschnitt 11.7 durchgeführt.
- Prognosen

Systematische Fortsetzung der **Empirische Analyse von Handelsströmen: Teil 1** im Kapitel zur Modellspezifikation mit Teil 2 in Abschnitt 10.3.

7. Der Kleinst-Quadrate-Schätzer und dessen geometrische Interpretation

Multiples lineares Regressionsmodell in der Matrixdarstellung für die gesamte Stichprobe

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}. \quad (6.1)$$

KQ-Schätzer

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (6.5)$$

Sehr hilfreich für das Verständnis des KQ-Schätzers ist die Geometrie des KQ-Schätzers zu betrachten. Dies erfolgt in zwei Schritten:

1. Interpretation der Normalgleichungen (6.10) als Orthogonalitätsbedingungen \implies Abschnitt 7.1.1.
2. Interpretation der so genannten **Projektionsmatrizen** $\mathbf{P}_\mathbf{X}$ und $\mathbf{M}_\mathbf{X}$ \implies Abschnitt 7.1.2

Die **Projektionsmatrizen** $\mathbf{P}_\mathbf{X}$ und $\mathbf{M}_\mathbf{X}$ ergeben sich bei der Prognose der abhängigen Variable y und bei der Berechnung der KQ-Residuen:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X} \underbrace{(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}}_{\hat{\boldsymbol{\beta}}} := \mathbf{P}_\mathbf{X} \mathbf{y}, \quad (7.1)$$

$$\begin{aligned} \hat{\mathbf{u}} &= \mathbf{y} - \hat{\mathbf{y}} \\ &= \mathbf{y} - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} = (\mathbf{I}_n - \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T) \mathbf{y} := \mathbf{M}_\mathbf{X} \mathbf{y}. \end{aligned} \quad (7.2)$$

Definition der Projektionsmatrizen:

$$\mathbf{P}_\mathbf{X} := \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T, \quad (7.3)$$

$$\mathbf{M}_\mathbf{X} := \mathbf{I} - \mathbf{P}_\mathbf{X}. \quad (7.4)$$

Anwendungsbeispiele der Projektionsmatrizen $\mathbf{P}_\mathbf{X}$ und $\mathbf{M}_\mathbf{X}$ in diesem Kapitel:

- Zerlegung (7.15) der Total Sum of Squares: $\mathbf{y}^T \mathbf{y} = \hat{\mathbf{y}}^T \hat{\mathbf{y}} + \hat{\mathbf{u}}^T \hat{\mathbf{u}}$
- Skalierung von \mathbf{X}_t für gefittete Werte unerheblich.
- Frisch-Waugh-Lovell-Theorem und partialling-out

- Bestimmtheitsmaße
- Analyse des Einflusses von möglichen Ausreißern auf den KQ-Schätzer (6.5)

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

Anwendungen der Projektionsmatrizen in folgenden Kapiteln:

- Berechnung der Varianz eines Schätzers eines einzelnen Parameters β_j (9.15) in Abschnitt 9.3
- χ^2 -Verteilung in Abschnitt 2.9.2.
- Ableitung der Verteilung der t -Statistik (11.14) in Abschnitt 11.3.1
- Ableitung der Verteilung der F -Statistik (11.28) in Abschnitt 11.3.2
- Ableitung Fixed-Effects-Schätzer für Paneldaten

7.1. Die Geometrie des KQ-Schätzers

Zur Erinnerung: **Vektordarstellung** des multiplen linearen Regressionsmodells:

- Das Regressionsmodell (6.1) entspricht einer **Vektoraddition**

$$\mathbf{y} = \mathbf{x}_1 \beta_1 + \mathbf{x}_2 \beta_2 + \cdots + \mathbf{x}_k \beta_k + \mathbf{u} \quad (6.2)$$

- Entsprechend gilt für das Regressionsmodell der Stichprobe

$$\mathbf{y} = \mathbf{x}_1 \hat{\beta}_1 + \mathbf{x}_2 \hat{\beta}_2 + \cdots + \mathbf{x}_k \hat{\beta}_k + \hat{\mathbf{u}}, \quad (7.5)$$

wobei für die KQ-Residuen $\hat{\mathbf{u}}$ gezeigt werden wird:

$$\mathbf{x}_i^T \hat{\mathbf{u}} = 0, \quad i = 1, \dots, k, \quad (7.6b)$$

7.1.1. Orthogonalitätsbedingungen: Beweis, Interpretation

Orthogonalitätsbedingungen:

$$\mathbf{X}^T \hat{\mathbf{u}} = \mathbf{0}, \quad (7.6a)$$

$$\mathbf{x}_i^T \hat{\mathbf{u}} = \langle \mathbf{x}_i, \hat{\mathbf{u}} \rangle = 0, \quad i = 1, \dots, k. \quad (7.6b)$$

Beweis: Aus den Normalgleichungen (6.10)

$$\mathbf{X}^T \mathbf{X} \hat{\beta} = \mathbf{X}^T \mathbf{y}$$

folgt

$$\mathbf{X}^T \underbrace{(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})}_{\hat{\mathbf{u}}} = \mathbf{0}$$

und damit (7.6a).

Für die i -te Zeile gilt entsprechend

$$\mathbf{x}_i^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) = 0, \quad i = 1, \dots, k,$$

und somit (7.6b). □

Beachte: Im Gegensatz hierzu gilt für die Störterme \mathbf{u} allgemein (vgl. (1.2))

$$\mathbf{x}_i^T \mathbf{u} = \|\mathbf{x}_i\| \|\mathbf{u}\| \cos(\theta), \quad i = 1, 2, \dots, k,$$

wobei $\|\cdot\|$ die Länge (Euklidische Norm) eines Vektors und θ den Winkel zwischen den beiden Vektoren \mathbf{x}_i und \mathbf{u} misst. Letzterer ist im Allgemeinen nicht 90 Grad, so dass das Produkt im Allgemeinen nicht Null ist.

Welche Größen liegen in welchen Räumen?

- Jede **Linearkombination der Regressoren** $\mathbf{X}\mathbf{d}$ mit einem $(k \times 1)$ -Vektor \mathbf{d} liegt im **Unterraum der Regressoren** $\delta(\mathbf{X})$, also auch der **Vektor der Prognosewerte** $\mathbf{X}\boldsymbol{\beta}$ bei bekanntem $\boldsymbol{\beta}$ und der **Vektor der geschätzten Prognosewerte** $\mathbf{X}\hat{\boldsymbol{\beta}}$.

– Zur Erinnerung:

Jede Linearkombination der Spalten einer Matrix \mathbf{X} liegt im Unterraum $\delta(\mathbf{X})$, der von den Spalten der Matrix \mathbf{X} aufgespannt wird, vgl. (1.6).

– Dies gilt auch für die Regressormatrix \mathbf{X} , so dass

$$\mathbf{X}\boldsymbol{\beta} = \sum_{i=1}^k \mathbf{x}_i \beta_i \in \delta(\mathbf{X}) \subset E^n$$

für beliebige $\boldsymbol{\beta}$ im **Unterraum der Regressoren** $\delta(\mathbf{X})$ enthalten ist.

– Dies gilt auch für die Prognosewerte (fitted values) $\hat{\mathbf{y}}$

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \in \delta(\mathbf{X}).$$

- Der **Vektor der KQ-Residuen** $\hat{\mathbf{u}}$ liegt aufgrund (7.6a) im **orthogonalen Komplement zum Unterraum der Regressoren** $\delta^\perp(\mathbf{X})$, vgl. (1.7),

$$\hat{\mathbf{u}} \in \delta^\perp(\mathbf{X}) \subset E^n.$$

Die Gleichungen (7.6a) bzw. (7.6b) werden deshalb als **Orthogonalitätsbedingungen** bezeichnet.

- Der **KQ-Residuenvektor** $\hat{\mathbf{u}}$ steht **senkrecht** auf den erklärten/ prognostizierten Werten $\mathbf{X}\hat{\boldsymbol{\beta}} \in \delta(\mathbf{X})$. ($\mathbf{x}_i^T \hat{\mathbf{u}} = 0$ impliziert, dass in (1.2) $\cos(\theta) = 0$).
- $\hat{\mathbf{u}}$ entspricht dem Lot von \mathbf{y} auf $\mathbf{X}\boldsymbol{\beta}$, das durch Minimierung der Euklidischen Norm von $\mathbf{u}(\boldsymbol{\beta}) = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$ bezüglich $\boldsymbol{\beta}$ bestimmt ist:

$$\min_{\boldsymbol{\beta}} \|\mathbf{u}(\boldsymbol{\beta})\|.$$

Der KQ-Schätzer minimiert also die Euklidische Norm des Residuenvektors!

- **Beachte:** Die Minimierung einer anderen Norm (was eine andere Verlustfunktion implizieren würde) würde zu einem anderen Schätzer führen und der Residuenvektor nicht mehr senkrecht auf \mathbf{X} stehen!
- **Definition**
 - **Einheitsbasisvektor:** $\mathbf{e}_t = (0 \ 0 \ \dots \ 0 \ 1 \ 0 \ \dots \ 0)^T$, wobei in der t -ten Zeile eine 1 steht. Alle n Einheitsbasisvektoren $\mathbf{e}_t, t = 1, \dots, n$ bilden eine **Basis** für E^n , wobei jeder Basisvektor Norm $\|\mathbf{e}_t\| = 1$ hat.

- Zusätzliche Eigenschaft, falls **Konstante im Modell:**
Die Regressionsgerade, bzw. im Fall von $k > 2$ Regressoren die **Regressionshyperebene**, verläuft durch den **Schwerpunkt**, d. h. durch \bar{y} und die Mittelwerte der Regressoren $\bar{x}_i, i = 1, \dots, k$

$$\bar{y} = \hat{\beta}_1 + \hat{\beta}_2 \bar{x}_2 + \dots + \hat{\beta}_k \bar{x}_k. \quad (7.7)$$

Enthält die Regression eine Konstante, entspricht \mathbf{x}_1 einem Vektor $\boldsymbol{\iota}$ mit Einsen

$$\boldsymbol{\iota} := \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}. \quad (7.8)$$

Beweis: Ersetzen von \mathbf{x}_1 in (7.6b) durch $\boldsymbol{\iota}$ liefert

$$\boldsymbol{\iota}^T \hat{\mathbf{u}} = 0 \quad \text{bzw.} \quad \boldsymbol{\iota}^T \hat{\mathbf{u}} = \sum_{t=1}^n \hat{u}_t = 0, \quad (7.9)$$

d. h. die Abweichungen von der Regressionsgerade heben sich im Mittel auf.

$$\begin{aligned} \boldsymbol{\iota}^T \mathbf{y} &= \boldsymbol{\iota}^T \mathbf{x}_1 \hat{\beta}_1 + \boldsymbol{\iota}^T \mathbf{x}_2 \hat{\beta}_2 + \dots + \boldsymbol{\iota}^T \mathbf{x}_k \hat{\beta}_k + \underbrace{\boldsymbol{\iota}^T \hat{\mathbf{u}}}_{=0 \text{ siehe oben}} \\ n\bar{y} &= n\bar{x}_1 \hat{\beta}_1 + n\bar{x}_2 \hat{\beta}_2 + \dots + n\bar{x}_k \hat{\beta}_k + 0 \end{aligned}$$

liefert nach Multiplikation mit $1/n$ (7.7). □

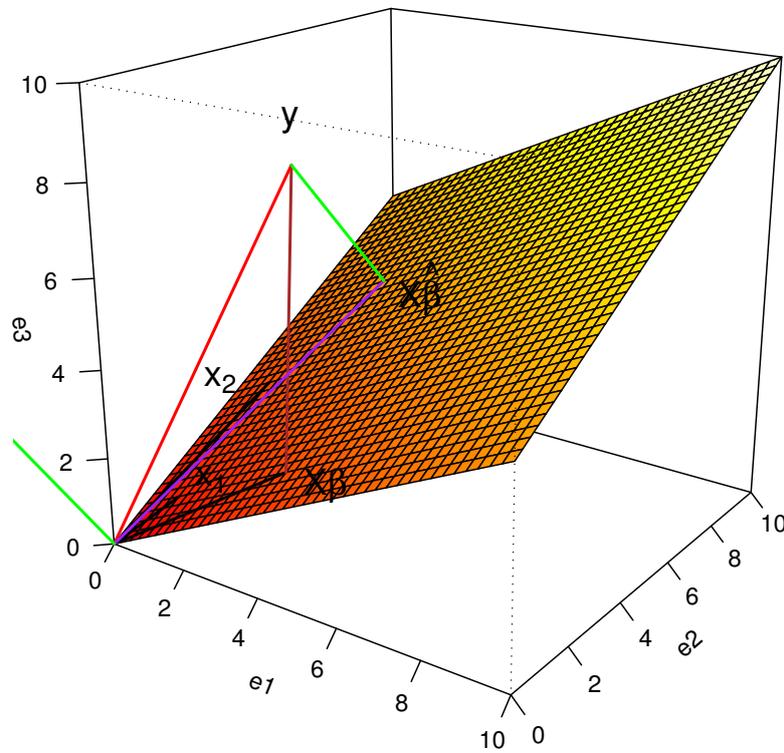


Abbildung 7.1.: Geometrie im E^3 des KQ-Schätzers,
 $n = 3$ (R-Programm (erlaubt Drehen und Kippen der Graphik) und Berechnungshinweise
siehe Abschnitt A.5, Seite 345

7.1.2. Orthogonale Projektionen und ihre Eigenschaften

Projektion in der Alltagssprache:

- Durch Lichteinwirkung wird von einem dreidimensionalen Gegenstand ein zweidimensionales Bild auf einer Wand erzeugt: Der dreidimensionale Gegenstand wird auf eine Fläche, also einem zweidimensionalen Gegenstand projiziert.
- Bei der Projektion aus dem dreidimensionalen Raum in den zweidimensionalen 'Raum' geht Information verloren.
- Je nach Standpunkt der Lichtquelle verändert sich die Projektion auf der Wand.

Definitionen

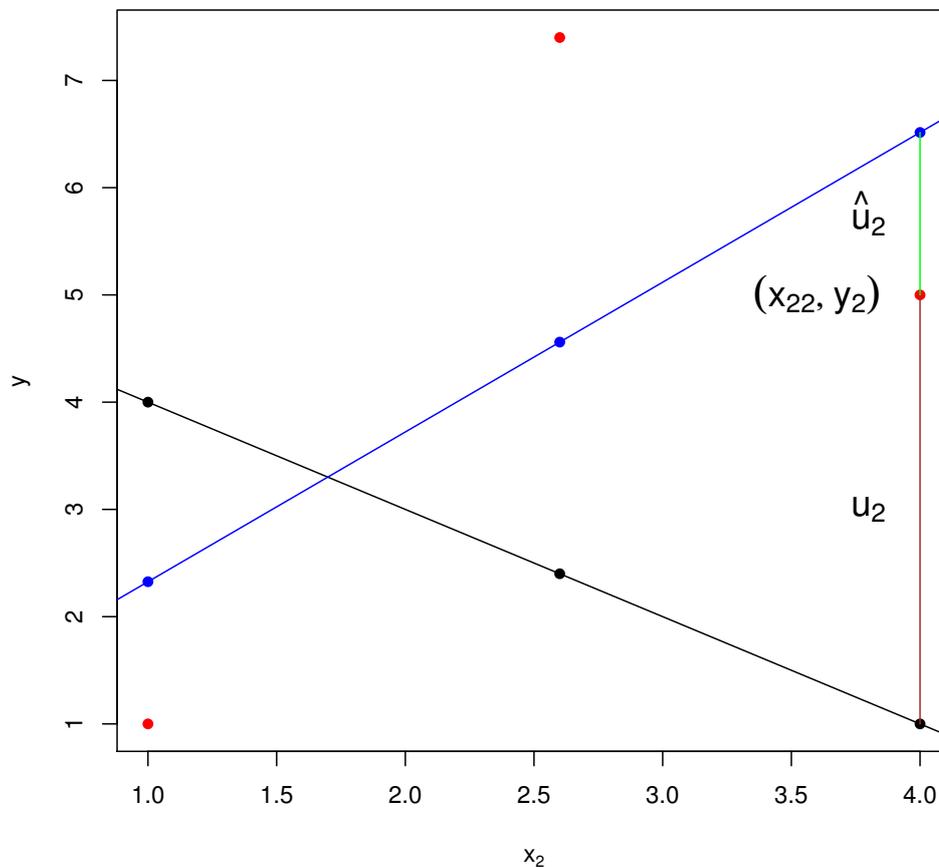


Abbildung 7.2.: Scatterplot $((x_{t2}, y_t)$ in Rot, $(x_{t2}, \mathbf{X}_t\boldsymbol{\beta})$ in Schwarz, (x_{t2}, \hat{y}_t) in Blau) (R-Programm siehe Abschnitt A.5, Seite 345)

- Eine **Projektion** ist ein Mapping von einem n -dimensionalen Raum in einen k -dimensionalen Unterraum, $k < n$. Innerhalb des Unterraums ist die Projektion invariant, da die Punkte sich durch das Mapping innerhalb des Unterraumes nicht verändern. (Vgl. Eigenschaft der Idempotenz bei Projektionsmatrizen.)
- Eine **orthogonale Projektion** ist ein Mapping, bei dem die Abstände zwischen den Punkten in E^n und den Projektionen im Unterraum minimiert werden. Also: Die Vektoren, die die Punkte in E^n und dem orthogonalen Unterraum verbinden, stehen senkrecht auf dem Unterraum.

Projektion in der Ökonometrie: Die n Stichprobenbeobachtungen \mathbf{y} legen einen Punkt in einem n -dimensionalen Euklidischen Raum fest. Ein Euklidischer Unterraum wird durch die $k \leq n$ Regressorvariablen festgelegt. Die Prognosewerte $\hat{\mathbf{y}} = \mathbf{X}\boldsymbol{\beta}$ liegen in dem genannten Unterraum, da, wie im folgenden gezeigt wird, der KQ-Schätzer eine Projektion von \mathbf{y} in diesen Unterraum darstellt. Siehe hierzu Abschnitt 7.1.

Wiederholung: Die Projektionsmatrizen für die KQ-Schätzer lauten

$$\mathbf{P}_{\mathbf{X}} := \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T, \quad (7.3)$$

$$\mathbf{M}_{\mathbf{X}} := \mathbf{I} - \mathbf{P}_{\mathbf{X}} \quad (7.4)$$

Die Projektionen in einen k -dimensionalen Unterraum erfordern, dass alle Regressoren linear unabhängig sind, also die Dimension von $\delta(\mathbf{X})$ gerade k ist. Dies begründet Annahme **(B3)** in Abs. 9.1.1).

KQ-Projektionen:

Der KQ-Schätzer umfasst zwei Projektionen:

- Der **KQ-Schätzer der gefitteten Werte** $\hat{\mathbf{y}}$

$$\hat{\mathbf{y}} = \mathbf{P}_{\mathbf{X}} \mathbf{y} \quad (7.1)$$

entspricht einer Projektion von $\mathbf{y} \in E^n$ auf $\hat{\mathbf{y}} \in \delta(\mathbf{X})$,

also vom n -dimensionalen Raum in den k -dimensionalen Unterraum $\delta(\mathbf{X})$, der von den Regressoren \mathbf{X} aufgespannt wird.

- Der **KQ-Schätzer der Residuen** $\hat{\mathbf{u}}$

$$\hat{\mathbf{u}} = \mathbf{M}_{\mathbf{X}} \mathbf{y} \quad (7.2)$$

entspricht einer Projektion von $\mathbf{y} \in E^n$ auf $\hat{\mathbf{u}} \in \delta^\perp(\mathbf{X})$,

also vom n -dimensionalen Raum in das orthogonale Komplement zu dem Unterraum, der von den Regressoren aufgespannt wird. Die Dimension des Unterraums $\delta^\perp(\mathbf{X})$ ist gerade $n - k$.

Eigenschaften der KQ-Projektionen und der dazugehörigen Projektionsmatrizen $\mathbf{P}_{\mathbf{X}}$ und $\mathbf{M}_{\mathbf{X}}$, (vgl. Abschnitt 1.6):

- Die Projektionsmatrizen $\mathbf{P}_{\mathbf{X}}$ und $\mathbf{M}_{\mathbf{X}}$ sind **idempotent**:

$$\mathbf{P}_{\mathbf{X}} \mathbf{P}_{\mathbf{X}} = \mathbf{P}_{\mathbf{X}}, \quad \mathbf{M}_{\mathbf{X}} \mathbf{M}_{\mathbf{X}} = \mathbf{M}_{\mathbf{X}}$$

und damit

$$\mathbf{P}_{\mathbf{X}} \cdot \dots \cdot \mathbf{P}_{\mathbf{X}} \cdot \mathbf{P}_{\mathbf{X}} = \mathbf{P}_{\mathbf{X}} \quad \text{bzw.} \quad \mathbf{M}_{\mathbf{X}} \cdot \dots \cdot \mathbf{M}_{\mathbf{X}} \cdot \mathbf{M}_{\mathbf{X}} = \mathbf{M}_{\mathbf{X}}.$$

- Die Projektionsmatrizen $\mathbf{P}_{\mathbf{X}}$ und $\mathbf{M}_{\mathbf{X}}$ sind **symmetrisch**, d. h. $\mathbf{P}_{\mathbf{X}}^T = \mathbf{P}_{\mathbf{X}}$ und $\mathbf{M}_{\mathbf{X}}^T = \mathbf{M}_{\mathbf{X}}$.
- $\mathbf{P}_{\mathbf{X}} \mathbf{M}_{\mathbf{X}} = \mathbf{0}$.

Geometrische Interpretation: die erste Projektion (d. h. einmalige Vormultiplikation mit $\mathbf{P}_{\mathbf{X}}$ bzw. $\mathbf{M}_{\mathbf{X}}$) liefert einen Vektor im invarianten Unterraum, den eine weitere Projektion nicht mehr verändern kann.

- \mathbf{P}_X und \mathbf{M}_X implizieren **komplementäre Projektionen**.

Denn wegen $\mathbf{M}_X = \mathbf{I} - \mathbf{P}_X$ ergibt ihre Summe den Ausgangsvektor:

$$\mathbf{P}_X \mathbf{y} + \mathbf{M}_X \mathbf{y} = \mathbf{y}. \quad (7.10)$$

- Die KQ-Methode entspricht **orthogonalen Projektionen**.

Beweis: Für zwei komplementäre Projektionen gilt

$$\mathbf{P}_X \mathbf{M}_X = \mathbf{P}_X (\mathbf{I} - \mathbf{P}_X) = \mathbf{P}_X - \mathbf{P}_X = \mathbf{O}. \quad (7.11)$$

Für beliebige Vektoren in den beiden Unterräumen $\mathbf{z} \in \delta(\mathbf{X})$ und $\mathbf{w} \in \delta^\perp(\mathbf{X})$ gilt, dass $\mathbf{z} = \mathbf{P}_X \mathbf{z}$ und $\mathbf{w} = \mathbf{M}_X \mathbf{w}$. Da \mathbf{P}_X symmetrisch ist, stehen \mathbf{z} und \mathbf{w} orthogonal zueinander, da

$$\mathbf{z}^T \mathbf{w} = \mathbf{z}^T \mathbf{P}_X^T \mathbf{M}_X \mathbf{w} = 0 \quad \text{bzw.} \quad \langle \mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{P}_X \mathbf{z}, \mathbf{M}_X \mathbf{w} \rangle = 0.$$

□

Allgemein gilt: Falls zwei Projektionen **komplementär** und die entsprechenden Projektionsmatrizen **symmetrisch** sind, so definieren sie eine **orthogonale Zerlegung**.

Geometrische Interpretation: \mathbf{P}_X und \mathbf{M}_X definieren eine **orthogonale Zerlegung** von E^n , also liegen die beiden Vektoren $\mathbf{P}_X \mathbf{y}$ und $\mathbf{M}_X \mathbf{y}$ in zwei orthogonalen Unterräumen.

Möchte man einen Vektor in $\delta(\mathbf{X})$ auf $\delta^\perp(\mathbf{X})$ projizieren, so muss das Lot in den Unterraum $\delta^\perp(\mathbf{X})$ gebildet werden. Das führt genau auf den Ursprung. Die beiden Projektionen löschen sich also gegenseitig aus. \mathbf{M}_X eliminiert alle Vektoren in $\delta(\mathbf{X})$ auf den Ursprung und entsprechend eliminiert \mathbf{P}_X alle Vektoren in $\delta^\perp(\mathbf{X})$.

Konsequenzen (der Orthogonalität) der KQ-Projektionen

- **Notation:**

Total Sum of Squares

$$TSS := \|\mathbf{y}\|^2 = \sum_{t=1}^n (y_t - \bar{y})^2 := SST, \quad (7.12)$$

Explained Sum of Squares

$$ESS := \|\hat{\mathbf{y}}\|^2 = \|\mathbf{P}_X \mathbf{y}\|^2 = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2 := SSE, \quad (7.13)$$

Sum of Squared Residuals

$$SSR := \|\hat{\mathbf{u}}\|^2 = \|\mathbf{M}_X \mathbf{y}\|^2. \quad (7.14)$$

SST, SSE wurden in [Wooldridge \(2009, Section 2.3\)](#) oder im Kursmaterial für die BAVorlesung [Einführung in die Ökonometrie](#) definiert.

$\mathbf{P}_{X,W}$ projiziert in den invarianten Unterraum $\delta(\mathbf{X}, \mathbf{W})$.

- Zerlegung der Total Sum of Squares (TSS)

$$\begin{aligned} \|\mathbf{y}\|^2 &= \|\mathbf{X}\hat{\boldsymbol{\beta}}\|^2 + \|\hat{\mathbf{u}}\|^2 \\ TSS &= ESS + SSR \end{aligned} \quad (7.15)$$

Der Zerlegung der TSS (7.15) entspricht dem **Satz von Pythagoras**.

Beweis:

$$\begin{aligned} \|\mathbf{y}\|^2 &= \|\mathbf{P}_{\mathbf{X}}\mathbf{y} + \mathbf{M}_{\mathbf{X}}\mathbf{y}\|^2 = \langle \mathbf{y}, \mathbf{y} \rangle \\ &= \langle \mathbf{P}_{\mathbf{X}}\mathbf{y} + \mathbf{M}_{\mathbf{X}}\mathbf{y}, \mathbf{P}_{\mathbf{X}}\mathbf{y} + \mathbf{M}_{\mathbf{X}}\mathbf{y} \rangle \\ &= \mathbf{y}^T \mathbf{P}_{\mathbf{X}}^T \mathbf{P}_{\mathbf{X}} \mathbf{y} + \mathbf{y}^T \mathbf{P}_{\mathbf{X}}^T \mathbf{M}_{\mathbf{X}} \mathbf{y} + \mathbf{y}^T \mathbf{M}_{\mathbf{X}}^T \mathbf{P}_{\mathbf{X}} \mathbf{y} \\ &\quad + \mathbf{y}^T \mathbf{M}_{\mathbf{X}}^T \mathbf{M}_{\mathbf{X}} \mathbf{y}. \end{aligned} \quad (7.16)$$

Man erhält

$$\begin{aligned} \|\mathbf{y}\|^2 &= \mathbf{y}^T \mathbf{P}_{\mathbf{X}} \mathbf{y} + \mathbf{y}^T \mathbf{M}_{\mathbf{X}} \mathbf{y} \\ &= \|\mathbf{P}_{\mathbf{X}}\mathbf{y}\|^2 + \|\mathbf{M}_{\mathbf{X}}\mathbf{y}\|^2, \end{aligned}$$

und damit (7.15). □

Aber:

$$\begin{aligned} \|\mathbf{P}_{\mathbf{X}}\mathbf{y}\|^2 &\leq \|\mathbf{y}\|^2 \quad \text{sowie} \\ \|\mathbf{y}\|^2 &\leq \|\mathbf{X}\hat{\boldsymbol{\beta}}\|^2 + \|\mathbf{u}\|^2. \end{aligned}$$

- **Gefittete Werte und Residuen sind unabhängig von Skalierung der Regressoren und unabhängigen Linearkombinationen** der Regressoren mittels einer nicht singulären $(k \times k)$ -Matrix \mathbf{A} , denn $\delta(\mathbf{X}) = \delta(\mathbf{X}\mathbf{A})$, da

$$\begin{aligned} \mathbf{P}_{\mathbf{X}\mathbf{A}} &= \mathbf{X}\mathbf{A} \left((\mathbf{X}\mathbf{A})^T \mathbf{X}\mathbf{A} \right)^{-1} (\mathbf{X}\mathbf{A})^T \\ &= \mathbf{X}\mathbf{A} \left(\mathbf{A}^T \mathbf{X}^T \mathbf{X}\mathbf{A} \right)^{-1} \mathbf{A}^T \mathbf{X}^T \\ &= \mathbf{X}\mathbf{A}\mathbf{A}^{-1} (\mathbf{X}^T \mathbf{X})^{-1} (\mathbf{A}^T)^{-1} \mathbf{A}^T \mathbf{X}^T \\ &= \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \\ &= \mathbf{P}_{\mathbf{X}} \end{aligned}$$

und entsprechend für $\mathbf{M}_{\mathbf{X}\mathbf{A}}$, d. h.

$$\begin{aligned} \mathbf{y} &= \mathbf{P}_{\mathbf{X}}\mathbf{y} + \mathbf{M}_{\mathbf{X}}\mathbf{y} \\ \mathbf{y} &= \mathbf{P}_{\mathbf{X}\mathbf{A}}\mathbf{y} + \mathbf{M}_{\mathbf{X}\mathbf{A}}\mathbf{y}. \end{aligned}$$

- **Frisch-Waugh-Lovell Theorem**, siehe nächsten Abschnitt.

Zu Lesen: Davidson & MacKinnon (2004, Section 2.3)

7.1.3. Partitionierte Regression und Frisch-Waugh-Lovell Theorem

- Ausgangspunkt ist wieder das multiple lineare Regressionsmodell (6.1)

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

- Ist man insbesondere an β_k interessiert, lässt sich (6.1) wie folgt schreiben:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{x}_k\beta_k + \mathbf{u} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{x}_k \end{pmatrix} \begin{pmatrix} \boldsymbol{\beta}_1 \\ \beta_k \end{pmatrix} + \mathbf{u}, \quad (7.17)$$

wobei

- \mathbf{X}_1 eine $(n \times (k-1))$ -Matrix und \mathbf{x}_k ein $(n \times 1)$ -Vektor ist,
- $\boldsymbol{\beta}_1$ ein $((k-1) \times 1)$ -Vektor und β_k ein Skalar ist.

In Abschnitt 9.6 wird gezeigt, dass der **KQ-Schätzer von β_k** mittels

$$\mathbf{y} = \mathbf{x}_k\beta_k + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u}$$

verzerrt ist, außer die empirische Korrelation zwischen \mathbf{x}_k und allen anderen Regressoren $\mathbf{x}_1, \dots, \mathbf{x}_{k-1}$ ist Null oder $\boldsymbol{\beta} = \mathbf{0}$, vgl. (9.32). Die empirische Korrelation ist Null, wenn in der Regression

$$\mathbf{x}_k = \mathbf{X}_1\boldsymbol{\delta} + \boldsymbol{\eta}$$

gilt:

$$\begin{aligned} \hat{\boldsymbol{\delta}} = \left(\mathbf{X}_1^T \mathbf{X}_1\right)^{-1} \mathbf{X}_1^T \mathbf{x}_k = \mathbf{0} &\Leftrightarrow \mathbf{X}_1^T \mathbf{x}_k = \mathbf{0} \\ &\Leftrightarrow \mathbf{x}_1^T \mathbf{x}_k = \mathbf{x}_2^T \mathbf{x}_k = \dots = \mathbf{x}_{k-1}^T \mathbf{x}_k = 0. \end{aligned} \quad (7.18)$$

Geometrische Interpretation von (7.18): \mathbf{x}_k steht orthogonal auf $\mathbf{x}_1, \dots, \mathbf{x}_{k-1}$.

- Was tun, wenn (7.18) nicht gilt? **Orthogonalisieren!**
Gleich Betrachtung des allgemeinen Falls: Das Regressionsmodell lautet dann

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u} \quad (7.19)$$

mit Partitionierung der Regressormatrix

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix}$$

in die $(n \times k_1)$ -Matrix \mathbf{X}_1 und die $(n \times k_2)$ -Matrix \mathbf{X}_2 ($k = k_1 + k_2$).

- Wie Orthogonalisieren? **Verwendung von orthogonalen Projektionen.**
Orthogonalisieren durch

$$\mathbf{Z} = \mathbf{M}_{\mathbf{X}_1} \mathbf{X}_2.$$

Test mit $\mathbf{M}_{\mathbf{X}_1} := \mathbf{M}_1$:

$$\mathbf{X}_1^T \mathbf{Z} = \mathbf{X}_1^T (\mathbf{M}_1 \mathbf{X}_2) = \mathbf{X}_1^T (\mathbf{I} - \mathbf{P}_1) \mathbf{X}_2 = \mathbf{X}_1^T \mathbf{X}_2 - \mathbf{X}_1^T \mathbf{X}_2 = \mathbf{0}.$$

- Zur Schätzung von β_2 kann man also folgende Regressionen durchführen:
 - eine KQ-Regression für $\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{u}$ oder
 - eine KQ-Regression für $\mathbf{y} = \mathbf{M}_1\mathbf{X}_2\beta_2 + \mathbf{v}$.

Mögliches **Problem**: Die Residuenvektoren sind nicht gleich (verifizieren!).
Ausweg: Multiplikation aller Variablen mit \mathbf{M}_1 . Man erhält

$$\mathbf{M}_1\mathbf{y} = \mathbf{M}_1\mathbf{X}_1\beta_1 + \mathbf{M}_1\mathbf{X}_2\beta_2 + \mathbf{M}_1\mathbf{u}, \quad (7.20)$$

$$\mathbf{M}_1\mathbf{y} = \mathbf{M}_1\mathbf{X}_2\beta_2 + \boldsymbol{\varepsilon}. \quad (7.21)$$

Frisch-Waugh-Lovell Theorem (FWL-Theorem)

1. Die KQ-Schätzer für β_2 für die Regressionsmodelle

$$\mathbf{y} = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{u} \quad (7.19)$$

und

$$\mathbf{M}_1\mathbf{y} = \mathbf{M}_1\mathbf{X}_2\beta_2 + \boldsymbol{\varepsilon} \quad (7.21)$$

sind numerisch identisch.

2. Die KQ-Residuen der Regressionen für (7.19) und (7.21) sind numerisch identisch.

Regeln zum Rechnen mit Projektionsmatrizen bei partitionierten Regressionen (7.19):

$$\mathbf{P}_X\mathbf{P}_1 = \mathbf{P}_1\mathbf{P}_X = \mathbf{P}_1 \quad (7.22)$$

$$\mathbf{M}_X\mathbf{M}_1 = \mathbf{M}_1\mathbf{M}_X = \mathbf{M}_X \quad (7.23)$$

Die Multiplikation von zwei unterschiedlichen Projektionsmatrizen, wobei der Unterraum der einen Projektionsmatrix im Unterraum der anderen Projektionsmatrix enthalten ist, entspricht der Projektionsmatrix, die in den kleineren Unterraum projiziert.

Beweis des FWL-Theorems: Vgl. Davidson & MacKinnon (2004, Section 2.5, p. 68-69). Aussage 1.: Der KQ-Schätzer für (7.21) lautet

$$\hat{\beta}_2 = (\mathbf{X}_2^T\mathbf{M}_1\mathbf{X}_2)^{-1}\mathbf{X}_2^T\mathbf{M}_1\mathbf{y}. \quad (7.24)$$

Einsetzen der KQ-Schätzer für die vollständige Regression (7.19) ergibt

$$\mathbf{y} = \mathbf{X}_1\hat{\beta}_1 + \mathbf{X}_2\hat{\beta}_2 + \hat{\mathbf{u}} \quad (7.25)$$

Multiplikation von links mit $\mathbf{X}_2^T\mathbf{M}_1$ liefert

$$\mathbf{X}_2^T\mathbf{M}_1\mathbf{y} = \mathbf{X}_2^T\mathbf{M}_1\mathbf{X}_2\hat{\beta}_2, \quad (7.26)$$

da $\mathbf{X}_2^T \mathbf{M}_1 \mathbf{X}_1 = \mathbf{0}$ und $\mathbf{X}_2^T \mathbf{M}_1 \hat{\mathbf{u}} = \mathbf{X}_2^T \mathbf{M}_1 \mathbf{M}_X \mathbf{y} = \underbrace{\mathbf{X}_2^T \mathbf{M}_X \mathbf{y}}_{=0} = \mathbf{0}$. Auflösen von (7.26) liefert gerade (7.24).

Aussage 2.: Multiplikation von (7.25) mit \mathbf{M}_1 liefert

$$\mathbf{M}_1 \mathbf{y} = \mathbf{M}_1 \mathbf{X}_2 \hat{\beta}_2 + \underbrace{\mathbf{M}_1 \hat{\mathbf{u}}}_{=\mathbf{M}_1 \mathbf{M}_X \mathbf{y}} = \mathbf{M}_1 \mathbf{X}_2 \hat{\beta}_2 + \underbrace{\mathbf{M}_X \mathbf{y}}_{\hat{\mathbf{u}}} = \mathbf{M}_1 \mathbf{X}_2 \hat{\beta}_2 + \hat{\mathbf{u}}.$$

Damit entsprechen die KQ-Residuen $\hat{\boldsymbol{\varepsilon}}$ für (7.21) gerade den KQ-Residuen $\hat{\mathbf{u}}$ für die volle Regression (7.19). \square

Interpretation des Frisch-Waugh-Lovell Theorems: Der KQ-Schätzer für β_2 kann auch sequentiell durch verschiedene KQ-Regressionen mit weniger Variablen ausgeführt werden. Die Regression

$$\mathbf{M}_1 \mathbf{y} = \mathbf{M}_1 \mathbf{X}_2 \beta_2 + \boldsymbol{\varepsilon} \quad (7.21)$$

entspricht dabei einer Regression von Residuen auf Residuen folgender KQ-Schätzungen:

- $\mathbf{M}_1 \mathbf{y}$ entspricht gerade den Residuen einer Regression von \mathbf{y} auf \mathbf{X}_1 .
- \mathbf{X}_2 enthält gerade die Regressoren $\mathbf{x}_{k_1+1}, \dots, \mathbf{x}_{k_1+k_2}$. Damit entspricht für jedes $j = k_1 + 1, \dots, k_1 + k_2$ der Vektor $\mathbf{M}_1 \mathbf{x}_j$ gerade den Residuen einer Regression von \mathbf{x}_j auf \mathbf{X}_1 .

Durch Vormultiplizieren von \mathbf{M}_1 in (7.21) werden von der jeweiligen Variable Residuen erzeugt, die senkrecht auf dem Unterraum stehen, der von den Regressoren in \mathbf{X}_1 aufgespannt wird, so dass für die KQ-Schätzung von (7.21) die Einflüsse der Regressoren in \mathbf{X}_1 keine Rolle spielen, da sie jeweils orthogonal zu den Variablen in (7.21) sind.

Zu Lesen: Davidson & MacKinnon (2004, Section 2.4)

7.2. Anwendungen des Frisch-Waugh-Lovell Theorems

1. Bereinigung von nicht interessierenden Regressoren

Beispiele:

- **Konstante:** Sei o.B.d.A. $\mathbf{x}_1 = \boldsymbol{\iota} = (1, 1, \dots, 1)^T$ und damit $\mathbf{M}_{\boldsymbol{\iota}} := \mathbf{I}_n - \frac{1}{n} \boldsymbol{\iota} \boldsymbol{\iota}^T$. $\mathbf{M}_{\boldsymbol{\iota}}$ heißt zentrierende Matrix, da

$$\mathbf{M}_{\boldsymbol{\iota}} = \mathbf{I}_n - \frac{1}{n} \boldsymbol{\iota} \boldsymbol{\iota}^T = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & & \\ \vdots & & \ddots & \\ 0 & & & 1 \end{pmatrix} - \frac{1}{n} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & & \\ \vdots & & \ddots & \\ 1 & & & 1 \end{pmatrix} = \begin{bmatrix} 1 - \frac{1}{n} & & & -\frac{1}{n} \\ & 1 - \frac{1}{n} & & \\ & & \ddots & \\ -\frac{1}{n} & & & 1 - \frac{1}{n} \end{bmatrix}.$$

Vormultiplikation eines Vektors mit $\mathbf{M}_{\boldsymbol{\iota}}$ berechnet die **Abweichungen vom Mittelwert des Vektors**. $\mathbf{M}_{\boldsymbol{\iota}} \mathbf{X}$ liefert **zentrierte Regressoren**. Der Vektor

der Steigungsparameter β_2 lässt sich mit dem Frisch-Waugh-Lovell-Theorem schätzen:

$$\begin{aligned}\mathbf{M}_l \mathbf{y} &= \mathbf{M}_l \mathbf{X}_2 \beta_2 + \mathbf{M}_l \mathbf{u}, \\ \hat{\beta}_2 &= (\mathbf{X}_2^T \mathbf{M}_l \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{M}_l \mathbf{y}.\end{aligned}$$

Interpretation: Die Punktwolke in einem Scatterdiagramm wird durch Zentrierung von \mathbf{x} bzw. \mathbf{y} verschoben, die Steigung der Regressionsgerade ändert sich nicht.

- ‡ **Saisondummies:** Bei Zeitreihen können regelmäßig wiederkehrende Schwankungen häufig durch Saisondummies modelliert werden. Fasst man Saisondummies und Konstante, sofern vorhanden, in der Matrix \mathbf{S} zusammen und ist man lediglich am Parametervektor β interessiert, kann man

$$\begin{aligned}\mathbf{y} &= \mathbf{S}\alpha + \mathbf{X}\beta + \mathbf{u} \quad \text{oder} \\ \mathbf{M}_S \mathbf{y} &= \mathbf{M}_S \mathbf{X}\beta + \mathbf{M}_S \mathbf{u}\end{aligned}$$

schätzen, wobei $\mathbf{M}_S = \mathbf{I} - \mathbf{S}(\mathbf{S}^T \mathbf{S})^{-1} \mathbf{S}^T$. Bei Quartalsdaten, die mit dem ersten Quartal eines Jahres beginnen und mit dem letzten Quartal eines Jahres aufhören, kann man \mathbf{S} folgendermaßen wählen:

$$\mathbf{S} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad \text{oder} \quad \mathbf{S} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad \text{oder ...}$$

- ‡ **Zeittrend**

2. ‡ Darstellung des unzentrierten Bestimmtheitsmaßes

- Beachte Definitionen von SSE, SST, TSS, ESS in (7.12) und (7.13), SSR in (7.14) sowie

$$\|\mathbf{y}\|^2 = \|\hat{\mathbf{y}}\|^2 + \|\hat{\mathbf{u}}\|^2 \quad (7.15)$$

- **Unzentriertes R^2 :**

$$R_u^2 := \frac{ESS}{TSS} = \frac{\|\hat{\mathbf{y}}\|^2}{\|\mathbf{y}\|^2} = \frac{\|\mathbf{P}_{\mathbf{X}} \mathbf{y}\|^2}{\|\mathbf{y}\|^2} = \cos^2 \theta \quad \Rightarrow \quad 0 \leq R_u^2 \leq 1. \quad (7.27)$$

Beweisskizze: Das letzte Gleichheitszeichen in (7.27) folgt aus der Definition des Kosinus: $\cos \theta = \text{Ankathete}/\text{Hypotenuse} = \|\mathbf{P}_{\mathbf{X}} \mathbf{y}\|/\|\mathbf{y}\|$. \square

Aus (7.15) folgt auch

$$R_u^2 = 1 - \frac{SSR}{TSS} = 1 - \frac{\|\hat{\mathbf{u}}\|^2}{\|\mathbf{y}\|^2} = 1 - \frac{\|\mathbf{M}_{\mathbf{x}}\mathbf{y}\|^2}{\|\mathbf{y}\|^2}. \quad (7.28)$$

Nachteil von R_u^2 : Ist eine Konstante im Regressionsmodell, $\mathbf{x}_1 = \mathbf{1}$, und sind die Daten nicht zentriert, hängt R_u^2 von der Größe der Konstante ab (Davidson & MacKinnon 2004, Section 2.5), da bei einer Erhöhung von β_1 der Nenner sich verändert, während der Zähler konstant bleibt.

3. Darstellung des (zentrierten) Bestimmtheitsmaßes (Coefficient of Determination)

(Zentriertes) Bestimmtheitsmaß R^2 : Wird vom Bestimmtheitsmaß gesprochen, wird darunter im Allgemeinen das zentrierte R^2 verstanden. In der Literatur gebräuchliche Definitionen:

$$R^2 = \frac{SSE}{SST} = \frac{\sum_{t=1}^n (\hat{y}_t - \bar{y})^2}{\sum_{t=1}^n (y_t - \bar{y})^2}. \quad (7.29)$$

$$R^2 = 1 - \frac{SSR}{SST} = 1 - \frac{\|\hat{\mathbf{u}}\|^2}{\|\mathbf{M}_i\mathbf{y}\|^2} = 1 - \frac{\|\mathbf{M}_{\mathbf{x}}\mathbf{y}\|^2}{\|\mathbf{M}_i\mathbf{y}\|^2}. \quad (7.30)$$

$$R^2 = \frac{\|\mathbf{M}_i\hat{\mathbf{y}}\|^2}{\|\mathbf{M}_i\mathbf{y}\|^2} = \frac{\|\mathbf{M}_i\mathbf{P}_{\mathbf{x}}\mathbf{y}\|^2}{\|\mathbf{M}_i\mathbf{y}\|^2}. \quad (7.31)$$

$$R^2 = \frac{\|\mathbf{P}_{\mathbf{x}}\mathbf{M}_i\mathbf{y}\|^2}{\|\mathbf{M}_i\mathbf{y}\|^2} \Rightarrow 0 \leq R^2 \leq 1 \quad (\text{wegen (7.10) mit } \mathbf{M}_i\mathbf{y}). \quad (7.32)$$

$$\begin{aligned} R^2 &= \widehat{Corr}(\hat{\mathbf{y}}, \mathbf{y})^2 = \frac{\left(\sum_{t=1}^n (\hat{y}_t - \bar{\hat{y}})(y_t - \bar{y})\right)^2}{\left(\sum_{t=1}^n (\hat{y}_t - \bar{\hat{y}})^2\right) \left(\sum_{t=1}^n (y_t - \bar{y})^2\right)} \\ &= \frac{(\hat{\mathbf{y}}^T \mathbf{M}_i \mathbf{y})^2}{(\hat{\mathbf{y}}^T \mathbf{M}_i \hat{\mathbf{y}}) (\mathbf{y}^T \mathbf{M}_i \mathbf{y})} \Rightarrow 0 \leq R^2 \leq 1. \end{aligned} \quad (7.33)$$

Anmerkungen:

- Alle Definitionen sind identisch, falls Konstante im Modell.
- **Warnung:** Wenn keine Konstante im Modell enthalten, garantieren nicht alle Definitionen, dass $R^2 \in [0, 1]$, siehe folgende Tabelle. Software liefert je nach verwendeter Definition unterschiedliche Ergebnisse.
- **Eigenschaften** verschiedener Definitionen bei KQ:

Definition	verwendet z. B. von	Wertebereich ohne Konstante in \mathbf{X}
(7.29)	Wooldridge (2009, Equation (2.38))	≥ 0
(7.30)	Davidson & MacKinnon (2004, Equation (2.55)), Wooldridge (2009, Equation (2.38))	≤ 1
(7.31)	Greene (2008, Equation (3-26))	≥ 0
(7.32)	Davidson & MacKinnon (2004, Equation (2.55))	$[0, 1]$
(7.33)	Greene (2008, Equation (3-27))	$[0, 1]$

- Gültige Umformungen, falls \mathbf{X} mit Konstante:

$$\mathbf{P}_i \mathbf{P}_\mathbf{X} = \mathbf{P}_i. \quad (7.34a)$$

$$\mathbf{M}_i \mathbf{M}_\mathbf{X} = \mathbf{M}_\mathbf{X}. \quad (7.34b)$$

$$\hat{\mathbf{y}}^T \mathbf{M}_i \hat{\mathbf{y}} = \hat{\mathbf{y}}^T \mathbf{M}_i \mathbf{y}. \quad (7.34c)$$

$$\iota \bar{\hat{\mathbf{y}}} = \mathbf{P}_i \hat{\mathbf{y}} = \mathbf{P}_i \mathbf{P}_\mathbf{X} \mathbf{y} = \mathbf{P}_i \mathbf{y} = \iota \bar{\mathbf{y}} \Leftrightarrow \bar{\hat{\mathbf{y}}} = \bar{\mathbf{y}}. \quad (7.34d)$$

Beweisskizze: Anwenden der Regeln zum Rechnen mit Projektionsmatrizen (7.22) und (7.23) für $\mathbf{x}_1 = \iota$. (7.34c) gilt da $\hat{\mathbf{y}}^T \mathbf{M}_i \hat{\mathbf{u}} = \hat{\mathbf{y}}^T \mathbf{M}_i \mathbf{M}_\mathbf{X} \hat{\mathbf{u}} = \mathbf{y}^T \mathbf{P}_\mathbf{X}^T \mathbf{M}_\mathbf{X} \hat{\mathbf{u}} = 0$. \square

Allgemeine Anmerkungen

- Alle Definitionen von R^2 (alle außer (7.33)), die auf dem Satz von Pythagoras basieren, sind nur bei Verwendung des KQ-Schätzers aussagekräftig. Ansonsten können Werte kleiner Null oder größer Eins auftreten.
- Da für (7.33) $0 \leq \widehat{Corr}(\hat{\mathbf{y}}, \mathbf{y})^2 \leq 1$ gilt, aber der Satz des Pythagoras nicht verwendet wurde, kann das **Quadrat des empirischen Korrelationskoeffizienten** als **Goodness-of-Fit**-Maß immer verwendet werden. Es wird dann häufig als **Pseudo- R^2** bezeichnet.

Zu Lesen: Davidson & MacKinnon (2004, Section 2.5)

4. ‡ Leverage-Effekt

- Um den Effekt einer möglicherweise einflussreichen Stichprobenbeobachtung (y_t, \mathbf{X}_t) abzuschätzen, werden die KQ-Schätzer für die komplette Stichprobe mit dem KQ-Schätzer für die Stichprobe ohne Beobachtung t verglichen. Letztere erhält man durch Aufnahme einer Impulsdummy \mathbf{e}_t in (6.1)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}_t \alpha + \mathbf{u}, \quad (7.35)$$

da $\mathbf{M}_{\mathbf{e}_t} \mathbf{y} = \mathbf{M}_{\mathbf{e}_t} \mathbf{X}\boldsymbol{\beta} + \text{Residuen}$ (Frisch-Waugh-Lovell Theorem) gilt und wegen $\mathbf{M}_{\mathbf{e}_t} = \mathbf{I} - \mathbf{e}_t \mathbf{e}_t^T$ die t -te Beobachtung wegfällt.

- $\mathbf{P}_\mathbf{X}$ wird manchmal als **Hat-Matrix (hat matrix)** und ihr t -tes Diagonalelement deswegen als h_t bezeichnet. Letzteres entspricht

$$0 \leq h_t = \mathbf{e}_t^T \mathbf{P}_\mathbf{X} \mathbf{e}_t = \|\mathbf{P}_\mathbf{X} \mathbf{e}_t\|^2 \leq \|\mathbf{e}_t\|^2 = 1. \quad (7.36)$$

Es gilt, dass $\sum_{t=1}^n h_t = \text{tr}(\mathbf{P}_\mathbf{X}) = k$, siehe Übung oder (Davidson & MacKinnon 2004, Section 2.6), und somit

$$\bar{h} = \frac{k}{n} \quad (7.37)$$

und falls \mathbf{X} eine Konstante enthält, dass

$$h_t \geq 1/n \quad \Leftrightarrow \quad h_t = \|\mathbf{P}_\mathbf{X} \mathbf{e}_t\|^2 \geq \|\mathbf{P}_t \mathbf{P}_\mathbf{X} \mathbf{e}_t\|^2 = \|\mathbf{P}_t \mathbf{e}_t\|^2 = 1/n.$$

- Wird der KQ-Schätzer für $\boldsymbol{\beta}$ auf Basis von (7.35) (ohne die t -te Beobachtung) mit $\hat{\boldsymbol{\beta}}^{(t)}$ bezeichnet, lässt sich die Differenz der KQ-Schätzer angeben als

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{(t)} = \hat{\alpha} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_\mathbf{X} \mathbf{e}_t = \frac{1}{1 - h_t} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_t^T \hat{u}_t. \quad (7.38)$$

Die t -te Beobachtung ist möglicherweise **einflussreich (influential)** und damit ein **Leverage-Punkt**, falls

- h_t groß (nahe 1) ist (bezieht sich auf x -Koordinaten),
- \hat{u}_t groß ist (bezieht sich auf y -Koordinate).

Beweis: Nachweis von (7.38) über mehrmalige Anwendungen der Eigenschaften von Projektionsmatrizen etc. (Details in Davidson & MacKinnon (2004, Section 2.6)):

$$\begin{aligned} \mathbf{y} &= \mathbf{P}_{\mathbf{X}, \mathbf{e}_t} \mathbf{y} + \mathbf{M}_{\mathbf{X}, \mathbf{e}_t} \mathbf{y}, \\ \mathbf{y} &= \mathbf{X} \hat{\boldsymbol{\beta}}^{(t)} + \hat{\alpha} \mathbf{e}_t + \mathbf{M}_{\mathbf{X}, \mathbf{e}_t} \mathbf{y}, \\ \mathbf{P}_\mathbf{X} \mathbf{y} &= \mathbf{X} \hat{\boldsymbol{\beta}}^{(t)} + \hat{\alpha} \mathbf{P}_\mathbf{X} \mathbf{e}_t + \mathbf{0} \\ \mathbf{X} (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{(t)}) &= \hat{\alpha} \mathbf{P}_\mathbf{X} \mathbf{e}_t, \\ \hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{(t)} &= \hat{\alpha} (\mathbf{X}^T \mathbf{X})^{-1} \underbrace{\mathbf{X}^T \mathbf{P}_\mathbf{X} \mathbf{e}_t}_{\mathbf{x}_t^T} = \frac{1}{1 - h_t} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_t^T \hat{u}_t, \end{aligned}$$

wobei mit dem FWL-Theorem $\hat{\alpha} = \frac{\mathbf{e}_t^T \mathbf{M}_\mathbf{X} \mathbf{y}}{\mathbf{e}_t^T \mathbf{M}_\mathbf{X} \mathbf{e}_t} = \frac{\hat{u}_t}{1 - h_t}$. □

•

R-Befehle

In R erhält man die h_t 's und $\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}^{(t)}$, $t = 1, \dots, n$ gegeben durch (7.36) und (7.38) mit `influence(...)`.

Zu lesen: Davidson & MacKinnon (2004, Section 2.6)

Noch mehr zur Geometrie des KQ-Schätzers findet sich z. B. in Ruud (2000).

8. Multiple Regression: Interpretation

8.1. Parameterinterpretation, funktionale Form und Datentransformation

- Die Bezeichnung *linear* im 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:

$$\begin{aligned}y_t &= \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t, \\y_t &= \beta_1 + \beta_2 \ln x_{t2} + u_t, \\ \ln y_t &= \beta_1 + \beta_2 \ln x_{t2} + \beta_3 x_{t3}^2 + u_t, \\ \ln y_t &= \beta_1 + \beta_2 x_t + u_t, \\y_t &= \beta_1 + \beta_2 x_t^2 + u_t.\end{aligned}$$

- Beispiele für Modelle, die *nichtlinear* in den Parametern sind:

$$\begin{aligned}y_t &= \beta_1 + \beta_2 x_{t2}^\gamma + u_t && \text{mit Parametern } \beta_1, \beta_2, \gamma, \\y_t^\gamma &= \beta_1 + \beta_2 \ln x_{t2} + u_t && \text{mit Parametern } \gamma, \beta_1, \beta_2, \\y_t &= \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + \frac{1}{1 + \exp(\lambda(x_{t2} - \pi))} (\delta_1 + \delta_2 x_{t2} + \delta_3 x_{t3}) + u_t \\ &&& \text{mit Parametern } \beta_1, \beta_2, \beta_3, \lambda, \pi, \delta_1, \delta_2, \delta_3.\end{aligned}$$

- Das letzte Beispiel ermöglicht das sanfte Umschalten zwischen zwei linearen Systemen/Regimen. Es sind natürlich nahezu unendlich viele beliebige Formen von Nichtlinearität denkbar. Die Schätzung erfordert z. B. den **nichtlinearen KQ-Schätzer**, der in dem MA-Modul **Fortgeschrittene Ökonometrie** behandelt wird.

Mit linearen Regressionsmodellen können jedoch nichtlineare Zusammenhänge zwischen abhängiger und unabhängiger Variable gut approximiert werden, wenn erstere durch Variablentransformation und/oder Berücksichtigung von Termen mit Potenzen höherer Ordnung eine gute (Taylor)approximation des nichtlinearen Zusammenhangs liefern.

Taylorentwicklung zweiter Ordnung:

$$\begin{aligned}g(x, z) &= g(x_0, z_0) + g_x(x_0, z_0)(x - x_0) + g_z(x_0, z_0)(z - z_0) \\ &+ \frac{1}{2} \left[g_{xx}(x_0, z_0)(x - x_0)^2 + 2g_{xz}(x_0, z_0)(x - x_0)(z - z_0) + g_{zz}(x_0, z_0)(z - z_0)^2 \right] \\ &+ \text{Rest}(x, z, x_0, z_0),\end{aligned}\tag{8.1}$$

mit folgender Notation der partiellen Ableitungen:

$$g_x(x_0, z_0) = \left. \frac{\partial g(x, z)}{\partial x} \right|_{x=x_0, z=z_0},$$

$$g_{xz}(x_0, z_0) = \left. \frac{\partial^2 g(x, z)}{\partial x \partial z} \right|_{x=x_0, z=z_0}.$$

- **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: $\ln(1+z) \approx z$, falls z nahe 0.

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

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

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.

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**

$$\eta(x) = \frac{dy}{dx} \frac{x}{y}. \quad (8.2)$$

- **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.

- **Notation:**

$$\begin{aligned} \Delta E[y|x_1, \dots, x_j, \dots, x_k] &:= E[y|x_1, \dots, x_j + \Delta x_j, \dots, x_k] - E[y|x_1, \dots, x_j, \dots, x_k] \\ &\approx \frac{\partial E[y|x_1, \dots, x_j, \dots, x_k]}{\partial x_j} \Delta x_j \end{aligned}$$

Verschiedene Modelle und Interpretationen von β_j

Für jedes Modell wird im Folgenden angenommen, dass es korrekt spezifiziert ist und der bedingte Erwartungswert der Fehler gegeben die Regressoren Null ist.

- In Folgenden erscheint der Index t nicht, da das Modell der Grundgesamtheit betrachtet wird.
- **Modelle, die linear in den Variablen sind (level-level oder Niveau)**

$$y = \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k + u.$$

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

$$\frac{\partial E[y|x_1, \dots, x_k]}{\partial x_j} = \beta_j$$

und somit näherungsweise

$$\Delta E[y|x_1, \dots, x_k] = \beta_j \Delta x_j.$$

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

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

$$y = \beta_1 \ln x_1 + \dots + \beta_j \ln x_j + \dots + \beta_k \ln x_k + u.$$

Es gilt

$$\frac{\partial E[y|x_1, \dots, x_k]}{\partial x_j} = \beta_j \frac{1}{x_j}$$

und somit näherungsweise

$$\Delta E[y|x_1, \dots, x_k] \approx \beta_j \Delta \ln x_j = \frac{\beta_j}{100} 100 \Delta \ln x_j \approx \frac{\beta_j}{100} \% \Delta x_j.$$

In Worten: Der bedingte Erwartungswert von y verändert sich um $\beta_j/100$ Einheiten, wenn sich x_j um 1% ändert.

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

$$\ln y = \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k + u$$

bzw.

$$y = e^{\ln y} = e^{\beta_1 x_1 + \dots + \beta_k x_k + u} = e^{\beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_k x_k} e^u.$$

Somit gilt

$$E[y|x_1, \dots, x_k] = e^{\beta_1 x_1 + \dots + \beta_k x_k} E[e^u|x_1, \dots, x_k].$$

Ist $E[e^u|x_1, \dots, x_k]$ konstant, gilt

$$\frac{\partial E[y|x_1, \dots, x_k]}{\partial x_j} = \beta_j \underbrace{e^{\beta_1 x_1 + \dots + \beta_k x_k} E[e^u|x_1, \dots, x_k]}_{E[y|x_1, \dots, x_k]} = \beta_j E[y|x_1, \dots, x_k].$$

Man erhält näherungsweise

$$\frac{\Delta E[y|x_1, \dots, x_k]}{E[y|x_1, \dots, x_k]} \approx \beta_j \Delta x_j, \quad \text{bzw.} \quad \% \Delta E[y|x_1, \dots, x_k] \approx 100 \beta_j \Delta x_j$$

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

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

$$\ln y = \beta_1 \ln x_1 + \dots + \beta_2 \ln x_k + u.$$

Ähnlich wie oben lässt sich zeigen, dass gilt:

$$\frac{\partial E[y|x_1, \dots, x_k]}{\partial x_j} = \beta_j \frac{E[y|x_1, \dots, x_k]}{x_j}, \quad \text{und somit} \quad \beta_j = \eta(x_1, \dots, x_k),$$

falls $E[e^u|x_1, \dots, x_k]$ konstant ist.

In diesem Modell entspricht der Steigungsparameter des log-log-Modells gerade der Elastizität für die ursprünglichen Niveauvariablen $E[y|x_1, \dots, x_k]$ und x_j . In Worten: Der bedingte Erwartungswert von y verändert sich um $\beta_j \%$, wenn sich x_j um 1% verändert.

- Die Transformationen von Regressoren können für verschiedene Regressoren unterschiedlich sein.

Beispiel: $y = \beta_1 + \beta_2 \ln x_{t2} + \beta_3 x_{t3}^2 + u$

Handelsströme: (Fortsetzung der empirischen Analyse von Abschnitt 6.3)

R-Code (Ausschnitt aus dem R-Programm in Abschnitt A.4)

```
summary(lm(trade_0_d_o ~ wdi_gdpusdcr_o))           #level - level Modell
summary(lm(trade_0_d_o ~ log(wdi_gdpusdcr_o)))      #level - log Modell
summary(lm(log(trade_0_d_o) ~ wdi_gdpusdcr_o))      #log - level Modell
summary(lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o))) #log - log Modelle
summary(lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o)+log(cepii_dist)))
```

Listing 8.1: ./R_code/4_ff_Beispiel_Handelsstroeme.R

R-Output für log-level Modell

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

Residuals:
    Min       1Q   Median       3Q      Max
-5.6770 -1.4776  0.3231  2.1255  3.4143

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.086e+01  3.776e-01  55.248 < 2e-16 ***
wdi_gdpusdcr_o 5.466e-13  2.010e-13   2.719  0.00915 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.505 on 47 degrees of freedom
(1 observation deleted due to missingness)
Multiple R-squared:  0.1359,    Adjusted R-squared:  0.1175
F-statistic: 7.392 on 1 and 47 DF,  p-value: 0.009148
```

Interpretation: Eine Erhöhung des BIP im Exportland um 1 Mrd. US-Dollar (= 10^9 US-Dollar) führt zu einem durchschnittlichen Anstieg der Importe in Höhe von $100 \hat{\beta}_2 10^9\% = 5.466 \cdot 10^{-13} \cdot 10^{11}\% = 0.055\%$. Entsprechend führt ein Anstieg um 100 Mrd., der ungefähr einem 1%-Anstieg entspricht, um einen durchschnittlichen Anstieg um 5.5%.

R-Output für log-log Modell

```
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
(1 observation deleted due to missingness)
```

Multiple R-squared: 0.7654, Adjusted R-squared: 0.7604
 F-statistic: 153.4 on 1 and 47 DF, p-value: < 2.2e-16

Interpretation: Ein 1%iger Anstieg des BIPs des exportierenden Landes geht mit einem durchschnittlichen Anstieg der Importe um 1,077% einher.

8.2. Datenskalisierung

- **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:

$$\begin{aligned} \underbrace{a\mathbf{y}}_{\mathbf{y}^*} &= \mathbf{X} \underbrace{a\boldsymbol{\beta}}_{\boldsymbol{\beta}^*} + \underbrace{a\mathbf{u}}_{\mathbf{u}^*} \\ \mathbf{y}^* &= \mathbf{X}\boldsymbol{\beta}^* + \mathbf{u}^* \end{aligned} \quad (8.3)$$

- KQ-Schätzer für $\boldsymbol{\beta}^*$ aus (8.3):

$$\begin{aligned} \hat{\boldsymbol{\beta}}^* &= (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}^* \\ &= a (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y} = a\hat{\boldsymbol{\beta}}. \end{aligned}$$

- Fehlervarianz bei homoskedastischen (siehe (9.10)) und unkorrelierten Fehlern:

$$\text{Var}(\mathbf{u}^*|\mathbf{X}) = \text{Var}(a\mathbf{u}|\mathbf{X}) = a^2 \text{Var}(\mathbf{u}|\mathbf{X}) = a^2 \sigma^2 \mathbf{I}.$$

- Varianz-Kovarianz-Matrix:

$$\text{Var}(\hat{\boldsymbol{\beta}}^*|\mathbf{X}) = \sigma^{*2} (\mathbf{X}'\mathbf{X})^{-1} = a^2 \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} = a^2 \text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X})$$

- 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: $\mathbf{X}^* = \mathbf{X}\mathbf{A}$, wobei \mathbf{A} quadratisch und bei Variablenskalisierung diagonal ist. \mathbf{A} muss invertierbar sein, vgl. Abschnitt 7.1.2. Neue Regressionsgleichung:

$$\mathbf{y} = \mathbf{X}\mathbf{A}\mathbf{A}^{-1}\boldsymbol{\beta} + \mathbf{u} = \mathbf{X}^*\boldsymbol{\beta}^* + \mathbf{u}. \quad (8.4)$$

- KQ-Schätzer für β^* aus (8.4):

$$\begin{aligned}\hat{\beta}^* &= (\mathbf{X}^{*T} \mathbf{X}^*)^{-1} \mathbf{X}^{*T} \mathbf{y} = (\mathbf{A}^T \mathbf{X}^T \mathbf{X} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{X}^T \mathbf{y} \\ &= \mathbf{A}^{-1} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y} = \mathbf{A}^{-1} \hat{\beta}.\end{aligned}$$

- 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 (8.1) wurde ein einfaches log-level-Modell für den Einfluss des BIPs auf die Importe geschätzt. Die Parameterschätzung $\hat{\beta}_{BIP} = 5.466 \cdot 10^{-13}$ 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 Milliarden Dollar reskalieren (mittels $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 10^{-9} \end{pmatrix}$) erhalten wir $\hat{\beta}_{BIP}^* = 5.466 \cdot 10^{-4}$.

- **Skalierung von Variablen in logarithmischer Form** verändert lediglich die Konstante β_1 , da $\ln y^* = \ln ay = \ln a + \ln y$.
- **Standardisierte Koeffizienten:** siehe [Wooldridge \(2009, Section 6.1\)](#) oder **Einführung in die Ökonometrie**, Abschnitt 6.2.

8.3. Qualitative Daten als Regressoren

8.3.1. 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 $D = 0$ und $D = 1$ kodiert. Dies ermöglicht eine sehr **einfache Interpretation**:

$$\begin{aligned}y &= \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1} + \delta D + u, \\ E[y|x_1, \dots, x_{k-1}, D = 1] - E[y|x_1, \dots, x_{k-1}, D = 0] &= \\ &= \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1} + \delta \\ &\quad - (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1}) = \delta\end{aligned}$$

Der Koeffizient δ einer Dummyvariablen gibt somit an, um wie viel sich der Achsenabschnitt verschiebt, wenn $D = 1$ statt $D = 0$ vorliegt. *Alle* Steigungsparameter β_i bleiben *unverändert*, wobei $i = 1, \dots, k-1$ (ohne Konstante) bzw. $i = 2, \dots, k-1$ (mit Konstante).

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.

Beispiel: Lohnregression:

- 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)).

Daten:

- *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.

R-Code (Ausschnitt aus R-Programm in Abschnitt A.6)

```
# Festlegung des Arbeitsverzeichnisses (working directory)
# in welchem sich das R-Programm und die Daten befinden

WD          <- getwd() # Bestimme Verzeichnis der R-Datei und
setwd(WD)    # setze es als Working Directory

# Einlesen der Daten
# Die Datendatei "wage1.txt" muss in demselben Verzeichnis wie die
# R-Datei liegen
wage_data   <- read.table("wage1.txt", header = TRUE)
attach(wage_data)

# Lohnregression mit Dummyvariable, siehe Abschnitt 8.4.1
wage_mod_1_kq <- lm(log(wage) ~ female +
                    educ + exper + I(exper^2) + tenure + I(tenure^2))
summary(wage_mod_1_kq)
```

Listing 8.2: ./R_code/8_4_Interpretationen_Wage.R

R-Output

```
Call:
lm(formula = log(wage) ~ female + educ + exper + I(exper^2) +
    tenure + I(tenure^2))

Residuals:
    Min       1Q   Median       3Q      Max
-1.83160 -0.25658 -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

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.).

Zur Interpretation der Parameter von Regressoren, die auch quadratisch eingehen, siehe Abschnitt 8.4.

- **Approximativer Partialeffekt bei log-level-Modellen** Der Parameter δ entspricht einer **approximativen Prognose** der Differenz in y , wenn $\ln y$ modelliert wird und sich die Dummyvariable *ceteris paribus* ändert.

Beispiel: Lohnregression: Der approximative durchschnittliche Einkommensunterschied zwischen weiblichen und männlichen Arbeitnehmern ist 1976 -29.65%.

Wie groß ist der exakte durchschnittliche Einkommensunterschied?

- **Erwartungswert einer log-normalverteilten Zufallsvariable:** Gilt $\ln z \sim N(\mu, \sigma^2)$, dann ist z log-normalverteilt mit

$$E[z] = E[e^{\ln z}] = e^{\mu + \sigma^2/2}. \quad (8.5)$$

Liegt eine **bedingt log-normalverteilte Zufallsvariable**

$$\ln z|x \sim N(g(x), \sigma^2(x))$$

vor, dann gilt

$$E[z|x] = E[e^{\ln z}|x] = e^{g(x) + \sigma^2(x)/2}. \quad (8.6)$$

- **Exakter Partialeffekt bei log-level-Modellen**

$$\ln y = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1} + \delta D + u,$$

Voraussetzung für Berechnung: $u|x_1, \dots, x_{k-1}, D \sim N(0, \sigma^2)$

Dann gilt $E[e^u|x_1, \dots, x_{k-1}, D] = e^{\sigma^2/2}$ und

$$\frac{E[y|x_1, \dots, x_{k-1}, D = 1] - E[y|x_1, \dots, x_{k-1}, D = 0]}{E[y|x_1, \dots, x_{k-1}, D = 0]} = (e^\delta - 1) \quad (8.7)$$

Beweis:

$$\begin{aligned}
 & E[y|x_1, \dots, x_{k-1}, D = 1] - E[y|x_1, \dots, x_{k-1}, D = 0] \\
 &= e^{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1} + \delta} E[e^u|x_1, \dots, x_{k-1}, D = 1] \\
 &- e^{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1}} E[e^u|x_1, \dots, x_{k-1}, D = 0] \\
 &= e^{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1}} e^\delta e^{\sigma^2/2} \\
 &- e^{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_{k-1} x_{k-1}} e^{\sigma^2/2} \\
 &= E[y|x_1, \dots, x_{k-1}, D = 0] (e^\delta - 1).
 \end{aligned}$$

Dividieren der Differenz durch $E[y|x_1, \dots, x_{k-1}, D = 0]$ liefert (8.7). \square

Beispiel: Lohnregression: Frage: Wie groß ist der exakte Lohnunterschied?
Antwort: $100(e^{-0.2965} - 1) = -25.66\%$, wenn man normalverteilte Fehler annimmt.

- **Beachte:** Wenn Δx_j nicht nahe Null ist, sollte immer der exakte Partialeffekt

$$\frac{E[y|x_1, \dots, x_j + \Delta x_j, \dots, x_k] - E[y|x_1, \dots, x_j, \dots, x_k]}{E[y|x_1, \dots, x_j, \dots, x_k]} = (e^{\beta_j \Delta x_j} - 1) \quad (8.8)$$

berechnet werden, da dann die Taylorapproximation die Logarithmusfunktion nicht gut approximiert und damit der Wert des approximativen Partialeffekts wenig verlässlich ist.

- **Wichtig:** Bei Vergleichen zwischen Gruppen ist der Vergleich bedingter Mittelwerte viel aussagekräftiger als der Vergleich unbedingter Mittelwerte.

Beispiel: Lohnregression: Vergleich der Löhne von Männern und Frauen: Nimmt man normalverteilte Fehler an, ist der exakte Partialeffekt -25.66%. Frauen verdienen durchschnittlich nach Berücksichtigung der Ausbildung, Berufserfahrung und Zeit in einem Unternehmen um ca. 26% weniger als Männer.

Vergleicht man hingegen die unbedingten Mittelwerte, z. B. mit dem

R-Code (Ausschnitt aus R-Programm in Abschnitt A.6)

```

# Relative Differenz der unbedingten Lohnmittelwerte von Frauen und Männern
(mean(wage[female==1]) - mean(wage[female==0])) / mean(wage[female==0])
# alternative Berechnungsmöglichkeit
wage_mean <- lm(wage~0+female+I(1-female))
(wage_mean$coef[1] - wage_mean$coef[2]) / wage_mean$coef[2]

```

Listing 8.3: ./R_code/8_4_Interpretationen_Wage.R

dann beträgt der Unterschied 35.38%, ist also wesentlich größer, weil sich Männer und Frauen offensichtlich auch in der Ausbildung, Berufserfahrung und Zeit in einem Unternehmen unterscheiden.

Es ist also essentiell, relevante Einflussfaktoren zu berücksichtigen!

- **Exakte und approximative Prognose bei log-level-Modell:** Erwartungswert von y gegeben die Regressoren x_1, \dots, x_k ist gegeben durch

$$E[y|x_1, \dots, x_k] = e^{\beta_1 x_1 + \dots + \beta_k x_k} \cdot E[e^u|x_1, \dots, x_k].$$

Der wahre Wert von $E[e^u|x_1, \dots, x_k]$ hängt von der Wahrscheinlichkeitsverteilung von u ab.

Gilt $u|x_1, \dots, x_k \sim N(0, \sigma^2)$, dann ist $E[e^u|x_1, \dots, x_k] = e^{E[u|x_1, \dots, x_k] + \sigma^2/2}$. Die **exakte Prognose** ist somit

$$E[y|x_1, \dots, x_k] = e^{\beta_1 x_1 + \dots + \beta_k x_k + \sigma^2/2}.$$

Beispiel: Lohnregression — exakte Prognose: Wie viel verdient ein Frau mit 12 Jahren Ausbildung, 10 Jahren Erfahrung und einem Jahr Beschäftigungsdauer? Die exakte Vorhersage des Stundenlohns ist

$$\begin{aligned} E[\text{wage} | \text{female} = 1, \text{educ} = 12, \text{exper} = 10, \text{tenure} = 1] \\ &= \exp(0.4167 - 0.2965 \cdot 1 + 0.0802 \cdot 12 + 0.02943 \cdot 10 \\ &\quad - 0.0006 \cdot (10^2) + 0.0317 \cdot 1 - 0.0006 \cdot (1^2) + 0.3998^2/2) \\ &= 4.18, \end{aligned}$$

wobei σ^2 mit s^2 (9.25) geschätzt wird. Der genaue Wert des mittleren Stundenlohns der beschriebenen Person ist also etwa 4.18\$.

Lässt man den Term $e^{s^2/2}$ bei der Prognose weg, dann erhält man eine **approximative Prognose**.

Beispiel: Lohnregression — approximative Prognose:

$$\begin{aligned} E[\ln(\text{wage}) | \text{female} = 1, \text{educ} = 12, \text{exper} = 10, \text{tenure} = 1] \\ &= 0.4167 - 0.2965 \cdot 1 + 0.0802 \cdot 12 + 0.0294 \cdot 10 \\ &\quad - 0.0006 \cdot (10^2) + 0.0317 \cdot 1 - 0.0006 \cdot (1^2) \\ &= 1.35 \end{aligned}$$

Demnach ist der erwartete Stundenlohn **approximativ** $\exp(1.35) = 3.86$ US-\$ und damit 30 Cent kleiner als der exakte Wert.

Fazit: bei **log-log** bzw. **log-level-Modellen**:

- für exakte Prognosen benötigt man die empirische Varianz (bei Normalverteilungsannahme),
- bei approximativen Prognosen reicht schon “Einsetzen” in die Regressionsgleichung.

8.3.2. Mehrere Untergruppen

Illustration anhand eines Beispiels:

Beispiel: Lohnregression: (Fortsetzung von Abschnitt 8.3.1)

Ein Arbeitnehmer ist weiblich oder männlich und verheiratet oder ledig \implies 4 Untergruppen

1. weiblich und ledig
2. weiblich und verheiratet
3. männlich und ledig
4. männlich und verheiratet

Vorgehen

- Wir bestimmen eine Gruppe als **Referenzgruppe**, etwa: weiblich und ledig
- Definiere Dummyvariablen für die anderen Untergruppen.

R-Code (Ausschnitt aus R-Programm in Abschnitt A.6)

```
femmarr <- female * married
malesing <- (1 - female) * (1 - married)
malemarr <- (1 - female) * married

wage_mod_2_kq <- lm(log(wage) ~ femmarr + malesing + malemarr +
                    educ + exper + I(exper^2) + tenure + I(tenure^2))
summary(wage_mod_2_kq)
```

Listing 8.4: ./R_code/8_4_Interpretationen_Wage.R

R-Output

```
Call:
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:

- Ceteris paribus verdienen verheiratete Frauen im Durchschnitt approximativ um 8.8% weniger als ledige Frauen. Dieser Effekt ist jedoch nur auf dem 10% Niveau signifikant (bei einem zweiseitigen Test).
- Der erwartete Lohnunterschied zwischen verheirateten Männern und Frauen

beträgt *ceteris paribus* ungefähr $32.3 - (-8.8) = 41.1\%$. Hierfür kann keine *t*-Statistik direkt berechnet werden, jedoch eine *F*-Statistik. (Um einen *t*-Test durchführen zu können, 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 hat \mathbf{X} reduzierten Spaltenrang. Warum?

Ordinale Daten in der Regression

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 Dummyvariable 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 Dummies 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..

8.3.3. 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 Lohnleichung

$$y = \beta_1 + \delta_1 \text{male} + \delta_2 \text{married} + \delta_3 \text{male} \cdot \text{married} + \dots$$

hinzu. Ein Vergleich zwischen verheirateten und ledigen Männern ist somit gegeben durch

$$\begin{aligned} E[y|\text{male} = 1, \text{married} = 1] - E[y|\text{male} = 1, \text{married} = 0] \\ = \beta_1 + \delta_1 + \delta_2 + \delta_3 + \dots - (\beta_1 + \delta_1 + \dots) = \delta_2 + \delta_3. \end{aligned}$$

- **Interaktionen zwischen Dummies und quantitativen Variablen:**

- Dies ermöglicht für unterschiedliche Gruppen unterschiedliche Steigungsparameter

$$y = \beta_1 + \beta_2 D + \beta_3 x + \beta_4 (x \cdot D) + u.$$

Beachte: β_2 bezeichnet hier *lediglich* für den Fall $x = 0$ die Unterschiede zwischen beiden Gruppen. Falls $x \neq 0$, ist diese Differenz

$$\begin{aligned} & E[y|D = 1, x] - E[y|D = 0, x] \\ &= \beta_1 + \beta_2 \cdot 1 + \beta_3 x + \beta_4 (x \cdot 1) - (\beta_1 + \beta_3 x) \\ &= \beta_2 + \beta_4 x. \end{aligned}$$

Selbst, wenn β_2 negativ ist, ist der Gesamteffekt möglicherweise positiv!

Beispiel: Lohnregression (Fortsetzung von Abschnitt 8.3.1)

Interaktion des Dummies mit dem Regressor educ:

R-Code (Ausschnitt aus R-Programm in Abschnitt A.6)

```
wage_mod_3_kq <- lm(log(wage) ~ female +
                    educ + exper + I(exper^2) + tenure + I(tenure^2) +
                    I(female*educ))
summary(wage_mod_3_kq)
```

Listing 8.5: ./R_code/8_4_Interpretationen_Wage.R

R-Output

```
Call:
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
```

Die *Rendite der Ausbildung*, also die durchschnittliche Stundenlohndifferenz für ein zusätzliches Ausbildungsjahr, ist nicht geschlechtsspezifisch, da der p -Wert des entsprechenden Interaktionsterms über jedem gebräuchlichen Signifikanzniveau liegt.

- **Fazit:** Kommt eine Regressionsvariable in mehreren Termen (Interaktionen, quadratische Terme) im Modell vor, sind im Allgemeinen mehr Parameter zur Interpretation eines Zusammenhangs zu berücksichtigen.
- **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:

$$y = \beta_1 + \beta_2 D + \beta_3 x_1 + \beta_4 (x_1 \cdot D) + \beta_5 x_2 + \beta_6 (x_2 \cdot D) + u. \quad (8.9)$$

Hypothesenpaar:

$$H_0 : \beta_2 = \beta_4 = \beta_6 = 0 \quad \text{vs.}$$

$$H_1 : \beta_2 \neq 0 \text{ und/oder } \beta_4 \neq 0 \text{ und/oder } \beta_6 \neq 0.$$

8.4. Modelle mit quadratischen Regressoren

- **Modelle mit quadratischen Regressoren:**
 - Als Beispiel sei folgendes multiples Regressionsmodell

$$y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_3^2 + u$$

angenommen. Der marginale Einfluss einer Veränderung von x_3 auf den bedingten Erwartungswert y ist

$$\frac{\partial E[y|x_1, \dots, x_3]}{\partial x_3} = \beta_3 + 2\beta_4 x_3.$$

Somit beeinflusst eine Veränderung von x_3 um Δx_3 ceteris paribus die unabhängige Variable y im Durchschnitt um

$$\Delta E[y|x_1, \dots, x_3] = (\beta_3 + 2\beta_4 x_3) \Delta x_3.$$

Der Einfluss hängt also offensichtlich vom Niveau von x_3 ab (und somit ist eine Interpretation von β_3 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

$$\ln y = \beta_1 + \beta_2 x_2 + \beta_3 \ln x_3 + \beta_4 (\ln x_3)^2 + u.$$

Die Elastizität von y bezüglich x_3 ist demnach

$$\beta_3 + 2\beta_4 \ln x_3$$

und ist dann und nur dann konstant, wenn $\beta_4 = 0$.

- **Beispiel: Handelsströme:** 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: (Die Modelle 2 und 3a wurden in Abschnitt 6.3 geschätzt. Die Modelle 1, 3b und 4 werden erst in Abschnitt 10.3 eingeführt.)

$$\begin{aligned} \ln(\text{Importe}) = & \beta_1 + \beta_2 \ln(\text{BIP}) + \beta_3 (\ln(\text{BIP}))^2 + \beta_4 \ln(\text{Entfernung}) \\ & + \beta_5 \text{Offenheit} + \beta_6 \ln \text{Flaeche} + u. \end{aligned}$$

Eben wurde gezeigt, dass dann für die Elastizität der *Importe* bezüglich *BIP* gilt:

$$\beta_2 + 2\beta_3 \ln(\text{BIP}). \quad (8.10)$$

Die Schätzung des Modells 5 mit folgendem

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# Modell 5: Verwende auch log(BIP)^2 als Regressor
mod_5_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)

mod_5_kq <- lm(mod_5_formula)
summary(mod_5_kq)
```

Listing 8.6: ./R_code/4_ff_Beispiel_Handelsstroeme.R

R-Output

```
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
(1 observation deleted due to missingness)
Multiple R-squared: 0.9155, Adjusted R-squared: 0.9056
F-statistic: 93.12 on 5 and 43 DF, p-value: < 2.2e-16
```

Wer bereits mit Signifikanztests vertraut ist, sieht, dass der quadratische Term auf dem 5% Niveau signifikant ist. Damit ist der statistische Nachweis für eine nicht-lineare Elastizität erbracht. Setzt man die Parameterschätzungen in (8.10) ein, erhält man

$$\eta(BIP) = 3.908811 - 0.057108 \ln(BIP).$$

In Abbildung 8.1 ist die Elastizität von $\eta(BIP)$ für jeden beobachteten Wert von BIP gegen BIP dargestellt (mit folgendem R-Code erzeugt).

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# Generiere Plot der Elastizitäten für verschiedene BIPs
elast_gdp <- mod_5_kq$coef[2] + 2* mod_5_kq$coef[3]*log(wdi_gdpusdcr_o)
# Erstelle Scatterplot
if (save.pdf) pdf("plot_modell15_elast.pdf.pdf", height=6, width=6)
```

Listing 8.7: ./R_code/4_ff_Beispiel_Handelsstroeme.R

Die BIP-Elastizität der Importe ist für kleine Volkswirtschaften (gemessen am BIP) viel größer als für große Volkswirtschaften. Mit anderen Worten: Bei kleinen Volkswirtschaften wirkt sich eine Steigerung des BIPs viel stärker auf die Importe aus, als bei großen 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:

$$y = \beta_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_3 x_2 + u.$$

Der marginale Effekt einer Veränderung von x_3 ist gegeben durch

$$\Delta E[y|x_2, x_3] = (\beta_3 + \beta_4 x_2) \Delta x_3.$$

Somit hängt der marginale Effekt auch vom Niveau von x_2 ab!

Zu Lesen: Kapitel 6 (ohne Abschnitt 6.4) und Kapitel 7 (ohne die Abschnitte 7.5 und 7.6) in Wooldridge (2009).

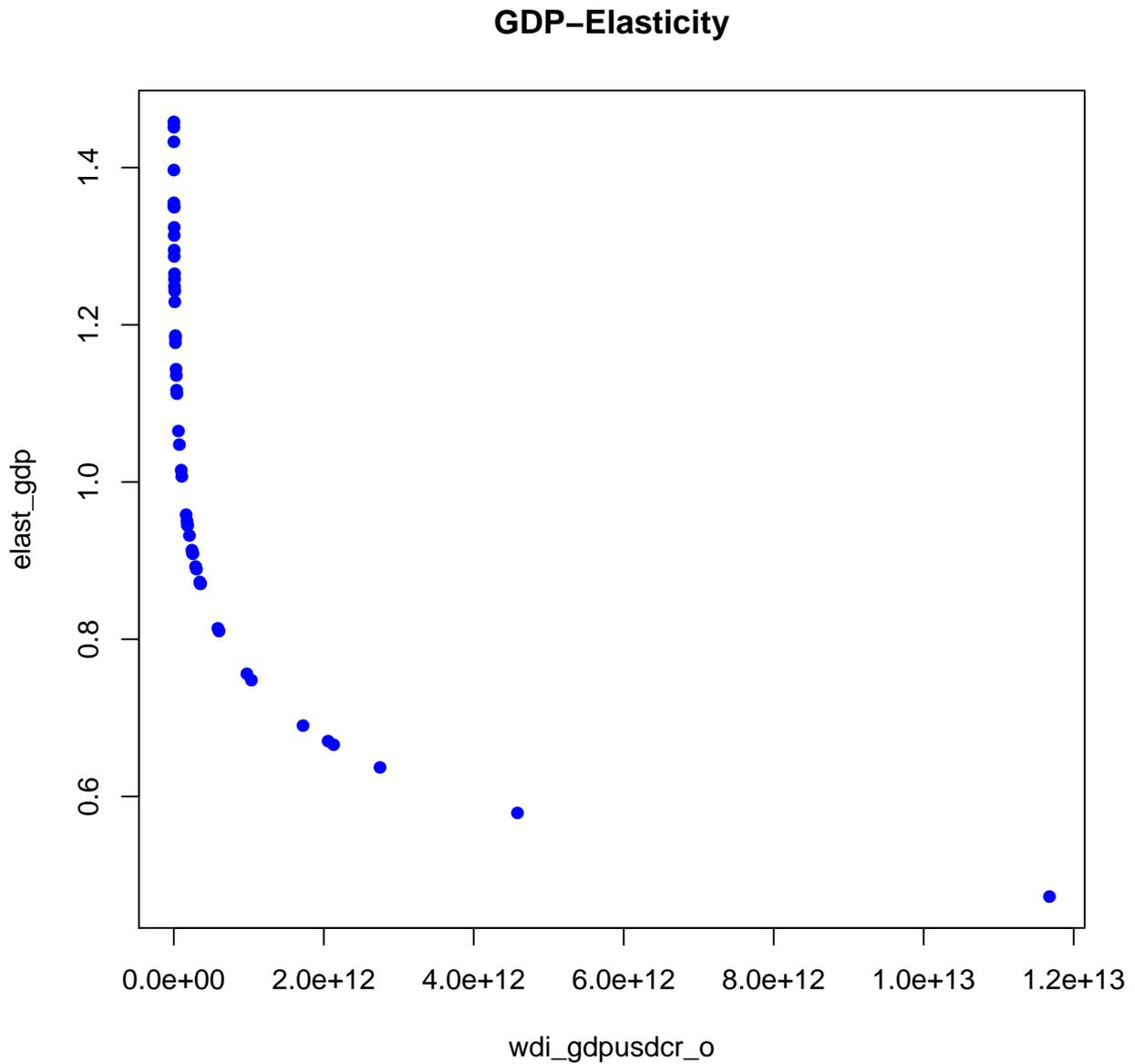


Abbildung 8.1.: Elastizität von $\eta(BIP)$ (R-Code siehe Beispiel zu Handelsströmen)

9. Statistische Eigenschaften des KQ-Schätzers: Erwartungswert und Kovarianz

- Die algebraischen und geometrischen Eigenschaften des **KQ-Schätzers**

$$\hat{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (6.5)$$

für das **multiple lineare Regressionsmodell**

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \cdots + \beta_k x_{tk} + u_t, \quad t = 1, \dots, n, \quad (5.24)$$

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t, \quad t = 1, \dots, n, \quad (5.25)$$

$$\mathbf{y} = \mathbf{X} \boldsymbol{\beta} + \mathbf{u}. \quad (6.1)$$

wurden in Kapitel 7 analysiert.

- Zur Beantwortung vieler Fragen ist die Kenntnis der algebraischen und geometrischen Eigenschaften des KQ-Schätzers nicht ausreichend, sondern die **Kenntnis der statistischen Eigenschaften des KQ-Schätzers notwendig**.

Beispiele:

- Beispiel auf Basis von Handelsströmen
- Was lässt sich auf Basis des geschätzten Parametervektors $\hat{\beta}$ über die Werte des Parametervektors β für den DGP (für die Grundgesamtheit) aussagen?
- Inwieweit lässt sich überprüfen, ob das gewählte ökonometrische Modell den DGP enthält?
- Angenommen, Ihnen liegt eine weitere Stichprobe mit k Regressoren zu derselben Fragestellung vor.
 - * Warum unterscheiden sich die beiden KQ-Schätzungen vermutlich?
 - * Welche der beiden KQ-Schätzungen wählen Sie?
 - * Sollen Sie die KQ-Ergebnisse beider Stichproben zusammenfügen?

Sollen auf Basis der Stichprobe Aussagen über den DGP getroffen werden, sind induktive statistische Methoden notwendig. Um etwas über die Eigenschaften solcher Aussagen sagen zu können, sind Annahmen über den DGP und das ökonometrische Modell notwendig.

Welche Annahmen zu welchen (statistischen) Eigenschaften des KQ-Schätzers führen, ist Gegenstand dieses Kapitels.

Würde die Grundgesamtheit der Stichprobe entsprechen und wären wir nur an Kennzahlen wie Stichprobenkorrelation oder Bestimmtheitsmaß interessiert, wären wir bereits fertig.

- **Wichtige Eigenschaften eines Schätzers:**
 - Erwartungstreue, Unverzerrtheit
 - Varianz
 - Mittlerer quadratischer Fehler (mean squared error) (MSE)
 - Konsistenz
 - Effizienz
 - exakte Verteilung in endlichen Stichproben
 - asymptotische Verteilung

Analyse der Eigenschaften des KQ-Schätzers Übersicht über Abschnitte

	β	σ^2	KQ-Schätzer für Kovarianzmatrix
Erwartungstreue, Unverzerrtheit	9.1.1	9.5	9.3
Varianz	9.3		
MSE	9.6		
Konsistenz	9.2		
Effizienz	9.4		
exakte Verteilung in endlichen Stichproben	11.1		
asymptotische Verteilung	11.2		

9.1. Erwartungstreue des KQ-Schätzers

Wiederholung von Abschnitt 5.4:

- Die **Verzerrung** (bias) eines Parameterschätzers $\hat{\theta}$ für θ ist definiert als

$$E[\hat{\theta}] - \theta_0,$$

wobei θ_0 der wahre Parameterwert, d. h. der Parameterwert des DGPs ist (vgl. (5.51)).

- Ein Schätzer $\hat{\theta}$ heißt **erwartungstreu** oder **unverzerrt**, wenn er für alle zulässigen Werte von θ_0 keine Verzerrung aufweist.
- **Interpretation:** Erwartungstreue impliziert, dass bei einer großen Anzahl an Stichproben der Durchschnittswert aller Schätzungen sehr nahe am wahren Wert liegt.

- Sind zwei Schätzer in allen Eigenschaften gleich bis auf die Unverzerrtheit, ist der unverzerrte Schätzer vorzuziehen. Warum?

9.1.1. Bedingungen für die Erwartungstreue des KQ-Schätzers

Ableitung: Es gilt, sofern \mathbf{X} vollen Rang hat und das multiple lineare Regressionsmodell korrekt spezifiziert ist,

$$\begin{aligned}\hat{\boldsymbol{\beta}} &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ &= \boldsymbol{\beta}_0 + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}\end{aligned}$$

und so

$$E[\hat{\boldsymbol{\beta}}] - \boldsymbol{\beta}_0 = E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}].$$

Unverzerrtheit des KQ-Schätzers gilt, wenn mindestens eine der folgenden Annahmen bezüglich der **Regressoren** und Fehler erfüllt ist:

- alle Regressoren sind **nicht-stochastisch** und $E[\mathbf{u}] = \mathbf{0}$:

$$E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\mathbf{u}] = \mathbf{0}.$$

- Regressoren \mathbf{X} sind stochastisch, aber **stochastisch unabhängig** von dem Fehlervektor \mathbf{u} mit $E[\mathbf{u}] = \mathbf{0}$. Dann gilt

$$E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}] = E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T] E[\mathbf{u}] = \mathbf{0}.$$

- Eine schwächere Annahme als stochastische Unabhängigkeit ist

$$E[\mathbf{u}|\mathbf{X}] = \mathbf{0}. \quad (9.1a)$$

Damit gilt

$$E[(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} | \mathbf{X}] = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\mathbf{u} | \mathbf{X}] = \mathbf{0}.$$

Alternativ lässt sich Annahme (9.1a) schreiben als

$$E[u_t | \mathbf{X}] = E[u_t | \mathbf{X}_1, \dots, \mathbf{X}_t, \dots, \mathbf{X}_n] = 0, \quad t = 1, \dots, n \quad (9.1b)$$

Erklärende Variablen, die (9.1) erfüllen, werden als **exogen** bezeichnet. Sehr häufig werden Variablen, die Annahme (9.1) erfüllen, als **streng exogen (strictly exogenous)** bezeichnet (z. B. Wooldridge (2009, Chapter 10)), siehe auch BA-Modul **Zeitreihenökonomie, Kapitel 2**.

- Beachte: Aus (9.1) folgt durch Anwendung des iterierten Erwartungswertes

$$E[E[u_t | \mathbf{X}_1, \dots, \mathbf{X}_t, \dots, \mathbf{X}_n] | x_{sj}] = E[u_t | x_{sj}] = 0 \implies \text{Cov}(u_t, x_{sj}) = 0$$

für alle $s = 1, \dots, n$ und alle $j = 1, \dots, k$. (9.2)

Strenge Exogenität impliziert also, dass der Fehler u_t mit vergangenen, gegenwärtigen oder zukünftigen Regressoren unkorreliert ist.

- Beachte: Die Annahme (9.1) ist ohne Spezifikation eines Modells für die Fehler \mathbf{u} , wie beispielsweise $\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, ohne Aussage und gewinnt erst durch einen Bezug auf ein (parametrisches) Modell Bedeutung. Somit bezieht die Bedingung (strenger) Exogenität implizit immer ein (parametrisches) Modell mit ein.

Beispiel: Für das einfache (normale) lineare Regressionsmodell, das sich aus (5.32) ergibt, ist (9.1) erfüllt, da für das Paar $\beta_1, \beta_2 \in \mathbb{R}$ des DGP gilt:

$$E[\ln(\text{Importe}_t) | BIP_1, BIP_2, \dots, BIP_n] = \beta_1 + \beta_2 BIP_t.$$

- **Zusammenfassung der Annahmen bzw. Voraussetzungen für die Unverzerrtheit des KQ-Schätzers $\hat{\boldsymbol{\beta}}$ für den Parametervektor $\boldsymbol{\beta}$:**

- **(B1) Korrekt spezifiziertes Modell** Der DGP ist für $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ im multiplen linearen Regressionsmodell (6.1)

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

enthalten (MLR.1 in Wooldridge (2009)).

- **(B2a) Exogenität bzw. Strenge Exogenität (9.1):** (folgt aus MLR.2 und MLR.4 in Wooldridge (2009)).

$$E[\mathbf{u} | \mathbf{X}] = \mathbf{0}.$$

- Annahme **(B2b)** wird erst später benötigt.

- **(B3) Keine perfekte Kollinearität \mathbf{X}** (bzw. $\mathbf{X}^T \mathbf{X}$) hat vollen Rang (MLR.3 in Wooldridge (2009)).

- Erwartungstreue kann mit Monte-Carlo-Simulation „überprüft“ werden

Beispiel: Generieren von 1000 Stichproben mit $n = 50$ und Schätzen eines korrekt spezifizierten einfachen linearen Regressionsmodell. Der DGP lautet

$$y_t = 1 + 0.9x_t + u_t, \quad u_t \sim NID(0, 4), \quad t = 1, 2, \dots, n. \quad (9.3)$$

Siehe Abschnitt 2.9.1 zu Definition von *NID*. Die x_t werden aus der Menge $1, 2, \dots, 20$ mit Zurücklegen zufällig gezogen. Mit dem **R-Programm**, siehe **Abschnitt A.7, Seite 350** ergeben sich über die 1000 Replikationen für β_1 der Mittelwert 0.9973185 und für β_2 der Mittelwert 0.9004453. D. h. die Mittelwerte als Schätzer des Erwartungswertes liegen sehr nahe an den wahren Werten. Die Histogramme für $\hat{\beta}_1$ und $\hat{\beta}_2$ in Abbildung 9.1 zeigen, dass die KQ-Schätzungen um die wahren Parameter herum streuen.

Histogram of beta_hat_store[, 1] Histogram of beta_hat_store[, 2]

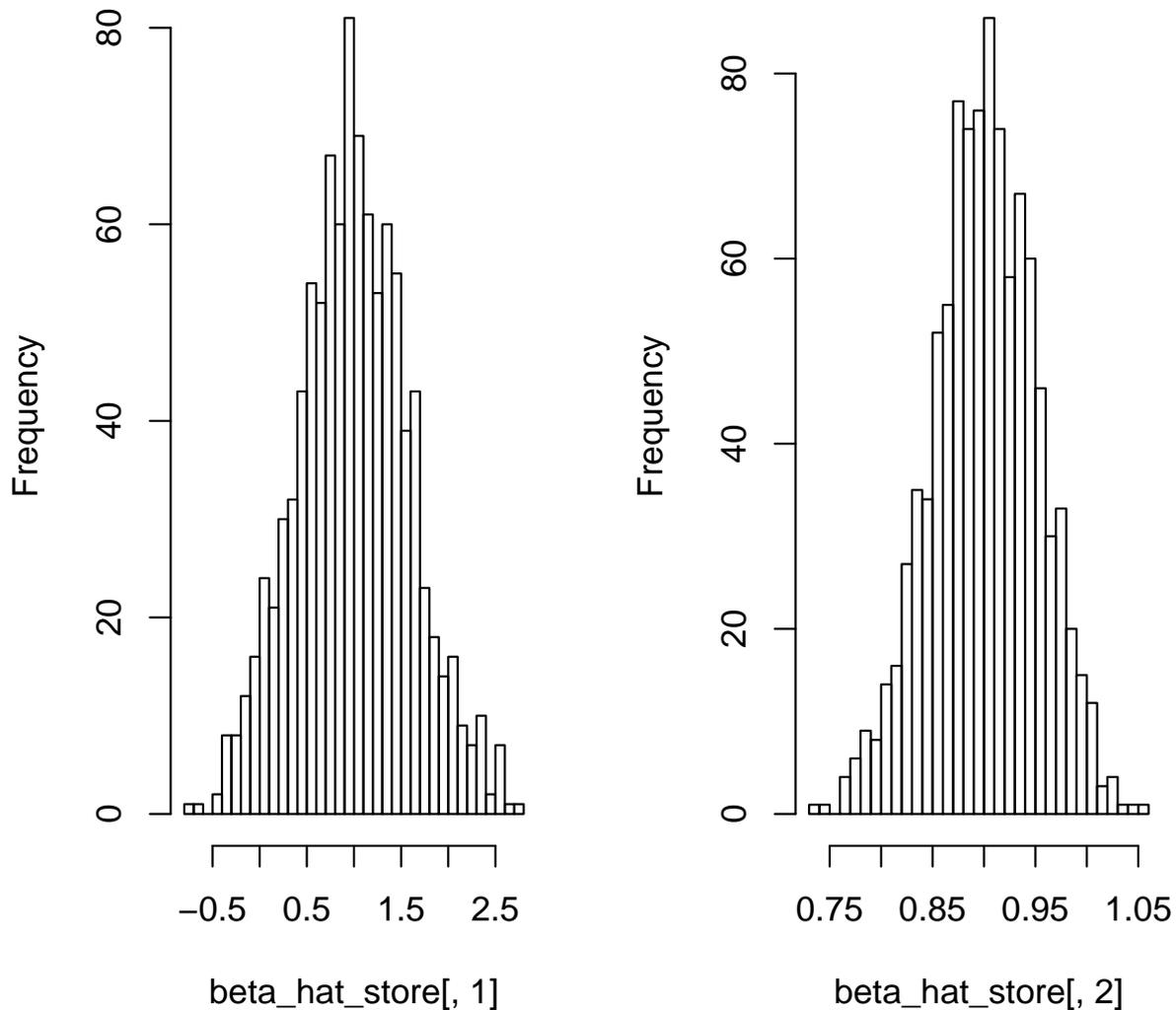


Abbildung 9.1.: Histogramme von KQ-Schätzungen für β_1 und β_1 auf Basis von 1000 Replikationen (**R-Programm**, siehe Abschnitt A.7, Seite 350)

9.1.2. Vorherbestimmte Regressoren

- Eine schwächere Annahme als strenge Exogenität (9.1) ist

$$E[u_t | \mathbf{X}_t] = 0 \quad \text{für } t = 1, \dots, n, \quad (9.4)$$

weil der Fehler u_t lediglich *nicht* von den Regressoren \mathbf{X}_t der t -ten Stichprobenbeobachtung abhängen darf.

- Regressoren \mathbf{X}_t , die die Bedingung (9.4) erfüllen, werden als **vorherbestimmt** bezüglich

des Fehlerterms u_t bezeichnet.

- Bei Regressionsmodellen für Zeitreihendaten werden die Fehler u_t auch als **Innovationen** oder **Schocks** bezeichnet.
- **Wooldridge (2009, Chapter 10)** bezeichnet die Annahme (9.4) auch als **contemporaneous exogeneity**,
- Strenge Exogenität (9.1) folgt aus der Voraussetzung vorherbestimmter Regressoren (9.4) (entspricht **Wooldridge 2009, MLR.4**) und der Annahme einer Zufallsstichprobe (**Wooldridge 2009, MLR.2**), da dann

$$E[u_t | \mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_t, \dots, \mathbf{X}_n] = E[u_t | \mathbf{X}_t].$$

- **Modelle, deren Regressoren die Annahme strenger Exogenität verletzen, aber vorherbestimmt bezüglich u_t sind:**
 - **autoregressive Modelle**, siehe Abschnitt 12.3.1.
 - **dynamische lineare Regressionsmodelle**, siehe Abschnitt 13.4.

Beide Modelle enthalten verzögert abhängige Variablen als Regressoren.

- Ist die Annahme strenger Exogenität (9.1) verletzt, ist der KQ-Schätzer **verzerrt**. Um einen unverzerrten Schätzer zu erhalten, reicht es nicht aus, wenn Regressoren vorherbestimmt (9.4) sind.
- Wiederhole **Zusammenhang zwischen bedingtem Erwartungswert und Kovarianz** (2.29b), (2.29c), (2.29f)

9.2. Konsistenz des KQ-Schätzers

- Siehe Abschnitt 5.4 zur Definition und Bedeutung der Konsistenz eines Schätzers.
- **Konsistenz des KQ-Schätzers:** Es gelten zusätzlich zu (B1) die **Annahmen:**
 - (A1) Es gilt ein LLN für $\mathbf{X}^T \mathbf{X} / n$

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \left(\frac{\mathbf{X}^T \mathbf{X}}{n} \right) &= \text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E \left[\mathbf{X}_t^T \mathbf{X}_t \right] = \mathbf{S}_{\mathbf{X}^T \mathbf{X}} \end{aligned}$$

und $\mathbf{S}_{\mathbf{X}^T \mathbf{X}}$ hat vollen Rang.

(entspricht **Davidson & MacKinnon 2004**, Gleichungen (3.17) bzw. (4.49))

- (A2) Es gilt ein LLN für $\mathbf{X}^T \mathbf{u}/n$

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T u_t = \mathbf{0}.$$

Dann ist $\text{plim}_{n \rightarrow \infty} \hat{\beta}_n = \beta_0$ und der KQ-Schätzer ist **konsistent**.

- **Typische Vorgehensweise zur theoretischen Ableitung von Konsistenzbedingungen** am Beispiel des KQ-Schätzers:

$$\begin{aligned} \hat{\beta}_n &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \\ &= \beta_0 + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \\ &= \beta_0 + \underbrace{\left(\frac{\mathbf{X}^T \mathbf{X}}{n} \right)^{-1}}_{:= \mathbf{A}_n} \underbrace{\frac{\mathbf{X}^T \mathbf{u}}{n}}_{:= \mathbf{a}_n}. \end{aligned}$$

Anwenden der Rechenregeln für plim's (3.1) in Abschnitt 3.4 ergibt unter der Annahme (B1) eines korrekt spezifizierten Modells

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \hat{\beta}_n &= \beta_0 + \text{plim}_{n \rightarrow \infty} \left(\frac{\mathbf{X}^T \mathbf{X}}{n} \right)^{-1} \text{plim}_{n \rightarrow \infty} \frac{\mathbf{X}^T \mathbf{u}}{n} \\ &= \beta_0 + \left(\underbrace{\text{plim}_{n \rightarrow \infty} \left(\frac{\mathbf{X}^T \mathbf{X}}{n} \right)}_{\text{existiert und nichtsingulär wegen (A1)}} \right)^{-1} \\ &\quad \underbrace{\text{plim}_{n \rightarrow \infty} \frac{\mathbf{X}^T \mathbf{u}}{n}}_{= \mathbf{0}, \text{ da wegen (A2) ein LLN gilt}} \\ &= \beta_0 \end{aligned}$$

• Diskussion der Annahmen

- Einfachster Fall für Gültigkeit der Annahmen (A1) und (A2): $\mathbf{X} = \mathbf{1}$, eine Konstante ist einziger Regressor, und $u_t \sim IID(0, \sigma^2)$. Dann gilt das WLLN von Chintschin (siehe Abschnitt 5.5.1), so dass (A2) gilt. (A1) ist leicht zu zeigen.

Beispiel: arithmetisches Mittel bei IID-Zufallsvariablen

DGP: $y_t = \mu_0 + u_t$, $u_t \sim IID(0, \sigma_0^2)$. Dann ist $\mathbf{S}_{\mathbf{X}^T \mathbf{X}} = n/n = 1$.

- Liegt eine **Zufallsstichprobe** vor und gilt (9.4), dann gilt Annahme (B2a) und es sind (A1) und (A2) erfüllt.

Beweis: Da die Stichprobenelemente unabhängig und identisch verteilt sind, gilt $E[\mathbf{X}_t^T \mathbf{X}_t] = \mathbf{M}$, $t = 1, 2, \dots, n$, so dass automatisch $\mathbf{M} = \mathbf{S}_{\mathbf{X}^T \mathbf{X}}$ und damit (A1) folgt. Weiter gilt wegen (9.4) und dem Gesetz der iterierten Erwartungen

$E[\mathbf{X}_t u_t] = \mathbf{0}$. Aufgrund der Zufallsstichprobe lässt sich für jedes Vektorelement $z_t = \mathbf{X}_{tj} u_t$ Chintschin's schwaches Gesetz der großen Zahlen anwenden, woraus (A2) folgt. \square

- Auch wenn keine Zufallsstichprobe vorliegt, z. B. weil eine Stichprobe mit Zeitreihendaten vorliegt, existieren Annahmen, die sich leichter als (A2) überprüfen lassen. Diese finden sich in Abschnitt 13.4.
- Es gibt einfache Fälle, für die (A1) verletzt ist.

Beispiel: $x_t = t$.

- In Abschnitt 13.4 wird auch deutlich, dass Annahme (A2) schwächer als die Annahme (B2a) ist.
- Die Annahme (B3) ist nicht genannt, weil erlaubt ist, dass diese für einzelne Realisationen von Stichproben verletzt sein kann. Lediglich im Limit wird gefordert, dass keine lineare Abhängigkeit zwischen den Regressoren vorliegt, da $\mathbf{S}_{\mathbf{X}^T \mathbf{X}}$ in (A1) vollen Rang haben muss.

Beispiel: Monte-Carlo-Simulation zu Schätzeigenschaften des KQ-Schätzers bei einer Zufallsstichprobe

- DGP (wie in der Monte-Carlo-Simulation im vorherigen Abschnitt):

$$y_t = 1 + 0.9x_t + u_t, \quad u_t \sim NID(0, 4), \quad t = 1, 2, \dots, n. \quad (9.3)$$

Siehe Abschnitt 2.9.1 zu Definition von *NID*. Die x_t werden aus der Menge $1, 2, \dots, 20$ mit Zurücklegen zufällig gezogen.

- Stichprobengrößen: $n = 50, 100, 500, 1000, 10000, 100000$.
- $R = 10000$ Replikationen

R-Code, siehe Abschnitt A.8, Seite 352

R-Output

	N	beta_1_hat_mean	beta_1_hat_sd	beta_2_hat_mean	beta_2_hat_sd
[1,]	5e+01	0.9939493	0.59113380	0.9008026	0.049219333
[2,]	1e+02	0.9979973	0.41867138	0.9005215	0.035010494
[3,]	5e+02	0.9979537	0.18655091	0.9001597	0.015467546
[4,]	1e+03	0.9983807	0.13101124	0.9001677	0.010893364
[5,]	1e+04	0.9996438	0.04134015	0.9000331	0.003431829
[6,]	1e+05	1.0001878	0.01323944	0.8999901	0.001098157

Man sieht deutlich die Erwartungstreue des KQ-Schätzers und den Rückgang der Standardabweichung des KQ-Schätzers mit zunehmender Stichprobengröße. Die Histogramme für in den Abbildungen 9.2 und 9.3 für die Parameterschätzer und Stichprobengrößen $n = 500, 100, 500, 1000$ deuten auf die Gültigkeit des zentralen Grenzwertsatzes hin. Dazu mehr in Abschnitt 11.2. Histogramme für $n = 10000, 100000$ werden mit dem R-Code erzeugt, aber nicht hier abgebildet.

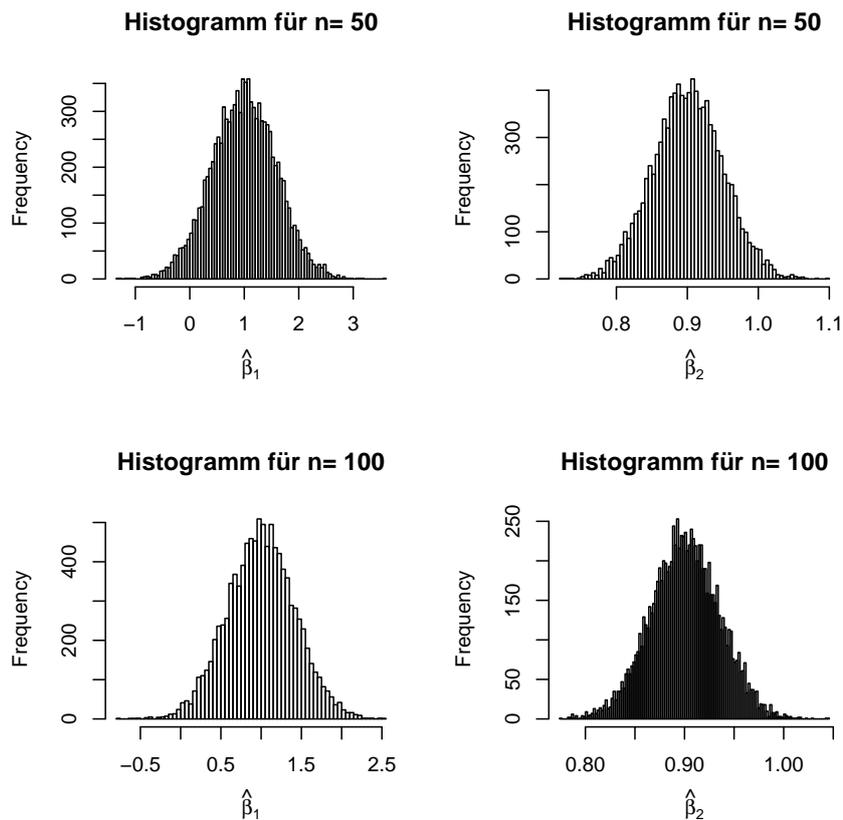


Abbildung 9.2.: Histogramme des KQ-Schätzers für β für $n = 50, 100$ (R-Programm siehe Abschnitt A.8, Seite 352) DGP siehe Gleichung (9.3)

Beispiel: Monte-Carlo-Simulation zu Schätzeigenschaften des KQ-Schätzers bei AR-Prozessen In Abschnitt 13.5 wird der KQ-Schätzer zur Schätzung von Zeitreihendaten verwendet. Erhöhen Sie in der MC-Studie zur Bestimmung der Verzerrung des KQ-Schätzers im AR(1)-Modell, Abschnitt 12.3.1, die Stichprobengröße N und notieren Sie Ihre Ergebnisse. Berechnen Sie auch die Varianz der Schätzungen über alle Replikationen.

9.3. Die Kovarianzmatrix der Parameterschätzer

- **Kovarianzmatrix / Varianz-Kovarianzmatrix / Varianzmatrix:** siehe Gleichung (5.52)
- **Bedingte Varianz-Kovarianzmatrix:** Die bedingte Varianz-Kovarianzmatrix gibt die

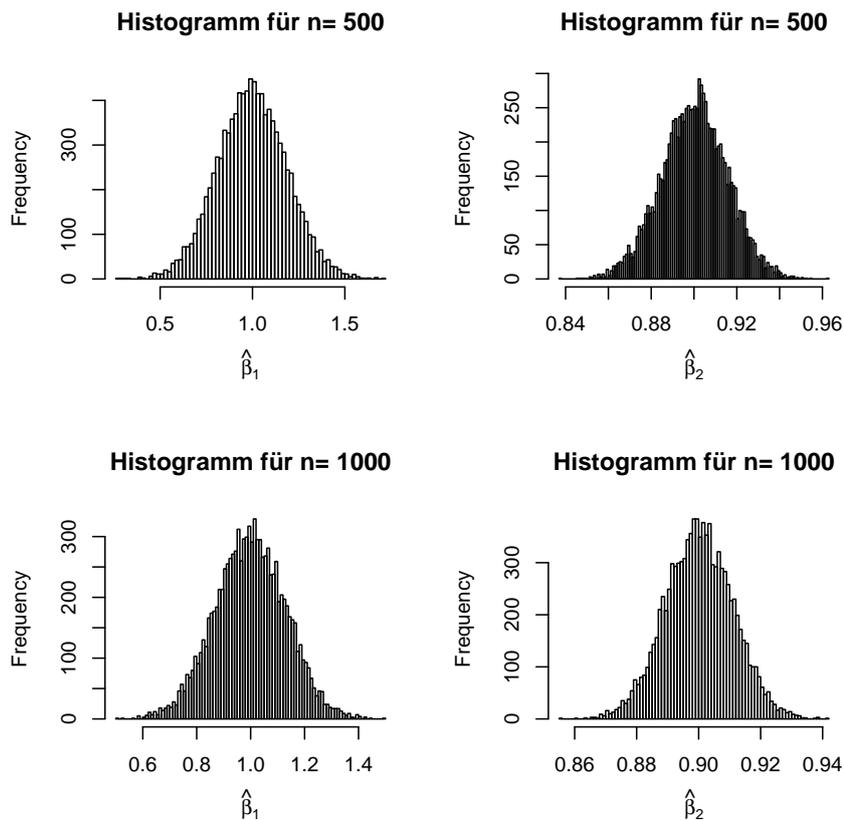


Abbildung 9.3.: Histogramme des KQ-Schätzers für β für $n = 500, 1000$ (R-Programm siehe Abschnitt A.8, Seite 352) DGP siehe Gleichung (9.3)

Varianz von $\hat{\boldsymbol{\theta}}$ an, die mit der bedingten Verteilung von $\tilde{\boldsymbol{\theta}}$ gegeben \mathbf{X} assoziiert ist:

$$\text{Var}(\tilde{\boldsymbol{\theta}}|\mathbf{X}) = E \left[\left(\tilde{\boldsymbol{\theta}} - E[\tilde{\boldsymbol{\theta}}|\mathbf{X}] \right) \left(\tilde{\boldsymbol{\theta}} - E[\tilde{\boldsymbol{\theta}}|\mathbf{X}] \right)^T \middle| \mathbf{X} \right] \quad (9.5a)$$

$$= E \left[\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T | \mathbf{X} \right] - E[\tilde{\boldsymbol{\theta}}|\mathbf{X}] E[\tilde{\boldsymbol{\theta}}|\mathbf{X}]^T. \quad (9.5b)$$

- **Zusammenhang zwischen unbedingten und bedingten Varianzen** (siehe (2.28) für skalaren Fall)

$$\text{Var}(\tilde{\boldsymbol{\theta}}) = E \left[\text{Var}(\tilde{\boldsymbol{\theta}}|\mathbf{X}) \right] + \text{Var} \left(E(\tilde{\boldsymbol{\theta}}|\mathbf{X}) \right). \quad (9.6)$$

Beweis: ‡ Ableitung:

$$\begin{aligned}
& E \left[\left(\tilde{\boldsymbol{\theta}} - E(\tilde{\boldsymbol{\theta}}) \right) \left(\tilde{\boldsymbol{\theta}} - E(\tilde{\boldsymbol{\theta}}) \right)^T \right] \\
&= E \left[\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T \right] - E(\tilde{\boldsymbol{\theta}}) E(\tilde{\boldsymbol{\theta}}^T) \\
&= E \left[E \left(\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T \mid \mathbf{X} \right) \right] - E \left[E(\tilde{\boldsymbol{\theta}} \mid \mathbf{X}) \right] E \left[E(\tilde{\boldsymbol{\theta}}^T \mid \mathbf{X}) \right] \\
&= E \left[E \left(\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T \mid \mathbf{X} \right) - E(\tilde{\boldsymbol{\theta}} \mid \mathbf{X}) E(\tilde{\boldsymbol{\theta}}^T \mid \mathbf{X}) + E(\tilde{\boldsymbol{\theta}} \mid \mathbf{X}) E(\tilde{\boldsymbol{\theta}}^T \mid \mathbf{X}) \right] \\
&\quad - E \left[E(\tilde{\boldsymbol{\theta}} \mid \mathbf{X}) \right] E \left[E(\tilde{\boldsymbol{\theta}}^T \mid \mathbf{X}) \right] \\
&= E \left[\underbrace{E \left(\tilde{\boldsymbol{\theta}} \tilde{\boldsymbol{\theta}}^T \mid \mathbf{X} \right) - E(\tilde{\boldsymbol{\theta}} \mid \mathbf{X}) E(\tilde{\boldsymbol{\theta}}^T \mid \mathbf{X})}_{E \left[\text{Var}(\tilde{\boldsymbol{\theta}} \mid \mathbf{X}) \right]} \right] + \\
&\quad \underbrace{E \left[E(\tilde{\boldsymbol{\theta}} \mid \mathbf{X}) E(\tilde{\boldsymbol{\theta}}^T \mid \mathbf{X}) \right] - E \left[E(\tilde{\boldsymbol{\theta}} \mid \mathbf{X}) \right] E \left[E(\tilde{\boldsymbol{\theta}}^T \mid \mathbf{X}) \right]}_{\text{Var} \left(E(\tilde{\boldsymbol{\theta}} \mid \mathbf{X}) \right)}
\end{aligned}$$

□

- **Varianz-Kovarianzmatrix des unverzerrten KQ-Schätzers $\hat{\boldsymbol{\beta}}$** (Annahmen **(B1)**, **(B2a)**, **(B3)** erfüllt):

$$\begin{aligned}
\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) &= E \left[(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \mid \mathbf{X} \right] \\
&= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E(\mathbf{u} \mathbf{u}^T \mid \mathbf{X}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \\
&= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{u} \mid \mathbf{X}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}. \tag{9.7}
\end{aligned}$$

Dies ist die allgemeine Varianz-Kovarianzmatrix des KQ-Schätzers, bei der auch Heteroskedastie und Korrelation zwischen den Fehlern gegeben \mathbf{X} zugelassen ist, da die bedingte Varianz-Kovarianzmatrix der Fehler $\text{Var}(\mathbf{u} \mid \mathbf{X})$ nicht weiter spezifiziert ist. Dieser allgemeine Fall wird in Kapitel 14 behandelt.

- **Varianz-Kovarianzmatrix des KQ-Schätzers bei homoskedastischen und unkorrelierten Fehlern:**

Es gilt zusätzlich die **Annahme**

(B2b) Homoskedastie und Unkorreliertheit der Fehler

$$\text{Var}(\mathbf{u} \mid \mathbf{X}) = \sigma^2 \mathbf{I},$$

wobei für die Fehlervarianz des DGPs $\sigma^2 = \sigma_0^2$ gilt.

- Dann vereinfacht sich die **Varianz-Kovarianzmatrix des KQ-Schätzers** (9.7) zur bekannten Form

$$\text{Var}(\hat{\boldsymbol{\beta}} \mid \mathbf{X}) = \sigma_0^2 (\mathbf{X}^T \mathbf{X})^{-1}. \tag{9.8}$$

- Die **unbedingte Varianz-Kovarianzmatrix** ergibt sich mit Hilfe von (9.6):

$$\text{Var}(\hat{\boldsymbol{\beta}}) = \sigma_0^2 E \left[(\mathbf{X}^T \mathbf{X})^{-1} \right] \quad (9.9)$$

wegen $\text{Var} \left(E[\hat{\boldsymbol{\beta}}|\mathbf{X}] \right) = \text{Var}(\boldsymbol{\beta}) = \mathbf{0}$.

‡ Zur Existenz von $E \left[(\mathbf{X}^T \mathbf{X})^{-1} \right]$ siehe technische Ergänzung am Ende des Abschnitts 9.4.

- **Homoskedastische Fehler:** Die Fehler u_t werden als homoskedastisch bezeichnet, wenn deren Varianz für alle Stichprobenbeobachtungen konstant ist, also gilt:

$$\text{Var}(u_t|\mathbf{X}_t) = \sigma_t^2 = \sigma^2. \quad (9.10)$$

Eine strengere Form ist $\text{Var}(u_t|\mathbf{X}) = \sigma^2$.

- **Verhalten der Varianz-Kovarianzmatrix für ansteigende Stichprobengröße:** Eine äquivalente Darstellung zu (9.8) ist:

$$\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = \left(\frac{1}{n} \sigma_0^2 \right) \left(\frac{1}{n} \mathbf{X}^T \mathbf{X} \right)^{-1}. \quad (9.11)$$

Ist außerdem die Bedingung (A1)

$$\left(\frac{1}{n} \mathbf{X}^T \mathbf{X} \right)^{-1} \xrightarrow{p} \mathbf{S}_{\mathbf{X}^T \mathbf{X}}^{-1}$$

erfüllt, verringern sich im Allgemeinen die bedingten Varianzen $\text{Var}(\hat{\beta}_j|\mathbf{X})$ bzw. Kovarianzen $\text{Cov}(\hat{\beta}_j, \hat{\beta}_i|\mathbf{X})$, wenn

- * die Stichprobengröße n ansteigt,
- * die Fehlervarianz σ_0^2 kleiner wird.

- **Asymptotische Varianz-Kovarianzmatrix:** Gilt Bedingung (A1) für (9.11), dann erhält man

$$\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) \xrightarrow{p} \mathbf{0}, \quad (9.12)$$

da $\sigma_0^2/n \rightarrow 0$ für $n \rightarrow \infty$. Um eine für $n \rightarrow \infty$ feste Varianz-Kovarianzmatrix zu erhalten, muss genau dieser Term gegen einen festen Wert konvergieren, was genau dann der Fall ist, wenn man $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0$ mit \sqrt{n} multipliziert, denn dann gilt

$$\text{plim}_{n \rightarrow \infty} \text{Var}(\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)|\mathbf{X}) = \sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}}^{-1}. \quad (9.13)$$

Der Ausdruck auf der rechten Seite wird als **asymptotische Varianz-Kovarianzmatrix** des KQ-Schätzers bezeichnet.

In der Praxis wird $\mathbf{S}_{\mathbf{X}^T \mathbf{X}}$ durch $\mathbf{X}^T \mathbf{X}/n$ geschätzt, wodurch sich nach Kürzen von n die Approximation

$$\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) \approx \sigma_0^2 (\mathbf{X}^T \mathbf{X})^{-1} \quad (9.14)$$

ergibt, wobei σ_0^2 wieder durch s^2 geschätzt wird.

- **Varianz eines Schätzers eines einzelnen Parameters β_j** : Es gilt, wenn die Regression eine Konstante enthält,

$$\text{Var}(\hat{\beta}_j | \mathbf{X}) = \frac{\sigma_0^2}{SST_j(1 - R_j^2)}, \quad (9.15)$$

wobei R_j^2 das Bestimmtheitsmaß einer Regression von \mathbf{x}_j auf alle übrigen Regressoren bezeichnet.

Interpretation: Die Varianz von $\hat{\beta}_j$ ist umso größer,

- * je besser \mathbf{x}_j durch die verbleibenden Regressoren in \mathbf{X} erklärt wird, d. h. je größer das Bestimmtheitsmaß der Regression von \mathbf{x}_j auf die verbleibenden Regressoren in \mathbf{X} ist,
- * je kleiner die Streuung des Regressors \mathbf{x}_j ist,
- * je größer die Fehlervarianz σ_0^2 ist.

Beweis: Ableitung von (9.15) (wobei zur Vereinfachung o.E.d.A. $j = 1$ gewählt wird): Damit ist folgende Partitionierung

$$\mathbf{y} = \mathbf{x}_1\beta_1 + \mathbf{X}_2\beta_2 + \mathbf{u}$$

möglich und β_1 kann mit dem Frisch-Waugh-Lovell-Theorem (vgl. Abschnitt 7.1) auf Basis der Regression

$$\mathbf{M}_2\mathbf{y} = \mathbf{M}_2\mathbf{x}_1\beta_1 + \text{Residuen}$$

geschätzt werden, wobei $\mathbf{M}_2 = \mathbf{I} - \mathbf{X}_2(\mathbf{X}_2^T\mathbf{X}_2)^{-1}\mathbf{X}_2^T$ ist. Man erhält den KQ-Schätzer:

$$\hat{\beta}_1 = (\mathbf{x}_1^T\mathbf{M}_2\mathbf{x}_1)^{-1} \mathbf{x}_1^T\mathbf{M}_2\mathbf{y}.$$

Es lässt sich (leicht) zeigen, dass

$$\text{Var}(\hat{\beta}_1 | \mathbf{X}) = \sigma_0^2(\mathbf{x}_1^T\mathbf{M}_2\mathbf{x}_1)^{-1} = \frac{\sigma_0^2}{\mathbf{x}_1^T\mathbf{M}_2\mathbf{x}_1}.$$

Beachte, dass der Ausdruck $\mathbf{x}_1^T\mathbf{M}_2\mathbf{x}_1 = \|\mathbf{M}_2\mathbf{x}_1\|^2 = SSR_1$ (vgl. (7.14)) der quadrierten Länge des Residuenvektors der Regression von \mathbf{x}_1 auf \mathbf{X}_2 entspricht, bzw. der Residuenquadratsumme der Regression von \mathbf{x}_1 auf \mathbf{X}_2 . Da $R_1^2 = SSE_1/SST_1$ und, falls \mathbf{X}_2 eine Konstante enthält, $SST_1 = SSE_1 + SSR_1$ gilt, erhält man $SSE_1 = R_1^2 SST_1$ und damit via $SST_1 - R_1^2 SST_1 = SSR_1$ auch

$$\|\mathbf{M}_2\mathbf{x}_1\|^2 = SST_1(1 - R_1^2)$$

und somit (9.15) für $j = 1$. □

- **Multikollinearität** oder kurz **Kollinearität**:

Wie gerade festgestellt, folgt aus (9.15): Ist der Vektor \mathbf{x}_j 'nahezu' von mindestens einer anderen Spalte in \mathbf{X} linear abhängig, ist die Länge des Residuenvektors kurz und die

Varianz für $\hat{\beta}_j$ groß. In diesem Fall sagt man, dass die Variable j mit einer oder mehreren Variablen **multikollinear** ist. Es liegt dann **Multikollinearität** oder kurz **Kollinearität** vor.

Das Problem von Multikollinearität kann nur gelöst werden, indem die Stichprobengröße n erhöht wird. Weglassen der Variable j wird dagegen im Allgemeinen zu einem fehlspezifizierten Modell führen, siehe Abschnitt 9.6. Allerdings ist es möglich, den Gesamteffekt zu betrachten, indem der **mittlere quadratische Fehler** (5.50) bzw. (9.36) betrachtet wird.

In der Praxis ist es nicht notwendig, für jede Variable R_j^2 zu berechnen. Stattdessen betrachtet man die **Korrelationsmatrix** $\text{Corr}(\hat{\beta}|\mathbf{X})$. Ist die Korrelation zwischen $\hat{\beta}_i$ und $\hat{\beta}_j$ im Absolutbetrag recht nahe 1, weist dies auf Multikollinearität hin.

- **Varianz von linearen Funktionen von Parameterschätzern**

Ist die zu schätzende Größe γ eine lineare Funktion der geschätzten Parameter

$$\hat{\gamma} = \mathbf{w}^T \hat{\beta},$$

wobei \mathbf{w} ein geeignet dimensionierter Spaltenvektor ist, dann lässt sich die Varianz von $\hat{\gamma}$ sehr einfach bestimmen durch

$$\begin{aligned} \text{Var}(\hat{\gamma}|\mathbf{X}) &= \text{Var}(\mathbf{w}^T \hat{\beta}|\mathbf{X}) \\ &= E \left[\mathbf{w}^T (\hat{\beta} - E[\hat{\beta}|\mathbf{X}]) (\hat{\beta} - E[\hat{\beta}|\mathbf{X}])^T \mathbf{w} | \mathbf{X} \right] \\ &= \mathbf{w}^T E \left[(\hat{\beta} - E[\hat{\beta}|\mathbf{X}]) (\hat{\beta} - E[\hat{\beta}|\mathbf{X}])^T | \mathbf{X} \right] \mathbf{w} \\ &= \mathbf{w}^T \text{Var}(\hat{\beta}|\mathbf{X}) \mathbf{w}. \end{aligned} \quad (9.16)$$

Und bei homoskedastischen und unkorrelierten Fehlern (Annahme **(B2b)**):

$$\text{Var}(\hat{\gamma}|\mathbf{X}) = \sigma_0^2 \mathbf{w}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{w}. \quad (9.17)$$

Beispiel Skalanelastizität: $\gamma = \alpha_1 + \alpha_2$ in Cobb-Douglas-Produktionsfunktion:

$$\begin{aligned} Y &= AL^{\alpha_1} K^{\alpha_2} e^u \\ \ln Y &= \beta_1 + \alpha_1 \ln L + \alpha_2 \ln K + u \end{aligned} \quad (9.18)$$

- **Varianz des Vorhersagefehlers bei unverzerrter Prognose** (Anwendung von (9.17))

Sind die Annahmen **(B1)**, **(B2a)**, **(B3)** erfüllt und damit auch das Modell korrekt spezifiziert, ist die Prognose $\hat{y}_s = \mathbf{X}_s \hat{\beta}$ für (y_s, \mathbf{X}_s) aus der Grundgesamtheit unverzerrt, da

$$E[\hat{y}_s | \mathbf{X}, \mathbf{X}_s] = \mathbf{X}_s \beta_0. \quad (9.19)$$

Daraus ergibt sich der Prognosefehler

$$y_s - \mathbf{X}_s \hat{\beta} = \mathbf{X}_s (\beta_0 - \hat{\beta}) + u_s,$$

dessen Erwartungswert Null ist. Die Varianz des Prognosefehlers lautet deshalb

$$\begin{aligned} \text{Var}(y_s - \mathbf{X}_s \hat{\boldsymbol{\beta}} | \mathbf{X}_s, \mathbf{X}) &= E \left[\left\{ \mathbf{X}_s (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) + u_s \right\} \left\{ (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})^T \mathbf{X}_s^T + u_s \right\} \middle| \mathbf{X}_s, \mathbf{X} \right] \\ &= \mathbf{X}_s \text{Var}(\hat{\boldsymbol{\beta}} | \mathbf{X}) \mathbf{X}_s^T + E[u_s^2 | \mathbf{X}_s] - 2 \mathbf{X}_s \underbrace{\text{Cov}(\hat{\boldsymbol{\beta}}, u_s | \mathbf{X}_s, \mathbf{X})}_{=0, \text{ bei Unkorreliertheit}} \\ &= \sigma_0^2 \mathbf{X}_s (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}_s^T + \sigma_0^2 \quad (\text{gegeben Annahme (B2b)}) \end{aligned}$$

→ Prognosefehlervarianz = Varianz des Schätzers der abhängigen Variablen + Varianz von u_s .

• **Zusammenfassung der Annahmen des multiplen linearen Regressionsmodells mit streng exogenen Regressoren**

– (B1) Korrekt spezifiziertes Modell: Der DGP ist für $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ im multiplen linearen Regressionsmodell enthalten.

– (B2): $\mathbf{u} | \mathbf{X} \sim (\mathbf{0}, \sigma^2 \mathbf{I}) \iff \left\{ \begin{array}{l} \text{(B2a): } E[\mathbf{u} | \mathbf{X}] = \mathbf{0} \text{ (} \mathbf{X} \text{ ist (streng) exogen) \& \\ \text{(B2b): } \text{Var}(\mathbf{u} | \mathbf{X}) = \sigma^2 \mathbf{I} \text{ (Fehler sind auf } \mathbf{X} \text{ bedingt} \\ \text{homoskedastisch und unkorreliert).} \end{array} \right.$

– (B3) \mathbf{X} hat vollen Spaltenrang

9.4. Die Effizienz unverzerrter KQ-Schätzer

• Vgl. zur Definition der **Effizienz** eines Schätzers Abschnitt 5.4 und (5.56). Im Folgenden wird die Klasse der linearen Schätzer betrachtet.

• **Linearer Schätzer:** Ein Schätzer $\tilde{\boldsymbol{\beta}}$ für den Parametervektor $\boldsymbol{\beta}$ in einem multiplen linearen Regressionsmodell heißt **linear**, wenn $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$ gilt, wobei die $(k \times n)$ -Matrix $\mathbf{A} := \mathbf{A}(\mathbf{X})$ **ausschließlich von den Regressoren \mathbf{X} abhängen darf**, nicht jedoch von \mathbf{y} , also $E[\mathbf{A} | \mathbf{X}] = \mathbf{A}$ gilt.

• Der KQ-Schätzer ist ein linearer Schätzer, da $\mathbf{A} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T$ gilt.

• Ein linearer Schätzer $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$ ist **unverzerrt**, wenn die Annahmen (B1), (B2a) gelten, sowie

$$\mathbf{A}\mathbf{X} = \mathbf{I}, \quad \text{da } E[\tilde{\boldsymbol{\beta}} | \mathbf{X}] = \mathbf{A}\mathbf{X}\boldsymbol{\beta}_0 + \mathbf{A}E[\mathbf{u} | \mathbf{X}]. \quad (9.20)$$

• **Vergleich des KQ-Schätzers $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$ mit beliebigen linearen und erwartungstreuen Schätzern $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$ mit $\mathbf{A}\mathbf{X} = \mathbf{I}$**

– **Gauss-Markov-Theorem:** Unter den Annahmen (B1), (B2), (B3) ist der KQ-Schätzer $\hat{\boldsymbol{\beta}}$ unter allen linearen und unverzerrten Schätzern $\tilde{\boldsymbol{\beta}}$ der effiziente Schätzer (**best**

linear unbiased estimator (BLUE)). Das bedeutet, dass die Matrix der Differenz der Varianz-Kovarianzmatrizen $Var(\tilde{\beta}) - Var(\hat{\beta})$ **positiv semidefinit** ist.

Beispiele ineffizienter linearer unverzerrter Schätzer:

- * Schätzer des Erwartungswertes (mean) $(y_1 + y_n)/2$.
- * Jeder KQ-Schätzer, der auf ein Regressionsmodell mit redundanten unabhängigen Variablen angewendet wird, siehe Abschnitt 9.6.
- * Instrumentvariablenschätzer, siehe z. B. das BA-Modul **Weiterführende Fragen der Ökonometrie** oder das MA-Modul **Fortgeschrittene Ökonometrie**.

Beweisskizze: Da $\tilde{\beta} - \hat{\beta} = \underbrace{(\mathbf{A} - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T)}_{\mathbf{C}} \mathbf{y} = \mathbf{C} \mathbf{X} \boldsymbol{\beta} + \mathbf{C} \mathbf{u} = \mathbf{C} \mathbf{u}$, gilt,

dass

$$Var(\tilde{\beta}) = Var(\hat{\beta} + \mathbf{C} \mathbf{u}) = Var(\hat{\beta}) + Var(\mathbf{C} \mathbf{u}),$$

da unter Berücksichtigung von (9.20) und (B2b) $E[(\hat{\beta} - \beta_0)(\mathbf{C} \mathbf{u})^T] = \mathbf{0}$ gezeigt werden kann. Da jede Varianz-Kovarianzmatrix positiv semidefinit ist, gilt dies auch für $Var(\mathbf{C} \mathbf{u})$. \square

- Ursprünglich wurde das Gauss-Markov-Theorem für nicht-stochastische Regressoren \mathbf{X} bewiesen.
- ‡ **Technische Ergänzung:** Ist \mathbf{X} stochastisch, ist es prinzipiell möglich, dass z. B. Annahme (B3) bzw. (9.20) für eine spezifische Realisation von \mathbf{X} verletzt ist, also \mathbf{X} nicht vollen Rang hat und damit $(\mathbf{X}^T \mathbf{X})$ nicht invertierbar ist. Sind die Regressoren stetig verteilt, dann ist die Wahrscheinlichkeit hierfür 0.
 - Gilt für ein Ereignis C , dass $P(C) = 1$, dann gilt für das Komplement C^c , dass $P(C^c) = 0$. Man sagt dann, dass das Ereignis C **fast sicher (almost surely (a.s.))** eintritt.
 - Beispiel für ein fast sicheres Ereignis: Gegeben sei eine stetige Zufallsvariable $X \in \mathbb{R}$. Das Ereignis $C = \{X \in (-\infty, a) \cup (a, \infty)\}$ hat das komplementäre Ereignis $C^c = \{X = a\}$. Da $P(X = a) = P(C^c) = 0$, gilt für C , dass $P(C) = 1$.
 - Enthält \mathbf{X} nur diskrete Regressoren, beispielsweise eine Konstante und eine Dummyvariable, dann besteht eine positive Wahrscheinlichkeit, dass eine Stichprobe gezogen wird, in der die Dummyvariable für alle Beobachtungen den Wert 1 annimmt und damit \mathbf{X} reduzierten Rang hat und $\mathbf{X}^T \mathbf{X}$ nicht invertierbar ist. Die Annahme (B3) ist also für dieses Beispiel nicht *fast sicher* erfüllt. In diesem Fall existiert auch $E[(\mathbf{X}^T \mathbf{X})^{-1}]$ nicht, da eine positive Wahrscheinlichkeit vorliegt, dass die Matrix $\mathbf{X}^T \mathbf{X}$ nicht invertierbar ist.
 - Die Existenz des unbedingten Erwartungswertes und der unbedingten Varianz des KQ-Schätzers setzt also voraus, dass die Annahmen (B1) bis (B3) fast sicher gelten.
 - Für die **Praxis** ist es im Allgemeinen ausreichend, die Verteilungseigenschaften **gegeben**

die **Regressoren** zu kennen. Dann braucht man sich über diese Problematik keine Gedanken zu machen.

- Möchte man jedoch Monte-Carlo-Simulationen durchführen, in denen auch \mathbf{X} bei jeder Realisation neu gezogen wird, aber mit positiver Wahrscheinlichkeit \mathbf{X} reduzierten Rang hat, wird immer wieder der Fall einer singulären $\mathbf{X}^T \mathbf{X}$ Matrix auftreten und der KQ-Schätzer nicht berechenbar sein.

9.5. Schätzen der Fehlervarianz

- In diesem Abschnitt werden die Annahmen **(B1)** bis **(B3)** vorausgesetzt.
- Im korrekt spezifizierten KQ-Modell gilt

$$\begin{aligned}\hat{\mathbf{u}} &= \mathbf{M}_X \mathbf{y} \\ &= \mathbf{M}_X \mathbf{X} \boldsymbol{\beta}_0 + \mathbf{M}_X \mathbf{u} \\ &= \mathbf{M}_X \mathbf{u},\end{aligned}\tag{9.21}$$

da $\mathbf{M}_X \mathbf{X} = \mathbf{0}$.

Das Residuum \hat{u}_t entspricht einer **Linearkombination** des Fehlervektors \mathbf{u} .

- **Varianz** des Residuenvektors:

$$\begin{aligned}Var(\hat{\mathbf{u}}|\mathbf{X}) &= Var(\mathbf{M}_X \mathbf{u}|\mathbf{X}) \\ &= E[\mathbf{M}_X \mathbf{u} \mathbf{u}^T \mathbf{M}_X^T | \mathbf{X}] \\ &= \mathbf{M}_X (\sigma_0^2 \mathbf{I}) \mathbf{M}_X^T \\ &= \sigma_0^2 \mathbf{M}_X.\end{aligned}\tag{9.22}$$

- **Eigenschaften der Residuen** \hat{u}_t : Diese ergeben sich aus der Varianz-Kovarianzmatrix der Residuen $Var(\hat{\mathbf{u}}|\mathbf{X})$.

Die Residuen sind im Allgemeinen

- korreliert und
- heteroskedastisch mit $Var(\hat{u}_t|\mathbf{X}) \leq Var(u_t) = \sigma_0^2$.

Beweis: Wie in Abschnitt 7.2 bezeichnet \mathbf{e}_t einen **Einheitsbasisvektor**. Dann ist

$$\hat{u}_t = \mathbf{e}_t^T \hat{\mathbf{u}}$$

und

$$Var(\hat{u}_t|\mathbf{X}) = Var(\mathbf{e}_t^T \hat{\mathbf{u}}|\mathbf{X}) = \mathbf{e}_t^T Var(\hat{\mathbf{u}}|\mathbf{X}) \mathbf{e}_t = \sigma_0^2 \mathbf{e}_t^T \mathbf{M}_X \mathbf{e}_t = \sigma_0^2 \|\mathbf{M}_X \mathbf{e}_t\|^2.$$

Aufgrund der orthogonalen Zerlegung gilt

$$\|\mathbf{e}_t\|^2 = \underbrace{\|\mathbf{P}_X \mathbf{e}_t\|^2}_{h_t} + \underbrace{\|\mathbf{M}_X \mathbf{e}_t\|^2}_{1-h_t},$$

so dass $\|\mathbf{M}_X \mathbf{e}_t\|^2 \leq \|\mathbf{e}_t\|^2 = 1$. \square

- **Maximum-Likelihood-Schätzer für die Fehlervarianz:**

– Der Schätzer

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^n \hat{u}_t^2 \quad (9.23)$$

wird als **Maximum-Likelihood-Schätzer** für die Fehlervarianz σ^2 bezeichnet, da er sich aus dem Maximum-Likelihood-Ansatz ergibt, siehe MA-Veranstaltung **Fortgeschrittene Ökonometrie**.

– **Eigenschaft:** $\hat{\sigma}^2$ ist **verzerrt**.

Beweis:

$$\begin{aligned} E[\hat{\sigma}^2 | \mathbf{X}] &= \frac{1}{n} \sum_{t=1}^n E[\hat{u}_t^2 | \mathbf{X}] \\ &= \frac{1}{n} \sum_{t=1}^n \text{Var}(\hat{u}_t | \mathbf{X}) \\ &= \sigma_0^2 \frac{1}{n} \sum_{t=1}^n \|\mathbf{M}_X \mathbf{e}_t\|^2. \end{aligned}$$

Aus $\|\mathbf{P}_X \mathbf{e}_t\|^2 = h_t$ folgt schließlich

$$E[\hat{\sigma}^2 | \mathbf{X}] = \sigma_0^2 \frac{1}{n} \sum_{t=1}^n \underbrace{(1 - h_t)}_{\leq 1} \leq \sigma_0^2.$$

Mit Hilfe des Spur-Operators kann man zeigen, dass

$$\sum_{t=1}^n (1 - h_t) = n - k.$$

Daraus folgt

$$E[\hat{\sigma}^2 | \mathbf{X}] = \frac{n - k}{n} \sigma_0^2. \quad (9.24)$$

\square

- **Unverzerrter Schätzer für die Fehlervarianz:** Berücksichtigen von (9.24) in (9.23) liefert den unverzerrten Schätzer

$$s^2 = \frac{1}{n - k} \sum_{t=1}^n \hat{u}_t^2. \quad (9.25)$$

(Beachte die Notation: in vielen anderen Ökonometriebüchern, z. B. [Wooldridge \(2009\)](#), wird dieser Schätzer mit $\hat{\sigma}^2$ bezeichnet.)

- Die Wurzel von s^2 wird als **Standardfehler einer Regression (standard error of regression)** bezeichnet.
- Ein **unverzerrter Schätzer der Kovarianzmatrix des KQ-Schätzers** ist dann

$$\widehat{\text{Var}}(\hat{\beta}|\mathbf{X}) = s^2(\mathbf{X}^T\mathbf{X})^{-1}. \quad (9.26)$$

Beispiel: Handelsströme Für die KQ-Schätzungen von Modell 3 (6.17) sind im folgenden R-Output die Varianz-Kovarianzmatrix und Korrelationsmatrix der Parameterschätzer angegeben

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# Schätze Varianz-Kovarianzmatrix der KQ-Schätzer für Modell 3a
summary(mod_3a_kq)$cov

# Schätze Korrelationsmatrix der KQ-Schätzer für Modell 3a
cov2cor(summary(mod_3a_kq)$cov)

# Schätze Kovarianzmatrix der Stichprobenbeobachtungen für Modell 3a
cor(data.frame(log_wdi_gdpusdcr_o = log(wdi_gdpusdcr_o),
               log_cepil_dist=log(cepil_dist), ebrd_tfes_o))
```

Listing 9.1: ./R_code/4_ff_Beispiel_Handelsstroeme.R

R-Output

```
(Intercept) log(wdi_gdpusdcr_o) log(cepil_dist) ebrd_tfes_o
(Intercept)      6.2069332      -0.124749916      -0.315428513      -0.183737444
log(wdi_gdpusdcr_o) -0.1247499      0.004936052      0.002017428      -0.003353511
log(cepil_dist)    -0.3154285      0.002017428      0.030582699      0.009851900
ebrd_tfes_o       -0.1837374      -0.003353511      0.009851900      0.048163990

(Intercept) log(wdi_gdpusdcr_o) log(cepil_dist) ebrd_tfes_o
(Intercept)      1.0000000      -0.7127084      -0.7239766      -0.3360454
log(wdi_gdpusdcr_o) -0.7127084      1.0000000      0.1641989      -0.2174947
log(cepil_dist)    -0.7239766      0.1641989      1.0000000      0.2566970
ebrd_tfes_o       -0.3360454      -0.2174947      0.2566970      1.0000000

               log_wdi_gdpusdcr_o log_cepil_dist ebrd_tfes_o
log_wdi_gdpusdcr_o      1.0000000      -0.2332410  0.2723423
log_cepil_dist          -0.2332410      1.0000000 -0.3037030
ebrd_tfes_o             0.2723423      -0.3037030  1.0000000
```

Beachte, dass die Korrelationen zwischen den Variablen im Absolutbetrag nicht größer als 0.26, also relativ gering sind, und keine Anzeichen für Multikollinearität vorliegen.

9.6. Über- oder fehlspezifizierte lineare Regressionsmodelle

Zur Definition der **Informationsmenge** siehe Abschnitt 5.3.

Überspezifizierung (overspecification)

- Ein Modell \mathbb{M} ist **überspezifiziert**, wenn es Variablen enthält, die zur Informationsmenge Ω_t gehören, aber nicht im DGP enthalten sind. (Beachte: Überspezifizierte Modelle sind nicht fehlspezifiziert.)

Beispiel: Der DGP sei in

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{u}, \quad \mathbf{u}|\mathbf{X} \sim (\mathbf{0}, \sigma_0^2\mathbf{I}), \quad (9.27)$$

enthalten ((B1), (B2) gelten), geschätzt wird aber

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \mathbf{u}, \quad \mathbf{u}|\mathbf{X}, \mathbf{Z} \sim (\mathbf{0}, \sigma^2\mathbf{I}). \quad (9.28)$$

Das ‘unrestringierte’ Modell (9.28) enthält ebenfalls den DGP ($DGP \in \mathbb{M}$), da ja die Parameter $\boldsymbol{\beta} = \boldsymbol{\beta}_0$, $\boldsymbol{\gamma} = \mathbf{0}$ und $\sigma^2 = \sigma_0^2$ möglich sind.

- **Eigenschaften des KQ-Schätzers $\tilde{\boldsymbol{\beta}}$** des überspezifizierten Modells (9.28):
 - (i) **unverzerrt**, da nach dem Frisch-Waugh-Lovell-Theorem, siehe Seite 7.1.3, der KQ-Schätzer $\tilde{\boldsymbol{\beta}}$ der Regression

$$\mathbf{M}_Z\mathbf{y} = \mathbf{M}_Z\mathbf{X}\boldsymbol{\beta} + \text{Residuen}$$

mit $\mathbf{M}_Z = \mathbf{I} - \mathbf{Z}(\mathbf{Z}^T\mathbf{Z})^{-1}\mathbf{Z}^T$ mit dem KQ-Schätzer für $\boldsymbol{\beta}$ in dem überspezifizierten Modell (9.28) identisch ist. Deshalb gilt

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + (\mathbf{X}^T\mathbf{M}_Z\mathbf{X})^{-1}\mathbf{X}^T\mathbf{M}_Z\mathbf{u} \quad \Rightarrow \quad E(\tilde{\boldsymbol{\beta}}) = \boldsymbol{\beta}_0.$$

- (ii) im Allgemeinen im Vergleich zum KQ-Schätzer $\hat{\boldsymbol{\beta}}$ des ‘kleinsten’ korrekt spezifizierten Modells (9.27) **nicht effizient**. Dies gilt aufgrund des **Gauss-Markov-Theorems**, vgl. Abschnitt 9.4. Daraus folgt u.a., vgl. (5.56),

$$\text{Var}(\tilde{\beta}_j|\mathbf{X}, \mathbf{Z}) \geq \text{Var}(\hat{\beta}_j|\mathbf{X}), \quad j = 1, \dots, k.$$

Diese Ungleichung ergibt sich, vgl. (9.15), auch direkt aus

$$\frac{\sigma_0^2}{SST_j(1 - R_{j,\mathbf{X},\mathbf{Z}}^2)} \geq \frac{\sigma_0^2}{SST_j(1 - R_{j,\mathbf{X}}^2)}, \quad j = 1, \dots, k.$$

Auch wird durch zusätzliche, nicht benötigte Variablen die Gefahr von **Multikollinearität** erhöht.

- Diese Ergebnisse gelten unabhängig von der Stichprobengröße. Es lässt sich deshalb zeigen, dass der Schätzer eines überspezifizierten Modells **asymptotisch ineffizient** ist.

Fehlspezifizierung (auch Unterspezifizierung) (mis-/underspecification)

- Ein Modell \mathbb{M} ist **unterspezifiziert** bzw. **fehlspezifiziert**, wenn der DGP nicht im Modell enthalten ist.

Beispiel: DGP ist in

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_0 + \mathbf{Z}\boldsymbol{\gamma}_0 + \mathbf{u}, \quad \mathbf{u}|\mathbf{X}, \mathbf{Z} \sim (\mathbf{0}, \sigma_0^2\mathbf{I}), \quad \boldsymbol{\gamma}_0 \neq \mathbf{0}, \quad (9.29)$$

mit $(n \times k_1)$ -Regressormatrix \mathbf{X} und $(n \times k_2)$ -Regressormatrix \mathbf{Z} enthalten, es wird jedoch das Regressionsmodell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} \quad (9.30)$$

geschätzt. Dann ergibt sich für den KQ-Schätzer für (9.30)

$$\begin{aligned} \hat{\boldsymbol{\beta}} &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}\boldsymbol{\beta}_0 + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}\boldsymbol{\gamma}_0 + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u} \\ &= \boldsymbol{\beta}_0 + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}\boldsymbol{\gamma}_0 + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}. \end{aligned}$$

- **Beachte:** Der erste Teil des zweiten Terms auf der rechten Seite (vgl. zur Notation (6.2)) lautet:

$$(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z} = \left((\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{z}_1 \quad (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{z}_2 \quad \cdots \quad (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{z}_{k_2} \right).$$

Die l -te Spalte von $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}$ enthält also gerade den KQ-Schätzer $\hat{\boldsymbol{\delta}}_l = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{z}_l$ der (Hilfs-)Regression

$$\mathbf{z}_l = \mathbf{X}\boldsymbol{\delta}_l + \text{Fehler}. \quad (9.31)$$

Damit lässt sich der KQ-Schätzer $\hat{\boldsymbol{\beta}}$ schreiben als

$$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}_0 + \left(\hat{\boldsymbol{\delta}}_1 \quad \hat{\boldsymbol{\delta}}_2 \quad \cdots \quad \hat{\boldsymbol{\delta}}_{k_2} \right) \boldsymbol{\gamma}_0 + (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}. \quad (9.32)$$

Je nach Wahl der Bedingung im (bedingten) Erwartungswert erhält man unterschiedliche Verzerrungen:

- Somit ist der KQ-Schätzer **für gegebene Stichprobenwerte aller im DGP relevanter Regressoren verzerrt**, wenn

$$E[\hat{\boldsymbol{\beta}}|\mathbf{X}, \mathbf{Z}] = \boldsymbol{\beta}_0 + \left(\hat{\boldsymbol{\delta}}_1 \quad \hat{\boldsymbol{\delta}}_2 \quad \cdots \quad \hat{\boldsymbol{\delta}}_{k_2} \right) \boldsymbol{\gamma}_0 \neq \boldsymbol{\beta}_0, \quad (9.33)$$

also die Regressoren \mathbf{X} und \mathbf{Z} in einer gegebenen Stichprobe nicht orthogonal sind.

- Somit ist der KQ-Schätzer **für gegebene Stichprobenwerte aller Regressoren in \mathbf{X} verzerrt**, falls

$$E[\hat{\boldsymbol{\beta}}|\mathbf{X}] = E \left[E[\hat{\boldsymbol{\beta}}|\mathbf{X}, \mathbf{Z}|\mathbf{X}] \right] = \boldsymbol{\beta}_0 + E \left[\left(\hat{\boldsymbol{\delta}}_1 \quad \hat{\boldsymbol{\delta}}_2 \quad \cdots \quad \hat{\boldsymbol{\delta}}_{k_2} \right) |\mathbf{X} \right] \boldsymbol{\gamma}_0 \neq \boldsymbol{\beta}_0, \quad (9.34)$$

also $E \left[\left(\hat{\boldsymbol{\delta}}_1 \quad \hat{\boldsymbol{\delta}}_2 \quad \cdots \quad \hat{\boldsymbol{\delta}}_{k_2} \right) |\mathbf{X} \right] \neq \mathbf{0}$ ist.

- Somit ist der KQ-Schätzer **verzerrt**, falls

$$E[\hat{\boldsymbol{\beta}}] = E \left[E[\hat{\boldsymbol{\beta}}|\mathbf{X}, \mathbf{Z}] \right] = \boldsymbol{\beta}_0 + E \left[\left(\hat{\boldsymbol{\delta}}_1 \quad \hat{\boldsymbol{\delta}}_2 \quad \cdots \quad \hat{\boldsymbol{\delta}}_{k_2} \right) \right] \boldsymbol{\gamma}_0 \neq \boldsymbol{\beta}_0, \quad (9.35)$$

also der unbedingte Erwartungswert $E \left[\left(\hat{\boldsymbol{\delta}}_1 \quad \hat{\boldsymbol{\delta}}_2 \quad \cdots \quad \hat{\boldsymbol{\delta}}_{k_2} \right) \right] \neq \mathbf{0}$ ist. Mit anderen Worten, mindestens ein z_{ti} und x_{tj} sind miteinander **korreliert**.

- **Wichtig:** Gilt für den Erwartungswert $E\left[\begin{pmatrix} \hat{\delta}_1 & \hat{\delta}_2 & \dots & \hat{\delta}_{k_2} \end{pmatrix}\right] \neq 0$ unabhängig von der Stichprobengröße n , also eben auch für $n \rightarrow \infty$, dann ist der KQ-Schätzer für β_0 **inkonsistent!**

Fazit:

	Fehlspezifiziertes Modell	Überspezifiziertes Modell
	KQ-Schätzer ist	
endliche Stichprobe	i. Allg. verzerrt	ineffizient
asymptotisch	i. Allg. inkonsistent	asymptotisch ineffizient

Offensichtlich ist die Wahl eines korrekt, aber nicht überspezifizierten Modells sehr wichtig. Dies ist die Aufgabe von Verfahren zur Modellselektion, die im nächsten Kapitel dargestellt werden.

Mittlerer quadratischer Fehler:

- Die Matrix des **mittleren quadratischen Fehlers (MSE)**, vgl. (5.50), lautet gegeben alle Regressoren \mathbf{X}, \mathbf{Z} :

$$MSE(\hat{\beta}|\mathbf{X}, \mathbf{Z}) = E\left[\left(\hat{\beta} - \beta_0\right)\left(\hat{\beta} - \beta_0\right)^T \middle| \mathbf{X}, \mathbf{Z}\right]. \quad (9.36)$$

Wie bei der Verzerrung kann man hier hinsichtlich der Bedingungen unterscheiden (was hier aber nicht weiter gemacht wird).

- Beachte: nur für unverzerrte Schätzer ist die Matrix des mittleren quadratischen Fehlers gleich der Varianz-Kovarianzmatrix.
- Man kann zeigen (ggf. Übungsaufgabe), dass gilt

$$MSE(\hat{\beta}|\mathbf{X}, \mathbf{Z}) = \underbrace{\sigma_0^2(\mathbf{X}^T\mathbf{X})^{-1}}_{\text{Varianz}} + \underbrace{(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Z}\gamma_0\gamma_0^T\mathbf{Z}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}}_{\text{Verzerrung quadriert}}. \quad (9.37)$$

Eine eindeutige Aussage zum Vergleich dieser MSE-Matrix mit der des unverzerrten KQ-Schätzer in (9.29), d. h. $MSE(\tilde{\beta}|\mathbf{X}, \mathbf{Z}) = \sigma_0^2(\mathbf{X}^T\mathbf{M}_Z\mathbf{X})^{-1}$, ist nicht möglich, sondern hängt von der Größe der Verzerrung ab.

- Der MSE (9.36) ist damit geeignet, als Beurteilungskriterium zwischen verschiedenen Modellen zu dienen, da
 - im Fall fehlspezifizierter Modelle, die quadrierte Verzerrung berücksichtigt wird und
 - im Fall überspezifizierter die zu große KQ-Schätzvarianz
 gegeneinander abgewogen werden.
- Um die **Genauigkeit** des KQ-Schätzers des fehlspezifizierten Modells zu bestimmen, ist es aufgrund der Verzerrung des Schätzers nicht mehr sinnvoll, die Kovarianzmatrix heranzuziehen.

Zu lesen: Davidson & MacKinnon (2004), Kapitel 3.

10. Modellspezifikation

Ziele der Modellselektion:

- **Wesentliche Aspekte ökonometrischer Modellierung:**

1. Suche nach dem Modell, das den DGP enthält, der die Stichprobenbeobachtungen generiert hat.
2. Vermeiden zu großer Modelle.
3. Suche nach einem effizienten Schätzverfahren.

Mit anderen Worten: Verfahren zur **Modellspezifikation** dienen dazu, die **Wahrscheinlichkeiten** bei der Wahl des Modells

- für ein **fehlspezifiziertes Modell** und
- für ein **überspezifiziertes Modell**

geeignet zu begrenzen bzw. asymptotisch, wenn möglich, gegen Null gehen zu lassen.

Die **asymptotischen Anforderungen** implizieren:

1. **Konsistente Schätzung** interessierender Größen.
 2. **Effiziente Schätzung** interessierender Größen.
 3. Vorhandensein (**asymptotischer**) **Testverteilungen** zum Durchführen von **Hypothesentests**.
- **In der Praxis** wird es selten ein korrekt spezifiziertes Modell geben. Stattdessen Suche nach dem „besten“ Modell für die beabsichtigte Untersuchung, z. B. zum Ermitteln der BIP-Elastizität der Exporte oder von $E[y_t|\Omega_t]$. Was heißt „bestes“ Modell? Die Qualität des Modells hängt davon ab, wie bei Modellen in reduzierter Form ein darin enthaltenes Element den DGP approximieren kann. Bei Modellen in struktureller Form muss hierzu das implizierte Modell in reduzierter Form betrachtet werden. Siehe zur strukturellen und reduzierten Form Abschnitt 5.2. Die Bewertung der Approximation hängt von der Verlustfunktion, beispielsweise (5.45), ab, die für die Fragestellung gewählt wurde.
 - Allerdings ist die ökonometrische Theorie für approximierende Modelle für diesen Kurs zu kompliziert. Deshalb unterstellen wir in **diesem Kurs**, dass es ein korrekt spezifiziertes Modell gibt.

Anmerkungen:

Die **C**-Annahmen sind bisher noch nicht behandelt worden. Sie schließen Zeitreihendaten mit ein und werden in Abschnitt 13.4 detailliert besprochen. Der Vollständigkeit halber werden sie im Folgenden jedoch mit genannt.

Zu 1. **Konsistenz** erfordert unter anderem, dass das Modell korrekt spezifiziert ist, d. h. es müssen die

- Annahmen **(B1)** und **(B2a)**
 - bzw. allgemeiner bei Zeitreihen die Annahmen **(C1)** und **(C2a)**
- erfüllt sein. Dann gilt

$$E[y_t|\Omega_t] = \mathbf{X}_t\boldsymbol{\beta}_0. \quad (10.1)$$

Zu 2. **Effizienz** erfordert unter anderem, dass

- ein effizientes Schätzverfahren gewählt wird und
- vermieden wird, dass im Modell überflüssige Variablen enthalten sind.

Z. B. ist der KQ-Schätzer nur

- effizient, wenn unter anderem Annahmen **(B2b)** gilt, bzw.
- asymptotisch effizient, wenn unter anderem die Annahme **(C2b)** gilt, also die Fehler homoskedastisch sind.

Zu 3. Das Ableiten von **Testverteilungen** erfordert zusätzliche Annahmen, z. B. **(B4)** für exakte Tests oder **(C4a)** oder **(C4b)** für asymptotische Tests. (Vgl. Kapitel 11.)

- Die Verwendung von **Modellselektionskriterien** soll sicherstellen, dass
 - a) keine überflüssige Variablen im Modell enthalten sind und damit die Effizienz des Schätzers reduziert wird,
 - b) alle relevanten Variablen im Modell enthalten sind, also (10.1) gilt und damit eine Voraussetzung für Konsistenz vorliegt.

In kleineren Stichproben kann es sein, dass es nicht möglich ist, alle relevanten Variablen in das Modell aufzunehmen ohne dass die Schätzvarianz zu groß wird. Modellselektionskriterien erlauben hier einen „Trade-off“ zwischen a) und b).

- **Geschachtelte Modelle (nested models)**: \mathbb{M}_1 und \mathbb{M}_2 sind geschachtelt, wenn entweder $\mathbb{M}_1 \subset \mathbb{M}_2$ oder $\mathbb{M}_2 \subset \mathbb{M}_1$ gilt.

10.1. Modellselektionskriterien

- **Grundidee von Modellselektionskriterien:**

$$\text{Selektionskriterium} = \text{Fitmaß} + \text{Zahl der Parameter} \cdot \text{Straffunktion}(n) \quad (10.2)$$

- **Erster Term: Fitmaß:** Misst, wie gut sich das geschätzte Modell an die Daten anpasst. Dabei werden Fitmaße ausgewählt, die mit einem zusätzlichen Parameter im Allgemeinen zu einer Fitverbesserung führen, niemals jedoch zu einer Fitverschlechterung. Typischerweise werden hier entweder der Maximum-Likelihood-Schätzer $\hat{\sigma}^2 = \hat{\mathbf{u}}^T \hat{\mathbf{u}}/n$ der Fehlervarianz (9.23) oder Minus zweimal die Log-Likelihood-Funktion gewählt, wobei letztere sich für eine gegebene Stichprobengröße nur um eine Konstante von $\hat{\sigma}^2$ unterscheidet, siehe (10.3).

Es lässt sich zeigen, dass das Fitmaß von einem Modell, in dem relevante Regressoren fehlen, asymptotisch größer ist, als ein korrekt spezifiziertes Modell. Dies legt nahe, das Modell zu wählen, dessen Fitmaß am kleinsten ist. Ist allerdings das Modell überspezifiziert und benutzt man $\hat{\sigma}^2$, wird die wahre Fehlervarianz typischerweise unterschätzt. Deshalb besteht die Gefahr, ein überspezifiziertes Modell zu wählen. Um die Wahrscheinlichkeit hierfür zu reduzieren, verwendet man einen Strafterm, der die Hinzunahme von zusätzlichen irrelevanten Regressoren erschweren soll.

- **Zweiter Term: Strafterm:** Produkt aus Zahl der geschätzten Parameter k in β und Straffunktion:
 - * Der **Strafterm** bestraft die Anzahl der Parameter, um zu vermeiden, dass überflüssige Variablen mit ins Modell aufgenommen werden und somit das Schätzverfahren ineffizient ist.
 - * Der Strafterm steigt mit steigendem k und die Straffunktion muss so gewählt werden, dass sie mit steigendem n fällt. In letzterem Fall bedeutet dies, dass zusätzliche Parameter in größeren Stichproben relativ weniger bestraft werden, allerdings darf diese Strafe nicht zu schnell gegen Null gehen!
- Dies impliziert einen **Trade-off**: Regressoren werden dann in das Modell aufgenommen, wenn die Strafe geringer ausfällt als die Verbesserung des Fits.

Durch die Wahl der Straffunktion (und damit des Kriteriums) legt man fest, wie dieser Trade-off quantitativ vorgenommen wird. Gängig sind drei verschiedene Kriterien: AIC, HQ und SC/BIC, siehe unten.

- **Regel:** Unter allen in Erwägung gezogenen Kandidaten wird die Spezifikation gewählt, für die das Kriterium den *kleinsten* Wert annimmt.
- Es ist ratsam, jeweils AIC, HQ und SC/BIC zu prüfen. In günstigen Fällen liefern alle Kriterien dasselbe Resultat. Beachte, dass SC für Stichprobengrößen ($n > 8$) zusätzliche Parameter stärker bestraft als HQ, und HQ wiederum stärker als AIC.
- Es ist möglich, mit Hilfe von Selektionskriterien auch aus nicht-geschachtelten Modellen

auszuwählen, solange die abhängige Variable identisch ist, siehe empirisches Beispiel in Abschnitt 10.3.

- Zur Information: **Log-Likelihood-Funktion**, genauer konzentrierte Log-Likelihood-Funktion

$$l(\hat{\beta}, \hat{\sigma} | \mathbf{y}, \mathbf{X}) = -\frac{n}{2} (1 + \ln(2\pi)) - \frac{n}{2} \ln \hat{\sigma}^2 \quad (10.3)$$

vgl. zur Erklärung und Ableitung Davidson & MacKinnon (2004, Equation (10.12)) oder Veranstaltung Fortgeschrittene Ökonometrie, Abschnitt 5.5

- **Alternative Definitionen von Modellselektionskriterien:**

Kriterium Fitmaß Zahl Par. Straffunktion(n)

$$AIC = \ln \hat{\sigma}^2 + k \cdot \frac{2}{n}, \quad (10.4)$$

$$HQ = \ln \hat{\sigma}^2 + k \cdot \frac{\ln(\ln(n))}{n}, \quad (10.5)$$

$$SC = \ln \hat{\sigma}^2 + k \cdot \frac{\ln(n)}{n}, \quad (10.6)$$

$$AIC = -\frac{2}{n} \left(\underbrace{-\frac{n}{2} (1 + \ln(2\pi)) - \frac{n}{2} \ln \hat{\sigma}^2}_{=\text{Log-Likelihood-Funktion}} \right) + k \cdot \frac{2}{n} \quad (10.7)$$

$$HQ = -\frac{2}{n} \left(-\frac{n}{2} (1 + \ln(2\pi)) - \frac{n}{2} \ln \hat{\sigma}^2 \right) + k \cdot \frac{\ln(\ln(n))}{n} \quad (10.8)$$

$$SC = -\frac{2}{n} \left(-\frac{n}{2} (1 + \ln(2\pi)) - \frac{n}{2} \ln \hat{\sigma}^2 \right) + k \cdot \frac{\ln(n)}{n} \quad (10.9)$$

In R berechnet der Befehl `AIC()` Modellselektionskriterien, die sich in zweifacher Hinsicht von obigen Berechnungen unterscheiden:

- Es wird nicht durch n dividiert.
- Es wird zusätzlich zu den geschätzten Parametern in β auch die Varianz als geschätzter Parameter hinzugenommen.

Für einen Vergleich verschiedener Modelle für gegebenes n spielt dies keine Rolle.

Fitmaß Zahl Par. Straffunktion(n)

$$AIC = -2 \left(-\frac{n}{2} (1 + \ln(2\pi)) - \frac{n}{2} \ln \hat{\sigma}^2 \right) + (k + 1) \cdot 2 \quad (10.10)$$

$$HQ = -2 \left(-\frac{n}{2} (1 + \ln(2\pi)) - \frac{n}{2} \ln \hat{\sigma}^2 \right) + (k + 1) \cdot \ln(\ln(n)) \quad (10.11)$$

$$SC = -2 \left(-\frac{n}{2} (1 + \ln(2\pi)) - \frac{n}{2} \ln \hat{\sigma}^2 \right) + (k + 1) \cdot \ln(n) \quad (10.12)$$

Es gibt auch Definitionen, in denen die Modellselektionskriterien maximiert werden, z. B. in [Davidson & MacKinnon \(2004, Section 15.4\)](#). Also immer die genauen Definitionen in der verwendeten Software beachten!

Formel	Software - Befehl
Akaike Information Criterion (AIC)	
(10.4) R: <code>extractAIC()</code>	
(10.7) EViews, R: eigenes Programm <code>SelectCritEViews()</code> , siehe Abschnitt B.2	
(10.10) R: <code>AIC()</code>	
Hannan-Quinn (HQ)	
(10.5) R: <code>extractAIC(,k = log(log(n)))</code>	
(10.8) EViews, R: eigenes Programm <code>SelectCritEViews()</code> , siehe Abschnitt B.2	
(10.11) R: <code>AIC(,k = log(log(n)))</code>	
Bayesian Information Criterion (BIC)/Schwarz Kriterium (SC)	
(10.6) R: <code>extractAIC(,k = log(n))</code>	
(10.9) EViews, R: eigenes Programm <code>SelectCritEViews()</code> , siehe Abschnitt B.2	
(10.12) R: <code>AIC(,k = log(n))</code>	

- Alternative zur Verwendung von Modellselektionskriterien: Sequentielles Testen. Hierzu sind t -Tests oder F -Tests notwendig, die in Kapitel [11](#) behandelt werden.
- ‡ Der Vergleich zweier Modelle mittels eines Modellselektionskriteriums kann auch als Test interpretiert werden, wobei das Signifikanzniveau durch den Strafterm bestimmt wird.

10.2. Tests für nichtgeschachtelte Modelle

Siehe Abschnitt 9.3.1 in Kursmaterial zu BA-Modul **Einführung in die Ökonometrie** oder [Wooldridge \(2009, Chapter 9\)](#) oder [Davidson & MacKinnon \(2004, Section 15.3\)](#).

Dort werden behandelt:

- **Encompassing-Test**, R-Befehl: `encomptest(model_1,model_2)` (erfordert R-Paket `lmtest`)
- **J-Test**, R-Befehl: `jtest(model_1,model_2)` (erfordert R-Paket `lmtest`)

10.3. Empirische Analyse von Handelsströmen: Teil 2

Fortsetzung der **Empirische Analyse von Handelsströmen: Teil 1** in Abschnitt [6.3](#).

Zu Schritt II.3: Spezifizieren, Schätzen und Auswählen eines ökonomischen Modells

- **Spezifizieren und Schätzen** verschiedener Modelle: Es werden jetzt fünf verschiedene Modelle spezifiziert und geschätzt: Gleich in R-Formeln:

Model 1

```
mod_1_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o)
```

Model 2

```
mod_2_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist)
```

Model 3a

```
mod_3a_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist)
+ ebrd_tfes_o
```

Model 3b

```
mod_3b_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist)
+ log(cepii_area_o)
```

Model 4

```
mod_4_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist)
+ ebrd_tfes_o + log(cepii_area_o)
```

Berechnen der Modelle via

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# Berechnen der Werte der Tabelle.
# Anwenden der Funktion "SelectCritEviews" auf vier verschiedene Modelle

mod_1_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o))
summary(mod_1_kq)
deviance(mod_1_kq) # Berechnet SSR
SelectCritEviews(mod_1_kq) # Berechnet AIC, HQ, SC

mod_2_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist))
summary(mod_2_kq)
deviance(mod_2_kq) # Berechnet SSR
SelectCritEviews(mod_2_kq) # Berechnet AIC, HQ, SC

mod_3a_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o)
summary(mod_3a_kq)
deviance(mod_3a_kq) # Berechnet SSR
SelectCritEviews(mod_3a_kq) # Berechnet AIC, HQ, SC

mod_3b_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  log(cepii_area_o))
summary(mod_3b_kq)
deviance(mod_3b_kq) # Berechnet SSR
SelectCritEviews(mod_3b_kq) # Berechnet AIC, HQ, SC

mod_4_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o + log(cepii_area_o))
```

```
summary(mod_4_kq)
deviance(mod_4_kq)          # Berechnet SSR
SelectCritEviews(mod_4_kq)  # Berechnet AIC, HQ, SC
```

Listing 10.1: ./R_code/4_ff_Beispiel_Handelsstroeme.R

liefert Output, aus dem folgende Tabelle erzeugt werden kann:

Abhängige Variable: $\ln(\text{Importe nach Deutschland})$					
Unabhängige Variable/Modell	(1)	(2)	(3a)	(3b)	(4)
Konstante	-5.77 (2.184)	4.676 (2.178)	2.741 (2.175)	3.409 (2.098)	2.427 (2.132)
$\ln(BIP)$	1.077 (0.087)	0.975 (0.063)	0.940 (0.0613)	1.080 (0.071)	1.025 (0.076)
$\ln(\text{Entfernung})$	—	-1.074 (0.156)	-0.970 (0.152)	-915 (0.159)	-0.888 (0.156)
<i>Offenheit</i>	—	—	0.507 (0.191)	—	0.353 (0.206)
$\ln(\text{Flaeche})$	—	—	—	-0.213 (0.089)	-0.151 (0.085)
Stichprobengröße	49	49	49	49	49
R^2	0.765	0.883	0.900	0.900	0.906
Standardfehler der Regression	1.304	0.928	0.873	0.871	0.853
Residuenquadratsumme	80.027	39.644	34.302	34.148	32.017
AIC	3.4100	2.7484	2.6445	2.6400	2.6164
HQ	3.4393	2.7924	2.7031	2.6986	2.6896
SC	3.4872	2.8642	2.7989	2.7945	2.8094

- **Modellwahl:** Aus der Tabelle ist ersichtlich, dass Modell 4 zu wählen ist, wenn das Akaike-Kriterium (AIC) oder das Hannan-Quinn (HQ) gewählt werden, jedoch Modell 3b, wenn das Schwarz (SC)-Kriterium gewählt wird.

Fortsetzung der **Empirische Analyse von Handelsströmen: Teil 3** in Abschnitt [11.7](#).

11. (Asymptotische) Verteilung des KQ-Schätzers und Testen im multiplen linearen Regressionsmodell

11.1. Exakte Verteilung des KQ-Schätzers

- Mit bisherigen Annahmen gilt für den KQ-Schätzer

$$\hat{\beta}_n \stackrel{(B3)}{=} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \stackrel{(B1)}{=} \beta_0 + (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}.$$

- Ohne eine Verteilungsannahme für den Fehlervektor \mathbf{u} lässt sich offensichtlich nichts weiter über die Verteilung von $\hat{\beta}_n$ sagen, selbst wenn die \mathbf{X} gegeben sind.

Wir treffen die Annahme (vgl. zur Schreibweise Davidson (2000, Section 2.4.1))

(B4) Multivariat normalverteilte Fehler gegeben \mathbf{X}

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

wobei für die Fehlervarianz des DGPs $\sigma^2 = \sigma_0^2$ gilt.

Die gemeinsame (auf \mathbf{X} bedingte) Dichte lautet (vgl. (2.32))

$$f(u_1, u_2, \dots, u_n | \mathbf{X}; \sigma^2) = f(\mathbf{u} | \mathbf{X}; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2} \mathbf{u}^T \mathbf{u}\right). \quad (11.1)$$

- Wendet man (2.31) auf $\hat{\beta}_n$ an, erhält man aufgrund von Annahme (B4), sowie den bisherigen Annahmen (B2a), (B2b), dass für jede(!) Stichprobengröße n

$$\hat{\beta}_n | \mathbf{X} \sim N\left(\beta_0, \sigma_0^2 (\mathbf{X}^T \mathbf{X})^{-1}\right), \quad (11.2)$$

also der **KQ-Schätzer gegeben \mathbf{X} exakt multivariat normalverteilt** ist.

- Wendet man (2.31) auf $\mathbf{y} = \mathbf{X}\beta_0 + \mathbf{u}$ an, erhält man

$$\mathbf{y} | \mathbf{X} \sim N\left(\mathbf{X}\beta_0, \sigma_0^2 \mathbf{I}\right) \iff y_t | \mathbf{X} \sim NID\left(\mathbf{X}_t \beta_0, \sigma_0^2\right), t = 1, \dots, n. \quad (11.3)$$

Für beliebige Parameter erhält man das **normale multiple lineare Regressionsmodell**

$$y_t | \mathbf{X}_t \sim NID(x_{t1}\beta_1 + x_{t2}\beta_2 + \dots + x_{tk}\beta_k, \sigma^2), \quad \beta_1, \dots, \beta_k \in \mathbb{R}, \sigma^2 \in \mathbb{R}^+. \quad (11.4)$$

- Beachte, dass eine einfache exakte Verteilung wie (11.2) nur unter der multivariaten Normalverteilungsannahme möglich ist. Wieso?

- **Zusammenfassung der Annahmen des normalen multiplen linearen Regressionsmodells**

- **(B1)** Korrekt spezifiziertes Modell: Der DGP ist für $\beta = \beta_0$ im multiplen linearen Regressionsmodell enthalten.
- **(B3)** \mathbf{X} hat vollen Spaltenrang und
- **(B4)** $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$.

Beachte, dass die Annahme **(B4)** die Annahme **(B2)** enthält.

- Liegt eine von der Normalverteilung verschiedene bedingte Verteilung für den Fehlervektor \mathbf{u} vor, lässt sich die exakte Verteilung des KQ-Schätzers im Allgemeinen nur mit Hilfe von Simulationsmethoden bestimmen.
- Weiß man nichts über die Art der bedingten Verteilung der Fehler, dann ist die exakte Verteilung für endliche n unbekannt, also $\hat{\beta}_n|\mathbf{X} \sim \text{unbekannte Verteilung}$. Wie im folgenden gezeigt wird, ist es unter bestimmten Voraussetzungen jedoch möglich, die asymptotische Verteilung zu bestimmen.

11.2. Asymptotische Verteilung des KQ-Schätzers

- **Ableitung**

- Wie im Fall des Schätzers des Erwartungswertes muss man auch den KQ-Schätzer mit \sqrt{n} multiplizieren, um eine nicht singuläre asymptotische Varianz-Kovarianzmatrix zu erhalten. Man erhält unter den Annahmen **(B1)** und **(B3)**

$$\sqrt{n}(\hat{\beta}_n - \beta_0) = \underbrace{\sqrt{n}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T}_{:=\mathbf{A}_n} \underbrace{\mathbf{u}}_{:=\mathbf{a}_n} = \left(\frac{\mathbf{X}^T\mathbf{X}}{n}\right)^{-1} \frac{\mathbf{X}^T\mathbf{u}}{\sqrt{n}}.$$

- Jetzt muss man Slutsky's Theorem (3.4) aus Abschnitt 3.5 anwenden: Falls

i) $\mathbf{A}_n \xrightarrow{P} \mathbf{A}$ und

ii) $\mathbf{a}_n \xrightarrow{d} \mathbf{a}$ gilt,

dann gilt $\mathbf{A}_n\mathbf{a}_n \xrightarrow{d} \mathbf{A}\mathbf{a}$.

- Damit i) gilt, muss weiterhin **(A1)** gelten, so dass

$$\text{plim}_{n \rightarrow \infty} \left(\mathbf{X}^T\mathbf{X}/n\right)^{-1} = \mathbf{S}_{\mathbf{X}^T\mathbf{X}}^{-1}$$

gilt.

- Damit ii) gilt, muss Annahme **(A2)** “verstärkt” werden. Nunmehr muss ein Zentraler Grenzwertsatz für $\mathbf{X}^T \mathbf{u} / \sqrt{n}$ gelten:

$$\text{(A3)} \quad \frac{1}{\sqrt{n}} \mathbf{X}^T \mathbf{u} \xrightarrow{d} \mathbf{w}_\infty \sim N(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}})$$

- **Asymptotische Verteilung des KQ-Schätzers** Es gelten für das multiple lineare Regressionsmodell die Annahmen **(B1)**, **(B3)**, sowie die Annahmen **(A1)** und **(A3)**. Dann gilt

$$\begin{aligned} \sqrt{n} (\hat{\beta}_n - \beta_0) &= \left(\frac{\mathbf{X}^T \mathbf{X}}{n} \right)^{-1} \frac{1}{\sqrt{n}} \mathbf{X}^T \mathbf{u} \\ &\xrightarrow{d} \mathbf{S}_{\mathbf{X}^T \mathbf{X}}^{-1} \mathbf{w}_\infty \sim N(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}}^{-1}). \end{aligned} \quad (11.5)$$

- In der Praxis können die Annahmen **(A1)** und **(A3)** (high level assumptions) nicht direkt überprüft werden. Deshalb werden diese Annahmen im Allgemeinen durch Annahmen ersetzt, die anschaulicher und leichter überprüfbar sind. Dazu in Kürze mehr.
- **Anwendung der asymptotischen Verteilung in der Praxis:**

- In heuristischer Schreibweise lässt sich die asymptotische Verteilung auch schreiben als

$$\hat{\beta}_n \overset{\text{approximativ}}{\sim} N\left(\beta_0, \frac{\sigma_0^2}{n} \mathbf{S}_{\mathbf{X}^T \mathbf{X}}^{-1}\right),$$

da sich für gegebene Stichprobengröße n herauskürzt.

- Da $\mathbf{S}_{\mathbf{X}^T \mathbf{X}}$ und σ_0^2 unbekannt sind, ist die asymptotische Verteilung so nicht anwendbar. Die Fehlervarianz σ_0^2 kann mit s^2 geschätzt werden und $\mathbf{S}_{\mathbf{X}^T \mathbf{X}}$ durch

$$\frac{1}{n} \mathbf{X}^T \mathbf{X} = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t. \quad (11.6)$$

Damit erhält man in heuristischer Schreibweise

$$\hat{\beta}_n \overset{\text{approximativ}}{\sim} N\left(\beta_0, s^2 (\mathbf{X}^T \mathbf{X})^{-1}\right).$$

Der **zentrale Unterschied** zur exakten Verteilung ist, dass die Normalverteilung nur approximativ gilt, jedoch die Approximation mit zunehmender Stichprobengröße n immer genauer wird.

- Möchte man analysieren, wie gut die Approximation der asymptotischen Normalverteilung ist, muss man dies im Allgemeinen mit Hilfe von Computersimulationen, sogenannten **Monte-Carlo-Simulationen** machen.
- **Wann ist Annahme (A3) erfüllt?**

Zum Beispiel, wenn eine **Zufallsstichprobe** vorliegt und Annahme **(B2)** gilt. Diese Annahmen können abgeschwächt werden, siehe Abschnitt 13.4.

Beweisskizze:

– Es gilt $\mathbf{X}^T \mathbf{u} = \sum_{t=1}^n \underbrace{\mathbf{X}_t^T u_t}_{:=\mathbf{v}_t}$. Zunächst werden $E[\mathbf{v}_t]$ und $Var(\mathbf{v}_t)$ bestimmt.

– Aus Annahme **(B2a)** $E[\mathbf{u}|\mathbf{X}] = \mathbf{0}$ (strenge Exogenität) folgt, dass

$$\begin{aligned} E[u_t|\mathbf{X}] &= 0 \quad \text{für alle } t = 1, 2, \dots, n. \\ E[E[u_t|\mathbf{X}|\mathbf{X}_t]] &= E[u_t|\mathbf{X}_t] = 0. \\ E[\mathbf{X}_t^T u_t] &= E[E[\mathbf{X}_t^T u_t|\mathbf{X}_t]] = \mathbf{0}. \end{aligned}$$

Somit ist der Erwartungswert von $\mathbf{v}_t = \mathbf{X}_t^T u_t$ ein Nullvektor.

– Wegen Annahme **(B2b)** gilt $Var(\mathbf{u}|\mathbf{X}) = \sigma_0^2 \mathbf{I}$, sowie

$$\begin{aligned} Var(u_t|\mathbf{X}) &= E[u_t^2|\mathbf{X}] = \sigma_0^2 \quad \text{für alle } t = 1, 2, \dots, n. \\ E[E[u_t^2|\mathbf{X}|\mathbf{X}_t]] &= E[u_t^2|\mathbf{X}_t] = Var(u_t|\mathbf{X}_t) = \sigma_0^2. \\ Var(\mathbf{v}_t) &= Var(\mathbf{X}_t^T u_t) = E[\mathbf{X}_t^T u_t^2 \mathbf{X}_t] = E[E[u_t^2 \mathbf{X}_t^T \mathbf{X}_t|\mathbf{X}_t]] = \sigma_0^2 E[\mathbf{X}_t^T \mathbf{X}_t]. \end{aligned}$$

Da $\mathbf{v}_t \sim (\mathbf{0}, Var(\mathbf{v}_t))$ und damit $\mathbf{X}_t^T u_t \sim (\mathbf{0}, Var(\mathbf{X}_t^T u_t))$ gilt, sowie eine Zufallsstichprobe angenommen wurde, kann auf den Schätzer des Erwartungswertes

$$\hat{\boldsymbol{\mu}}_{\mathbf{v},n} = \frac{1}{n} \mathbf{X}^T \mathbf{u} = \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T u_t$$

der Multivariate Zentrale Grenzwertsatz (5.73) angewendet werden. Man erhält

$$\sqrt{n} \hat{\boldsymbol{\mu}}_{\mathbf{v},n} \xrightarrow{d} N\left(\mathbf{0}, \sigma_0^2 \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E[\mathbf{X}_t^T \mathbf{X}_t]\right).$$

Es lässt sich zeigen, dass aufgrund von Annahme **(A1)** gilt:

$$\mathbf{S}_{\mathbf{X}^T \mathbf{X}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E[\mathbf{X}_t^T \mathbf{X}_t].$$

Damit erhält man

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{X}_t^T u_t \xrightarrow{d} N(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}}). \quad (11.7)$$

– ‡ Verwendung des Cramér-Wold Devices: Wähle beliebigen $(k \times 1)$ -Vektor $\boldsymbol{\lambda}$. Mit den bisherigen Ergebnissen gilt

$$\boldsymbol{\lambda}^T \mathbf{X}_t^T u_t \sim (0, \sigma_0^2 \boldsymbol{\lambda}^T E[\mathbf{X}_t^T \mathbf{X}_t] \boldsymbol{\lambda}).$$

Man betrachtet dann die asymptotischen Eigenschaften des Schätzers des Erwartungswertes

$$\hat{v}_n = \frac{1}{n} \sum_{t=1}^n \boldsymbol{\lambda}^T \mathbf{X}_t^T u_t$$

(= (skalare) Zufallsfolge). Unter der **zusätzlichen Voraussetzung**, dass die **Stichprobenbeobachtungen stochastisch unabhängig** sind und den üblichen Regularitätsbedingungen, lässt sich der Zentrale Grenzwertsatz für heterogene, aber unabhängige Zufallsvariable (5.72) anwenden und es gilt

$$\sqrt{n}\hat{\nu}_n \xrightarrow{d} N\left(0, \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sigma_0^2 \boldsymbol{\lambda}^T E[\mathbf{X}_t^T \mathbf{X}_t] \boldsymbol{\lambda}\right).$$

Da dies für alle $\boldsymbol{\lambda}$ mit $\|\boldsymbol{\lambda}\| > 0$ gilt, kann man aufgrund des Cramér-Wold Devices $\boldsymbol{\lambda}$ weglassen und man erhält

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{X}_t^T u_t \xrightarrow{d} N\left(0, \sigma_0^2 \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E[\mathbf{X}_t^T \mathbf{X}_t]\right)$$

bzw. wieder

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \mathbf{X}_t^T u_t \xrightarrow{d} N\left(0, \sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}}\right). \quad (11.7)$$

□

R-Befehle

Berechnen der Varianz-Kovarianzmatrix von zwei Variablen mit `cov()`. Umrechnen der Varianz-Kovarianzmatrix in eine **Korrelationsmatrix** mit `cov2cor()`.

11.3. Exakte Tests

Anwendungen von exakten Tests:

- **Spezifikation des normalen linearen Regressionsmodells und Überprüfen der Annahmen**, vgl. Abschnitt 11.1
 - **(B1)** und $E[\mathbf{u}|\mathbf{X}] = \mathbf{0}$ (**(B2a)**): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ enthält DGP
 - * t -Tests, siehe Abschnitt 11.3.1; F -Tests, siehe Abschnitt 11.3.2.
 - * Testen der korrekten funktionalen Form, z. B. mit **RESET-Test**, siehe Abschnitt 15.3.
 - * Testen auf Parameterstabilität, z. B. mit **Chow-Test**, siehe (11.34) in Abschnitte 11.3.2.
 - **(B3)**: $\mathbf{X}^T \mathbf{X}$ hat Rang k : Verletzung führt zu Fehlermeldung „singuläre Matrix“.
 - **(B4)**: $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$:
 - * Setzt $E[\mathbf{u}|\mathbf{X}] = \mathbf{0}$ voraus, siehe oben.

- * Setzt voraus: $\text{Var}(\mathbf{u}|\mathbf{X}_t) = \sigma^2$ (**Homoskedastie**): **Tests auf Heteroskedastie**, siehe Abschnitt 15.2.
- * Erfordert normalverteilte Fehler: **Lomnicki-Jarque-Bera-Test**, siehe Abschnitt 15.4.

• Überprüfen ökonomischer Hypothesen

11.3.1. t -Tests: Testen einer einzelnen Restriktion

- Der zu testende Parameter wird als β_2 bezeichnet. Das normale multiple lineare Regressionsmodell lautet dann:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{x}_2\beta_2 + \mathbf{u}, \quad \mathbf{u}|\mathbf{X}_1, \mathbf{x}_2 \sim N(\mathbf{0}, \sigma^2\mathbf{I}). \quad (11.8)$$

- Hypothesenpaar: $H_0 : \beta_2 = \beta_{2,H_0}$ versus $H_1 : \beta_2 \neq \beta_{2,H_0}$
- t -Test bei bekannter Fehlervarianz σ_0^2 :

– **Teststatistik:**

$$z_{\beta_2} = \frac{\hat{\beta}_2 - \beta_{2,H_0}}{\sigma_{\hat{\beta}_2}}. \quad (11.9)$$

– **Exakte Verteilung:** Unter den Annahmen **(B1)**, **(B3)**, **(B4)** und bekannter Fehlervarianz gilt unter H_0 :

$$z_{\beta_2}|\mathbf{X} \sim N(0, 1). \quad (11.10)$$

Die Testverteilung ist unter H_0 vollständig bekannt.

Beweis:

Überblick über Vorgehensweise (Die Vorgehensweise ist analog zur Ableitung des Tests bzgl. des Erwartungswertes (5.80))

1. Mit Hilfe des Frisch-Waugh-Lovell Theorems, Seite 164, lässt sich die Teststatistik z_{β_2} als Linearkombination normalverteilter Fehler schreiben.
2. Da eine Linearkombination multivariat normalverteilter Zufallsvariablen wieder normalverteilt ist, ist die Teststatistik z_{β_2} normalverteilt.
3. Die Standardisierung in (11.9) wurde so gewählt, dass unter H_0 (11.10) gilt

Die Schritte im Einzelnen

1. Berechnung der Teststatistik: Anwenden des FWL-Theorems auf β_2 in $\mathbf{M}_1\mathbf{y} = \mathbf{M}_1\mathbf{x}_2\beta_2 + \mathbf{M}_1\mathbf{u}$ ergibt

$$\hat{\beta}_2 = \frac{\mathbf{x}_2^T \mathbf{M}_1 \mathbf{y}}{\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2}, \quad \sigma_{\hat{\beta}_2}^2 = \sigma_0^2 (\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)^{-1}$$

Einsetzen in (11.9) ergibt

$$z_{\beta_2} = \frac{\frac{\mathbf{x}_2^T \mathbf{M}_1 \mathbf{y}}{\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2} - \beta_{2,H_0}}{\sigma_0 (\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)^{-1/2}}.$$

Einsetzen von (11.8) mit $\beta_2 = \beta_{2,H_0}$, da H_0 vorausgesetzt wird, liefert für $z_{\hat{\beta}_2}$ eine Linearkombination von \mathbf{u}

$$z_{\beta_2} = \frac{\mathbf{x}_2^T \mathbf{M}_1 \mathbf{u}}{\sigma_0 (\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)^{1/2}} = \mathbf{B} \mathbf{u}.$$

2. und 3. Ableitung der Verteilung: Aufgrund von (2.33) erhält man

$$z_{\beta_2} | \mathbf{X} \sim N(0, 1)$$

da $E[z_{\beta_2} | \mathbf{X}] = E[\mathbf{B} \mathbf{u} | \mathbf{X}] = 0$ und

$$\text{Var}(z_{\beta_2} | \mathbf{X}) = \text{Var}(\mathbf{B} \text{Var}(\mathbf{u} | \mathbf{X}) \mathbf{B}^T | \mathbf{X}) = \frac{E(\mathbf{x}_2^T \mathbf{M}_1 \mathbf{u} \mathbf{u}^T \mathbf{M}_1 \mathbf{x}_2 | \mathbf{X}_1, \mathbf{x}_2)}{\sigma_0^2 (\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)} = \frac{\sigma_0^2 (\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)}{\sigma_0^2 (\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)} = 1.$$

□

– Gilt H_1 , so ist die Teststatistik ebenfalls normalverteilt, aber mit einem Mittelwert, der von Null verschieden ist. Vgl. Berechnung der Powerfunktion in Abschnitt 5.6.

• **t-Test bei geschätzter Fehlervarianz $\hat{\sigma}^2$:**

– **Idee:** (Vgl. Ableitung von (5.74) in Abschnitt 5.6) Man ersetzt in Teststatistik (11.9) σ durch s . Damit ergibt sich für $\sigma_{\hat{\beta}_2}$ der Schätzer

$$s_{\hat{\beta}_2}^2 = s^2 (\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)^{-1} = \frac{\mathbf{y}^T \mathbf{M}_{\mathbf{X}} \mathbf{y}}{n - k} (\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)^{-1}.$$

– **Teststatistik:**

$$t_{\beta_2} = \frac{\hat{\beta}_2 - \beta_{2,H_0}}{s_{\hat{\beta}_2}}. \quad (11.11)$$

– **Exakte Verteilung:** Unter den Annahmen (B1), (B3), (B4) und bekannter Fehlervarianz gilt unter H_0 :

$$t_{\beta_2} | \mathbf{X} \sim t_{n-k}. \quad (11.12)$$

Die Testverteilung ist unter H_0 vollständig bekannt.

Beweis:

Überblick über Vorgehensweise

1. Umformulieren der Teststatistik als Quotient (11.13) der Teststatistik z_{β_2} und einer Zufallsvariable, für die in Schritt 2 die χ^2 -Verteilung gezeigt wird.

2. Zeigen, dass Nenner in (11.13) χ^2 -verteilt ist.
3. Zeigen, dass in (11.13) die normalverteilte Zufallsvariable im Zähler und die χ^2 -verteilte Zufallsvariable im Nenner stochastisch unabhängig sind
4. Gemäß (2.36) gilt dann die t -Verteilung.

Die Schritte im Einzelnen:

1. Berechnung: $\hat{\beta}_2$ bleibt gleich und die Varianz des Parameterschätzers $\sigma_{\hat{\beta}_2}^2$ wird durch $s_{\hat{\beta}_2}^2$ geschätzt, so dass sich unter H_0 ergibt:

$$\begin{aligned} t_{\beta_2} &= \left(\underbrace{\frac{\mathbf{y}^T \mathbf{M}_X \mathbf{y}}{(n-k)}}_{s^2} \right)^{-1/2} \frac{\mathbf{x}_2^T \mathbf{M}_1 \mathbf{u}}{(\mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)^{1/2}} \\ &= \left(\frac{\mathbf{y}^T \mathbf{M}_X \mathbf{y}}{\sigma_0^2 (n-k)} \right)^{-1/2} \frac{\mathbf{x}_2^T \mathbf{M}_1 \mathbf{u}}{(\sigma_0^2 \mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)^{1/2}} = \frac{z_{\beta_2}}{\left(\frac{s^2}{\sigma_0^2} \right)^{1/2}}. \end{aligned} \quad (11.13)$$

2. Ableitung der Verteilung der Zufallsvariablen im Nenner:

Es gilt $\frac{\mathbf{y}^T}{\sigma_0} \mathbf{M}_X \frac{\mathbf{y}}{\sigma_0} = \frac{\mathbf{u}^T}{\sigma_0} \mathbf{M}_X \frac{\mathbf{u}}{\sigma_0} = \frac{(n-k)s^2}{\sigma_0^2} \sim \chi^2(n-k)$, da $\mathbf{u}/\sigma_0 \sim N(\mathbf{0}, \mathbf{I})$ und in dem Ausdruck $\frac{\mathbf{u}^T}{\sigma_0} \mathbf{M}_X \frac{\mathbf{u}}{\sigma_0}$ die Projektionsmatrix \mathbf{M}_X gerade Rang $n-k$ hat. Damit ergibt sich aufgrund von (2.35) eine χ^2 -Verteilung mit $n-k$ Freiheitsgraden.

3. Stochastische Unabhängigkeit von Zähler und Nenner

– Zähler:

$$\mathbf{x}_2^T \mathbf{M}_1 \mathbf{y} = \mathbf{x}_2^T \mathbf{P}_X \mathbf{M}_1 \mathbf{y} = \mathbf{x}_2^T \mathbf{M}_1 \mathbf{P}_X \mathbf{y}$$

da \mathbf{x}_2 bereits im Unterraum von \mathbf{P}_X liegt und

$$\mathbf{P}_X \underbrace{(\mathbf{I} - \mathbf{P}_1)}_{\mathbf{M}_1} = \mathbf{P}_X - \mathbf{P}_X \mathbf{P}_1 = \mathbf{P}_X - \mathbf{P}_1 \mathbf{P}_X = \mathbf{M}_1 \mathbf{P}_X$$

gilt. Zusammen mit $\mathbf{P}_X \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{P}_X \mathbf{u}$ ergibt sich für den Zähler

$$\mathbf{x}_2^T \mathbf{M}_1 \mathbf{y} = \mathbf{x}_2^T \mathbf{M}_1 \mathbf{X}\boldsymbol{\beta} + \mathbf{x}_2^T \mathbf{M}_1 \mathbf{P}_X \mathbf{u},$$

dass dieser gegeben \mathbf{X} ausschließlich vom Zufallsvektor $\mathbf{P}_X \mathbf{u}$ abhängt.

- Nenner: basiert auf der Wurzel aus der quadratischen Form von $\mathbf{M}_X \mathbf{u}/\sigma_0$
- Gegeben \mathbf{X} sind die Zufallsvektoren im Zähler $\mathbf{P}_X \mathbf{u}$ und im Nenner $\mathbf{M}_X \mathbf{u}$. Deren Kovarianz ist Null, da

$$E(\mathbf{P}_X \mathbf{u} \mathbf{u}^T \mathbf{M}_X | \mathbf{X}_1, \mathbf{x}_2) = \mathbf{P}_X \sigma_0^2 \mathbf{I} \mathbf{M}_X = \sigma_0^2 \mathbf{P}_X \mathbf{M}_X = \mathbf{0},$$

da die jeweiligen Unterräume orthogonal zueinander stehen.

- Da $\mathbf{P}_\mathbf{X}\mathbf{u}$ und $\mathbf{M}_\mathbf{X}\mathbf{u}$ beide auf Basis *desselben* Vektors \mathbf{u} multivariat normalverteilt sind, ergibt sich aus der Unkorreliertheit Unabhängigkeit (vgl. Davidson (2000, Theorem C.4.1, S. 466)).

4. Gültigkeit der t -Verteilung:

Damit ist die t -Statistik (11.11) gemäß (2.36) unter H_0 exakt t -verteilt mit $n - k$ Freiheitsgraden, da Zähler und Nenner stochastisch unabhängig sind, der Zähler standardnormalverteilt ist, sowie im Nenner $\frac{\mathbf{y}^T}{\sigma_0} \mathbf{M}_\mathbf{X} \frac{\mathbf{y}}{\sigma_0}$ gerade $\chi^2(n - k)$ verteilt ist und nach Division durch die Zahl der Freiheitsgrade gerade s^2/σ_0^2 ergibt:

$$t_{\beta_2} | \mathbf{X} = \frac{\hat{\beta}_2 - \beta_{2,H_0}}{s \hat{\beta}_2} | \mathbf{X} \sim t_{n-k}. \quad (11.14)$$

□

- Mit dem t -Test können auch kompliziertere einzelne Restriktionen getestet werden.

Skalenelastizität einer Cobb-Douglas Produktionsfunktion:

$$\log Y = \beta_1 + \beta_2 \log K + \beta_3 \log L + u$$

wobei Y , K und L jeweils Output, Kapital und Arbeit bezeichnen. Die Null- bzw. Alternativhypothese einer linearen Skalenelastizität

$$H_0 : \beta_2 + \beta_3 = 1 \quad \text{versus} \quad H_1 : \beta_2 + \beta_3 \neq 1$$

lassen sich mit $\theta = \beta_2 + \beta_3$ schreiben als

$$H_0 : \theta = 1 \quad \text{versus} \quad H_1 : \theta \neq 1,$$

wobei dann mit $\beta_3 = \theta - \beta_2$

$$\log Y = \beta_1 + \beta_2(\log K - \log L) + \theta \log L + u$$

geschätzt wird. Alternativ kann auch ein F -Test durchgeführt werden.

11.3.2. F -Tests: Testen mehrerer Restriktionen

Häufig impliziert eine (ökonomische) Theorie mehrere Restriktionen bezüglich der Parameter eines Regressionsmodells.

- Beispiele möglicher linearer Restriktionen:

1. $H_0 : \beta_2 = \beta_k$

2. $H_0 : \beta_1 = 1, \beta_k = 0$

3. $H_0 : \beta_1 = \beta_3, \beta_2 = \beta_3$
 4. $H_0 : \beta_j = 0, j = 2, \dots, k$
 5. $H_0 : \beta_j + 2\beta_{j+1} = 1, \beta_k = 2.$
- Alle $q \leq k$ linearen Restriktionen können in folgender Form dargestellt werden:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \quad \text{vs.} \quad H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r} \quad (11.15)$$

wobei die $(q \times k)$ -Matrix \mathbf{R} und der $(q \times 1)$ -Vektor \mathbf{r} gegeben und fest sind. Bei der Formulierung muss natürlich sichergestellt werden, dass alle Restriktionen in (11.15) widerspruchsfrei und nicht redundant sind.

Darstellungen der Beispiele:

1. $H_0 : \beta_2 = \beta_k \Leftrightarrow \beta_2 - \beta_k = 0$:

$$(0 \quad 1 \quad 0 \quad \dots \quad 0 \quad -1) \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_{k-1} \\ \beta_k \end{pmatrix} = 0.$$

2. $H_0 : \beta_1 = 1, \beta_k = 0$:

$$\begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

3. $H_0 : \beta_1 = \beta_3, \beta_2 = \beta_3$:

$$\begin{pmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

4. $H_0 : \beta_j = 0, j = 2, \dots, k$:

$$\underbrace{\begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}}_{(\mathbf{0} \quad \mathbf{I}_{k-1})} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \vdots \\ \beta_k \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}}_{((k-1) \times 1)}.$$

5. $H_0 : \beta_j + 2\beta_{j+1} = 1, \beta_k = 2:$

$$\begin{pmatrix} 0 & \cdots & 1 & 2 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_j \\ \beta_{j+1} \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}.$$

Fortsetzung Handelsströme: Betrachte Regressionsmodell

$$\begin{aligned} \ln(\text{Importe}_i) &= \beta_1 + \beta_2 \ln(\text{BIP}_i) + \beta_3 \ln(\text{Entfernung}_i) \\ &+ \beta_4 \text{Offenheit}_i + \beta_5 \ln(\text{Fläche}) + u_i. \end{aligned}$$

Frage: Spielen die Variablen **Offenheit** und **Fläche** gemeinsam eine Rolle? Mit anderen Worten: Sind beide Parameter **gemeinsam** statistisch signifikant? Das **Hypothesenpaar** lautet:

$$\begin{aligned} H_0 : \beta_4 = 0 \quad \text{und} \quad \beta_5 = 0 \quad \text{versus} \\ H_1 : \beta_4 \neq 0 \quad \text{und/oder} \quad \beta_5 \neq 0. \end{aligned}$$

Schreiben der Nullhypothese in Matrixform $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$

$$H_0 : \underbrace{\begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{R}} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{\mathbf{r}}. \quad (11.16)$$

F-Test: Überblick und Zusammenfassung: (vgl. Abschnitt 5.6):

1. **Hypothesenpaar mit disjunkter Null- und Alternativhypothese:** Es können $q \leq k$ lineare Restriktionen überprüft werden, die in folgender Form dargestellt werden können:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \quad \text{vs.} \quad H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r} \quad (11.15)$$

wobei die $(q \times k)$ -Matrix \mathbf{R} und der $(q \times 1)$ -Vektor \mathbf{r} gegeben und fest sind.

2. **Teststatistik:** Die F -Teststatistik lautet:

$$F = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T \left[\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) / q}{s^2} \sim F_{q, n-k} \quad (11.17)$$

Die F -Statistik (11.28) ist F -verteilt mit q und $n - k$ Freiheitsgraden.

3. **Entscheidungsregel** für F -Test: Verwerfe $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, falls

$$F > F_{q, n-k, 1-\alpha}.$$

Alternativ: Lehne H_0 ab, falls p -Wert kleiner Signifikanzniveau ist. (Falls σ_0^2 bekannt ist, wird in F s^2 durch σ_0^2 ersetzt und das $1 - \alpha$ -Quantil der χ^2 -Verteilung mit $n - k$ Freiheitsgraden verwendet, siehe (11.21).)

Ableitung der F -Teststatistik (11.17):

- Wie kann man eine skalare Teststatistik für mehrere Hypothesen bilden?

Grundidee: Durch Summieren der quadrierten Abweichungen

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) > \text{kritischer Wert.}$$

Ist es möglich, für die quadrierten Abweichungen die Wahrscheinlichkeitsverteilung zu bestimmen?

- **Verteilung von $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}$:**

– Sind die Annahmen **(B1)** und **(B3)** erfüllt, gilt

$$\mathbf{R}\hat{\boldsymbol{\beta}} = \mathbf{R}\boldsymbol{\beta}_0 + \mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u}$$

bzw.

$$\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = \mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{u},$$

– Sind die Fehler außerdem **multivariat normalverteilt gegeben \mathbf{X}** , d. h. gilt Annahme **(B4)** ist $\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ wegen (2.33) ebenfalls multivariat normalverteilt:

$$\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) | \mathbf{X} \sim N\left(\mathbf{0}, \sigma_0^2 \mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1} \mathbf{R}^T\right), \quad (11.18)$$

wobei $\mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1} \mathbf{R}^T$ Rang q hat, da $\text{rk}(\mathbf{AB}) = \text{rk}(\mathbf{A})$, wenn \mathbf{B} nicht singulär ist (vgl. Schmidt & Trenkler 2006, Regel 3.2.7).

Beweis: Ableiten der Varianz:

$$\begin{aligned} \text{Var}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}_0 | \mathbf{X}) &= \text{Var}(\mathbf{R}\hat{\boldsymbol{\beta}} | \mathbf{X}) = \mathbf{R} \text{Var}(\hat{\boldsymbol{\beta}} | \mathbf{X}) \mathbf{R}^T \\ &= \sigma_0^2 \mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1} \mathbf{R}^T \end{aligned}$$

□

– Addieren und Subtrahieren von \mathbf{r} in $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{R}\boldsymbol{\beta}_0 = \mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} + \mathbf{r} - \mathbf{R}\boldsymbol{\beta}_0$ liefert:

$$\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} | \mathbf{X} \sim N\left(\mathbf{R}\boldsymbol{\beta}_0 - \mathbf{r}, \sigma_0^2 \mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1} \mathbf{R}^T\right), \quad (11.19)$$

– Unter $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, wobei $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ enthalten ist, vereinfacht sich (11.19) zu

$$\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} | \mathbf{X} \sim N\left(\mathbf{0}, \sigma_0^2 \mathbf{R}(\mathbf{X}^T\mathbf{X})^{-1} \mathbf{R}^T\right), \quad (11.20)$$

- **Verteilung der gewichteten Quadratsumme:**

- **Fehlervarianz σ_0^2 bekannt:** Aufgrund der Eigenschaften der χ^2 -Verteilung (2.34) gilt unter H_0 für die gewichtete Summe der Quadrate des $(q \times 1)$ -normalverteilten Vektors $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}$ (11.20), dass

$$(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T \left[\sigma_0^2 \mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \sim \chi_q^2. \quad (11.21)$$

Als Teststatistik sollte also eine gewichtete anstatt einer ungewichteten Summe der quadrierten Abweichungen von $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}$ verwendet werden, da hierfür die Verteilung unter H_0 bei Kenntnis der Fehlervarianz σ_0^2 bekannt ist.

- **Fehlervarianz σ_0^2 unbekannt:** In der Teststatistik (11.21) steht die Fehlervarianz σ_0^2 im Nenner. Ersetzt man σ_0^2 durch den Schätzer s^2 , steht im Nenner nun auch eine Zufallsvariable. Damit ist folgende Statistik ein Kandidat für die F -Verteilung

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T \left[\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{s^2}. \quad (11.22)$$

Überprüfen der Voraussetzungen für die F -Verteilung, vgl. (2.37): Damit eine Zufallsvariable F -verteilt ist, müssen u. a. Zähler und Nenner χ^2 -verteilt sein.

1. Zähler: Da (11.21) χ^2 -verteilt ist, dividiert man Zähler und Nenner von (11.22) durch σ_0^2 , damit der neue Zähler wieder genau (11.21) entspricht und somit χ^2 -verteilt ist:

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T \left[\sigma_0^2 \mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{s^2 / \sigma_0^2}. \quad (11.23)$$

2. Nenner: Der Nenner s^2 / σ_0^2 muss noch mit $n - k$ multipliziert werden, denn gemäß vorherigen Abschnitt 11.3.1 zum t -Test gilt

$$(n - k) s^2 / \sigma_0^2 = \mathbf{y}^T \mathbf{M}_{\mathbf{X}} \mathbf{y} / \sigma_0^2 = \mathbf{u}^T \mathbf{M}_{\mathbf{X}}^T \mathbf{M}_{\mathbf{X}} \mathbf{u} / \sigma_0^2 \sim \chi_{n-k}^2. \quad (11.24)$$

Man erhält:

$$\frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T \left[\sigma_0^2 \mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})}{(n - k) s^2 / \sigma_0^2}. \quad (11.25)$$

3. Zähler und Nenner in (11.25) stochastisch unabhängig? Ja.

Beweis:

* Zähler und Nenner lassen sich jeweils als eine quadratische Form (vgl. Abschnitt 9.3) $\mathbf{x}_i^T \mathbf{A}_i \mathbf{x}_i$, $i = Z, N$, schreiben:

* Zähler:

$$\begin{aligned} \mathbf{x}_Z &= \mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} = \mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}, \\ \mathbf{A}_Z &= \left[\sigma_0^2 \mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} \end{aligned}$$

* Nenner: Der Ausdruck in (11.24) ist auch eine quadratische Form mit

$$\mathbf{x}_N = \mathbf{M}_X \mathbf{u}, \quad (11.26)$$

$$\mathbf{A}_N = \frac{1}{\sigma_0^2} \mathbf{I}. \quad (11.27)$$

* Da \mathbf{A}_Z und \mathbf{A}_N gegeben \mathbf{X} bekannt sind, sind die Verteilungseigenschaften von \mathbf{x}_Z und \mathbf{x}_N entscheidend. Wegen Annahme (B4) ist \mathbf{u} gegeben \mathbf{X} multivariat normalverteilt mit Erwartungswert Null. Deshalb sind auch \mathbf{x}_Z und \mathbf{x}_N gegeben \mathbf{X} multivariat normalverteilt mit Erwartungswert Null. Da \mathbf{x}_Z und \mathbf{x}_N von demselben multivariat normalverteilten Fehlervektor \mathbf{u} abhängen, sind sie stochastisch unabhängig, wenn sie unkorreliert sind, d. h. $Cov(\mathbf{x}_Z, \mathbf{x}_N | \mathbf{X}) = E[\mathbf{x}_Z \mathbf{x}_N^T | \mathbf{X}] = \mathbf{0}$ gilt. Einsetzen ergibt:

$$\begin{aligned} E[\mathbf{x}_Z \mathbf{x}_N^T | \mathbf{X}] &= E\left[\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} (\mathbf{M}_X \mathbf{u})^T | \mathbf{X}\right] \\ &= \mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T E[\mathbf{u} \mathbf{u}^T | \mathbf{X}] \mathbf{M}_X = \sigma_0^2 \mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \underbrace{\mathbf{X}^T \mathbf{M}_X}_{=\mathbf{0}} = \mathbf{0} \end{aligned}$$

Zähler und Nenner sind χ^2 -verteilt, weil \mathbf{x}_Z und \mathbf{x}_N multivariat normalverteilt sind. Sind letztere stochastisch unabhängig, müssen dies auch davon abhängige Funktionen, wie die hier vorliegenden quadratischen Formen sein. Deshalb sind Zähler und Nenner in (11.25) stochastisch unabhängig.

□

4. Division durch korrekte Zahl an Freiheitsgraden? Muss noch erfolgen. D. h. der Zähler in (11.25) muss durch die Anzahl der Restriktionen q dividiert werden. Der Nenner in (11.25) muss durch die korrekte Anzahl an Freiheitsgraden $n - k$ dividiert werden. Man erhält:

$$F = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T \left[\sigma_0^2 \mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) / q}{[(n - k) s^2 / \sigma_0^2] / (n - k)}.$$

Beachte: σ_0^2 und $n - k$ kürzen sich heraus, so dass man (11.17) erhält. Damit ist F unter H_0 eine F -verteilte Teststatistik

$$F \sim F_{q, n-k}. \quad (11.28)$$

Die F -Statistik (11.28) ist also F -verteilt mit q und $n - k$ Freiheitsgraden.

Die Teststatistik F wird als **F -Statistik** bezeichnet.

Alternative Schreibweisen der F -Statistik:

$$F = \frac{(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T \left[\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) / q}{s^2} \quad (11.17)$$

$$= \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T \mathbf{R}^T \left[\mathbf{R} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T \right]^{-1} \mathbf{R} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) / q}{\mathbf{y}^T \mathbf{M}_X \mathbf{y} / (n - k)} \quad (11.29)$$

Güte von F -Tests:

- Es lässt sich zeigen, dass unter der Alternativhypothese alle Quantile der Verteilung der F -Statistik rechts von denen der F -Verteilung unter H_0 liegen. Je weiter rechts die Quantile unter H_1 im Vergleich zu den Quantilen unter H_0 liegen (beispielsweise durch eine steigende Stichprobengröße n), desto größer ist auch die Güte des F -Tests.

Gemeinsame Ausschluss/Nullrestriktionen (joint exclusion restrictions): weitere Berechnungsmöglichkeiten der F -Statistik

- Man kann immer die Variablen in einem multiplen Regressionsmodell so umordnen, dass alle **Ausschluss-/Nullrestriktionen** bezüglich β in dem Modell

$$\mathbf{y} = \underbrace{\mathbf{X}_1}_{(n \times k_1)} \beta_1 + \underbrace{\mathbf{X}_2}_{(n \times k_2)} \beta_2 + \mathbf{u},$$

$k = k_1 + k_2$, in β_2 zusammengefasst werden.

Das Hypothesenpaar lautet dann:

$$\begin{aligned} H_0 : \beta_j = 0, j = k_1 + 1, \dots, k_1 + k_2 &\Leftrightarrow \beta_2 = \mathbf{0} \quad \text{versus} \\ H_1 : \beta_{k_1+1} \neq 0 \text{ und/oder } \dots \text{ und/oder } \beta_{k_1+k_2} \neq 0 &\Leftrightarrow \beta_2 \neq \mathbf{0}. \end{aligned}$$

Die Nullrestriktionen können dann geschrieben werden als

$$\underbrace{\begin{pmatrix} 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{pmatrix}}_{\begin{pmatrix} \mathbf{0}_{k_2 \times k_1} & \mathbf{I}_{k_2} \end{pmatrix}} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{k_1-1} \\ \beta_{k_1} \\ \beta_{k_1+1} \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} \mathbf{0}_{k_2 \times k_1} & \mathbf{I}_{k_2} \end{pmatrix} \beta = \mathbf{0}_{k_2 \times 1}.$$

In diesem Fall existieren weitere Berechnungsmöglichkeiten der F -Statistik (vgl. auch BA-Modul **Einführung in die Ökonometrie**), indem man das restringierte und das unrestringierte Modell getrennt schätzt:

1. **Restringierte Regression:** Regressiere \mathbf{y} ausschließlich auf \mathbf{X}_1 und speichere die Residuenquadratsumme $SSR_1 = \tilde{\mathbf{u}}^T \tilde{\mathbf{u}}$ oder im Fall einer in \mathbf{X}_1 enthaltenen Konstanten auch R_1^2 .
2. **Unrestringierte Regression:** Regressiere \mathbf{y} auf $\mathbf{X} = (\mathbf{X}_1 \quad \mathbf{X}_2)$ und speichere $SSR = \hat{\mathbf{u}}^T \hat{\mathbf{u}}$ bzw. R^2 .

Die weiteren Berechnungsmöglichkeiten sind (beachte $q = k_2$):

$$F = \frac{(SSR_1 - SSR)/k_2}{SSR/(n - k)} \quad (11.30)$$

$$= \frac{(\tilde{\mathbf{u}}^T \tilde{\mathbf{u}} - \hat{\mathbf{u}}^T \hat{\mathbf{u}})/k_2}{\hat{\mathbf{u}}^T \hat{\mathbf{u}}/(n - k)}$$

$$= \frac{(R^2 - R_1^2)/k_2}{(1 - R^2)/(n - k)} \quad (11.31)$$

$$\sim F_{k_2, n-k}.$$

Fortsetzung Handelsströme: Überprüfen der Nullhypothese (11.16):

- Festlegen des kritischen Wertes c : Berechne $1 - \alpha$ -Quantil der $F_{2,44}$ -Verteilung mit R-Befehl `qf(1-alpha,2,44)`. Für $\alpha = 0.05$ erhält man 3.209278.
- **Berechnen der F -Statistik** und des **p -Werts** erfolgt am einfachsten mit dem R-Befehl `linearHypothesis` (erfordert R-Paket `car`):

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
alpha          <- 0.05          # Signifikanzniveau
# Schätzen von Modell 4
mod_4_kq       <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) +
                    log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))
summary(mod_4_kq)
qf(1-alpha,2,44)          # Kritischer Wert
library(car)              # Library car laden für F-Test
# F-Test
F_stat         <- linearHypothesis(mod_4_kq,
                                   c("ebrd_tfes_o=0", "log(cepii_area_o)=0"),
                                   test=c("F"))
F_stat
```

Listing 11.1: `./R_code/4_ff_Beispiel_Handelsstroeme.R`

und liefert

```
Hypothesis:
ebrd_tfes_o = 0
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 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
```

- Die Nullhypothese wird abgelehnt. Mindestens eine der beiden Variablen Offenheit bzw. $\log(\text{Fläche})$ ist auf dem 5%-Niveau signifikant von Null verschieden.

Beweis: ‡ **Beweismöglichkeiten** für (11.30) und (11.31)

- **1. Beweismöglichkeit:** mit Hilfe der Formel zur Inversion partitionierter Matrizen
- **2. Beweismöglichkeit:** mit Hilfe des Frisch-Waugh-Lovell Theorems:
 1. Man beachte, dass die Residuenquadratsumme des unrestringierten Modells

$$SSR = \mathbf{y}^T \mathbf{M}_X \mathbf{y}$$

mit Hilfe der Zerlegung der Residuenquadratsumme und des **Frisch-Waugh-Lovell Theorems**, siehe Abschnitt 7.1, auf Basis der Regression

$$\mathbf{M}_1 \mathbf{y} = \mathbf{M}_1 \mathbf{X}_2 \beta_2 + \text{Residuen}$$

auch geschrieben werden kann als

$$\begin{aligned} SSR &= TSS - ESS \\ &= \mathbf{y}^T \mathbf{M}_1 \mathbf{y} - \mathbf{y}^T \mathbf{M}_1 \mathbf{P}_{\mathbf{M}_1 \mathbf{X}_2} \mathbf{M}_1 \mathbf{y} \\ &= \mathbf{y}^T \mathbf{M}_1 \mathbf{y} - \mathbf{y}^T \mathbf{M}_1 \mathbf{M}_1 \mathbf{X}_2 \underbrace{\left(\mathbf{X}_2^T \mathbf{M}_1 \mathbf{M}_1 \mathbf{X}_2 \right)^{-1} \mathbf{X}_2^T \mathbf{M}_1 \mathbf{M}_1 \mathbf{y}}_{\mathbf{P}_{\mathbf{M}_1 \mathbf{X}_2}} \\ &= \mathbf{y}^T \mathbf{M}_1 \mathbf{y} - \mathbf{y}^T \mathbf{M}_1 \mathbf{X}_2 \left(\mathbf{X}_2^T \mathbf{M}_1 \mathbf{X}_2 \right)^{-1} \mathbf{X}_2^T \mathbf{M}_1 \mathbf{y}. \end{aligned}$$

2. Der Zähler in der F -Statistik (11.30) ist dann

$$\begin{aligned} SSR_1 - SSR &= \mathbf{y}^T \mathbf{M}_1 \mathbf{y} - \left[\mathbf{y}^T \mathbf{M}_1 \mathbf{y} - \mathbf{y}^T \mathbf{M}_1 \mathbf{X}_2 \left(\mathbf{X}_2^T \mathbf{M}_1 \mathbf{X}_2 \right)^{-1} \mathbf{X}_2^T \mathbf{M}_1 \mathbf{y} \right] \\ &= \mathbf{y}^T \mathbf{M}_1 \mathbf{X}_2 \left(\mathbf{X}_2^T \mathbf{M}_1 \mathbf{X}_2 \right)^{-1} \mathbf{X}_2^T \mathbf{M}_1 \mathbf{y} \quad (11.32) \\ &= \mathbf{u}^T \mathbf{P}_{\mathbf{M}_1 \mathbf{X}_2} \mathbf{u}. \end{aligned}$$

Das letzte Gleichheitszeichen gilt, da unter H_0 $\mathbf{M}_1 \mathbf{y} = \mathbf{M}_1 \mathbf{u}$ (verifizieren!).

3. Da $\mathbf{P}_{\mathbf{M}_1 \mathbf{X}_2}$ eine Projektionsmatrix mit Rang k_2 ist, folgt aus der Eigenschaft (2.35) der χ^2 -Verteilung, dass bei normalverteilten Fehlern unter H_0

$$SSR_1 - SSR = \frac{\mathbf{u}^T}{\sigma} \mathbf{P}_{\mathbf{M}_1 \mathbf{X}_2} \frac{\mathbf{u}}{\sigma} \sim \chi^2(k_2).$$

Für den Nenner gilt

$$SSR = \frac{\mathbf{u}^T}{\sigma} \mathbf{M}_X \frac{\mathbf{u}}{\sigma} \sim \chi^2(n - k).$$

Zähler und Nenner sind also jeweils χ^2 -verteilt.

Die Zufallsvektoren im Zähler $\mathbf{P}_{\mathbf{M}_1 \mathbf{X}_2} \mathbf{u}$ und Nenner $\mathbf{M}_X \mathbf{u}$ haben Kovarianz Null, da

$$\mathbf{M}_X \mathbf{M}_1 = \mathbf{M}_1 \mathbf{M}_X$$

und folglich

$$\mathbf{M}_X \mathbf{M}_1 \mathbf{X}_2 = \mathbf{M}_1 \mathbf{M}_X \mathbf{X}_2 = \mathbf{0}$$

(die Spalten von \mathbf{X}_2 sind im orthogonalen Raum zu \mathbf{M}_X enthalten) und somit $E[\mathbf{P}_{\mathbf{M}_1 \mathbf{X}_2} \mathbf{u} \mathbf{u}^T \mathbf{M}_X] = \mathbf{0}$. Aufgrund der multivariaten Normalverteilungsannahme sind die Zufallsvektoren damit auch stochastisch unabhängig.

Damit gilt aufgrund der Definition der F -Verteilung

$$F = \frac{(SSR_1 - SSR)/k_2}{SSR/(n-k)} \sim F_{k_2, n-k}$$

□

- Durch (11.32) ergibt sich noch eine weitere Schreibweise der F -Statistik (11.30)

$$F = \frac{\mathbf{y}^T \mathbf{M}_1 \mathbf{X}_2 (\mathbf{X}_2^T \mathbf{M}_1 \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{M}_1 \mathbf{y} / k_2}{\mathbf{y}^T \mathbf{M}_X \mathbf{y} / (n-k)} \quad (11.33)$$

- Die F -Statistik (11.30) kann auch für allgemeine lineare Restriktionen verwendet werden. Dazu muss jedoch das Modell unter H_0 geeignet umgeformt werden, siehe BA-Modul **Einführung in die Ökonometrie**.

Weitere bekannte F -Tests:

- **Einzelne Hypothese:** F -Statistik ist Quadrat der t -Statistik und entspricht einem zweiseitigem t -Test.
- **Chow-Strukturbruchtest:** Test auf Konstanz aller/einiger Parameter über 2 Teilstichproben, jeweils mit I und II indiziert, hinweg. Sind diese nicht konstant, muss man für jede Teilstichprobe eine eigene Schätzung durchführen

$$\mathbf{y}_I = \mathbf{X}_I \boldsymbol{\beta}_I + \mathbf{u}_I \quad (11.34a)$$

$$\mathbf{y}_{II} = \mathbf{X}_{II} \boldsymbol{\beta}_{II} + \mathbf{u}_{II}. \quad (11.34b)$$

Die Nullhypothese (Parameterkonstanz) lautet

$$H_0 : \boldsymbol{\beta}_I = \boldsymbol{\beta}_{II}.$$

Unter H_0 ist also das Modell

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

zu schätzen.

Unter H_1 hingegen können sich Elemente von $\boldsymbol{\beta}_{II}$ und $\boldsymbol{\beta}_I$ unterscheiden und man schätzt in Matrixschreibweise mit

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_I \\ \mathbf{y}_{II} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_I \\ \mathbf{X}_{II} \end{pmatrix}$$

das Modell

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \begin{pmatrix} \mathbf{0} \\ \mathbf{X}_{II} \end{pmatrix} \boldsymbol{\gamma} + \mathbf{u}.$$

Das Hypothesenpaar lautet:

$$H_0 : \boldsymbol{\gamma} = \mathbf{0} \quad \text{versus} \quad H_1 : \gamma_j \neq 0 \text{ für mindestens ein } j.$$

Sind zusätzlich zu (B1), (B3), (B4) die Teilstichproben stochastisch unabhängig, ist der **Chow-Test** exakt.

Pivote Tests

- **Definition:** Eine Zufallsvariable, z. B. eine Teststatistik unter H_0 , mit der Eigenschaft, dass ihre Verteilung für alle DGPs in einem Modell \mathbb{M} gleich ist, heißt **pivot** für das Modell \mathbb{M} .
- Die Nullhypothese spezifiziert selten den kompletten DGP. Ist dies der Fall, spricht man von einer **einfachen Hypothese (simple hypothesis)**.
- I. Allg. enthält das Modell unter der Nullhypothese mehrere verschiedene DGPs: **zusammengesetzte Hypothese (compound hypothesis)**. Hängt die **exakte** Verteilung eines Tests einer zusammengesetzten Nullhypothese vom DGP ab, der die Stichprobendaten generiert hat, ist die Teststatistik nicht pivot, da sich je nach spezifischen DGP bei gleicher Nullhypothese die Testverteilung ändert. Eine **Ausnahme** bilden hierzu **exakte Tests**.
- Mögliche Auswege für alle anderen Fälle:
 - Ohne Kenntnis des DGP: **asymptotisch pivote Tests**, siehe nächsten Abschnitt 11.4, d. h. die asymptotische Verteilung der Teststatistik ist pivot.
 - Bei Kenntnis des DGP: **Monte-Carlo-Tests**, siehe Abschnitt 11.5.1.
 - Ohne Kenntnis des DGP: **Bootstraptests**, siehe Abschnitt 11.5.2.

11.4. Asymptotische Tests

Das normale multiple lineare Regressionsmodell ist gegeben durch

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad \mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I}).$$

Ist die Annahme (B2) (streng) exogener Regressoren beispielsweise durch verzögerte endogene Variable als Regressoren oder die Annahme (B4) normalverteilter Fehler nicht erfüllt, dann ist die exakte Verteilung der t - und F -Statistiken aus Abschnitt 11.3 im Allgemeinen nicht analytisch bestimmbar. Selbst wenn dies möglich wäre, wäre die Verteilung der t -Statistik i. Allg. nicht pivot.

- Unter den Annahmen (B1), (B2), (B3), (A1) und (A3) gilt, dass die eingeführten t -Tests und F -Tests **asymptotisch** gültig sind, da der KQ-Schätzer asymptotisch normalverteilt ist.
- Die Ergebnisse bleiben auch unter den Annahmen des dynamisch linearen Regressionsmodells (C1), (C2), (C3) und (C4a) oder (C4b), siehe Abschnitt 13.4, gültig.

11.4.1. Asymptotischer t -Test

Hier: Der im linearen Regressionsmodell

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{x}_2\beta_2 + \mathbf{u}, \quad u_t | \mathbf{X}_t \sim (0, \sigma^2), t = 1, \dots, n.$$

zu testende Parameter ist β_2 .

Asymptotischer t -Test: Überblick

1. Das **Hypothesenpaar** lautet: $H_0 : \beta_2 = \beta_{2,H_0}$ versus $H_1 : \beta_2 \neq \beta_{2,H_0}$.
2. **Teststatistik und Testverteilung:** Unter H_0 gilt:

$$t_{\beta_2} = \frac{\hat{\beta}_2 - \beta_{2,H_0}}{s_{\hat{\beta}_2}} = \frac{z_{\beta_2}}{(s^2/\sigma_0^2)^{1/2}} \xrightarrow{d} N(0, 1).$$

3. **Entscheidungsregel: analog zu Entscheidungsregel bei zwei- oder einseitigen Tests.**

Hinweis: In der Praxis wird üblicherweise die t -Verteilung mit $n - k$ Freiheitsgraden verwendet, da diese häufig eine bessere Approximation der (unbekannten) exakten Verteilung als die Standardnormalverteilung liefert.

Ableitung der asymptotischen Verteilung

1. Nenner: Unter H_0 gilt

$$\text{plim}_{n \rightarrow \infty} (s^2/\sigma_0^2)^{1/2} = 1.$$

2. Zähler: Der Zähler in (11.13) wird mit $n^{-1/2}$ erweitert zu

$$z_{\beta_2} = \frac{n^{-1/2} \mathbf{x}_2^T \mathbf{M}_1 \mathbf{u}}{\sigma_0 (n^{-1} \mathbf{x}_2^T \mathbf{M}_1 \mathbf{x}_2)^{1/2}}$$

und hat offensichtlich Erwartungswert 0 und die Varianz 1, da die Varianz des Zählers gerade dem Quadrat des Nenners entspricht (beides verifizieren!).

3. Unter der Annahme, dass die Regularitätsbedingungen für einen multivariaten Zentralen Grenzwertsatz für $n^{-1/2} \mathbf{x}_2^T \mathbf{M}_1 \mathbf{u}$ (also für (A1), (A3)) erfüllt sind, resultiert

$$z_{\beta_2} \xrightarrow{d} N(0, 1).$$

4. Mit der Regel, siehe Abschnitt 3.4 „Falls $\mathbf{a}_n \xrightarrow{d} \mathbf{a}$ und $\text{plim } \mathbf{A}_n = \mathbf{A}$, dann gilt $\mathbf{A}_n \mathbf{a}_n \xrightarrow{d} \mathbf{A} \mathbf{a}$ “ folgt

$$t_{\beta_2} = \underbrace{(s^2/\sigma_0^2)^{-1/2}}_{\xrightarrow{P} 1} \underbrace{z_{\beta_2}}_{\xrightarrow{d} N(0,1)} \xrightarrow{d} N(0, 1) \quad (11.35)$$

Dann erhält man unter H_0 asymptotisch wieder die Standardnormalverteilung und alle Eigenschaften des t -Tests bleiben asymptotisch gültig.

11.4.2. Asymptotischer F -Test

Asymptotischer F -Test: Überblick

1. **Hypothesenpaar mit disjunkter Null- und Alternativhypothese:** wie bei exaktem F -Test.
2. **Teststatistik und Testverteilung:** Unter H_0 gilt:

$$qF_n = (\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0))^T [s^2 \mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T]^{-1} (\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)) \xrightarrow{d} \chi^2(q) \quad (11.36)$$

In kleinen Stichproben wird stattdessen häufig direkt die F -Statistik zusammen mit einer (approximativen) F -Verteilung mit q und $n - k$ Freiheitsgraden verwendet, da diese häufig eine bessere Approximation der (unbekannten) exakten Verteilung als die χ^2 -Verteilung liefert.

3. **Entscheidungsregel:**

$$qF > \chi_{q,1-\alpha}^2 \quad (11.37)$$

$$F > F_{q,n-k,1-\alpha} \quad (11.38)$$

Alternativ: Lehne H_0 ab, falls p -Wert (auf Basis der asymptotischen Verteilung) kleiner Signifikanzniveau ist.

Ableitung asymptotischer F -Test

- Wenn die relevanten Annahmen, vgl. Beginn des Abschnitts, erfüllt sind, so dass

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}}^{-1})$$

gilt, folgt aus dem Theorem über stetige Abbildungen (3.3) und den Eigenschaften der χ^2 -Verteilung (2.34) eine asymptotische χ^2 -Verteilung:

$$n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)^T [\sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}}^{-1}]^{-1} (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} \chi^2(k). \quad (11.39)$$

- Asymptotische Verteilung der F -Statistik (11.28): Unter Anwendung von $\mathbf{A}_n \mathbf{a}_n \xrightarrow{d} \mathbf{A} \mathbf{a}$, vgl. Abschnitt 3.4, erhält man aus (11.28) zusammen mit $\text{plim}_{n \rightarrow \infty} s^2 = \sigma_0^2$ und (A1) (bzw. (C3)) wieder (11.39) nachdem mit q multipliziert wurde. Damit gilt unter H_0

$$qF_n = (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r})^T [s^2 \mathbf{R}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{R}^T]^{-1} (\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \xrightarrow{d} \chi^2(q). \quad (11.36)$$

- Für den Fall von Ausschlussrestriktionen lässt sich unter H_0 die F -Statistik (11.28) alternativ schreiben als (11.33). In diesem Fall gilt natürlich auch

$$qF_n = q \frac{\mathbf{y}^T \mathbf{M}_1 \mathbf{X}_2 (\mathbf{X}_2^T \mathbf{M}_1 \mathbf{X}_2)^{-1} \mathbf{X}_2^T \mathbf{M}_1 \mathbf{y} / q}{\mathbf{y}^T \mathbf{M}_X \mathbf{y} / (n - k)} \xrightarrow{d} \chi^2(q). \quad (11.40)$$

- Da gilt (vgl. Abschnitt 2.9), dass für $n \rightarrow \infty$ eine Folge von F -verteilten Zufallsvariablen $X_n \sim F(q, n - k)$ gegen eine χ^2 -Verteilung konvergiert,

$$qX_n \xrightarrow{d} \chi^2(q), \quad (11.41)$$

kann F_n auch durch eine $F(q, n - k)$ -Verteilung approximiert werden, die in kleinen Stichproben sogar häufig eine bessere Approximation liefert als die χ^2 -Verteilung.

Güte: Es kann gezeigt werden, dass unter H_1 gilt

$$qF \xrightarrow{n \rightarrow \infty} \infty. \quad (11.42)$$

Damit geht für $n \rightarrow \infty$ die Güte asymptotisch gegen 1, da $\lim_{n \rightarrow \infty} P(qF > c) = 1$. In endlichen Stichproben ist die Güte typischerweise kleiner als 1.

Tatsächliche versus nominale Größe

- **Tatsächliche Größe:** Größe eines Tests (5.78), die sich auf Basis der exakten, aber möglicherweise unbekanntem Verteilung ergibt.
- **Nominale Größe:** Größe eines Tests, die sich auf Basis der asymptotischen Verteilung ergibt.
- Da bei exakten Tests die exakte Verteilung für jeden DGP und Stichprobengröße bekannt ist, stimmen nominales und tatsächliches Signifikanzniveau überein.
- Bei asymptotischen Tests ist die Übereinstimmung von nominalem und tatsächlichem Signifikanzniveau umso besser, je genauer die asymptotische Verteilung die tatsächliche Verteilung (die im Allgemeinen vom DGP und der Beobachtungszahl abhängt) approximiert. Für vorbestimmte DGPs lässt sich der Grad der Übereinstimmung mit Monte-Carlo-Simulationen feststellen.
- Bei asymptotischen Tests ist die tatsächliche Größe unbekannt. Deshalb wird der kritische Wert so gewählt, dass die nominale Größe dem gewählten Signifikanzniveau entspricht.
- Ein Test heißt "oversized", wenn die tatsächliche Größe (z. B. durch Simulationen bestimmt) größer als das Signifikanzniveau ist.

11.5. Monte-Carlo-Tests und Bootstraptests

11.5.1. Monte-Carlo-Tests

- **Empirische Verteilungsfunktion (empirical distribution function)** der beobachteten Stichprobenelemente x_t , $t = 1, \dots, n$:

$$\hat{F}(x) = \frac{1}{n} \sum_{t=1}^n 1(x_t \leq x), \quad (11.43)$$

wobei $1(\cdot)$ die **Indikatorfunktion**

$$1(A) = \begin{cases} 1 & \text{falls } A \text{ wahr} \\ 0 & \text{falls } A \text{ falsch} \end{cases} \quad (11.44)$$

bezeichnet.

Fundamental Theorem of Statistics Die empirische Verteilungsfunktion ist im Fall i.i.d.-verteilter Zufallsvariablen konsistent

$$\text{plim } \hat{F}(x) = F(x). \quad (11.45)$$

Die i.i.d.-Annahme kann abgeschwächt werden.

- Notation in diesem Abschnitt: τ bezeichnet eine beliebige Teststatistik und $\hat{\tau} = \hat{\tau}(\mathbf{X}, \mathbf{y})$ einen auf Basis von Stichprobenbeobachtungen berechneten Wert der Teststatistik τ .
- Der **exakte p -Wert** einer berechneten Teststatistik $\hat{\tau}$ **mit rechtsseitigem kritischem Wert** ergibt sich (vgl. (5.85)) aus

$$p(\hat{\tau}) := P(\tau > \hat{\tau} | \boldsymbol{\theta}_{H_0}) = 1 - P(\tau \leq \hat{\tau} | \boldsymbol{\theta}_{H_0}) = 1 - F(\hat{\tau} | \boldsymbol{\theta}_{H_0}), \quad (11.46)$$

wobei hier $F(\cdot | \boldsymbol{\theta}_{H_0})$ die exakte Verteilung der berechneten Teststatistik $\hat{\tau}$ unter H_0 bezeichnet.

Zur Erinnerung: Lehne H_0 ab, falls $p(\hat{\tau}) < \alpha$ bzw. $\hat{\tau} > c_\alpha$.

Ist $F(\cdot | \boldsymbol{\theta}_{H_0})$ unbekannt, lässt sich die Testverteilung durch die empirische Verteilungsfunktion beliebig genau approximieren, **sofern der DGP vollständig bekannt** ist oder der Test **pivot** ist. Je größer die Zahl der Replikationen (Monte-Carlo-Simulationen) B , desto genauer die Approximation. Der **computer-simulierte p -Wert** ist

$$\hat{p}(\hat{\tau}) = 1 - \hat{F}(\hat{\tau} | \boldsymbol{\theta}_{H_0}) = 1 - \frac{1}{B} \sum_{j=1}^B 1(\tau_j^* \leq \hat{\tau}), \quad (11.47)$$

wobei τ_j^* der Wert der Teststatistik in der j -ten Simulation unter H_0 ist.

- Die Durchführung eines Monte-Carlo-Tests erfordert die Generierung von Zufallszahlen mit Hilfe eines Zufallszahlengenerators, siehe hierzu z. B. [Davidson & MacKinnon \(2004, S. 157-159\)](#).

11.5.2. Bootstraptests

- Die **Idee** eines Bootstraptests ist, den unbekanntem DGP zu schätzen und dann die Technik des Monte-Carlo-Tests anzuwenden.
- Notwendige Voraussetzung: Alle notwendigen Eigenschaften des DGPs können konsistent mit geeigneter Konvergenzrate geschätzt werden.
- Beispiel: Multiples Regressionsmodell

$$y_t = \beta_1 x_{t1} + \cdots + \beta_k x_{tk} + u_t, \quad u_t | \mathbf{X} \sim IID(0, \sigma^2)$$

Es soll

$$H_0 : \beta_k = 0 \quad \text{versus} \quad H_1 : \beta_k \neq 0$$

getestet werden. Der Bootstraptest funktioniert, wenn u.a.

- β und σ^2 konsistent mit Rate \sqrt{n} geschätzt werden können und
- die Verteilung von \mathbf{u} gegeben \mathbf{X} bekannt ist oder entsprechend geschätzt werden kann.

Generieren von Bootstrapstichproben

- Schätzen von β mit einem konsistenten und möglichst effizienten Schätzer und berechnen der gewünschten Teststatistik $\hat{\tau}$.
 - Man kann β unter H_1 schätzen und erhält den KQ-Residuenvektor $\hat{\mathbf{u}}$.
 - Man kann $(\beta_1 \dots \beta_{k-1})^T$ unter H_0 schätzen und erhält den KQ-Residuenvektor $\tilde{\mathbf{u}}$.

Im Allgemeinen erhält man bessere Ergebnisse, wenn man die Schätzung unter H_0 durchführt.

- **Annahme i.i.d. normalverteilter Fehler** (Annahme **(B4)**): **Parametrischer Bootstrap** Man kann dann σ^2 schätzen und für jede Bootstrapstichprobe $(\mathbf{y}_j^*, \mathbf{X}_j)$ die n Fehlervariablen in \mathbf{u}_j^* aus $N(\mathbf{0}, s^2 \mathbf{I})$ generieren.

1. Dann lässt sich für die j -te Bootstrapstichprobe der Vektor der abhängigen Variablen unter H_0 iterativ generieren durch

$$y_{jt}^* = \tilde{\beta}_1 x_{t1} + \dots + \tilde{\beta}_{k-1} x_{t,k-1} + u_{jt}^*, \quad t = 1, 2, \dots, n.$$

2. Für die j -Stichprobe $(\mathbf{y}_j^*, \mathbf{X}_j)$ lässt sich dann die Teststatistik, hier der quadrierte t -Test (= F -Test) berechnen, indem das unrestringierte Modell geschätzt wird

$$\tau_j^* = (t_j^*)^2, \quad t_j^* = \hat{\beta}_{jk}^* / \hat{\sigma}_{j, \hat{\beta}_{jk}^*}^*.$$

Nach B Replikationen berechnet man dann die empirische Verteilungsfunktion und erhält den Bootstrap p -Wert gemäß (11.47) aus

$$\hat{p}(\hat{\tau}) = 1 - B^{-1} \sum_{j=1}^B 1(\tau_j^* \leq \hat{\tau}).$$

- **Annahme i.i.d. verteilter Fehler** (Annahme **(B2)**): **Nichtparametrischer / semi-parametrischer Bootstrap**

1. Unter H_0 sind die KQ-Parameterschätzer konsistent und damit auch die geschätzten Fehler

$$\begin{aligned} \text{plim}_{n \rightarrow \infty} \tilde{u}_t &= \text{plim}_{n \rightarrow \infty} (y_t - \tilde{\beta}_{n1} x_{t1} - \dots - \tilde{\beta}_{n,k-1} x_{t,k-1}) \\ &= y_t - x_{t1} \text{plim}_{n \rightarrow \infty} \tilde{\beta}_{n1} - \dots - x_{t,k-1} \text{plim}_{n \rightarrow \infty} \tilde{\beta}_{n,k-1} \\ &= y_t - \beta_1 x_{t1} - \dots - \beta_{k-1} x_{t,k-1} = u_t. \end{aligned}$$

2. 'Asymptotisch' kann man also auch aus den Fehlern mit Zurücklegen ziehen (**resampling**), denn aufgrund des Fundamental Theorems of Statistics approximiert die empirische Verteilung der u_t 's die wahre Fehlerverteilung.
3. Statt der unbekanntem Fehler lassen aufgrund der Konsistenz des Residuenschätzers auch die Residuen verwenden.
4. Verfeinerungen:
 - **reskalierte Residuen (rescaled residuals)**

$$\tilde{u}_t^+ = \tilde{u}_t \left(\frac{n}{n - k_{H_0}} \right)^{1/2}.$$

Damit wird die Varianz der Residuen, die ja kleiner ist als die Varianz der Fehler (vgl. Abschnitt 9.5), so korrigiert, dass sie der geschätzten Varianz der Fehler s^2 entspricht.

- **zentrierte und reskalierte Residuen (centered residuals)**

$$\tilde{u}_t^+ = (\tilde{u}_t - \bar{\tilde{u}}) \left(\frac{n}{n - k_{H_0}} \right)^{1/2}.$$

Dies ist notwendig, wenn z. B. das Regressionsmodell keine Konstante enthält, denn dann ist der Mittelwert der Residuen ungleich Null und damit wird der Bootstraptest verzerrt.

- **Wild Bootstrap** und **Block Bootstrap**: Im Fall heteroskedastischer und autokorrelierter Fehler funktionieren die obigen Verfahren nicht. Hier sind kompliziertere Verfahren notwendig.
- **Zahl der Bootstrapreplikationen**: Wähle B so, dass das Quantil, vgl. (2.8) in Abschnitt 2.5.1, für Fehler 1. Art exakt zu bestimmen ist:
 - Insgesamt gibt es $B + 1$ Rangpositionen r für die Teststatistik $\hat{\tau}$. Beispiel: $B = 2$, wobei die Ränge absteigend angeordnet werden (vgl. Davidson & MacKinnon (2004), S. 164):

$$r = 2 : \hat{\tau} < \min_j(\tau_j^*), \quad r = 1 : \min_j(\tau_j^*) < \hat{\tau} < \max_j(\tau_j^*), \quad r = 0 : \max_j(\tau_j^*) < \hat{\tau}$$

- Dividiert man die Rangposition r durch die Anzahl der Bootstrapreplikationen B erhält man den p -Wert für $\hat{\tau}$, denn $0 = \frac{0}{B} \leq \frac{r}{B} \leq \frac{B}{B} = 1$.
- Damit lehnt der Bootstraptest unter H_0 ab, wenn $r/B < \alpha$, wobei α das gewählte Signifikanzniveau bezeichnet. Es gilt also $r < B\alpha$.
- Es bezeichne $\lfloor x \rfloor$ die größte ganzzahlige Zahl, die kleiner x ist. Dann lässt sich für gegebenes $B\alpha$ die Anzahl an Rängen, für die H_0 abgelehnt wird, ausdrücken als $\lfloor B\alpha \rfloor + 1$. Beispiel: $B = 9$ und $\alpha = 0.5$. Damit wird für $r = 0, 1, 2, 3, 4$ die Nullhypothese abgelehnt. Es gibt $\lfloor B\alpha \rfloor + 1 = \lfloor 4.5 \rfloor + 1 = 5$ Rangpositionen mit Ablehnung.

- Da es insgesamt $B + 1$ Rangpositionen gibt, muss

$$\frac{\lfloor B\alpha \rfloor + 1}{B + 1}$$

gleich α sein. Gegeben α bestimmt man B also aus

$$\alpha(B + 1) = \lfloor \alpha B \rfloor + 1.$$

Für $\alpha = 0.05$ ist beispielsweise $B = 99$ sinnvoll.

Bemerkungen

- **Bootstraptest statt asymptotischem Test?**

Wenn

- die Verteilung der Teststatistik asymptotisch pivot ist und
- die Fehler des Modells i.i.d. sind (andernfalls müssen kompliziertere Bootstrapmethoden herangezogen werden, z. B. Block Bootstrap bei korrelierten Fehlern),

dann konvergiert die Verteilung des Bootstraptests mit wachsendem Stichprobenumfang schneller gegen die (unbekannte) exakte Verteilung der Teststatistik als die asymptotische Verteilung, genauer gesagt mit n^{-1} anstatt mit $n^{-1/2}$. Dies erklärt die weite Verbreitung von Bootstrap.

- **Achtung:** Ist die Teststatistik nicht asymptotisch pivot, dann haben der Bootstraptest und der asymptotische Test die gleiche Konvergenzrate, Bootstrap bringt dann also nichts.
- Bootstrapverfahren lassen sich auch unter bestimmten Bedingungen bei dynamischen Regressionsmodellen anwenden. Dann wird für die j -te Stichprobe (y_j^*, \mathbf{X}_j^*) auch \mathbf{X}_j^* generiert. Zur Durchführung in einem einfachen Beispiel siehe [Davidson & MacKinnon \(2004, p. 160\)](#).
- Weiterführende Literatur: z. B. [Horowitz \(2001\)](#), [Horowitz \(2003\)](#).

11.6. Konfidenzintervalle und -ellipsoide

11.6.1. Konfidenzintervalle

- Definition: **Konfidenzintervall:**

- Ein Zufallsintervall, das auf Basis von Stichprobeninformation (\mathbf{y}, \mathbf{X}) berechnet werden kann und mit Wahrscheinlichkeit $1 - \alpha$ den wahren Parameterwert θ_0 enthält, heißt Konfidenzintervall. (Daraus folgt, dass bei einer großen Anzahl von Stichproben, die alle von demselben DGP erzeugt wurden, der wahre Parameterwert ungefähr in $1 - \alpha$ aller berechneten Konfidenzintervalle enthalten sein sollte.)

- Davidson & MacKinnon (2004, Chapter 5) wählen eine alternative Definition: Fasst man alle Nullhypothesen (bzgl. eines Parameters)

$$H_0 : \theta = \theta_{H_0},$$

die zu einem gegebenen Signifikanzniveau von α nicht abgelehnt werden, in einem Intervall zusammen, erhält man ein Konfidenzintervall mit Konfidenzniveau

$$1 - \alpha.$$

- Formal: Gegeben eine nichtnegative Teststatistik $\tau(\mathbf{y}, \mathbf{X}, \theta_{H_0})$ und ein Signifikanzniveau α , enthält ein Konfidenzintervall alle θ_{H_0} , für die gilt

$$KI = \left\{ \theta_{H_0} \mid P_{\theta_{H_0}} (\tau(\mathbf{y}, \mathbf{X}, \theta_{H_0}) \leq c_\alpha) = 1 - \alpha \right\}, \quad (11.48)$$

wobei $P_{\theta_{H_0}}(\cdot)$ bedeutet, dass die Wahrscheinlichkeit unter der jeweiligen Nullhypothese H_0 berechnet wird und c_α der kritische Wert zum Signifikanzniveau α ist.

- Die Grenzen $[\theta_l, \theta_u]$ des Konfidenzintervalls ergeben sich durch Lösung von

$$\tau(\mathbf{y}, \mathbf{X}, \theta) = c_\alpha$$

nach θ und ergeben sich sozusagen durch „Invertierung“ der Teststatistik $\tau(\mathbf{y}, \mathbf{X}, \theta_{H_0})$.

- Die Länge und Grenzen von Konfidenzintervallen sind zufällig, da sie von der Stichprobe \mathbf{y}, \mathbf{X} abhängen.
- Die **Überdeckungswahrscheinlichkeit (coverage probability)** gibt die Wahrscheinlichkeit an, dass man eine Stichprobe zieht und auf deren Basis ein Konfidenzintervall berechnet, das den wahren Parameter θ_0 enthält.
- **Liegt bereits eine Stichprobe vor**, dann ist der wahre Parameter θ_0 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.
- **Exakte Konfidenzintervalle** überdecken den wahren Parameter θ_0 mit einer Überdeckungswahrscheinlichkeit von $1 - \alpha$. Dies ist dann der Fall, wenn die Teststatistik in (11.48) pivot ist.
- Ist die Teststatistik in (11.48) nicht pivot, aber **asymptotisch pivot**, d. h. deren **asymptotische Verteilung** ist für alle DGPs unter der Nullhypothese bekannt und unabhängig vom jeweiligen DGP im betrachteten Modell \mathbb{M} , dann erhält man ein **asymptotisches Konfidenzintervall**.
- Bei **asymptotischen** Konfidenzintervallen stimmen die tatsächliche und die nominal (gewählte) Überdeckungswahrscheinlichkeit im Allgemeinen nicht überein. Stehen mehrere Verfahren zur Berechnung von approximativen Konfidenzintervallen zur Verfügung, sollte man dasjenige wählen, für das der Unterschied zwischen tatsächlicher und nominaler Überdeckungswahrscheinlichkeit möglichst klein ist.

- Wird anstelle eines Parameters ein Parametervektor betrachtet, erhält man **mehrdimensionale Konfidenzellipsoide**, siehe Abschnitt 11.6.2.
- **Asymptotisches Konfidenzintervall für β_j im multiplen linearen Regressionsmodell** auf Basis der χ^2 -Statistik

$$\tau(\mathbf{y}, \mathbf{X}, \beta_{j,H_0}) = \left(\frac{\hat{\beta}_j - \beta_{j,H_0}}{s_{\hat{\beta}_j}} \right)^2$$

mit

$$s_{\hat{\beta}_j} = s(\mathbf{x}_j^T \mathbf{M}_{-j} \mathbf{x}_j)^{-1/2},$$

wobei $\mathbf{M}_{-j} = \mathbf{I} - \mathbf{X}_{-j} (\mathbf{X}_{-j}^T \mathbf{X}_{-j})^{-1} \mathbf{X}_{-j}^T$ und \mathbf{X}_{-j} enthält alle Regressoren außer den j -ten Regressor.

– Die Grenzen des Konfidenzintervalls ergeben sich aus

$$\left(\frac{\hat{\beta}_j - \beta_{j,H_0}}{s_{\hat{\beta}_j}} \right)^2 = c_\alpha = q_{1-\alpha}$$

(wie oben durch Auflösen nach β_{j,H_0}) als

$$[\hat{\beta}_j - s_{\hat{\beta}_j} c_\alpha^{1/2}, \hat{\beta}_j + s_{\hat{\beta}_j} c_\alpha^{1/2}].$$

- Für $\alpha = 0.05$ ergibt sich für das $(1 - \alpha)$ -Quantil $c_\alpha = q_{1-\alpha}$ der χ^2 -Verteilung $c_\alpha^{1/2} = \sqrt{3.84} = 1.96 = z_{1-\alpha/2}$, wobei z_β das β -Quantil der Standardnormalverteilung bezeichnet.
- Dieses Intervall ist identisch mit dem Intervall, das man auf Basis der t -Statistik erhält, wenn man deren asymptotische Standardnormalverteilung zugrunde legt.
- Asymmetrische Konfidenzintervalle sind z. B. auf Basis der t -Statistik möglich. Wann will man ein asymmetrisches Konfidenzintervall?
- Ein **exaktes Konfidenzintervall für β_j im normalen linearen Modell** wird auf Basis der t -Statistik und t -Verteilung mit $n - k$ Freiheitsgraden bestimmt:

$$P \left(t_{\alpha/2}(n - k) \leq \frac{\hat{\beta}_j - \beta_{j,H_0}}{s_{\hat{\beta}_j}} \leq t_{1-\alpha/2}(n - k) \right) = 1 - \alpha$$

liefert

$$[\hat{\beta}_j - s_{\hat{\beta}_j} t_{1-\alpha/2}(n - k), \hat{\beta}_j - s_{\hat{\beta}_j} t_{\alpha/2}(n - k)]$$

bzw.

$$[\hat{\beta}_j - s_{\hat{\beta}_j} t_{1-\alpha/2}(n - k), \hat{\beta}_j + s_{\hat{\beta}_j} t_{1-\alpha/2}(n - k)].$$

- **Zusammenhang t -Test und Konfidenzintervall:** Da ein zweiseitiger t -Test einem F -Test entspricht, wenn die t -Statistik quadriert wird, folgt aus der hier durchgeführten Konstruktion von Konfidenzintervallen, dass die Nullhypothese eines zweiseitigen t -Tests mit Signifikanzniveau α genau dann nicht abgelehnt werden kann, wenn die Nullhypothese innerhalb des Konfidenzintervalls mit Konfidenzniveau $1 - \alpha$ liegt.

- **Bootstrapkonfidenzintervalle**

- Berechnung der kritischen Werte durch Bootstrap, siehe Abschnitt 11.5.2.
 - Wichtig: Ein Bootstrapkonfidenzintervall kann im Vergleich zu einem asymptotischen Konfidenzintervall nur dann schneller gegen das exakte Konfidenzintervall konvergieren, wenn die damit assoziierte asymptotische Verteilung der Teststatistik pivot ist!
 - Es existieren verschiedene Methoden zum Durchführen des Bootstrap
- Unterschiede ergeben sich hinsichtlich
- * der Schätzmethode für die Parameter (β, σ_0) des DGP,
 - * des Bootstrapverfahrens zum Ziehen der Fehler,
 - * der Wahl der t -Statistik oder der F -Statistik als Grundlage.
- Wird die t -Statistik verwendet, ist die Bootstrapverteilung häufig asymmetrisch und man muss die Grenzen des Konfidenzintervalls sorgfältig bestimmen, siehe Davidson & MacKinnon (2004, Section 5.3).
 - Konfidenzintervalle auf Basis der t -Statistik werden häufig als **studentized bootstrap** confidence interval oder als **percentile- t** oder **bootstrap- t** confidence interval bezeichnet.

11.6.2. Konfidenzellipsoide

- Wenn (11.39) gilt und $\mathbf{R} = \mathbf{I}_k$ gewählt wird, ergibt sich die Begrenzung des approximativen Konfidenzellipsoids aus

$$\tau(\mathbf{y}, \mathbf{X}, \beta_0) = kF_n = c_\alpha = q_{1-\alpha}.$$

- Gilt die Normalverteilung für die KQ-Schätzer exakt, dann lassen sich auch exakte Konfidenzellipsoide auf Basis der F -Statistik und dem dazugehörigen kritischen Wert aus der F -Verteilung mit q und $n - k$ Freiheitsgraden bestimmen.
- Es kann passieren, dass ein Parametervektor β in einem Konfidenzellipsoid liegt, aber nicht in den einzelnen Konfidenzintervallen für die einzelnen Elemente von β und umgekehrt (bitte graphisch verifizieren!). Ursache hierfür ist i. Allg. eine starke Kollinearität zwischen den einzelnen Parameterschätzern. Vgl. Diskussion in BA-Modul **Einführung in die Ökonometrie**.
- **Konfidenzellipse**: zweidimensionales Konfidenzellipsoid, Beispiel in Abs. 11.7.
- Es lassen sich wie im eindimensionalen Fall Konfidenzellipsoide mit Bootstrapverfahren berechnen.

11.7. Empirische Analyse von Handelsströmen: Teil 3

Fortsetzung der **Empirische Analyse von Handelsströmen: Teil 2** in Abschnitt 10.3.

Wiederholen der Schätzung des im Schritt II.3 gewählten Modell 4 (auf Basis des AIC)

$$\ln(\text{Importe}_i) = \beta_1 + \beta_2 \ln(\text{BIP}_i) + \beta_3 \ln(\text{Entfernung}_i) + \beta_4 \text{Offenheit}_i + \beta_5 \ln(\text{Flaeche}) + u_i. \quad (11.49)$$

R-Code siehe Abschnitt 10.3

Output:

```
Call:
lm(formula = mod_formula)

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
(1 observation deleted due to missingness)
Multiple R-squared:  0.9062,    Adjusted R-squared:  0.8976
F-statistic: 106.2 on 4 and 44 DF,  p-value: < 2.2e-16
```

4. Überprüfen des gewählten Modells (Teil 1):

- Testen der zugrundegelegten Modellannahmen: Entweder
 - (B1), (B3), (B4) (vgl. Abschnitt 11.3), so dass **exakte** Tests durchgeführt werden können, oder
 - (B1), (B2), (B3), (A1) und (A3), so dass **asymptotische** Tests durchgeführt werden können.
- Beispiel einer Annahmeüberprüfung: Gilt die Annahme von **homoskedastisch verteilten Fehlern (B2b)**, auch Voraussetzung für (B4)?

Plot der Residuen gegen die gefitteten Werte $\hat{y} = \mathbf{X}\hat{\beta}$ mit

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# Modell 4 wurde in Abschnitt 10.3 berechnet
resid_mod_4_kq <- mod_4_kq$resid # Residuen von Modell 4
```

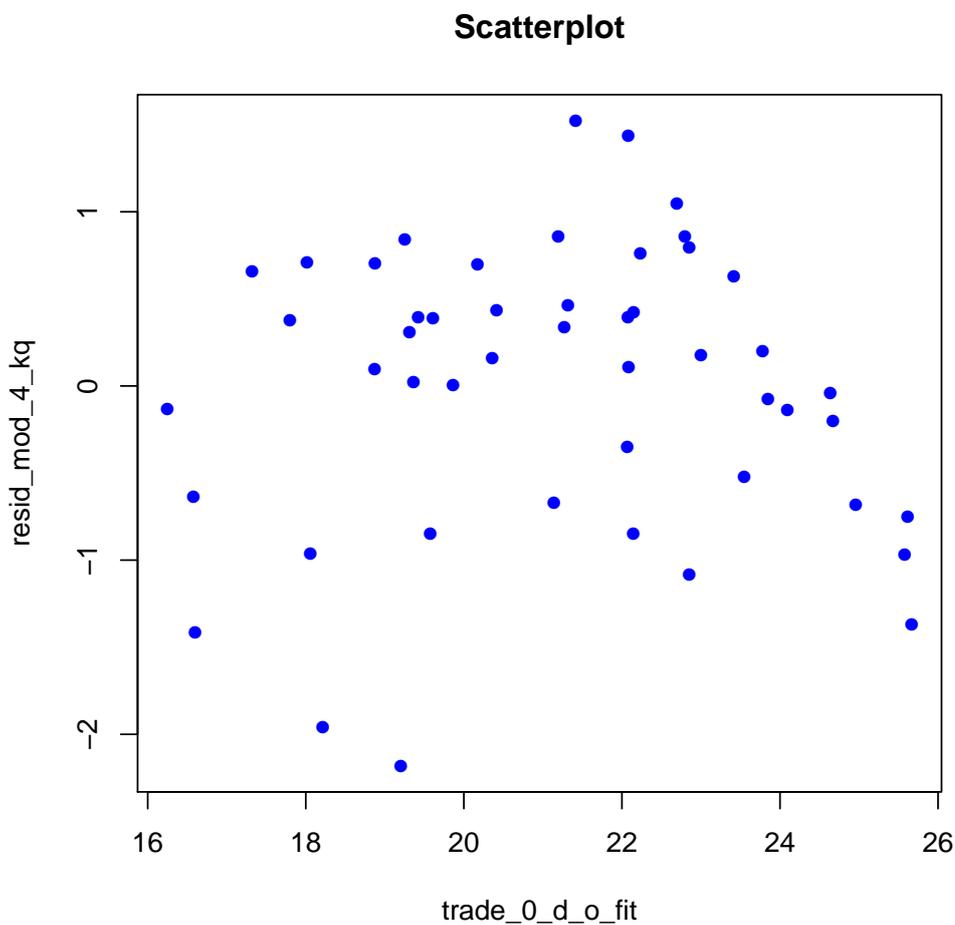
```
trade_0_d_o_fit <- mod_4_kq$fitted      # Gefittete Werte von Modell 4

# Plot der Residuen vs. der gefitteten Werte
if (save.pdf) pdf("plot_fits_vs_resids_mod_4.pdf", 6, 6)
plot(trade_0_d_o_fit, resid_mod_4_kq, col = "blue", pch = 16, main = "Scatterplot")
if (save.pdf) dev.off()
```

Listing 11.2: ./R_code/4_ff_Beispiel_Handelsstroeme.R

oder mit

```
plot(mod_4_kq, which=1)
```



Beachte:

- Residuen sind unter den oben genannten Annahmen konsistente Schätzer der Fehler, d. h.

$$\text{plim}_{n \rightarrow \infty} \hat{u}_t = u_t,$$

so dass in großen Stichproben Betrachtung der Residuen einer Betrachtung der unbekanntenen Fehler nahe kommt.

- $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{P}_{\mathbf{X}}\mathbf{y}$ liegt in dem Unterraum der \mathbf{X} .

Deshalb deutet eine Abhängigkeit der Streuung der Residuen \hat{u}_t von \hat{y}_t darauf hin, dass die Verteilung der Fehler von einem oder mehreren Regressoren abhängt.

Ursache kann

- eine Verletzung der Annahme homoskedastischer Fehler oder
- eine fehlspezifizierte Regressionsfunktion sein.

Im vorliegenden Fall optisch schwer zu sagen, daher besser: Verwendung von Heteroskedastietests, siehe Abschnitt 15.2, oder Tests auf korrekte Spezifikation der funktionalen Form, siehe Abschnitt 15.3.

- Prüfen einer möglichen Verletzung der **Annahme normalverteilter Fehler (B4)**.
 - Plot eines **Histogramms** und einer geschätzten Dichte der Residuen sowie einer Normalverteilungsdichte mit entsprechender Varianz und Berechnet verschiedener Kennzahlen mit

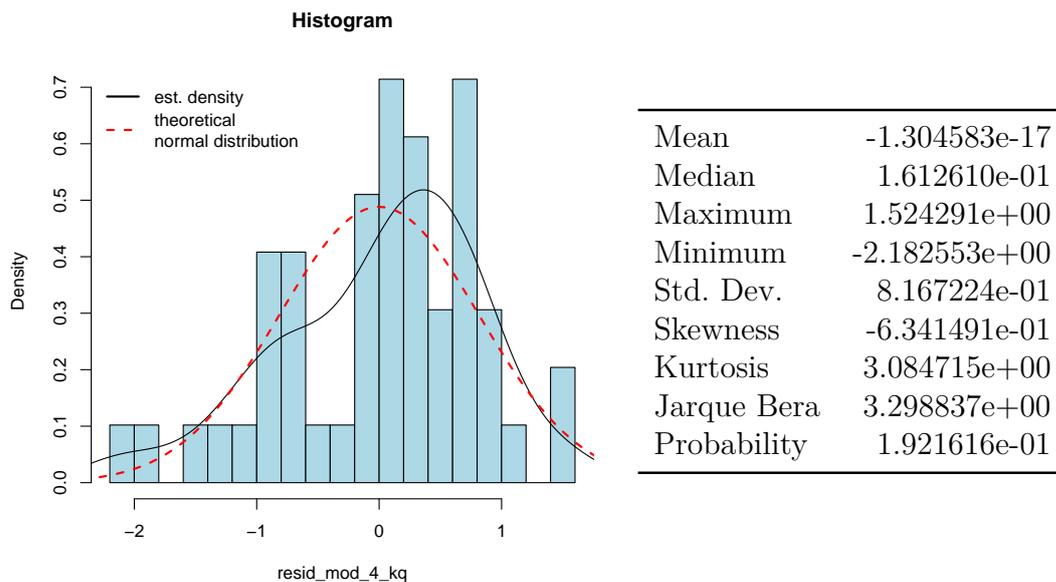
R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# Plot des Histogramms der Residuen
if (save.pdf) pdf("plot_hist_resids_mod_4.pdf", 6, 6)
hist(resid_mod_4_kq, breaks = 20, col = "lightblue", prob = T, main = "Histogram")
# Geschätzte Dichte der Residuen
lines(density(resid_mod_4_kq), col = "black", prob = T, add="T")
# Einzeichnen der zugehörigen theoretischen Normalverteilung
curve(dnorm(x, mean = mean(resid_mod_4_kq), sd = sd(resid_mod_4_kq)),
      from = -3, to = 3, add = T, col = "red", lty = 2, lwd = 2)
legend("topleft", c("est. density", "theoretical\nnormal distribution"),
      col = c("black", "red"), lwd = 2, lty = c(1,2), bty = "n")
if (save.pdf) dev.off()

# statistische Auswertung der Residuen
stats(resid_mod_4_kq)
```

Listing 11.3: ./R_code/4_ff_Beispiel_Handelsstroeme.R

- Berechnen von statistischen Kenngrößen der Residuen, inklusive eines **Normalitätstest** ((Lominicki-)Jarque-Bera-Test, siehe Abschnitt 15.4). Der (Lominicki-)Jarque-Bera-Test kann mit R-Befehl `jarque.test(model_kq)` (erfordert R-Paket `moments`) berechnet werden.



Je kleiner p -Wert des Normalitätstests, desto eher kann man bei streng exogenen Regressoren erwarten, dass Approximationsfehler der asymptotischen Normalverteilung gering ist.

p -Wert des (Lomnicki-)Jarque-Bera-Tests widerspricht optischem Eindruck: Annahme normalverteilter Fehler wird nicht abgelehnt, da p -Wert zu groß.

- noch besser als Histogramm: Plot einer geschätzten **Dichte** und Vergleich mit Dichte der Normalverteilung mit der geschätzten Fehlervarianz.

Siehe Abschnitt 15.7 für Fortsetzung der Modellüberprüfung **Überprüfen des gewählten Modells (Teil 2)**.

5. Verwenden des überprüften Modells: Konfidenzintervalle und Durchführen von Tests:

• Konfidenzintervalle

- Wahl eines **Konfidenzniveaus** $1 - \alpha$, im Folgenden 95%.
- Berechnen der Konfidenzintervalle aller geschätzten Regressionsparameter mit

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
#### Konfidenzintervalle
confint(mod_4_kq)
```

Listing 11.4: ./R_code/4_ff_Beispiel_Handelsstroeme.R

liefert:

```
2.5 %      97.5 %
(Intercept) -1.87014867  6.72570228
log(wdi_gdpusdcr_o) 0.87076944  1.17927579
log(cepii_dist) -1.20331827 -0.57397436
```

```
ebrd_tfes_o      -0.06286079  0.76916951
log(cepii_area_o) -0.32280233  0.02074077
```

- **Zweiseitiger Test**

- Statistisches **Hypothesenpaar**:

H_0 : Die BIP-Elastizität der Importe ist 1. versus H_1 : Die Elastizität ist ungleich 1.
 $H_0 : \beta_2 = 1$ versus $H_1 : \beta_2 \neq 1$.

- Wähle **Signifikanzniveau**, z. B. $\alpha = 0.05$.

Berechnen der (approximativen) **kritischen Werte**: $n - k = 49 - 5 = 44$ Freiheitsgrade. Da die t -Statistik unter strengen Annahmen exakt t -verteilt ist, aber unter schwächeren Annahmen die t -Verteilung eine gute Approximation darstellt, werden die (approximativen) kritischen Werte auf Basis der t -Verteilung bestimmt:

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# Bestimmen der kritischen Werte
alpha <- 0.05
qt(alpha/2, mod_4_kq$df)
qt(1-alpha/2, mod_4_kq$df)
```

Listing 11.5: ./R_code/4_ff_Beispiel_Handelsstroeme.R

$$-c = -2.015368, \quad c = 2.015368$$

- Berechne **t -Statistik** aus der passenden Zeile des Outputs

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
log(wdi_gdpusdcr_o)  1.02502    0.07654  13.392 < 2e-16 ***
```

mit R-Befehl

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
(t <- (coefficients(mod_4_kq)["log(wdi_gdpusdcr_o)"]-1)/
sqrt(vcov(mod_4_kq)["log(wdi_gdpusdcr_o)", "log(wdi_gdpusdcr_o)"]))
```

Listing 11.6: ./R_code/4_ff_Beispiel_Handelsstroeme.R

$$t_{\beta_2} = \frac{\hat{\beta}_2 - \beta_{2,H_0}}{s_{\hat{\beta}_2}} = \frac{1.02502 - 1}{0.07654} = 0.3269286$$

- **Entscheidung**: Da Teststatistik

$$-c < t_{\beta_2} < c \\ -2.015368 < 0.3269286 < 2.015368$$

außerhalb des (approximativen) kritischen Bereichs liegt, Nullhypothese nicht ablehnen.

- (Approximativer) **p-Wert** ist 0.7452378 berechnet mit

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# p-value
2*pt(-abs(t),mod_4_kq$df)
```

Listing 11.7: ./R_code/4_ff_Beispiel_Handelsstroeme.R

Da p -Wert größer als gewähltes Signifikanzniveau, lässt sich H_0 nicht ablehnen (selbst dann nicht, wenn ein Signifikanzniveau von 10% gewählt werden würde).

Der p -value besagt, dass man unter H_0 in etwa 75 von 100 Stichproben eine t -Statistik erhalten würde, deren Absolutbetrag mindestens 0.33 beträgt.

- Falls bereits (approximatives) **Konfidenzintervall** zu β_2 mit Konfidenzniveau $1 - \alpha$ vorliegt: Liegt Wert der Nullhypothese im Konfidenzintervall, wird H_0 nicht abgelehnt.

Alternativ und schneller: Verwenden des R-Befehls `linearHypothesis` (erfordert R-Paket `car`).

Beachte: es berechnet $F = t^2$, p -Werte auf Basis der $F_{1,n-k}$ -Verteilung

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
library("car")
(F_stat <- linearHypothesis(mod_4_kq,c("log(wdi_gdpusdcr_o)=1")))
```

Listing 11.8: ./R_code/4_ff_Beispiel_Handelsstroeme.R

```
Linear hypothesis test

Hypothesis:
log(wdi_gdpusdcr_o) = 1

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      45 32.095
2      44 32.018  1  0.077776 0.1069 0.7453
```

• Einseitiger Test

- Man kann bzgl. des Modell (11.49) auch eine Hypothese hinsichtlich eines negativen Einflusses von *Entfernung* auf *Importe* aufstellen. Da man Evidenz für $\beta_3 < 0$ erhält, indem $\beta_3 \geq 0$ statistisch verworfen wird, lautet das **Hypothesenpaar**:

$$H_0 : \beta_3 \geq 0 \quad \text{versus} \quad H_1 : \beta_3 < 0.$$

- Wahl eines **Signifikanzniveaus** von $\alpha = 0.05$ und Berechnen des (approximativen) **kritischen Werts**. Beachte, dass nur der **linke** kritische Wert benötigt wird, da

der Parameterbereich der Alternativhypothese links vom Parameterbereich der Nullhypothese liegt und damit auch der kritische Bereich links vom Nichtablehnbereich liegt:

R-Code (Ausschnitt aus R-Programm in Abschnitt [A.4](#))

```
alpha <- 0.05
(qt(alpha, mod_4_kq$df))
```

Listing 11.9: ./R_code/4_ff_Beispiel_Handelsstroeme.R

liefert -1.68023.

– Die **t-Statistik** ist im R-Output enthalten:

```
Coefficients:
Estimate Std. Error t value Pr(>|t|)
log(cepii_dist) -0.88865 0.15614 -5.691 9.57e-07 ***
```

bzw. ergibt sich im vorliegenden Fall folgendermaßen:

R-Code (Ausschnitt aus R-Programm in Abschnitt [A.4](#))

```
# t-Teststatistik
(t <- summary(mod_4_kq)$coefficients["log(cepii_dist)", 3])
```

Listing 11.10: ./R_code/4_ff_Beispiel_Handelsstroeme.R

$$t_{\beta_3} = -5.691.$$

– **Entscheidung:** Wegen

$$t_{\beta_3} = -5.691 < -1.68023 = c,$$

Ablehnen der Nullhypothese

– **p-Wert:**

R-Code (Ausschnitt aus R-Programm in Abschnitt [A.4](#))

```
# p-Value
(pt(t, mod_4_kq$df))
```

Listing 11.11: ./R_code/4_ff_Beispiel_Handelsstroeme.R

ergibt 4.783876e-07. Die Entfernung hat also selbst auf dem 1% Signifikanzniveau negativen Einfluss.

– **Interpretation:** Steigt die Entfernung um 1%, dann fallen ceteris paribus die erwarteten Importe nach Deutschland um ca. 0.9%.

• **F-Test: Testen gemeinsamer Hypothesen:**

- **Frage:** Im Modell (11.49) sind die Parameter der Variablen Offenheit und Fläche für das gewählte Signifikanzniveau von 5% nicht statistisch signifikant. Kann es jedoch sein, dass beide Parameter **gemeinsam** statistisch signifikant sind? Das **Hypothesenpaar** lautet:

$$H_0 : \beta_4 = 0 \quad \text{und} \quad \beta_5 = 0 \quad \text{versus} \\ H_1 : \beta_4 \neq 0 \quad \text{und/oder} \quad \beta_5 \neq 0.$$

- Wahl des **Signifikanzniveaus:** $\alpha = 0.05$ und der (approximativen) **kritischen Werte**. Der **kritische Bereich** liegt rechts vom kritischen Wert.

Bei Verwendung der F -Statistik, Bestimmen des (approximativen) kritischen Wertes auf Basis der $F_{2,44}$ -Verteilung ergibt 3.209278 mit

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# kritischer Wert für F-Statistik
(qf(1-alpha,2,mod_4_kq$df))
```

Listing 11.12: ./R_code/4_ff_Beispiel_Handelsstroeme.R

Bei Verwendung der χ^2 -Statistik, Bestimmen des (approximativen) kritischen Wertes auf Basis der $\chi^2(2)$ -Verteilung ergibt sich 5.991465 mit

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# kritischer Wert für chi^2-Statistik
(qchisq(1-alpha,2))
```

Listing 11.13: ./R_code/4_ff_Beispiel_Handelsstroeme.R

- **Berechnen der F -Statistik** und des **p -Werts** erfolgt am einfachsten mit dem R-Befehl `linearHypothesis` (erfordert R-Paket `car`):

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
# F-Statistik
(F2_stat <- linearHypothesis(mod_4_kq,c("ebrd_tfes_o=0","log(cepii_area_o)=0"),
```

Listing 11.14: ./R_code/4_ff_Beispiel_Handelsstroeme.R

und liefert

```
Linear hypothesis test

Hypothesis:
ebrd_tfes_o = 0
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 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

```

Berechnen der χ^2 -Statistik und des p -Werts erfolgt am einfachsten mit dem R-Befehl `linearHypothesis` (erfordert R-Paket `car`):

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```

# chi^2-Statistik
(Chisq_stat <- linearHypothesis(mod_4_kq,c("ebrd_tfes_o=0","log(cepii_area_o)=0"),

```

Listing 11.15: `./R_code/4_ff_Beispiel_Handelsstroeme.R`

und liefert

```

Linear hypothesis test

Hypothesis:
ebrd_tfes_o = 0
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  Chisq Pr(>Chisq)
1      46 39.645
2      44 32.018 2      7.6272 10.482  0.005296 **
----
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

In beiden Fällen liegt die jeweilige Teststatistik im kritischen Bereich, so dass die Nullhypothese auf dem 5%-Signifikanzniveau abgelehnt wird. Auf Basis beider p -Werte lässt sich sehen, dass die Nullhypothese auch auf dem 1%-Signifikanzniveau abgelehnt wird.

- **Interpretation:** Mindestens eine der beiden Variablen Offenheit oder logarithmierte Fläche hat einen Einfluss auf die Exporte nach Deutschland. Eine mögliche Ursache für die unterschiedlichen Testergebnisse der Einzeltests und der gemeinsamen Tests ist die Korrelation von 0.42 zwischen den Parameterschätzern, siehe unten.

• Korrelationsmatrix der Parameterschätzungen

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```

# Kovarianzmatrix
(cov_par <- vcov(mod_4_kq))
# Korrelationsmatrix
(corr_par <- cov2cor(cov_par))

```

Listing 11.16: `./R_code/4_ff_Beispiel_Handelsstroeme.R`

liefert

```

(Intercept) log(wdi_gdpusdcr_o) log(cepii_dist) ebrd_tfes_o log(cepii_area_o)
(Intercept)          1.00000000          -0.6077120          -0.71380469          -0.26879925
0.08289662
log(wdi_gdpusdcr_o) -0.60771198              1.0000000          0.30644626          -0.41648145
-0.62198317
log(cepii_dist)      -0.71380469              0.3064463          1.00000000          0.09807673
-0.29518939
ebrd_tfes_o          -0.26879925              -0.4164814          0.09807673          1.00000000
0.42127548
log(cepii_area_o)    0.08289662              -0.6219832          -0.29518939          0.42127548
1.00000000

```

- **Konfidenzellipse** für β_4 und β_5 :

- Wähle Konfidenzniveau, hier 95%.
- Wähle zwei Parameter, hier β_4 und β_5 .
- Verwende R-Befehl `confidenceEllipse()` (erfordert R-Paket `car`):

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

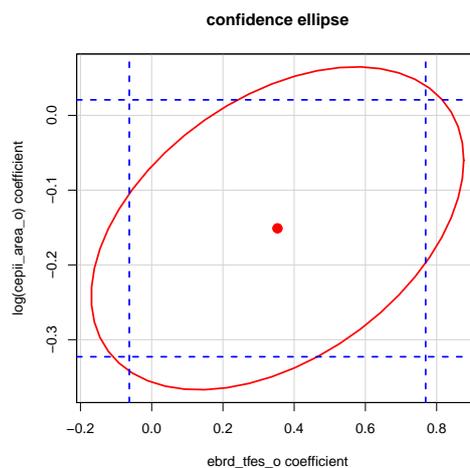
```

# Konfidenzellipse
if (save.pdf) pdf("plot_conf_ellipse.pdf", 6, 6)
confidenceEllipse(mod_4_kq, which.coef = c(4, 5), levels = 0.95,
                 main = "confidence ellipse", col = "red")
# Konfidenzintervall
abline(v = confint(mod_4_kq, "ebrd_tfes_o", level = 0.95), lty = 2,
       col = "blue", lwd = 2)
abline(h = confint(mod_4_kq, "log(cepii_area_o)", level = 0.95), lty = 2,
       col = "blue", lwd = 2)
if (save.pdf) dev.off()

```

Listing 11.17: ./R_code/4_ff_Beispiel_Handelsstroeme.R

liefert



12. Univariate Zeitreihenmodelle

In diesem Kapitel geht es hauptsächlich um Modelle für univariate Zeitreihen. Zeitreihen sind Stichproben, in denen die Beobachtungen über die Zeit hinweg vorliegen. Eine Zeitreihe ist univariat, wenn genau eine Variable über die Zeit hinweg vorliegt.

Beispiel: Das ifo Geschäftsklima: Abbildung 12.1 zeigt die Zeitreihen der monatlichen Indices der ifo Geschäftserwartungen, der ifo Geschäftsbeurteilung und des ifo Geschäftsklimas für den Zeitraum Januar 1991 bis September 2015.

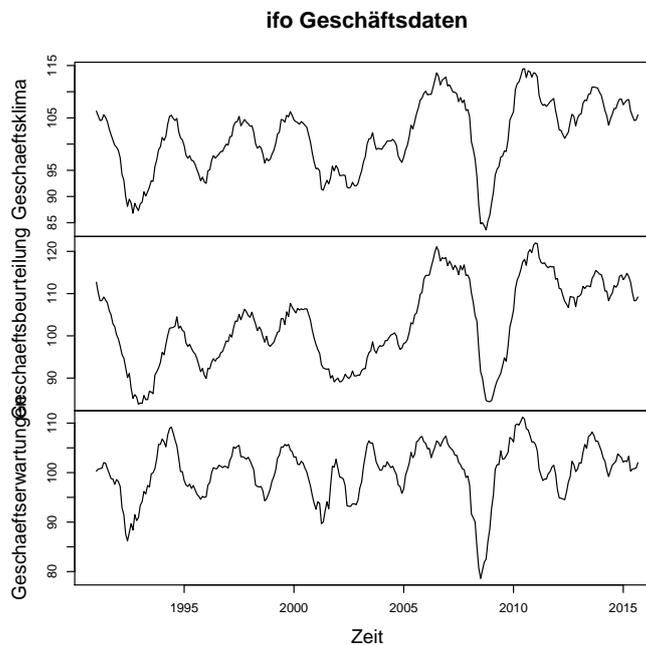


Abbildung 12.1.: Die ifo Geschäftserwartungen, die ifo Geschäftsbeurteilung und das ifo Geschäftsklima für die gewerbliche Wirtschaft (Januar 1991 bis November 2014) (R-Programm, siehe Abschnitt A.9, Seite 354)

Zeitreihen sind Realisationen von DGPs, die dann als stochastische Prozesse bezeichnet werden. Letztere werden im Folgenden genauer betrachtet. Modelle für univariate Zeitreihen sind Mengen, die stochastische Prozesse enthalten.

12.1. Stochastische Prozesse

Stochastischer Prozess

Ein stochastischer Prozess $\{y_t\}_{t \in \mathbb{T}}$ ist eine Folge von Zufallsvariablen

$$\{y_t | t \in \mathbb{T}\} \equiv \{y_t(\omega) | t \in \mathbb{T}, \omega \in \Omega\} \equiv \{y(t, \omega) | t \in \mathbb{T}, \omega \in \Omega\} : \Omega \times \mathbb{T} \rightarrow \mathbb{R}^{\mathbb{T}}, \quad (12.1)$$

die auf einer Ergebnismenge Ω und einer vorgegebenen Indexmenge \mathbb{T} definiert sind (Hassler (2007, Abschnitt 2.3), Mikosch (1998, Section 1.2)).

Bemerkungen:

- Englische Bezeichnungen: stochastic process, random process, random sequence (Davidson 2000, Section 4.1).
- Weitere gebräuchliche Schreibweisen sind: $\{y_t\}_{t \in \mathbb{T}}$ oder auch ohne Angabe der Indexmenge $\{y_t\}$.
- Stellt der Index t die Zeit dar, wird ein stochastischer Prozess auch als **Zeitreihenprozess** bezeichnet:
 - **Zeitstetige Prozesse:** \mathbb{T} ist ein Intervall in \mathbb{R} .
 - **Zeitdiskrete Prozesse:** \mathbb{T} ist eine endliche oder abzählbar unendliche Menge, typischerweise ist $\mathbb{T} = \mathbb{Z}$ oder $\mathbb{T} = \mathbb{N}$.

Bei zeitdiskreten Prozessen unterscheidet man zeitdiskrete stochastische Prozesse mit

* **regelmäßiger Beobachtungsfrequenz:**

Beispiele: monatliche Beobachtungen des ifo Geschäftsklimas, jährliche Wachstumsraten des BIP, wöchentliche Beobachtungen des DAX.

* **unregelmäßiger Beobachtungsfrequenz:**

Beispiel: Reuter's Tickerdaten.

- Univariate und multivariate stochastische Prozesse:
 - **Univariater stochastischer Prozess:** y_t ist eine skalare Zufallsvariable

Beispiel: Beobachtungen des ifo Geschäftsklimaindex.
 - **Multivariater stochastischer Prozess:** \mathbf{y}_t ist ein Zufallsvektor.

Beispiel: $\mathbf{y}_t = \begin{pmatrix} \text{ifo Geschäftsklima}_t \\ \text{ifo Geschäftsbeurteilung}_t \\ \text{ifo Geschäftserwartungen}_t \end{pmatrix}$.

In diesem Kapitel betrachten wir fast ausschließlich zeitdiskrete univariate stochastische Pro-

zesse mit regelmäßiger Beobachtungsfrequenz. In Kapitel 13 werden Modelle für multivariate stochastische Prozesse genauer betrachtet.

- Literatur zu allgemeinen **Existenzbedingungen** für stochastische Prozesse ist in Hassler (2007, Abschnitt 2.3, Fußnote 9) angegeben.
- Wichtig: Ein stochastischer Prozess ist eine Funktion von 2 Variablen:
 - Für eine gegebene Zeitperiode t_0 ist

$$y_{t_0} = y(t_0, \omega), \quad \omega \in \Omega,$$

eine Zufallsvariable. Die Erwartungswerte $E[y_t] = \mu_t, t \in \mathbb{T}$, heißen **Ensemblemittelwerte**. Abbildung 12.2 zeigt für jede Zeitperiode t verschiedene Realisationen.

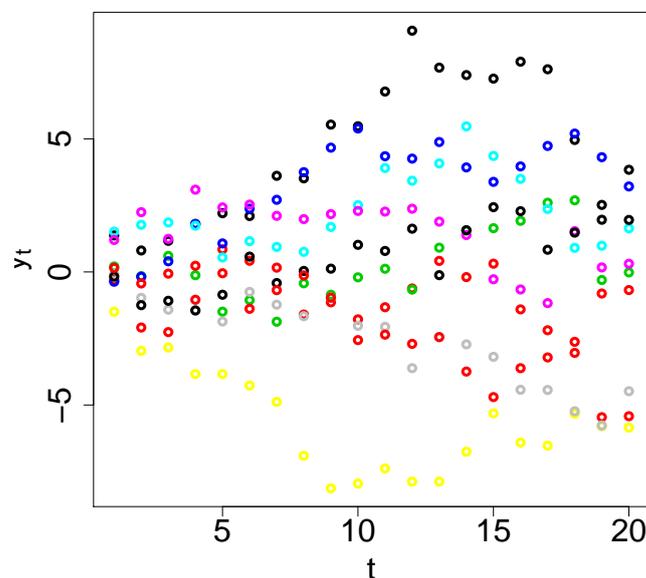


Abbildung 12.2.: Zehn verschiedene Realisationen für jede Zeitperiode t eines stochastischen Prozesses (**R-Programm**, siehe Abschnitt A.10, 355)

- Für ein gegebenes Elementarereignis ω_0 ist

$$y_t = y(t, \omega_0), \quad t \in \mathbb{T},$$

eine Funktion der Zeit.

Die Funktion wird dann eine **Realisation**, eine **Trajektorie** oder ein **Pfad** des stochastischen Prozesses $\{y_t\}$ genannt. Abbildung 12.3 zeigt verschiedene Trajektorien eines stochastischen Prozesses. Manche Autoren bezeichnen ausschließlich die Realisation eines stochastischen Prozesses als Zeitreihe (Hassler 2007, Abschnitt 2.3).

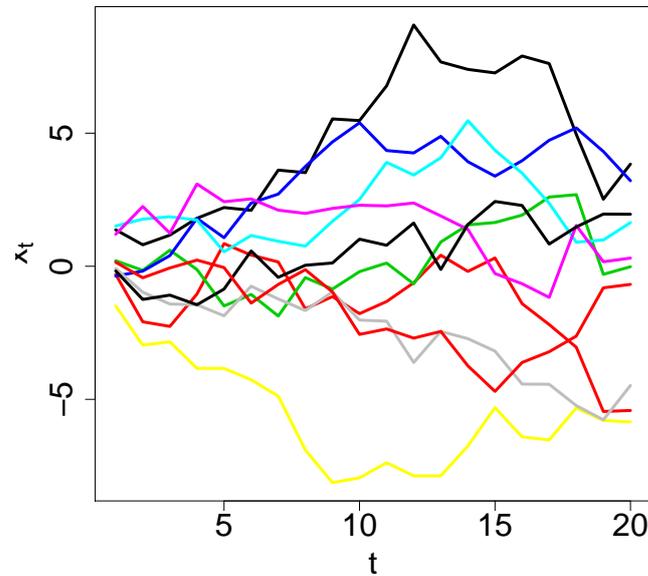


Abbildung 12.3.: Zehn verschiedene Trajektorien eines stochastischen Prozesses (**R-Programm**, siehe Abschnitt A.10, 355)

Zusammenfassung Realisation

- einer Zufallsvariable: Zahl
- eines stochastischen Prozesses: Trajektorie, Pfad: Funktion der Zeit t bzw. eine Folge von reellen Zahlen.

DGPs, gemeinsame und bedingte Dichten für stochastische Prozesse

• Univariate stochastische Prozesse

Der DGP eines univariaten stochastischen Prozesses $\{y_t | t \in \mathbb{T}\}$, $\mathbb{T} = \{1, 2, \dots, T\}$ für T mögliche Stichprobenbeobachtungen (y_1, y_2, \dots, y_T) ist vollständig bestimmt durch die gemeinsame Dichte $f_{Y_1, Y_2, \dots, Y_T}(y_1, y_2, \dots, y_T)$, die wiederum als Produkt von bedingten Dichten dargestellt werden kann (vgl. für multivariate stochastische Prozesse (5.2) in Abschnitt 5.1):

$$f_{Y_1, Y_2, \dots, Y_n}(y_1, y_2, \dots, y_n) = \prod_{t=1}^T f_{Y_t | Y_{t-1}, \dots, Y_1}(y_t | y_{t-1}, \dots, y_1). \quad (12.2)$$

- **Multivariate stochastische Prozesse**

(Vgl. (5.2) in Abschnitt 5.1):

$$f_{Y_1, Y_2, \dots, Y_T}(\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_T) = \prod_{t=1}^T f_{Y_t | Y_{t-1}, \dots, Y_1}(\mathbf{y}_t | \mathbf{y}_{t-1}, \dots, \mathbf{y}_1) \quad (5.2)$$

Vollständig und partiell spezifizierte Zeitreihenmodelle

- Sind die bedingten Dichten $f_{Y_t | Y_{t-1}, \dots, Y_1}(y_t | y_{t-1}, \dots, y_1)$ bzw. $f_{Y_t | Y_{t-1}, \dots, Y_1}(\mathbf{y}_t | \mathbf{y}_{t-1}, \dots, \mathbf{y}_1)$, $t = 1, \dots, T$, bekannt, kennt man den DGP des stochastischen Prozesses.

- **Vollständig spezifizierte Modelle**

Zeitreihenmodelle, bei denen die bedingten Dichten modelliert werden, sind **vollständig spezifiziert**.

- **Partiell spezifizierte Modelle**

Häufig ist man nur an einzelnen Charakteristika der bedingten Dichten interessiert, typischerweise dem bedingten Erwartungswert oder der bedingten Varianz. Dann ist es im Allgemeinen ausreichend, Modelle mit partiell spezifizierten stochastischen Prozessen zu verwenden:

– im univariaten Fall $E[y_t | y_{t-1}, \dots, y_1]$:

* Lineare stochastische Prozesse → Abschnitt 12.2

* Moving-Average-Prozesse → Abschnitt 12.2

* Autoregressive Prozesse → Abschnitt 12.3

* Autoregressive Integrierte Moving-Average-Prozesse → Abschnitt 12.3

* Nichtlineare autoregressive Prozesse → Beispiele in **Fortgeschrittene Ökonometrie**

* ...

– im multivariaten Fall $E[\mathbf{y}_t | \mathbf{y}_{t-1}, \dots, \mathbf{y}_1]$:

* Vektorautoregressive Prozesse → **Quantitative Wirtschaftsforschung II**

* Vektorfehlerkorrekturprozesse → **Quantitative Wirtschaftsforschung II**

* ...

– im univariaten Fall $Var(y_t | y_{t-1}, \dots, y_1)$:

- * Autoregressive Conditional Heteroskedastic Processes (ARCH-Prozesse) → **Applied Financial Econometrics**
- * Generalized Autoregressive Conditional Heteroskedastic Processes (GARCH-Prozesse) → **Applied Financial Econometrics**

Motivation der folgenden Definitionen:

Lässt sich Ensemblemittelwert μ_t schätzen?

- Würden R Realisationen $y_t^{(r)}$ für y_t vorliegen, könnten wir Schätzer des Erwartungswertes

$$\hat{\mu}_t = \frac{1}{R} \sum_{r=1}^R y_t^{(r)}$$

verwenden.

Problem: In der Praxis ist typischerweise $R = 1$.

- **Zentrale Frage:** Unter welchen Bedingungen lässt sich Ensemblemittelwert μ_t durch den Zeitmittelwert

$$\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t \quad (12.3)$$

schätzen? Beachte: hier $R = 1$.

- Antwort erfordert Maße zur Quantifizierung der stochastischen Abhängigkeiten zwischen Beobachtungen zu verschiedenen Zeitpunkten sowie Konzepte, die eine Konstanz von relevanten Eigenschaften stochastischer Prozesse über die Zeit hinweg definieren (u. a. des Ensemblemittelwertes $\mu_t = \mu$). Diese werden im Folgenden eingeführt. Die Antworten selbst finden sich in Abschnitt [12.4.1](#).

Messung der zeitlichen stochastischen Abhängigkeiten

Zur Messung der Abhängigkeitsstruktur, die durch die bedingten Dichten charakterisiert wird, werden typischerweise folgende Maße verwendet (Darstellung für univariate stochastische Prozesses):

- Autokovarianzfunktion
- Autokorrelationsfunktion
- Partielle Autokorrelationsfunktion

Konzepte, die eine Konstanz von relevanten Eigenschaften stochastischer Prozesse über die Zeit hinweg definieren

- Mittelwertstationarität
- schwache Stationarität

- strenge Stationarität

Autokovarianzfunktion

Die Autokovarianzfunktion eines stochastischen Prozesses $\{y_t | t \in \mathbb{T}\}$ ist für alle $t, t-k \in \mathbb{T}$, k ganzzahlig, definiert als

$$\begin{aligned} Cov(y_t, y_{t-k}) &\equiv E[(y_t - E[y_t])(y_{t-k} - E[y_{t-k}])] = E[y_t y_{t-k}] - E[y_t]E[y_{t-k}] \\ &= E[y_t y_{t-k}] - \mu_t \mu_{t-k}. \end{aligned} \quad (12.4)$$

Beachte: aus $E(y_t | y_{t-k}) = E(y_t)$ folgt $Cov(y_t, y_{t-k}) = 0$.

Autokorrelationsfunktion

Die Autokorrelationsfunktion eines stochastischen Prozesses $\{y_t | t \in \mathbb{T}\}$ ist für alle $t, t-k \in \mathbb{T}$, k ganzzahlig, definiert als

$$Corr(y_t, y_{t-k}) \equiv \frac{Cov(y_t, y_{t-k})}{\sqrt{Var(y_t) Var(y_{t-k})}}.$$

Partielle Autokorrelationsfunktion

Die partielle Autokorrelationsfunktion gibt die bedingte Autokorrelation zwischen y_t und y_{t-k} an, wobei die Bedingung sich aus allen Beobachtungen zusammensetzt, die zwischen den Perioden t und $t-k$ liegen, also $y_{t-1}, \dots, y_{t-k+1}$,

$$Corr(y_t, y_{t-k} | y_{t-1}, \dots, y_{t-k+1}).$$

Ein Beispiel für eine partielle Autokorrelationsfunktion ist (12.29) in Abschnitt 12.3.3.

Stationaritätskonzepte

Mittelwertstationärer Prozess

$\{y_t\}$ ist **mittelwertstationär**, wenn gilt

$$\mu_t = \mu \quad \text{für alle } t \in \mathbb{T}. \quad (12.5)$$

Über die Autokovarianzen wird nichts angenommen.

Schwach stationärer bzw. kovarianzstationärer Prozess

Ein univariater stochastischer Prozess $\{y_t | t \in \mathbb{T}\}$ wird als (**schwach**) **stationär** oder **kovarianzstationär** bezeichnet, wenn folgende Eigenschaften bzgl. der ersten beiden Momente erfüllt sind:

- $E[y_t] = \mu$ für alle $t \in \mathbb{T}$,

- $Cov(y_t, y_{t-k}) = \gamma_k$, für alle $t, t-k \in \mathbb{T}$,

d. h. der Mittelwert hängt nicht von der Zeitperiode t ab und die Autokovarianzfunktion hängt ausschließlich vom Lag k , nicht jedoch von der Zeitperiode t ab.

Folgerungen:

- Schwach stationäre Prozesse sind homoskedastisch, da für $k = 0$ gilt $Var(y_t) = \gamma_0$.
- Bei schwach stationären Prozessen gilt für die Autokorrelationsfunktion $\rho_k \equiv Corr(y_t, y_{t-k})$:

$$\rho_k = \gamma_k / \gamma_0.$$

Strenge Stationarität

Die Definition erfolgt hier für multivariate stochastische Prozesse: Ein multivariater stochastischer Prozess $\{\mathbf{y}_t\}$ wird als **streng stationär (strictly stationary)** bezeichnet, wenn für jede beliebige Menge an Zeitindizes $t_1 < t_2 < \dots < t_m$ die gemeinsame Wahrscheinlichkeitsverteilung für $(\mathbf{y}_{t_1}, \mathbf{y}_{t_2}, \dots, \mathbf{y}_{t_m})$ und die gemeinsame Wahrscheinlichkeitsverteilung für $(\mathbf{y}_{t_1+k}, \mathbf{y}_{t_2+k}, \dots, \mathbf{y}_{t_m+k})$ für beliebige ganzzahlige k gleich sind.

Beispiele für streng stationäre univariate Prozesse

-

$$\begin{pmatrix} y_t \\ y_{t-1} \end{pmatrix} \sim N \left(\underbrace{\begin{pmatrix} 0 \\ 0 \end{pmatrix}}_{\boldsymbol{\mu}}, \underbrace{\begin{pmatrix} 1 & 0.8 \\ 0.8 & 1 \end{pmatrix}}_{\boldsymbol{\Sigma}} \right)$$

und

$$\begin{pmatrix} y_{t+k} \\ y_{t+k-1} \end{pmatrix} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

haben die gleiche bivariate Normalverteilung für beliebige $t, t-1, t+k, t+k-1 \in \mathbb{T}$.

- ‡ Gegeben sei die Zufallsvariable z_1 mit einer beliebigen Verteilung. Dann ist

$$z_t = z_1, \quad t = 2, 3, \dots,$$

ein streng stationärer Prozess, wobei alle Autokorrelationen für $k \neq 0$ Eins sind (Hayashi 2000, Example 2.2).

Stochastische Prozesse ohne Autokorrelationen

Weißes Rauschen (White Noise)

$u_t \sim \text{WN}(0, \sigma^2)$ bedeutet für alle $t \in \mathbb{T}$:

- $E[u_t] = 0$,
- $\text{Var}(u_t) = E[u_t^2] = \sigma^2$,
- $\text{Cov}(u_t, u_{t-k}) = 0$ für $k \neq 0$.

Die Bedingungen bedeuten, dass der unbedingte Mittelwert von u_t für jede Periode Null ist und keine Heteroskedastie vorliegt.

Beachte: Es wird *keine* Annahme über die Verteilung der u_t 's gemacht, sondern es werden lediglich die ersten beiden Momente spezifiziert.

Unabhängiges Weißes Rauschen

Eine Folge von IID Zufallsvariablen wird als **IID-Prozess** oder **unabhängiges Weißes Rauschen** bezeichnet:

$$u_t \sim \text{IID}(0, \sigma^2), \quad t \in \mathbb{T}.$$

D. h., es hilft nichts, u_{t-k} zu beobachten, um die Wahrscheinlichkeit zu präzisieren, dass eine Realisation von u_t in einem bestimmten Intervall auftritt.

Beachte: Es wird auch keine Annahme über die Verteilung der u_t 's gemacht.

Gaußsches Weißes Rauschen (Gaussian White Noise)

Fügt man zum unabhängigen Weißen Rauschen eine Normalverteilungsannahme hinzu, so erhält man Gaußsches Weißes Rauschen:

$$u_t \sim \text{NID}(0, \sigma^2), \quad t \in \mathbb{T},$$

bzw.

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

Beachte: Unabhängiges Weißes Rauschen und Gaußsches Weißes Rauschen sind beide streng stationär.

R-Befehle

Erzeugen von Gaußschem Weißem Rauschem mit `rnorm()`.

‡ Beispiele für Weißes Rauschen, die jedoch kein unabhängiges Weißes Rauschen sind

- (Hayashi 2000, Example 2.4): Die Zufallsvariable w sei auf $[0, 2\pi]$ gleichverteilt und $u_t = \cos(tw)$, $t = 1, 2, \dots$. Damit ist $E[u_t | u_{t-k}] \neq 0$, $t - k$, $k > 0$, da alle u_t von ω beeinflusst werden. Es kann kein unabhängiges Weißes Rauschen vorliegen. Trotzdem gilt $E(u_t) = 0$, $Var(u_t) = 1/4$, $Cov(u_t, u_{t-k}) = 0$, $k \neq 0$, so dass Weißes Rauschen vorliegt.
- Ein einfacher Prozess mit bedingter Heteroskedastie, nämlich ein ARCH(1)-Prozess \rightarrow **Applied Financial Econometrics**.

Im Folgenden werden (partiell spezifizierte) univariate lineare stochastische Prozesse behandelt, die Autokorrelationen ungleich Null zulassen.

Dabei wird nur eine kurze Einführung gegeben. Sehr gute, detaillierte Lehrbücher sind: Hamilton (1994), Kirchgässner et al. (2013), Neusser (2009) und die anwendungsorientierte Darstellung in Lütkepohl & Kraetzig (2008).

12.2. Lineare stochastische Prozesse und Moving-Average-Prozesse

Linearer Prozess

Ein stochastischer Prozess $\{y_t\}_{t \in \mathbb{Z}}$ heißt **linearer Prozess**, falls er folgende Darstellung besitzt (Neusser (2006, Definition 2.4), Brockwell & Davis (1991, Section 11.1, p. 404))

$$y_t = \sum_{j=-\infty}^{\infty} \psi_j u_{t-j} \quad (12.6)$$

mit den Parametern $\psi_j \in \mathbb{R}$, $j \in \mathbb{Z}$, und

$$u_t \sim WN(0, \sigma^2), \quad (12.7a)$$

$$\sum_{j=-\infty}^{\infty} |\psi_j| < \infty. \quad (12.7b)$$

Beispiel: $y_t = \psi_0 u_t + \psi_{-1} u_{t+1} + \psi u_{t-1}$

Bemerkungen:

- Ohne Spezifizierung der Verteilung von u_t ist ein linearer stochastischer Prozess nur partiell spezifiziert (vgl. Weißes Rauschen).
- **Beachte: Bei unendlichen Summen von Zufallsvariablen ist das Vertauschen von Erwartungswert und Summierung im Allgemeinen nicht möglich.**

Ein Vertauschen von Erwartungswert und unendlicher Summe ist **nur dann möglich, wenn** für die unendliche Summe ein geeignet definierter Grenzwert existiert.

Obige Bedingung (12.7b) ist hinreichend dafür, dass die unendliche Summe von Zufallsvariablen $\sum_{j=-\infty}^{\infty} \psi_j u_{t-j}$ gegen einen wohldefinierten Grenzwert konvergiert, der mit y_t

bezeichnet wird. Die Konvergenz erfolgt **mit Wahrscheinlichkeit Eins** (Brockwell & Davis 1991, Proposition 3.3.1, S. 83).

Damit kann bei Gültigkeit von (12.7b) Erwartungswert und Summierung vertauscht werden. Diese Bedingung kann noch etwas abgeschwächt werden; siehe Appendix A.2 in **Fortgeschrittene Ökonometrie**.

- Unter allen genannten Bedingungen ist der lineare Prozess schwach stationär.
- Beachte, dass in dieser allgemeinen Definition y_t auch von zukünftigen Fehlern u_{t-j} , $j < 0$, beeinflusst sein darf. Wird dies ausgeschlossen, dann erhält man einen Moving Average Prozess, siehe (12.8) weiter unten.

Lag-Operator

Der Lag-Operator definiert eine Operation auf einer geordneten Menge (z. B. einem diskreten stochastischen Prozess), auf der er jedes Element auf das vorangegangene Element abbildet

$$Ly_t \equiv y_{t-1}$$

mit folgenden Eigenschaften:

$$\begin{aligned} L^0 &= 1 \\ L^2 y_t &= L(Ly_t) = Ly_{t-1} = y_{t-2} \\ L^{-1} &= y_{t+1} \\ Lc &= c \\ L^m L^n y_t &= y_{t-m-n}. \end{aligned}$$

Lag-Polynom, Filter

Das Lag-Polynom ist eine Linearkombination von verschiedenen Potenzen von Lag-Operatoren mit ganzzahligen Exponenten

$$\Psi(L) = \dots + \psi_{-2}L^{-2} + \psi_{-1}L^{-1} + \psi_0 + \psi_1L + \psi_2L^2 + \dots$$

und wird als **linearer Filter** bezeichnet (Neusser 2006, Definition 2.4).

Moving Average Prozess der Ordnung ∞ (MA(∞)-Prozess)

Ein Moving Average Prozess $\{y_t\}_{t \in \mathbb{Z}}$ der Ordnung ∞ ist ein linearer Prozess mit $\psi_j = 0$ für alle negativen j und $\sum_{j=0}^{\infty} |\psi_j| < \infty$

$$\begin{aligned} y_t &= \sum_{j=0}^{\infty} \psi_j u_{t-j}, \quad \psi_0 = 1, \\ &= (1 + \psi_1 L + \psi_2 L^2 + \dots) u_t = \Psi(L) u_t. \end{aligned} \tag{12.8}$$

u_t wird häufig als **Schock**, **Innovation** oder **Fehler** bezeichnet.

Der Ausdruck $\Psi(L) \equiv (1 + \psi_1 L + \psi_2 L^2 + \dots)$ wird als **MA(∞)-Polynom** bezeichnet.

Bemerkungen: Bei einem MA(∞)-Prozess hat die Zukunft keinen Einfluss auf die Gegenwart.

Eigenschaften

- Mittelwert: $E[y_t] = E[\sum_{j=0}^{\infty} \psi_j u_{t-j}] = \sum_{j=0}^{\infty} \psi_j E[u_{t-j}] = 0$. Das Vertauschen von Erwartungswert und unendlicher Summe ist möglich, da $\sum_{j=0}^{\infty} |\psi_j| < \infty$, so dass (12.7b) gilt.
- (Auto)Kovarianzfunktion:

$$\begin{aligned}
 \text{Cov}(y_t, y_{t-k}) &= E[y_t y_{t-k}] \\
 &= E \left[\left(\sum_{j=0}^{\infty} \psi_j u_{t-j} \right) \left(\sum_{l=0}^{\infty} \psi_l u_{t-k-l} \right) \right] \\
 &= \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \psi_j \psi_l E[u_{t-j} u_{t-k-l}] \\
 &= \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \psi_j \psi_l \begin{cases} \sigma^2 & \text{falls } t-j = t-k-l \\ 0 & \text{sonst} \end{cases} \quad \text{mit } j = k+l \\
 &= \sigma^2 \sum_{l=0}^{\infty} \psi_{k+l} \psi_l = \gamma_k. \tag{12.9}
 \end{aligned}$$

- Varianz

$$\gamma_0 = \text{Var}(y_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2 < \infty. \tag{12.10}$$

- ‡ Technische Bemerkung zur Ableitung von (12.9) und (12.10): Es kann gezeigt werden, dass unter (12.7b) der Grenzwert $(\sum_{j=0}^{\infty} \psi_j u_{t-j}) (\sum_{l=0}^{\infty} \psi_l u_{t-k-l})$ als $y_t y_{t-k}$ wohldefiniert ist, sodass das Vertauschen von Erwartungswert und Summierung möglich ist. Man beachte, dass aus $\sum_{j=0}^{\infty} |\psi_j| < \infty$ folgt: $\sum_{j=0}^{\infty} \psi_j^2 < \infty$. Die letztgenannte Bedingung ist eine notwendige Bedingung für die Existenz der Varianz in (12.10). Sie wird als quadratische Summierbarkeit bezeichnet.

Spezialfälle für die Praxis:

MA(q)-Prozesse

$$y_t = \sum_{l=0}^q \psi_l u_{t-l} \tag{12.11}$$

Eigenschaften wie bei MA(∞)-Prozessen außer:

$$\gamma_k = 0, \quad |k| > q. \tag{12.12}$$

R-Befehle

- Generieren von $MA(q)$ -Prozessen: mit Hilfe des Befehls `filter` mit der Spezifikation `sides=1,method="convolution"`.
- Berechnen der theoretischen Autokorrelationsfunktion. mit Befehl `ARMAacf`

Beispiel: Generieren eines $MA(2)$ -Prozesses und theoretische Autokorrelationsfunktion

- DGP: $y_t = u_t + \psi_1 u_{t-1} + 0.6u_{t-2}$, $u_t \sim NID(0, 4)$.
- Stichprobengröße, bzw. (-länge): $n = 1000$ **R-Programm**, siehe Abschnitt [A.11](#), Seite [356](#) liefert Abbildung [12.4](#) und Autokorrelationen $\rho_0 = 1, \rho_1 = 0.64, \rho_2 = 0.3$ und $\rho_k = 0$ für $k = 3, \dots$

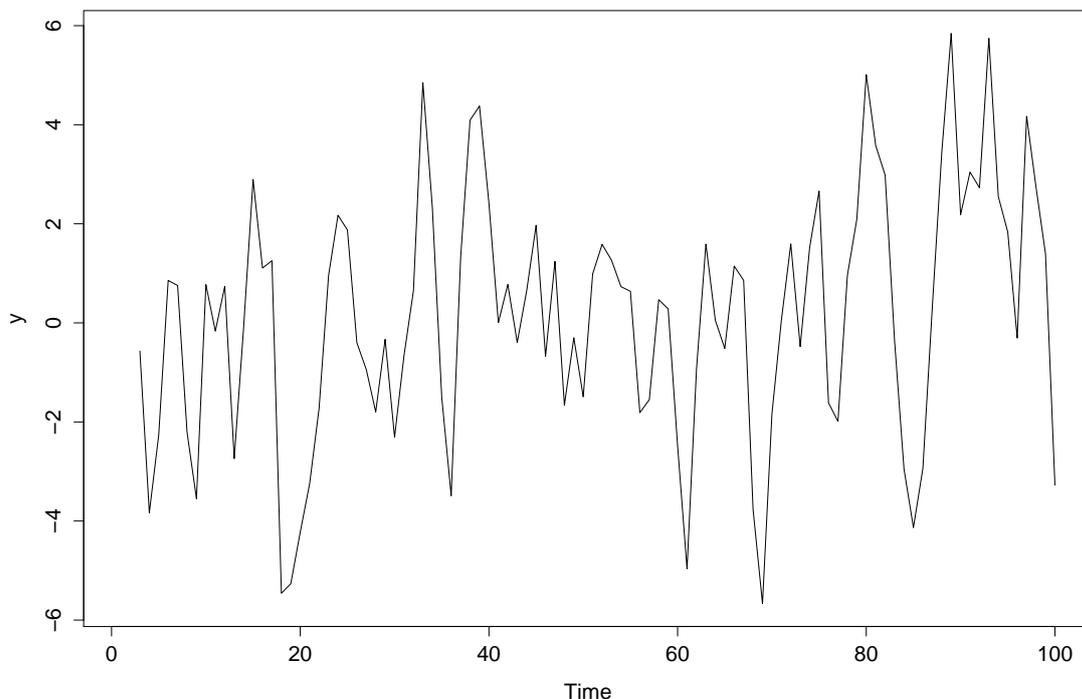


Abbildung 12.4.: Eine Realisierung für $n = 100$ eines $MA(2)$ -Prozesses mit $\psi_1 = 0.8$ und $\psi_2 = 0.6$ und $\sigma^2 = 4$ (**R-Programm**, siehe Abschnitt [A.11](#), Seite [356](#))

Problem bei $MA(q)$ -Prozessen: Sie können nicht mit OLS geschätzt werden, sondern erfordern Maximum-Likelihood-Schätzverfahren, siehe **Fortgeschrittene Ökonometrie**. Hingegen können die nun folgenden autoregressiven Prozesse mit OLS geschätzt werden.

12.3. Autoregressive Prozesse

Autoregressiver Prozess der Ordnung p (AR(p)-Prozess)

Ein stochastischer Prozess $\{y_t\}$ heißt **autoregressiver Prozess der Ordnung p (AR(p)-Prozess)**, wenn er folgende stochastische Differenzgleichung

$$y_t = \nu + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p} + u_t, \quad t \in \mathbb{T}, \quad (12.13)$$

bzw.

$$\begin{aligned} (1 - \alpha_1 L - \cdots - \alpha_p L^p) y_t &= \nu + u_t, \\ \alpha(L) y_t &= \nu + u_t \end{aligned}$$

erfüllt und $\{u_t\}$ Weißes Rauschen ist. Der Ausdruck $\alpha(L) \equiv (1 - \alpha_1 L - \cdots - \alpha_p L^p)$ wird als AR(p)-Polynom bezeichnet.

Bedingter Erwartungswert von AR(p)-Prozessen

Der bedingte Erwartungswert $E[y_t | y_{t-1}, \dots, y_1]$ lässt sich für AR(p)-Prozesse leicht berechnen:

$$E[y_t | y_{t-1}, \dots, y_1] = \nu + \alpha_1 y_{t-1} + \cdots + \alpha_p y_{t-p}. \quad (12.14)$$

Damit gilt

$$E[y_t | y_{t-1}, \dots, y_1] = E[y_t | y_{t-1}, \dots, y_{t-p}]. \quad (12.15)$$

R-Befehle

Generieren von AR(p)-Prozessen: mit Hilfe des Befehls `filter` mit der Spezifikation `method="recursive"`.

Die Ableitung der Eigenschaften von AR(p)-Prozessen ist aufwendiger als bei MA-Prozessen und unterbleibt deshalb. Allerdings lassen sich die wesentlichen Eigenschaften gut anhand von AR(1)-Prozessen analysieren.

12.3.1. AR(1)-Prozesse**Stochastische Eigenschaften eines AR(1)-Prozess**

$$y_t = \nu + \alpha_1 y_{t-1} + u_t, \quad u_t \sim WN(0, \sigma^2), \quad t \in \mathbb{T}, \quad (12.16)$$

- **Lösung:** k -maliges Einsetzen liefert:

$$\begin{aligned} y_t &= \nu + \nu \alpha_1 + \alpha_1 (\alpha_1 y_{t-2} + u_{t-1}) + u_t = \cdots \\ &= \nu \sum_{k=0}^{j-1} \alpha_1^k + \alpha_1^j y_{t-j} + \sum_{k=0}^{j-1} \alpha_1^k u_{t-k}. \end{aligned}$$

Für $j = t$

$$y_t = \nu \sum_{k=0}^{t-1} \alpha_1^k + \alpha_1^t y_0 + \sum_{k=0}^{t-1} \alpha_1^k u_{t-k}. \quad (12.17)$$

- **Stabilitätseigenschaften**

Für (12.17) gilt für beliebige α_1 und $\nu = 0$, $\mathbb{T} = \mathbb{N}$, Startwert y_0 und $j = t$:

– **AR-Prozess explosiv**, falls $|\alpha_1| > 1$.

– **Random Walk mit/ohne Drift**, falls $\alpha_1 = 1$:

* **Random Walk mit Drift**, falls $\alpha_1 = 1, \nu \neq 0$:

$$y_t = \nu t + y_0 + \sum_{j=0}^{t-1} u_{t-j}. \quad (12.18)$$

* **Random Walk ohne Drift**, falls $\alpha_1 = 1, \nu = 0$:

$$y_t = y_0 + \sum_{j=0}^{t-1} u_{t-j}. \quad (12.19)$$

Beispiel: Abbildung 12.3 zeigt verschiedene Realisationen eines Random Walks.

– **AR-Prozess stabil**, falls $|\alpha_1| < 1$.

Stationärer (und stabiler) AR(1)-Prozess: Wenn $t \in \mathbb{Z}$ und $j \rightarrow \infty$ (Prozess schon unendlich lange gelaufen), gilt für $|\alpha_1| < 1$

$$y_t = \nu / (1 - \alpha_1) + \sum_{j=0}^{\infty} \alpha_1^j u_{t-j}. \quad (12.20)$$

- **Unbedingter Erwartungswert** $\mu_t \equiv E[y_t]$ für Periode t :

$$\mu_t = \begin{cases} \nu(1 + \alpha_1 + \dots + \alpha_1^{t-1}) + \alpha_1^t E[y_0] & \text{falls } t = 1, 2, \dots \text{ — abhg.v. } t, \\ \nu / (1 - \alpha_1) & \text{falls } |\alpha_1| < 1 \text{ und } t \in \mathbb{Z} \text{ — unabh.v. } t, \\ t\nu + E[y_0] & \text{falls } \alpha_1 = 1 \text{ und } t = 1, 2, \dots \text{ — abhg.v. } t. \end{cases}$$

Für gegebenes μ_t erhält man den mittelwertbereinigten autoregressiven Prozess

$$y_t - \mu_t = \alpha_1(y_{t-1} - \mu_{t-1}) + u_t.$$

- **Unbedingte Varianz:**

$$\begin{aligned} \text{Var}(y_t) &= \alpha_1^2 \text{Var}(y_{t-1}) + \sigma^2 \\ &= \begin{cases} \sigma^2 \sum_{j=0}^{t-1} \alpha_1^{2j} & \text{falls } \text{Var}(y_0) = 0 \text{ und } t = 1, 2, \dots \text{ — abhg.v. } t, \\ \sigma^2 / (1 - \alpha_1^2) & \text{falls } |\alpha_1| < 1 \text{ und } t \in \mathbb{Z} \text{ — unabh.v. } t, \\ \sigma^2 t & \text{falls } \alpha_1 = 1, \text{Var}(y_0) = 0 \text{ und } t = 1, 2, \dots \text{ — abhg.v. } t. \end{cases} \end{aligned}$$

- **Autokovarianzfunktion** $Cov(y_t, y_s) \equiv E[(y_t - \mu_t)(y_s - \mu_s)]$:

$$Cov(y_t, y_{t-k}) = \alpha_1^k Var(y_{t-k})$$

$$= \begin{cases} \alpha_1^k \sigma^2 \sum_{j=0}^{t-1-k} \alpha_1^{2j} & \text{falls } Var(y_0) = 0 \text{ und } t = 1, 2, \dots \text{ — abhg.v. } t, \\ \alpha_1^k \sigma^2 / (1 - \alpha_1^2) & \text{falls } |\alpha_1| < 1 \text{ und } t \in \mathbb{Z} \text{ — unabhg.v. } t, \\ (t-k)\sigma^2 & \text{falls } \alpha_1 = 1, Var(y_0) = 0, t = 1, 2, \dots \text{ — abhg.v. } t. \end{cases}$$

- **Schwach stationärer AR(1)-Prozess** Falls $|\alpha_1| < 1$ und $t \in \mathbb{Z}$, ist ein AR(1)-Prozess schwach stationär, da die ersten beiden Momente unabhängig von t sind:

$$E[y_t] = \mu = \nu / (1 - \alpha_1)$$

$$Var(y_t) = \gamma_0 = \sigma^2 / (1 - \alpha_1^2)$$

$$Cov(y_t, y_{t-k}) = \gamma_k = \alpha_1^k \gamma_0$$

- **Autokorrelationsfunktion** Im Fall eines schwach stationären AR(1)-Prozesses gilt

$$\rho_k = Corr(y_t, y_{t-k}) = \alpha_1^k \quad (12.21)$$

- **Eigenschaften eines (schwach) stationären AR(1)-Prozesses:** Falls $|\alpha_1| \neq 0$:

- $\gamma_k \neq 0$ für alle $k \in \mathbb{Z}$,
- Autokovarianzen und Autokorrelationen konvergieren exponentiell schnell gegen Null:

$$\gamma_k = \alpha_1^k \gamma_0,$$

$$\rho_k = \alpha_1^k$$

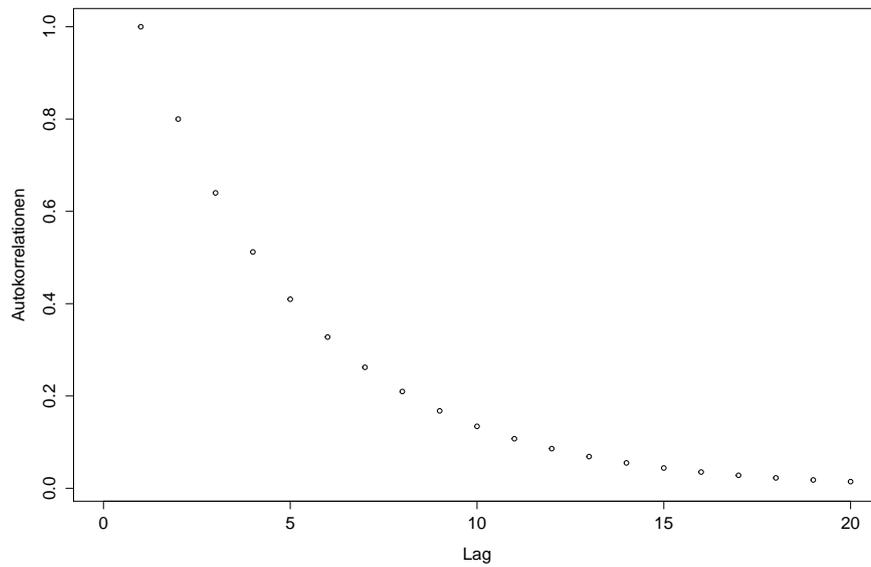
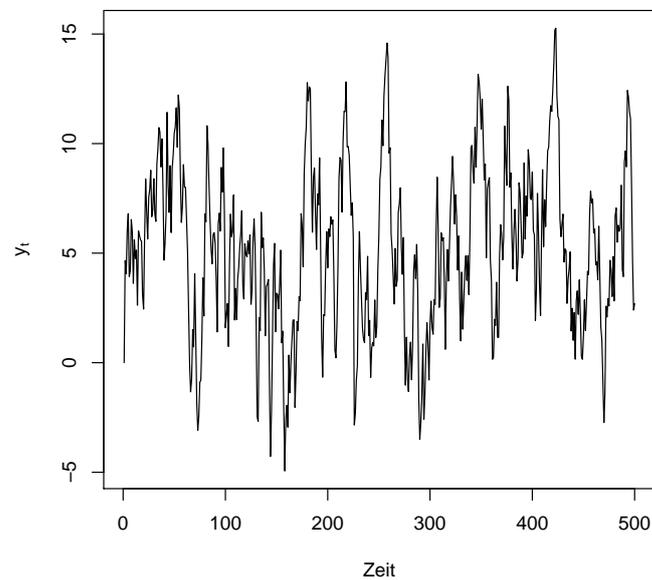
D.h. die Wirkung von Schocks wird relativ schnell 'vergessen'. Man spricht deshalb auch von **Modellen mit kurzem Gedächtnis (short memory)**. In extremem Gegensatz hierzu steht der Random Walk. Hier liegt ein perfektes Gedächtnis vor, da die Wirkung eines Schocks niemals vergessen wird. Random Walks sind ein Beispiel von **Modellen mit langem Gedächtnis (long memory)**. Siehe Bemerkung nach Gleichung (12.33).

Beispiel: Plot der Autokorrelationsfunktion (12.21) eines AR(1)-Prozesses für $\alpha_1 = 0.8$ und $k = 1, \dots, 20$

```
ar1_acf <- ARMAacf(ar=0.8,lag.max=20)
plot(ar1_acf,ylab="Autokorrelationen",xlab="Lag",cex=0.8,xlim=c(0,20))
```

Beispiel: Plot einer Realisation eines AR(1)-Prozesses

Parameter des DGPs: $\nu = 1, \alpha_1 = 0.8, \sigma^2 = 4$ mit $n = 500$. Abbildung 12.6 zeigt eine Realisation, generiert mit **R-Programm**, siehe Abschnitt A.12, Seite 356.

Abbildung 12.5.: Autokorrelationsfunktion eines AR(1)-Prozesses mit $\alpha_1 = 0.8$ Abbildung 12.6.: Realisation eines AR(1)-Prozesses mit $\nu = 1, \alpha_1 = 0.8, \sigma^2 = 4, y_0 = 0$ und $n = 500$ (**R-Programm** siehe Abschnitt [A.12](#), Seite [356](#))

- (Asymptotische) Stationarität:

- Falls $|\alpha_1| < 1$, gilt

$$\lim_{t \rightarrow \infty} E(y_t) = \mu,$$

$$\lim_{t \rightarrow \infty} Cov(y_t, y_{t-k}) = \gamma_k,$$

und der AR(1)-Prozess ist **asymptotisch stationär**.

- Jeder stationäre Prozess ist asymptotisch stationär.
- Ist ein Prozess nicht asymptotisch stationär, ist er **nichtstationär**.
- Welche Bedingungen sind für strenge Stationarität erforderlich?
- **Invertierbarer AR(1)-Prozess:** Definiert man $\psi_j = \alpha^j$, dann lässt sich die Darstellung (12.20) mit $\psi_0 = 1$ und $\psi_j = 0, j < 0$, auch als **MA(∞ -Prozess)** (12.8)

$$y_t = \mu + \sum_{j=0}^{\infty} \psi_j u_{t-j}, \quad t \in \mathbb{Z}, \quad (12.22)$$

schreiben. Der AR(1)-Prozess heißt dann **invertierbar**.

- Das AR(1)-Modell ist partiell spezifiziert. Um Realisationen mittels einer Monte-Carlo-(MC-)Studie generieren zu können, müssen zusätzliche Annahmen getroffen werden, z. B.:
 - ein Startwert $y_0 = 0$,
 - ein Parameterwert $\alpha = 0.9$ und
 - eine Verteilung für die Fehler $u_t \sim NID(0, 2)$. Siehe Abschnitt 2.9.1 zu Definition von *NID*.

Damit ist der DGP bekannt und folgende MC-Studie zur Überprüfung der Verzerrung des KQ-Schätzers kann durchgeführt werden.

12.3.2. Komplexe Zahlen

Um die Notwendigkeit des Studiums komplexer Zahlen zu motivieren, untersuchen wir die Stabilitätseigenschaften eines AR(2)-Prozesses.

Stabilitätseigenschaften eines AR(2)-Prozesses

- Darstellung als Verknüpfung von zwei AR(1)-Prozessen

$$(1 - \lambda_1 L)w_t = u_t,$$

$$(1 - \lambda_2 L)y_t = w_t.$$

Voraussetzung für Verknüpfung:

- $\{w_t\}$ schwach stationär: $|\lambda_1| < 1$,

– $\{u_t\}$ Weißes Rauschen.

Durchführen:

1. Invertiere AR(1)-Prozess $\{w_t\}$:

$$w_t = \frac{1}{1 - \lambda_1 L} u_t$$

2. und setze w_t in y_t -Gleichung ein:

$$\begin{aligned} (1 - \lambda_2 L)y_t &= \frac{u_t}{(1 - \lambda_1 L)} \\ (1 - \lambda_2 L)(1 - \lambda_1 L)y_t &= u_t \\ ((1 - \lambda_1 L - \lambda_2 L + \lambda_1 \lambda_2 L^2)y_t &= u_t \\ \underbrace{\quad}_{-\alpha_1 L} \quad \underbrace{\quad}_{-\alpha_2 L^2} & \\ y_t - \alpha_1 y_{t-1} - \alpha_2 y_{t-2} &= u_t \\ (1 - \alpha_1 L - \alpha_2 L^2)y_t &= u_t \end{aligned}$$

mit $\alpha_1 = \lambda_1 + \lambda_2$, $\alpha_2 = -\lambda_1 \lambda_2$.

Ergebnis: AR(2)-Prozess stationär, falls $|\lambda_1|, |\lambda_2| < 1$.

So eine Zerlegung existiert für jeden stationären AR(p)-Prozess, erfordert aber die Kenntnis komplexer Zahlen.

Literatur zu komplexen Zahlen: [Neusser \(2009, Appendix A\)](#), [Hamilton \(1994, Appendix A.2, S. 708-710\)](#)

- Motivation: $x^2 + 1 = 0 \iff x^2 = -1 \iff x = \pm\sqrt{-1}$ besitzt **in \mathbb{R} keine Lösung**.
- Idee: Erweitere \mathbb{R} mit einer imaginären Einheit („einer zweiten Dimension“) $i \equiv \sqrt{-1}$ um Wurzeln negativer, reeller Zahlen bilden zu können, d. h. definiere die

Menge der Komplexen Zahlen als $\mathbb{C} \equiv \mathbb{R}[i] \equiv \{z \equiv a + ib \mid a, b \in \mathbb{R}\}$

als Summe einer reellen und imaginären Zahl.

Dann gilt für beliebiges $a \in \mathbb{R}^+$: $\sqrt{-a} = \sqrt{-1 \cdot a} = \sqrt{-1} \cdot \sqrt{a} = i\sqrt{a} \in \mathbb{C}$.

- Wichtig für diesen Kurs: Was ist der **Betrag einer komplexen Zahl** $z \in \mathbb{C}$: $\|z\|_{\mathbb{C}}$?

Konjugierte und Betrag einer komplexen Zahl $z \in \mathbb{C}$:

- Zu z konjugiert komplexe Zahl: $\bar{z} \equiv a - ib$
- „Länge“, Betrag, Absolutwert von z : $\|z\|_{\mathbb{C}} \equiv r \equiv \sqrt{z\bar{z}} = \sqrt{a^2 + b^2}$
Dieser wird häufig mit $\|z\|_{\mathbb{C}}$, $\|z\|$, oder $|z|$ notiert - man beachte den Unterschied zu anderen

Beträgen

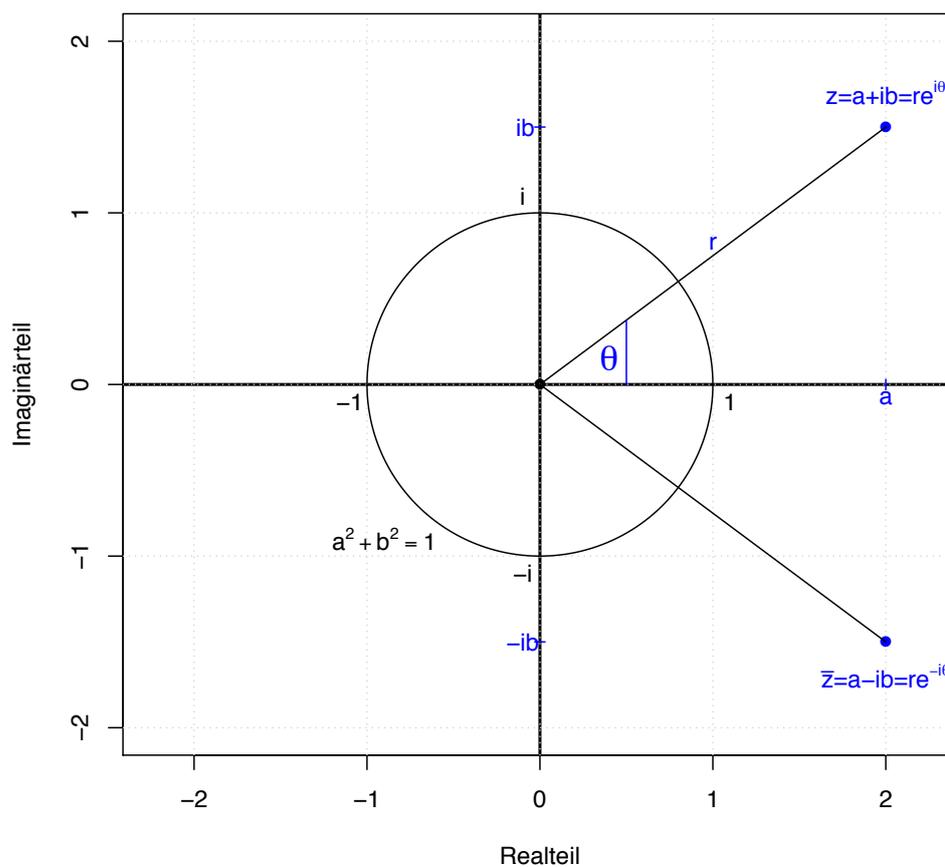
$$\|z\| \equiv |z| \equiv \begin{cases} |z|_{\mathbb{R}} \equiv \begin{cases} z & z \geq 0 \\ -z & z < 0 \end{cases} & \text{falls } z \in \mathbb{R} \\ |z|_{\mathbb{R}^n} \equiv \sqrt{z_1^2 + \dots + z_n^2} & \text{falls } z = (z_1, \dots, z_n)^t \in \mathbb{R}^n \\ |z|_{\mathbb{C}} \equiv \sqrt{z\bar{z}} = \sqrt{a^2 + b^2} & \text{falls } z = a + ib \in \mathbb{C} \end{cases}$$

Darstellung komplexer Zahlen $z \in \mathbb{C}$:

- In Cartesischen Koordinaten: $z = \underbrace{a}_{\text{reelle}} + i \underbrace{b}_{\text{komplexe Komponente}} = \text{Re}(z) + i \text{Im}(z)$
- In Polarkoordinaten: $z = \underbrace{r}_{\text{Länge}} \cdot \underbrace{e^{i\theta}}_{\text{Richtung}} = r(\cos \theta + i \sin \theta)$

Beachte: $\|e^{i\theta}\|_{\mathbb{C}} = \sqrt{e^{i\theta} \cdot e^{-i\theta}} = 1$

Folgende Abbildung ist an [Neusser \(2009, Bild A.1, S. 260\)](#) angelehnt:



Rechenregeln

- **Addition:** $(a + ib) + (c + id) = (a + c) + i(b + d)$
- **Subtraktion:** $(a + ib) - (c + id) = (a - c) + i(b - d)$
- **Multiplikation:**

$$(a + ib)(c + id) = (ac - bd) + i(ad + bc)$$

- **Division:**

$$\frac{a + ib}{c + id} = \frac{(ac + bd) + i(bc - ad)}{c^2 + d^2}$$

- **Betrag:** $r = |z| = z \cdot \bar{z} = (a + ib)(a - ib) = a^2 + b^2$
- $\cos \theta = a/r$
- $\sin \theta = b/r$
- **Satz von Moivre:**

$$\begin{aligned} z^n &= (re^{i\theta})^n \\ &= r^n e^{in\theta} \\ &= r^n (\cos n\theta + i \sin n\theta) \end{aligned}$$

Fundamentalsatz der Algebra

Jedes Polynom mit Koeffizienten $\phi_1, \dots, \phi_p \in \mathbb{R}$

$$\Phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p \quad (12.23)$$

vom Grade $p \geq 1$ zerfällt in der Menge (präzise: Körper) der komplexen Zahlen \mathbb{C} genau in p lineare Faktoren (besitzt also p komplexe Nullstellen, Wurzeln, roots $\lambda_1^{-1}, \dots, \lambda_p^{-1}$), wobei manche Nullstellen mehrfach vorkommen können (siehe [Neusser 2009](#), S. 261):

$$\Phi(z) = (1 - \lambda_1 z)(1 - \lambda_2 z) \cdots (1 - \lambda_p z).$$

Diese Wurzeln können reell oder komplex sein und treten im komplexen Fall als Wurzelpaare konjugierter Wurzeln auf. Liegen c komplexe Wurzelpaare und r reelle Wurzeln vor, so gilt: $p = 2c + r$.

Die $\lambda_1, \dots, \lambda_p$ heißen **Eigenwerte** des Polynoms $\Phi(z)$.

Beispiel: Das Polynom $\Phi(z) = z^3 - 2z^2 - 23z + 150$ hat in \mathbb{R} eine Nullstelle $\{z_1 = -6\}$, zerfällt aber in \mathbb{C} in alle Einzelbestandteile (Linearfaktoren) mit dem

komplexen Wurzelpaar $\{z_{2,3} = 4 \pm 3i\}$, kann also geschrieben werden:

$$\begin{aligned}\Phi(z) &= \underbrace{(z+6) \cdot (z^2 - 8z + 25)}_{\text{Faktorisierung über } \mathbb{R}} = \underbrace{(z+6) \cdot (z-4-3i) \cdot (z-4+3i)}_{\text{Linearfaktorisierung über } \mathbb{C}} \\ &= \left(1 - \underbrace{\frac{-1}{6}}_{\lambda_1} z\right) \cdot \left(1 - \underbrace{(0.16 - 0.12i)}_{\lambda_2} z\right) \cdot \left(1 - \underbrace{(0.16 + 0.12i)}_{\lambda_2} z\right)\end{aligned}$$

Der Fundamentalsatz der Algebra ermöglicht die Analyse der Stabilitätseigenschaften von AR(p)-Prozessen.

12.3.3. AR(p)-Prozesse und mehr

Eigenschaften von AR(p)-Prozessen:

Stabilitätsbedingung für AR(p)-Prozesse

Ein AR(p)-Prozess mit AR-Polynom, $z \in \mathbb{C}$,

$$\begin{aligned}\alpha(z) &= (1 - \alpha_1 z - \dots - \alpha_p z^p) \\ &= (1 - \lambda_1 z) \cdots (1 - \lambda_p z)\end{aligned}\tag{12.24}$$

mit

- Eigenwerten $\lambda_1, \dots, \lambda_p$ bzw.
- Wurzeln (roots) $\lambda_1^{-1}, \dots, \lambda_p^{-1}$

heißt stabil,

- wenn alle **Eigenwerte** im Absolutbetrag kleiner Eins sind

$$|\lambda_i| < 1, \quad i = 1, 2, \dots, p,\tag{12.25}$$

d. h. **innerhalb** des Einheitskreises liegen bzw.,

- wenn alle **Wurzeln/Nullstellen** z_i des Polynoms $\alpha(z)$, also der **charakteristischen Gleichung** des AR(p)-Polynoms

$$(1 - \alpha_1 z - \dots - \alpha_p z^p) = 0$$

außerhalb des Einheitskreises liegen, d. h.

$$|z_i| > 1, \quad i = 1, 2, \dots, p,\tag{12.26}$$

gilt.

R-Befehle

Berechnen der Wurzeln eines AR(p)-Polynoms: mit `polyroot()`. Deren Absolutwerte können mit `abs()` bestimmt werden.

Beispiel: AR(2)-Prozess: Die Absolutwerte der Wurzeln des AR(2)-Polynoms

$$\alpha(L) = 1 - 0.1L - 0.9L^2$$

sind $z = 1$ und $z = 1.111\dots$. Damit ist das Polynom nicht stabil.

```
abs(polyroot(c(1,-0.1,-0.9)))
```

Beachte folgende Eigenschaft des AR-Polynoms: $\alpha(1) = 1 - \alpha_1 \cdot 1 - \dots - \alpha_p \cdot 1^p = 1 - \alpha_1 - \dots - \alpha_p$.

Momente eines (schwach) stationären AR(p)-Prozesses

- **Mittelwert/Erwartungswert:**

$$E[y_t] = \mu = \nu / \alpha(1) = \mu / (1 - \alpha_1 - \alpha_2 - \dots - \alpha_p) \quad \text{für alle } t. \quad (12.27)$$

- **Varianz und Autokovarianzfunktion:**

Die Varianz und die Autokovarianzen eines schwach stationären AR(p)-Prozesses sind durch die folgenden **Yule-Walker-Gleichungen** bestimmt (vgl. [Hamilton \(1994, S. 59, Gl. \(3.4.36\)\)](#))

$$\gamma_k = \begin{cases} \alpha_1 \gamma_1 + \alpha_2 \gamma_2 + \dots + \alpha_p \gamma_p + \sigma^2 & \text{für } k = 0 \\ \alpha_1 \gamma_{k-1} + \alpha_2 \gamma_{k-2} + \dots + \alpha_p \gamma_{k-p} & \text{für } k = 1, 2, \dots \end{cases} \quad (12.28)$$

Es kann gezeigt werden, dass die Autokovarianzen eines stationären AR(p)-Prozesses exponentiell schnell gegen Null konvergieren ([Hamilton \(1994, S. 59\)](#), [Kirchgässner & Wolters \(2008, Example 2.4\)](#)). Siehe Abschnitt [12.3.1](#) für den Fall von AR(1)-Prozessen.

- **Partielle Autokorrelationsfunktion:**

Für einen schwach stationären AR(p)-Prozess gilt:

$$a_k = \text{Corr}(y_t, y_{t-k} | y_{t-1}, \dots, y_{t-k+1}). \quad (12.29)$$

D.h., alle partiellen Autokorrelationen für $k > p$ sind Null, da $a_k = \alpha_k = 0$ für $k > p$.

Invertierbarkeit eines stationären AR(p)-Prozesses

Ein stationärer AR(p)-Prozess kann als MA(∞)-Prozess ([12.8](#)) dargestellt werden:

$$\begin{aligned} y_t - \mu &= \psi(L)u_t & (12.30) \\ y_t - \mu &= u_t + \psi_1 u_{t-1} + \psi_2 u_{t-2} + \dots + \psi_i u_{t-i} + \dots, \end{aligned}$$

wobei die Koeffizienten des Lagpolynoms durch folgende Gleichung bestimmt sind:

$$\frac{1}{\alpha(L)} = \psi(L) \quad (12.31)$$

$$\alpha(L)\psi(L) = 1. \quad (12.32)$$

Die Parameter von $\psi(L)$ können mit Hilfe der Methode des Koeffizientenvergleichs bestimmt werden (Kirchgässner & Wolters 2008, Section 2.1.2):

$$\phi_j = \sum_{i=1}^j \phi_{j-i} \alpha_i, \quad j = 1, 2, \dots, \quad \nu = 1, \alpha_i = 0 \text{ für } i > p.$$

R-Befehle

Berechnen der MA-Parameter eines AR(p)-Prozesses: mit `ARMAtoMA()`.

Beispiel: AR(2)-Prozess: Realisation, theoretische ACF und PACF sowie MA-Parameter des invertierten Prozesses

Die Parameter sind: $\nu = 1$, $\alpha_1 = -0.5$, $\alpha_2 = -0.8$, $\sigma^2 = 4$ mit $n = 500$.

Die Abbildung 12.7 ist mit dem **R-Programm**, siehe Abschnitt A.13, Seite 357 erstellt.

Nichtstationäre Prozesse

- Definiere den Differenzenoperator

$$\Delta y_t \equiv (1 - L)y_t = y_t - y_{t-1}$$

- Die Stabilitätsbedingung (12.25) ist z.B. verletzt, wenn sich das AR(p)-Polynom $\alpha(z)$ folgendermaßen zerlegen lässt

$$\begin{aligned} 1 - \alpha_1 z - \dots - \alpha_p z^p &= (1 - z) \left(1 - \alpha_1^* z - \dots - \alpha_{p-1}^* z^{p-1} \right) \\ &= (1 - z) \alpha^*(z) = \Delta \alpha^*(z) \end{aligned}$$

wobei das AR($p - 1$)-Polynom $\alpha^*(L)$ die Stabilitätsbedingung (12.26) erfüllt. In diesem Fall enthält der AR(p)-Prozess eine **Random Walk-Komponente**. Man spricht auch von einem Prozess, der **integriert ist mit der Ordnung 1**, kurz

$$y_t \sim I(1).$$

Nach Anwenden des Differenzenoperators erhält man einen stabilen Prozess der Ordnung 0

$$\Delta y_t \sim I(0).$$

Die Random-Walk-Komponente wird häufig als **stochastischer Trend** bezeichnet, da sie häufig trendartige Trajektorien verursacht, siehe z. B. Abbildung 12.3.

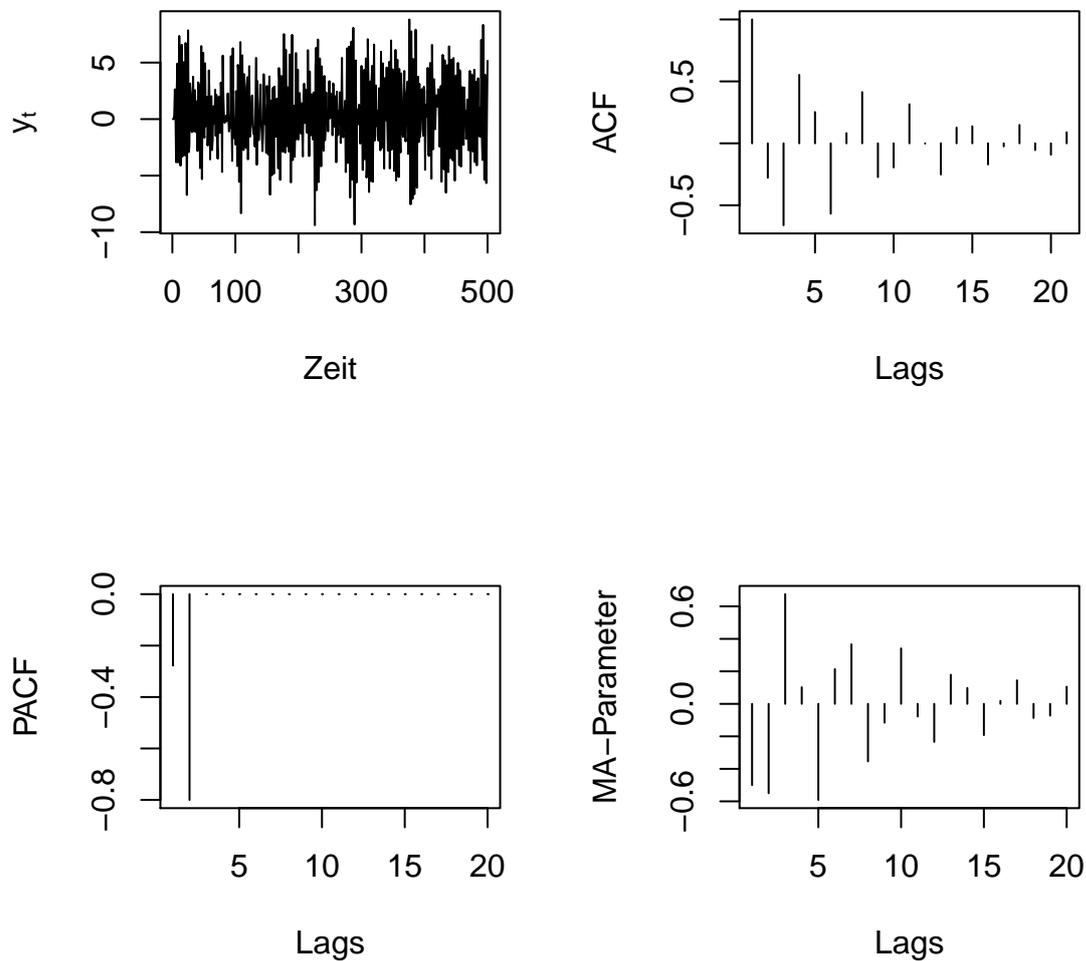


Abbildung 12.7.: Realisierung, ACF, MA-Parameter, PACF eines AR(2)-Prozesses mit $\nu = 1, \alpha_1 = -0.5, \alpha_2 = -0.8, \sigma^2 = 4, y_0 = 0$ und $n = 500$

Beispiel: Abbildung 12.3 zeigt verschiedene Realisationen eines Random Walks.

- Allgemein: Ein AR(p)-Prozess $\{y_t\}$ ist **integriert mit der Ordnung d** , kurz

$$y_t \sim I(d),$$

wenn gilt

$$\alpha(L) = (1 - L)^d \alpha^*(L), \quad (12.33)$$

wobei d eine ganze Zahl ist und $\alpha^*(L)$ die Stabilitätsbedingung erfüllt. Zur Stabilisierung eines integrierten Prozesses ist also die d -malige Anwendung des Differenzenoperators $(1 - L)$ notwendig.

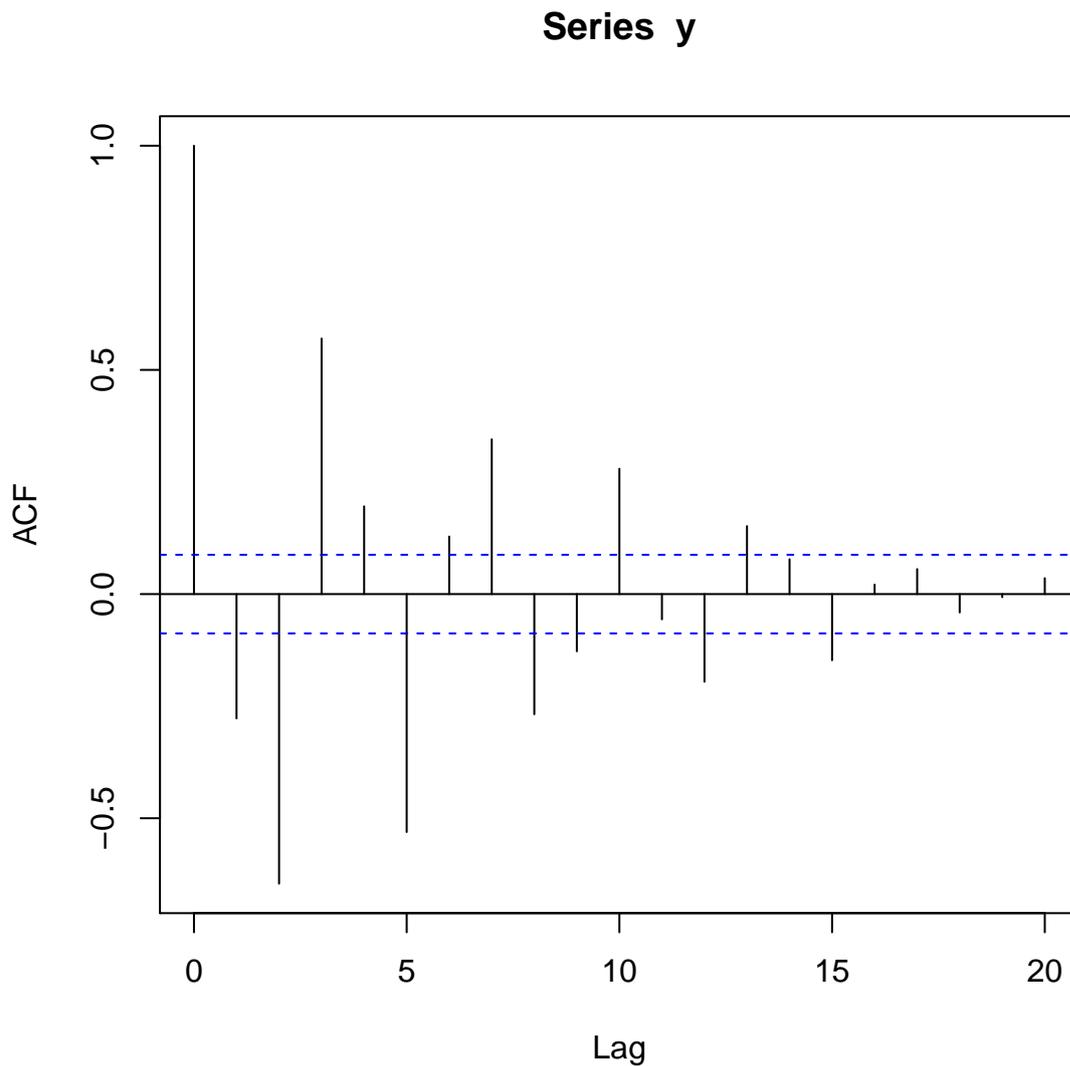


Abbildung 12.8.: Geschätzte Autokorrelationsfunktion einer Realisierung eines AR(2)-Prozesses mit $\nu = 1$, $\alpha_1 = -0.5$, $\alpha_2 = -0.8$, $\sigma^2 = 4$, $y_0 = 0$ und $n = 500$

- Es ist möglich, dass der Integrationsparameter d auch reelle Werte annimmt \rightarrow Long Memory Modelle/ Modelle mit langem Gedächtnis. ‡ Siehe hierzu als deutsche Einführung z. B. [Tschernig \(1994, Kapitel 3\)](#) und [Robinson \(2003\)](#) mit relevanten Aufsätzen zu Long Memory Modellen.
- Autoregressive Prozesse (und stochastische Prozesse allgemein) können einen **deterministischen Trend** bzw. **Zeittrend** enthalten. Sind solche Prozesse nach Beseitigung des Zeittrends stationär werden sie als **trendstationär** bezeichnet.

AR(p)-Prozessen eignen sich gut zum Erstellen von Prognosen. Diese können folgendermaßen berechnet werden:

h -Schritt-Prognose

$$\begin{aligned}
 y_{T+h|T} &\equiv E[y_{t+h}|y_t, \dots] \\
 y_{T+1|T} &= \alpha_1 y_T + \dots + \alpha_p y_{T-p} \\
 y_{T+2|T} &= \alpha_1 y_{T+1|T} + \dots + \alpha_p y_{T+1-p} \\
 &\vdots \\
 y_{T+h|T} &= \alpha_1 y_{T+h-1|T} + \alpha_p y_{T+h-p|T} \quad \text{mit } y_{T+h-p|T} = y_{T+h-p}, \text{ falls } h-p \leq 0. \quad (12.34)
 \end{aligned}$$

AR(p)- und Moving-Average-Prozesse lassen sich wie folgt kombinieren:

ARIMA(p, d, q)-Prozesse

$$\alpha(L)(1-L)^d y_t = \psi(L)u_t, \quad u_t \sim WN(0, \sigma^2) \quad (12.35)$$

wobei das AR-Polynom stabil ist, so dass $y_t \sim I(d)$ und $\Delta^d y_t \sim I(0)$ ist.

ARMA(p, q)-Prozesse

ARMA(p, q)-Prozess ist ein ARIMA($p, 0, q$)-Prozess:

$$\alpha(L)y_t = \psi(L)u_t, \quad u_t \sim WN(0, \sigma^2). \quad (12.36)$$

Eine Diskussion der Eigenschaften von ARMA- und ARIMA-Prozessen findet sich **Applied Financial Econometrics** oder in den genannten Lehrbüchern.

12.3.4. KQ-Schätzer für AR(p)-Modelle

Ein AR(p)-Modell (12.13) kann mit dem KQ-Schätzer geschätzt werden. Zum Bestimmen der Schätzeigenschaften müssen die Eigenschaften der Regressoren y_{t-1}, \dots, y_{t-p} überprüft werden. Zur Vereinfachung der Darstellung geschieht dies für ein AR(1)-Modell:

- **Überprüfen, ob Regressor $x_t = y_{t-1}$ vorherbestimmt bezüglich u_t :** Da die Fehler $u_t \sim IID(0, \sigma^2)$ sind, gilt (siehe Abschnitt 2.7)

$$E[u_t | u_{t-1}, u_{t-2}, \dots] = E[u_t] = 0.$$

Wegen (12.20) hängt y_{t-1} nur von vergangenen u_{t-1-j} , $j \geq 0$, ab. Damit ist y_{t-1} bestimmt, wenn die vergangenen Fehler bestimmt sind. Da der Erwartungswert von u_t jedoch unabhängig von vergangenen Fehlern ist und deshalb unabhängig von der Bedingung auf vergangene Fehler ist, ist er auch unabhängig von der Bedingung y_{t-1} . Deshalb gilt (9.4)

$$E[u_t | y_{t-1}] = 0$$

und y_{t-1} ist vorherbestimmt bezüglich der Fehler u_t .

- **Überprüfen strenger Exogenität:** Damit $x_t = y_{t-1}$ streng exogen ist, muss wegen (9.2) auch $Cov(u_t, x_{t+1}) = Cov(u_t, y_t) = 0$ gelten. Dies ist nicht der Fall, da

$$Cov(u_t, y_t) = Cov(u_t, \alpha y_{t-1} + u_t) = \alpha Cov(u_t, y_{t-1}) + Var(u_t) = \sigma^2 > 0.$$

Wegen (2.29c) folgt hieraus auch $E[u_t|y_t] \neq 0$. Damit ist $x_t = y_{t-1}$ nicht streng exogen und der **KQ-Schätzer für α ist nicht erwartungstreu**. Dies gilt generell für Modelle mit verzögert abhängigen Variablen.

Da AR(p)-Modelle ein Spezialfall von dynamischen linearen Regressionsmodellen sind, werden die Schätzeigenschaften dort in Abschnitt 13.5 detaillierter behandelt.

12.4. Schätzung erster und zweiter Momente im Fall stationärer Prozesse

Der Ensemblemittelwert und Varianz und Autokovarianzen können auch ohne Spezifikation eines Zeitreihenmodells geschätzt werden.

12.4.1. Schätzen des Mittelwertes

Konsistenz des Mittelwertschätzers

Sei $\{y_t\}$ ein schwach stationärer Prozess mit Mittelwert μ und Autokovarianzfunktion γ_h . Dann gilt für den Mittelwertschätzer

$$\bar{y}_T = \frac{1}{T} \sum_{t=1}^T y_t$$

für $T \rightarrow \infty$:

- Falls $\gamma_h \xrightarrow{h \rightarrow \infty} 0$, gilt (wie im IID-Fall)

$$\lim_{T \rightarrow \infty} Var(\bar{y}_T) = \lim_{T \rightarrow \infty} E[(\bar{y}_T - \mu)^2] = 0. \quad (12.37)$$

- Falls $\sum_{h=-\infty}^{\infty} |\gamma_h| < \infty$, gilt

$$\lim_{T \rightarrow \infty} T Var(\bar{y}_T) = \lim_{T \rightarrow \infty} T E[(\bar{y}_T - \mu)^2] = \sum_{h=-\infty}^{\infty} \gamma_h. \quad (12.38)$$

(Brockwell & Davis (1991, Theorem 7.1.1). Für einen Beweis siehe ebendort.)

Der Mittelwertschätzer \bar{y}_T

- ist unter der schwachen Bedingung $\gamma_h \xrightarrow{h \rightarrow \infty} 0$ gemäß (12.37) konsistent,

Beispiel: DGP ist stationärer AR(p)-Prozess.

- konvergiert mit \sqrt{T} gegen den wahren Mittelwert μ gemäß (12.38), falls die Autokovarianzfunktion absolut summierbar ist.
- Vgl. IID-Fall: $n\text{Var}(\bar{y}) = \sigma$ und $\text{Var}(\bar{y}) = \sigma/n$. **Im Zeitreihenfall müssen bei der Berechnung der Schätzvarianz alle Autokovarianzen berücksichtigt werden**

$$\text{Var}(\bar{y}_T) \approx \frac{\gamma_0 + 2 \sum_{h=1}^{\infty} \gamma_h}{T} \neq \underbrace{\frac{\gamma_0}{T}}_{\text{IID-Fall}}$$

Asymptotische Verteilung des Mittelwertschätzers

Theorem (Brockwell & Davis 1991, Theorem 7.1.2)

Ist $\{y_t\}$ ein stationärer linearer Prozess (vgl. (12.6)) mit Mittelwert $\mu = E(y_t)$ und unabhängigem Weißem Rauschen

$$y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j u_{t-j}, \quad u_t \sim \text{IID}(0, \sigma^2),$$

wobei $\sum_{j=-\infty}^{\infty} |\psi_j| < \infty$ und $\sum_{j=-\infty}^{\infty} \psi_j \neq 0$ gelten, dann gilt

$$\sqrt{T}(\bar{y}_T - \mu) \xrightarrow{d} N(0, v) \quad (12.39)$$

mit

$$v = \sum_{h=-\infty}^{\infty} \gamma_h = \sigma^2 \left(\sum_{j=-\infty}^{\infty} \psi_j \right)^2, \quad (12.40)$$

wobei γ_h die Autokovarianzfunktion von $\{y_t\}$ bezeichnet.

Beweis (mehrere Seiten) siehe Brockwell & Davis (1991, Section 7.3).

Bemerkungen:

- Anwendung von (12.39) in der Praxis: v wird geschätzt, indem nur $2p + 1$ Autokovarianzen geschätzt und aufsummiert werden, wobei üblicherweise zur Schätzung von γ_h (12.43) verwendet wird. Man erhält

$$\bar{y}_T \approx N \left(\mu, \sum_{h=-p}^p \hat{\gamma}_h \right), \quad (12.41)$$

wobei p mit Daumenregeln gewählt wird, die $p \sim cT^{1/4}$ erfüllen.

- ‡ Lässt sich ein linearer Prozess als ARMA(p, q)-Prozess (12.36) darstellen, lässt sich der Mittelwert mit Hilfe des BLUE-Schätzers (GLS-Schätzer, siehe Abschnitt 14.1)

$$\hat{\mu}_T = \left(\boldsymbol{\iota}' \Gamma_T^{-1} \boldsymbol{\iota} \right)^{-1} \boldsymbol{\iota}' \Gamma_T^{-1} \mathbf{y}_T, \quad \boldsymbol{\iota} = (1 \quad 1 \quad \dots \quad 1)'$$

berechnen, wobei sich die Kovarianzmatrix

$$\Gamma_T = \begin{pmatrix} \gamma_0 & \gamma_1 & \cdots & \gamma_{T-1} \\ \gamma_1 & \gamma_0 & \cdots & \gamma_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{T-1} & \gamma_{T-2} & \cdots & \gamma_0 \end{pmatrix} \quad (12.42)$$

aus den Parametern $\alpha_1, \dots, \alpha_p, m_1, \dots, m_q$ bestimmen lässt. Die asymptotische Varianz ist jedoch gleich (Brockwell & Davis 1991, S. 220, 236)

$$\lim_{T \rightarrow \infty} TVar(\bar{y}_T) = \lim_{T \rightarrow \infty} TVar(\hat{\mu}).$$

Exakte Normalverteilung des Mittelwertschätzers

Wird in obigem Theorem in (12.39) die IID-Bedingung zu Gaußischem Weißem Rauschen verstärkt, $u_t \sim NID(0, \sigma^2)$, dann ist der Mittelwertschätzer exakt normalverteilt

$$\bar{y}_T \sim N \left(\mu, \frac{1}{T} \sum_{|h| < T} \left(1 - \frac{|h|}{T} \right) \gamma_h \right).$$

Beweis: möglich über einfaches Regressionsmodell mit autokorrelierten Fehlern $Cov(u_t, u_{t-h}) = \gamma_h$, indem man als Regressor die Konstante (=streng exogen)

$$y_t = \mu \cdot 1 + u_t,$$

wählt. □

Wie können Autokovarianzen geschätzt werden?

A) Schätzen von parametrischen Zeitreihenmodellen.

Beispiel: AR(p)-Modelle, siehe Abschnitt 12.3.1.

B) Direkte (nichtparametrische) Schätzung der Autokovarianzen, siehe Abschnitt 12.4.2

Wie kann Konvergenz bzw. absolute Summierbarkeit der Autokovarianzen überprüft werden?

Nur einfach möglich bei A): Spezifizieren und Schätzen von parametrischen linearen Zeitreihenmodellen.

‡ Optional: Ein noch allgemeineres Ergebnis zur Schätzung des Mittelwertes ist das folgende Ergodentheorem.

Ergodischer stochastischer Prozess

Ein stationärer stochastischer Prozess heißt **ergodisch**, wenn ein Ereignis, das alle Zufallsvariablen y_t , $t \in \mathbb{T}$, beeinflusst, entweder Wahrscheinlichkeit 1 oder 0 hat (Davidson 2000, Section 4.4.3).

Beispiel für einen stationären, aber nicht ergodischen Prozess $\{y_t\}$ Es gelte $u_t \sim WN(0, \sigma^2)$ und für eine stetige Zufallsvariable z : $z \sim (0, Var(z))$.

$$\begin{aligned} y_t &= z + u_t, \quad t \in \mathbb{Z} \\ Cov(y_t, y_{t-j}) &= Var(z) \implies Cov(y_t, y_{t-j}) \not\rightarrow 0 \text{ für } j \rightarrow \infty \\ E[y_t] &= E[u_t] + E[z] = E[z] = 0 \end{aligned}$$

Der Prozess $\{y_t\}$ ist schwach stationär, da weder der Mittelwert, noch die Varianz oder die Autokovarianzen vom Zeitindex abhängen. Aber die (lineare) Abhängigkeit zwischen zwei Elementen des stochastischen Prozesses verschwindet nicht mit zunehmendem zeitlichem Abstand j . Deshalb ist $\{y_t\}$ nicht ergodisch. Falls $P(z = z_0) = 1$, ist z faktisch eine Konstante. Dann ist $\{y_t\}$ auch ergodisch, weil $P(z \neq z_0) = 0$.

Ergodensatz (Ergodic Theorem)

(Davidson 2000, Theorem 4.4.1)

Wenn $\{y_t\}$ stationär und ergodisch ist und $E[y_1]$ existiert, dann gilt

$$\bar{y}_T \xrightarrow{\text{a.s.}} E[y_1].$$

Man spricht dann auch von **Mittelwertergodizität**. Vgl. zu fast sicherer Konvergenz Abschnitt 3.3. Damit gilt auch

$$\bar{y}_T \xrightarrow{P} E[y_1].$$

Allgemein gilt: Wenn ein stochastische Prozess ergodisch und stationär ist, lässt sich der stationäre Ensemblemittelwert durch den Zeitmittelwert schätzen!

12.4.2. Schätzen der Autokovarianzfunktion

- $Cov(y_t, y_{t-h}) = E[(y_t - \mu_t)(y_{t-h} - \mu_{t-h})]$ ist ein Erwartungswert.
- Grundidee des Schätzens: Schätze Erwartungswert durch Durchschnittsbildung. Dies geht bei Zeitreihen wiederum nur, wenn zugrundeliegender DGP schwach stationär ist. Dann ist

$$Cov(y_T, y_{T-k}) = Cov(y_{T-1}, y_{T-1-k}) = \dots = Cov(y_{1+K}, y_1) = \gamma_k$$

und man hat auf einmal $T - k$ Beobachtungen

$$(y_T - \bar{y}_T)(y_{T-k} - \bar{y}_T), (y_{T-1} - \bar{y}_T)(y_{T-k-1} - \bar{y}_T), \dots, (y_{1+k} - \bar{y}_T)(y_1 - \bar{y}_T)$$

zur Verfügung, über die man mitteln kann. Möglicher Schätzer der Autokovarianzfunktion:

$$\hat{\gamma}_h = \frac{1}{T-h} \sum_{t=h+1}^T (y_t - \bar{y})(y_{t-h} - \bar{y}). \quad (12.43)$$

Ein Problem dieses Schätzers ist, dass man für h nahe T wieder nur über sehr wenig Beobachtungen mittelt und zwar unabhängig vom Stichprobenumfang.

Alternativer Schätzer:

$$\tilde{\gamma}_k = \frac{1}{T} \sum_{t=k+1}^T (y_t - \bar{y})(y_{t-k} - \bar{y}). \quad (12.44)$$

- Schätzeigenschaften (Brockwell & Davis 1991, Abschnitt 7.2)

– verzerrt

– Wird Autokovarianzmatrix (12.42) mit (12.44) geschätzt, ist die resultierende geschätzte Autokovarianzmatrix

$$\tilde{\Gamma}_T = \begin{pmatrix} \tilde{\gamma}_0 & \tilde{\gamma}_1 & \cdots & \tilde{\gamma}_{T-1} \\ \tilde{\gamma}_1 & \tilde{\gamma}_0 & \cdots & \tilde{\gamma}_{T-2} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\gamma}_{T-1} & \tilde{\gamma}_{T-2} & \cdots & \tilde{\gamma}_0 \end{pmatrix} \quad (12.45)$$

nicht negativ-definit. Dies gilt jedoch nicht den Schätzer $\hat{\gamma}_h$ (12.43).

– $\tilde{\Gamma}_T$ ist positiv definit, falls $\tilde{\gamma}_0 > 0$. Siehe Brockwell & Davis (1991, S. 221) zusammen mit Lütkepohl (1996, S. 151).

12.4.3. Schätzen der Autokorrelationsfunktion

- Die Autokorrelationsfunktion ρ_k lässt sich ebenfalls auf zwei Arten schätzen:

$$\hat{\rho}_k = \hat{\gamma}_k / \hat{\gamma}_0, \quad (12.46)$$

$$\tilde{\rho}_k = \tilde{\gamma}_k / \tilde{\gamma}_0. \quad (12.47)$$

– **Schätzeigenschaften Theorem** (Brockwell & Davis 1991, Theorem 7.2.1) Ist $\{y_t\}$ ein stationärer linearer Prozess (12.6) mit Mittelwert $\mu = E(y_t)$ und unabhängigem Weißem Rauschen

$$y_t = \mu + \sum_{j=-\infty}^{\infty} \psi_j u_{t-j}, \quad u_t \sim IID(0, \sigma^2),$$

wobei $\sum_{h=-\infty}^{\infty} |\gamma_h| < \infty$ und $E(u_t^4) < \infty$, dann gilt für jedes $h \in \mathbb{N}$

$$\sqrt{T}(\tilde{\rho}_h - \rho) \xrightarrow{d} N(\mathbf{0}, \mathbf{W}) \quad (12.48)$$

$$\tilde{\rho}'_h = (\tilde{\rho}_1 \quad \tilde{\rho}_2 \quad \cdots \quad \tilde{\rho}_h) \quad (12.49)$$

$$\rho'_h = (\rho_1 \quad \rho_2 \quad \cdots \quad \rho_h) \quad (12.50)$$

und \mathbf{W} ist eine Kovarianzmatrix mit ij -tem Element

$$w_{ij} = \sum_{k=-\infty}^{\infty} \left(\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} + 2\rho_i\rho_j\rho_k^2 - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} \right). \quad (12.51)$$

- Die Bedingung der Existenz vierter Momente in obigem Theorem kann ersetzt werden durch (Brockwell & Davis 1991, Theorem 7.2.2)

$$\sum_{j=-\infty}^{\infty} \psi_j^2 |j| < \infty. \quad (12.52)$$

- Ist $y_t \sim IID(0, \sigma^2)$, dann gilt $\rho_k = 0$ für $|k| > 0$ und $w_{ij} = 1$, falls $i = j$ und Null sonst. Man erhält damit eine asymptotisch pivote Verteilung für den Schätzer

$$\sqrt{T}(\tilde{\rho}_h - \mathbf{0}) \xrightarrow{d} N(\mathbf{0}, \mathbf{I}). \quad (12.53)$$

Daraus ergeben sich Konfidenzintervalle für die geschätzten Autokorrelationen von (unabhängigem) Weißem Rauschen. Für $\alpha = 0.95$ ergibt sich

$$[-1.96/\sqrt{T}, 1.96/\sqrt{T}].$$

- Die partielle Autokorrelation a_k lässt sich einfach mit Hilfe des OLS-Schätzers für a_k in dem autoregressiven Modell

$$y_t = \nu + a_1 y_{t-1} + \dots + a_k y_{t-k} + u_t$$

schätzen.

R-Befehle

Schätzen der Autokovarianzfunktion, der Autokorrelationsfunktion oder der partiellen Autokorrelationsfunktion einer Zeitreihe: mit `acf()`.

Beispiel: Schätzen der Autokorrelationsfunktion einer Realisation von Gauschem Weißem Rauschen mit $\sigma^2 = 4$ und $n = 100$.

In Abbildung 12.9 liegt keine ACF für die betrachteten Lags außerhalb des Konfidenzintervalls. Dies deutet auf das Vorliegen von Weißem Rauschen hin.

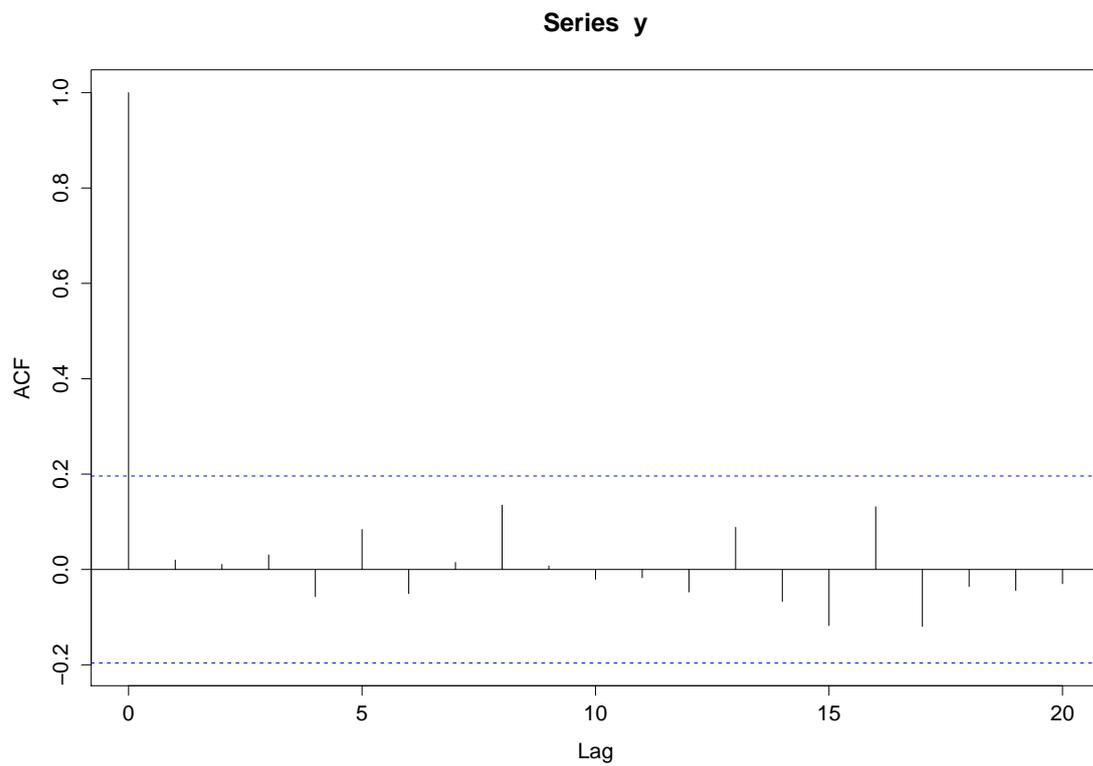


Abbildung 12.9.: Geschätzte Autokorrelationsfunktion mit 95%-Konfidenzintervallen einer Realisation von Gaußischem Weißen Rauschen mit $\sigma^2 = 4$ und $n = 100$ (**R-Programm**, siehe Abschnitt [A.14](#), Seite 358)

13. Modelle für multivariate Zeitreihen

13.1. Multivariate datengenerierende Prozesse

Ergänzungen zu und Wiederholung von Abschnitt 5.1:

- Es bezeichne \mathbf{s}_t einen $(m \times 1)$ -Vektor von (ökonomischen) Zufallsvariablen, die in Periode t generiert werden und gleichzeitig und über die Zeit hinweg in Beziehung stehen können.
- Zur Notation: Wie bereits in Abschnitt 4.1 kann der Vektor \mathbf{s}_t mehr Variablen enthalten, als letztlich modelliert werden müssen. Auch verzichten wir in den folgenden Abschnitten auf den Index bei den Dichtefunktionen.
- Die Kollektion $\{\mathbf{s}_t\}_{t \in \mathbb{T}}$ ist ein **vektorwertiger** bzw. **multivariater stochastischer Prozess**.
- Der **datengenerierende Prozess (data generating process, DGP)** eines m -dimensionalen multivariaten stochastischen Prozesses wird vollständig durch die bedingte Dichte

$$f_t(\mathbf{s}_t | \mathcal{S}_{t-1})$$

beschrieben, wobei \mathcal{S}_{t-1} die Informationsmenge aller verzögerten Vektoren \mathbf{s}_{t-j} , $j > 0$ bezeichnet

$$\mathcal{S}_{t-1} = \{\mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \mathbf{s}_{t-3}, \dots\}.$$

Diese Darstellung ist noch allgemeiner als die bedingten Dichten, die in (5.2) vorkommen, da bei allen bedingten Dichten Vorstichprobenwerte (presample values), $\mathbf{s}_0, \mathbf{s}_{-1}, \dots$, zugelassen sind.

- Natürlich könnte man anstelle der bedingten Dichte f auch die bedingte Wahrscheinlichkeitsfunktion F verwenden. Diese muss sogar verwendet werden, wenn nichtstetige Zufallsvariablen verwendet werden.

‡ Formal bezeichnet \mathcal{S}_{t-1} die (kleinste) σ -Algebra, also die kleinste Menge an Teilmengen, die es erlaubt, allen möglichen Ereignissen auf Basis der berücksichtigten erklärenden (Zufalls)vektoren $\mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \dots$, Wahrscheinlichkeiten zuzuordnen. Siehe zur Definition einer σ -Algebra Abschnitt 2.3. Korrekt müsste man anstelle von $\mathcal{S}_{t-1} = \{\mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \mathbf{s}_{t-3}, \dots\}$ also

$$\mathcal{S}_{t-1} = \sigma(\mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \mathbf{s}_{t-3}, \dots)$$

schreiben.

- Beachte, dass die Informationsmenge \mathcal{S}_t nicht kleiner wird, da gilt

$$\mathcal{S}_{t-2} \subseteq \mathcal{S}_{t-1} \subseteq \mathcal{S}_t \subseteq \dots,$$

also nichts vergessen bzw. Wissen akkumuliert wird. Die Informationsmengen sind über die Zeit hinweg geschachtelt (nested) (Davidson 2000, Sections 4.1, 5.3.1 und insbesondere 6.2.1).

13.2. Dynamische ökonometrische Modelle

- Im Folgenden verallgemeinern wir die bisherige Definition von ökonometrischen Modellen für Zufallsstichproben aus Abschnitt 5.2 für Zeitreihen.
- Ein **dynamisches ökonometrisches Modell** \mathbb{M} ist eine Familie von Funktionen $M(\cdot)$ in Abhängigkeit von den Daten und einem $p \times 1$ Parametervektor $\boldsymbol{\psi}$, dessen Elemente über die Zeit hinweg konstant sind. Die Funktionen beschreiben den gesamten DGP oder Teile davon, bzw. approximieren diesen zumindest (Davidson 2000, Section 4.1.1). Die Menge an möglichen und erlaubten Parametern ist der **Parameterraum** Ψ

$$\mathbb{M} = \{M(\mathbf{s}_t, \mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \dots, \mathbf{s}_2, \mathbf{s}_1, \dots, \mathbf{d}_t; \boldsymbol{\psi}), \boldsymbol{\psi} \in \Psi\}, \quad \Psi \subseteq \mathbb{R}^p \quad (13.1)$$

- In (13.1) bezeichnet der Vektor \mathbf{d}_t nicht-stochastische Variablen, aber möglicherweise zeitveränderliche Variablen, z.B. eine Konstante 1, einen Zeittrend t , Saisondummies, etc.
- Zeitabhängige Parameter werden über eine Funktion $\boldsymbol{\psi}_t = h(\mathbf{d}_t, \boldsymbol{\psi})$ erfasst.

Beispiel: AR(1)-Modell (12.16):

$$\text{Parametervektor } \boldsymbol{\psi} = \begin{pmatrix} \nu \\ \alpha_1 \\ \sigma^2 \end{pmatrix}, \text{ Parameterraum } \Psi = \mathbb{R} \times (-1, 1) \times \mathbb{R}^+.$$

Beispiel: Strukturelles vektorautoregressives Modell

Das Beispiel folgt Davidson (2000, Sections 4.5.5 und 4.7.2). Es sei $\mathbf{s}_t = \begin{pmatrix} y_t & z_t \end{pmatrix}^T$, wobei die stochastische Dynamik der Variablen x_t und z_t durch folgendes simultanes Gleichungssystem bestimmt wird:

$$y_t = \gamma_1 - \alpha_{12}z_t + \beta_{11}y_{t-1} + \beta_{12}z_{t-1} + u_{1t} \quad (13.2a)$$

$$z_t = \gamma_2 - \alpha_{21}y_t + \beta_{21}y_{t-1} + \beta_{22}z_{t-1} + u_{2t}, \quad (13.2b)$$

mit

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} \sim NID \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \underbrace{\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix}}_{\Sigma} \right). \quad (13.2c)$$

Das Modell (13.2) wird als ein **strukturelles vektorautoregressives Modell (SVAR-Modell)** bezeichnet. In Abschnitt 13.3 analysieren wir, welche Regressoren auf der rechten Seite von (13.2) exogen sind und in welchem Sinne. In der

Veranstaltung **Quantitative Wirtschaftsforschung II** werden Schätzmethoden und weitere Details behandelt. Der Parametervektor des Modells ist

$$\boldsymbol{\psi} = \left(\alpha_{12} \quad \alpha_{21} \quad \beta_{11} \quad \beta_{12} \quad \beta_{21} \quad \beta_{22} \quad \gamma_1 \quad \gamma_2 \quad \sigma_{11} \quad \sigma_{12} \quad \sigma_{22} \right)^T.$$

- Wie im Fall von Modellen für Zufallsstichproben (vgl. Abschnitt 5.2), sagen wir, dass das Modell \mathbb{M} **vollständig spezifiziert** ist, wenn sich aus \mathbb{M} ein Modell in reduzierter Form \mathbb{M}_D ableiten lässt, das als Elemente bedingte Dichten $f(\mathbf{s}_t | \mathcal{S}_{t-1}, \mathbf{d}_t; \boldsymbol{\theta}(\boldsymbol{\psi}))$ enthält (vgl. (5.19))

$$\mathbb{M}_D \equiv \{f(\mathbf{s}_t | \mathcal{S}_{t-1}, \mathbf{d}_t; \boldsymbol{\theta}), \boldsymbol{\theta} \in \boldsymbol{\Theta}\}.$$

- Ist ein strukturelles dynamisches Modell \mathbb{M} **vollständig und darüber hinaus korrekt** spezifiziert, existiert ein Parametervektor $\boldsymbol{\theta}_0 = \boldsymbol{\theta}(\boldsymbol{\psi}_0)$, für den die bedingte Dichte in \mathbb{M}_D dem DGP entspricht:

$$\begin{aligned} \mathbb{M}_D \supseteq M_D(\mathbf{s}_t, \mathbf{s}_{t-1}, \mathbf{s}_{t-2}, \dots, \mathbf{s}_2, \mathbf{s}_1, \dots, \mathbf{d}_t; \boldsymbol{\theta}_0) &\equiv f(\mathbf{s}_t | \mathcal{S}_{t-1}, \mathbf{d}_t; \boldsymbol{\theta}_0) \\ &= \underbrace{f_t(\mathbf{s}_t | \mathcal{S}_{t-1})}_{DGP}. \end{aligned} \quad (13.3)$$

Beispiel: SVAR-Modell – Reduzierte Form

Das SVAR-Modell (13.2) ist ein simultanes Gleichungsmodell. Deshalb muss das SVAR-Modell \mathbb{M} zu einem Modell \mathbb{M}_D in reduzierter Form umgeformt werden, damit die Menge der im Modell enthaltenen DGPs sichtbar wird. Hierfür ist es günstig, das SVAR-Modell (13.2) in Matrixschreibweise

$$\mathbf{B}\mathbf{s}_t = \mathbf{c} + \mathbf{C}\mathbf{s}_{t-1} + \mathbf{u}_t, \quad \mathbf{u}_t \sim NID(\mathbf{0}, \boldsymbol{\Sigma}). \quad (13.4)$$

mit

$$\mathbf{B} = \begin{pmatrix} 1 & \alpha_{12} \\ \alpha_{21} & 1 \end{pmatrix}, \quad \mathbf{x}_t = \begin{pmatrix} y_t \\ z_t \end{pmatrix}, \quad \mathbf{c} = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix}, \quad \mathbf{C} = \begin{pmatrix} \beta_{11} & \beta_{12} \\ \beta_{21} & \beta_{22} \end{pmatrix}, \quad \mathbf{u}_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

zu schreiben.

Um die **reduzierte Form** des SVAR-Modells (13.4) zu erhalten multipliziert man die Matrixgleichung mit der Inversen von \mathbf{B} (vorausgesetzt

$$\mathbf{B}^{-1} = \frac{1}{1 - \alpha_{12}\alpha_{21}} \begin{pmatrix} 1 & -\alpha_{12} \\ -\alpha_{21} & 1 \end{pmatrix}$$

existiert, d. h. $\alpha_{12} \neq \alpha_{21}$):

$$\begin{aligned} \mathbf{s}_t &= \underbrace{\mathbf{B}^{-1}\mathbf{c}}_{\mathbf{a}} + \underbrace{\mathbf{B}^{-1}\mathbf{C}}_{\mathbf{A}_1} \mathbf{s}_{t-1} + \underbrace{\mathbf{B}^{-1}\mathbf{u}_t}_{\boldsymbol{\varepsilon}_t}, \\ \mathbf{s}_t &= \mathbf{a} + \mathbf{A}_1 \mathbf{s}_{t-1} + \boldsymbol{\varepsilon}_t, \end{aligned} \quad (13.5a)$$

$$\boldsymbol{\varepsilon}_t = \mathbf{B}^{-1}\mathbf{u}_t = \frac{1}{1 - \alpha_{21}\alpha_{12}} \begin{pmatrix} u_{1t} - \alpha_{12}u_{2t} \\ u_{2t} - \alpha_{21}u_{1t} \end{pmatrix} \quad (13.5b)$$

$$\text{Var}(\boldsymbol{\varepsilon}_t) = \boldsymbol{\Omega} = \mathbf{B}^{-1}\boldsymbol{\Sigma}(\mathbf{B}')^{-1} \quad (13.5c)$$

mit bedingter Dichte

$$\mathbf{s}_t | \mathcal{S}_{t-1} \sim N(\mathbf{B}^{-1}\mathbf{c} + \mathbf{B}^{-1}\mathbf{C}\mathbf{x}_{t-1}, \mathbf{\Omega}) \quad (13.5d)$$

und mit bedingtem Erwartungswert

$$E(\mathbf{s}_t | \mathcal{S}_{t-1}) = \mathbf{B}^{-1}\mathbf{c} + \mathbf{B}^{-1}\mathbf{C}\mathbf{x}_{t-1}. \quad (13.5e)$$

Die Kovarianzmatrix von $\boldsymbol{\varepsilon}_t$ lautet

$$\mathbf{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} \\ \omega_{12} & \omega_{22} \end{pmatrix} = \frac{1}{(1 - \alpha_{21}\alpha_{12})^2} \begin{pmatrix} \sigma_{11} - 2\alpha_{12}\sigma_{21} + \alpha_{12}^2\sigma_{22} & (1 + \alpha_{12}\alpha_{21})\sigma_{21} - \alpha_{12}\sigma_{22} - \alpha_{21}\sigma_{11} \\ \omega_{12} & \sigma_{22} - 2\alpha_{21}\sigma_{21} + \alpha_{21}^2\sigma_{11} \end{pmatrix}. \quad (13.5f)$$

Die bedingte Normalverteilung für \mathbf{s}_t folgt aus der Linearität der multivariaten Normalverteilung.

Das Modell (13.5) ist die reduzierte Form eines SVAR-Modells wird i. Allg. als **VAR-Modell** bezeichnet.

- Die Elemente eines Modells in struktureller Form, also die Funktionen $M(\cdot)$, enthalten typischerweise mehr Parameter als durch das dazugehörige Modell in reduzierter Form festgelegt werden. Dann gibt es eine Funktion $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\psi})$, die nicht one-to-one ist, so dass aus Kenntnis von $\boldsymbol{\theta}_0$, dem wahren Parametervektor der reduzierten Form, nicht eindeutig $\boldsymbol{\psi}$ bestimmt werden kann.

Beispiel: SVAR-Modell Der Parametervektor $\boldsymbol{\psi}$ des SVAR-Modells (13.2) enthält 11 verschiedene Parameter. Der Parametervektor des reduzierten Modells (13.5) nur 9: $\boldsymbol{\theta} = (a_1 \ a_2 \ A_{11} \ A_{12} \ A_{21} \ A_{22} \ \omega_{11} \ \omega_{12} \ \omega_{22})$.

Problem: Ist dies der Fall, reicht die Information aus den Daten nicht aus, um die Parameter der strukturellen Form schätzen zu können. Zusätzlich sind hierfür sogenannte **Identifikationsannahmen** erforderlich, die aus der ökonomischen Theorie kommen müssen. Siehe **Quantitative Wirtschaftsforschung II**.

Denn Stichprobeninformation ermöglicht prinzipiell lediglich das Schätzen der Parameter von reduzierten Formen.

- Die strukturelle und reduzierte Form eines Modells können identisch sein.

Einfaches Beispiel: SVAR-Modell und VAR-Modell identisch Wird a priori in (13.2) angenommen, dass $\alpha_{12} = \alpha_{21} = 0$ gilt, dann verschwindet die simultane Beziehung zwischen z_t und y_t und z_t wird kausal für y_t . Dann entsprechen sich strukturelle und reduzierte Form.

Der Fall im vorhergehenden Beispiel wird allgemeiner in Abschnitt 13.3 betrachtet.

- Vgl. Davidson (2000, Section 4.1).

13.3. Bedingungen an exogene Variablen in dynamischen Modellen

- Selbst wenn man das korrekte und vollständige Modell \mathbb{M}_D (13.3) kennen würde, wäre es – gegeben typische Stichprobengrößen – unmöglich, den korrekten $(p^* \times 1)$ -Parametervektor ψ_0 verlässlich zu schätzen, wenn die Anzahl der Modellparameter p^* extrem groß ist. Das ist der Fall, wenn die Anzahl der betrachteten Variablen m sehr groß ist.
- Ist man nur an der Erklärung / Modellierung ausgewählter Variablen \mathbf{y}_t interessiert, so können wie in Abschnitt 5.2 bedingte Modelle verwendet werden. Für Zeitreihen muss jedoch die für Zufallsstichproben passende Definition (5.19) geeignet erweitert werden.
- Es gelte auch hier die Zerlegung des Vektors \mathbf{s}_t gemäß (5.5) in irrelevante Variablen \mathbf{w}_t , zu erklärende Variablen \mathbf{y}_t und erklärende Variablen \mathbf{z}_t .
- Aufgrund der Zeitstruktur der Daten und des DGP lassen sich unterschiedliche Arten von Exogenität unterscheiden, die von Engle et al. (1983) eingeführt wurden und sehr hilfreich sind:
 - **schwache Exogenität (weak exogeneity)**: z_t ist im Rahmen des betrachteten Modells kausal für y_t innerhalb der gleichen Zeitperiode
 - **starke Exogenität (strong exogeneity)**: z_t kann in Mehr-Schritt-Prognosen für y_{t+h} als gegeben betrachtet werden
 - **Super-Exogenität (super-exogeneity)**: z_t erfüllt die Voraussetzung, als wirtschafts-politische Steuerungsvariable benutzt zu werden.

Am Ende des Abschnitts werden diese in Bezug gesetzt zu den bisherigen Definitionen streng exogener und vorherbestimmter Variablen.

Vorgehen zum Definieren bedingter Modelle für Zeitreihen

- Partitioniere ursprünglichen Vektor \mathbf{s}_t und definiere dazugehörige Informationsmengen

$$\mathbf{s}_t = \begin{pmatrix} \mathbf{w}_t \\ \mathbf{y}_t \\ \mathbf{z}_t \end{pmatrix}, \quad \begin{aligned} \mathcal{W}_{t-1} &= \{\mathbf{w}_{t-1}, \mathbf{w}_{t-2}, \dots\} = \sigma(\mathbf{w}_{t-1}, \mathbf{w}_{t-2}, \dots) \\ \mathcal{Y}_{t-1} &= \{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots\} = \sigma(\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots) \\ \mathcal{Z}_{t-1} &= \{\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \dots\} = \sigma(\mathbf{z}_{t-1}, \mathbf{z}_{t-2}, \dots) \end{aligned},$$

so dass \mathbf{w}_t die irrelevanten Variablen enthält, \mathbf{y}_t die endogenen Variablen, die innerhalb des Modells erklärt werden müssen, und \mathbf{z}_t die Variablen, die für die vorliegende Fragestellungen nicht im Modell erklärt werden müssen, aber für Elemente von \mathbf{y}_t sind, also exogen sind.

- Kombination von Informationsmengen. Beachte die Verwendung des Symbols \vee (Davidson 2000, Section B.10):

$$\mathcal{S}_{t-1} = \sigma(\mathbf{w}_{t-1}, \mathbf{y}_{t-1}, \mathbf{z}_{t-1}, \mathbf{w}_{t-2}, \dots) \equiv \mathcal{W}_{t-1} \vee \mathcal{Y}_{t-1} \vee \mathcal{Z}_{t-1} \neq \mathcal{W}_{t-1} \cup \mathcal{Y}_{t-1} \cup \mathcal{Z}_{t-1}.$$

- Definiere

$$f(\mathbf{s}_t | \mathcal{W}_{t-1} \vee \mathcal{Y}_{t-1} \vee \mathcal{Z}_{t-1}) \equiv f(\mathbf{s}_t | \mathcal{W}_{t-1}, \mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}).$$

- Analog zu (5.6) Faktorisierung der (parametrischen) Dichte f für $\boldsymbol{\theta} = \boldsymbol{\theta}(\boldsymbol{\psi})$

$$\begin{aligned} f(\mathbf{s}_t | \mathcal{S}_{t-1}, \mathbf{d}_t, \boldsymbol{\theta}) &= f(\mathbf{w}_t | \mathbf{y}_t, \mathbf{z}_t, \mathcal{S}_{t-1}, \mathbf{d}_t, \boldsymbol{\theta}) f(\mathbf{y}_t | \mathbf{z}_t, \mathcal{S}_{t-1}, \mathbf{d}_t, \boldsymbol{\theta}) f(\mathbf{z}_t | \mathcal{S}_{t-1}, \mathbf{d}_t, \boldsymbol{\theta}) \\ \text{kurz: } f_{\mathbf{w}, \mathbf{y}, \mathbf{z}} &= f_{\mathbf{w} | \mathbf{y}, \mathbf{z}} f_{\mathbf{y} | \mathbf{z}} f_{\mathbf{z}}. \end{aligned} \quad (13.6)$$

Annahmen für schwache Exogenität

- Es existiert eine Partitionierung des Parametervektors $\boldsymbol{\psi}$

$$\boldsymbol{\psi} = \begin{pmatrix} \boldsymbol{\psi}_1 \\ \boldsymbol{\psi}_2 \end{pmatrix}, \quad \boldsymbol{\psi}_1 \in \boldsymbol{\Psi}_1, \boldsymbol{\psi}_2 \in \boldsymbol{\Psi}_2, \quad \boldsymbol{\Psi} = \boldsymbol{\Psi}_1 \times \boldsymbol{\Psi}_2, \quad (13.7)$$

- und für die bedingten Dichten gilt:

$$f_{\mathbf{w} | \mathbf{y}, \mathbf{z}} = f(\mathbf{w}_t | \mathbf{y}_t, \mathbf{z}_t, \mathcal{W}_{t-1}, \mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}, \mathbf{d}_t, \boldsymbol{\theta}_2, \quad) \quad (13.8a)$$

$$f_{\mathbf{y} | \mathbf{z}} = f(\mathbf{y}_t | \quad \mathbf{z}_t, \quad \mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}, \mathbf{d}_t, \quad \boldsymbol{\theta}_1) \quad (13.8b)$$

$$f_{\mathbf{z}} = f(\mathbf{z}_t | \quad \mathcal{W}_{t-1}, \mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}, \mathbf{d}_t, \boldsymbol{\theta}_2, \quad) \quad (13.8c)$$

mit $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1(\boldsymbol{\psi}_1)$ und $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2(\boldsymbol{\psi}_2)$.

- In Worten:
 - Die Funktionen der bedingten Dichten für die nicht zu erklärenden Variablen $\mathbf{w}_t, \mathbf{z}_t$ hängen nicht vom Parametervektor $\boldsymbol{\psi}_1$ ab.
 - Die Funktion der bedingten Dichte für die zu erklärenden \mathbf{y}_t hängt nicht von $\boldsymbol{\psi}_2$ und nicht von der Vergangenheit von \mathbf{w}_t ab. Der multivariate stochastische Prozess $\{\mathbf{w}_t\}$ ist für $f_{\mathbf{y} | \mathbf{z}}$ also irrelevant.
 - Es gilt nicht $\boldsymbol{\psi}_1 = \boldsymbol{\psi}_1(\boldsymbol{\psi}_2)$, so dass Kenntnis von $\boldsymbol{\psi}_2$ keine Verbesserung der Schätzeigenschaften für $\boldsymbol{\psi}_1$ haben kann. Man bezeichnet dann $\boldsymbol{\psi}_1$ und $\boldsymbol{\psi}_2$ als **variation free**.
- Gelten die Annahmen (13.7) und (13.8), ist es für die vollständige Modellierung von \mathbf{y}_t egal, ob $f_{\mathbf{w}, \mathbf{y}, \mathbf{z}}$ oder nur $f_{\mathbf{y} | \mathbf{z}}$ betrachtet wird.
- Wird $f_{\mathbf{y} | \mathbf{z}}$ betrachtet, sagt man, dass das Modell **bedingt auf \mathbf{z}_t (conditional model)** ist und **bezüglich \mathbf{w}_t marginalisiert** ist.
- Eine **Parametrisierung eines Modells ist niemals eindeutig**, da mittels einer beliebigen vektorwertigen Funktion $\boldsymbol{\phi} = f(\boldsymbol{\psi})$, die bijektiv ist, eine alternative Parametrisierung erzeugt werden kann, allerdings mit anderer **Interpretation**. Beispiel: $\boldsymbol{\phi} = \exp(\boldsymbol{\beta})$.
- Für die **Existenz** der bedingten Dichte $f_{\mathbf{y} | \mathbf{z}}$ in (13.8b) ist also wichtig, dass *irgendein*

Parametervektor ψ existiert, der (13.7) und (13.8) erfüllt.

- Man könnte u. U. ψ_2 weiter in Parameter für $f_{w|y,z}$ und f_z aufteilen. Da beide bedingten Dichten für die Analyse unter Annahmen (13.7) und (13.8) irrelevant sind, ist dies nicht nötig.

Schwache Exogenität

- Ist das **Ziel der Analyse** auf die Erklärung von \mathbf{y}_t (anstelle von allen Variablen in \mathbf{s}_t) beschränkt, möchte man nur das konditionale Modell für $f_{y|z}$ analysieren und auf die Analyse des marginalen Modells für f_z verzichten (vgl. Abschnitt 5.2).
- Dies ist genau dann möglich, wenn durch den Verzicht auf die Analyse des marginalen Modells, das f_z impliziert, keine Information für die Parameter ψ_1 des konditionalen Modells, das $f_{y|z}$ impliziert, verloren geht. Die Bedingungen hierfür sind (13.7) und (13.8). Die Variablen \mathbf{z}_t des marginalen Modells werden dann als **schwach exogen mit Bezug auf ψ** für das konditionale Modell bezeichnet, das $f_{y|z}$ impliziert. Die verbleibenden Variablen \mathbf{y}_t werden als **endogen** bezeichnet.
- Eine Beschränkung der Analyse auf das konditionale Modell (13.8b) hat genau dann Sinn, wenn die bedingte Dichte zur Erklärung von \mathbf{y}_t wesentlich weniger Parameter und bedingende Variablen enthält als die bedingte Dichte für \mathbf{w}_t .
- Natürlich kann sich das **Interesse bei der Erklärung von \mathbf{y}_t durch \mathbf{z}_t und $\mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}$** auf Basis der bedingten Dichte $f_{y|z}$ auch auf einen beliebigen **Parametervektor ϕ_1** beziehen, solange dieser durch $\phi_1 = g(\psi_1)$ bestimmt ist (eine Möglichkeit ist θ_1), wobei $g(\cdot)$ nicht umkehrbar sein muss und irgendeine Parametrisierung ψ existiert, für die (13.7) und (13.8) gilt.
- Dann werden die Variablen in \mathbf{y}_t als **endogen** und die Variablen in \mathbf{z}_t als **schwach exogen mit Bezug auf ϕ_1 (weakly exogenous for ϕ_1)** bezeichnet. Diese Bezeichnung wurde von [Engle et al. \(1983\)](#) eingeführt.

Bemerkungen:

- Teilung des Parametervektors ψ in ψ_1 und ψ_2 : 'sequential cut of the parameters' in 'die, die Auswirkungen für die Analyse haben' und 'den Rest' ([Engle et al. 1983](#)).
- Die Bedingung (13.7) garantiert, dass ϕ_1 weder direkt noch indirekt von ψ_2 abhängt. Damit ist es nicht möglich, dass z.B. die Kenntnis der Parameter des marginalen Modells für \mathbf{z}_t helfen würde, ϕ_1 genauer zu bestimmen.
- Beachte: Die Zerlegung (13.6) kann immer auch anders, z.B. $f_{w,y,z} = f_{w|y,z} f_{z|y} f_y$ erfolgen, aber möglicherweise ohne 'sequential cut of the parameters'.
- Häufig gesucht: Kleinste Teilmenge endogener Variablen, also wenn möglich y_t skalar, so dass (13.7) und (13.8) gerade noch gelten.
- Schwache Exogenität ist eine Eigenschaft,

- die sich auf **Variable und Parameter** innerhalb eines Modells bezieht,
- **nicht gleichbedeutend mit Kausalität** ist,
- jedoch simultane Beziehungen ausschließt (vgl. Abschnitt 4.1) und
- nur mit Bezug auf das 'größere' Modell $f_{w,y,z}$ beurteilt werden kann.

Der häufige Sprachgebrauch, \mathbf{z}_t ist exogen für \mathbf{y}_t , ist also ungenau, da die Abhängigkeit von den Modellparametern nicht deutlich wird.

- Vgl. Davidson (2000, Section 4.5.3), Hendry (1995, Chapter 5, insb. Sec. 5.3)
- Die Bedingungen für schwache Exogenität schließen nicht aus, dass es Rückkoppelungseffekte von \mathbf{y}_t auf \mathbf{z}_{t+1} (via \mathcal{Y}_t in (13.8c)) und damit auf zukünftige \mathbf{y}_t gibt. Genau deshalb wird die schwache Exogenität als „schwach“ bezeichnet. Um Rückkoppelungseffekte festzustellen, ist das Konzept der Granger-Kausalität zentral, das jedoch nicht identisch ist mit dem Konzept von Kausalität, das in Abschnitt 4.1 definiert wurde.

Für makroökonomische Fragestellungen lassen sich im Allgemeinen keine kontrollierten Zufallsexperimente durchführen und natürliche Experimente sind selten zu finden. → Verwendung eines schwächeren Konzepts:

Granger-Kausalität

Clive Granger (1969) (Nobelpreisträger 2003, zusammen mit Robert Engle).

- Eine Variable z_t ist **Granger-kausal** für y_t , wenn Kenntnis von z_t irgendwie hilft, die Vorhersage für y_{t+h} für mindestens ein $h > 0$ zu verbessern. Eine hinreichende Voraussetzung ist, dass für mindestens einen Prognosehorizont $h > 0$ *nicht gilt*

$$f(y_{t+h}|\{z_t, z_{t-1}, \dots\}, \tilde{\Omega}_t) = f(y_{t+h}|\tilde{\Omega}_t), \quad (13.9)$$

wobei $\tilde{\Omega}_t$ eine Informationsmenge bezeichnet, die alle beliebigen Variablen enthalten kann **außer** $\{z_t, z_{t-1}, \dots\}$.

- Gilt (13.9) für alle $h > 0$, ist z_t **nicht Granger-kausal** für y_t .
- Granger-Kausalität $\not\Rightarrow$ Existenz eines kausalen Mechanismus.
- Granger-Kausalität bezieht sich ausschließlich auf Prognosefähigkeit.
- Vgl. Davidson (2000, Section 4.5.4), Lütkepohl (2004, Section 3.7.1).

Starke Exogenität (strong exogeneity)

- Betrachte $\mathbf{z}_t = (\mathbf{z}'_{1t} \quad \mathbf{z}'_{2t})'$.

- y_t ist **nicht Granger-kausal** für z_{2t} , wenn gilt

$$f(\mathbf{z}_{1t}, \mathbf{z}_{2t} | \mathcal{W}_{t-1}, \mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}) = f(\mathbf{z}_{1t} | \mathbf{z}_{2t}, \mathcal{W}_{t-1}, \mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}) f(\mathbf{z}_{2t} | \mathcal{W}_{t-1}, \mathcal{Z}_{t-1}), \quad (13.10)$$

- z_{2t} **stark exogen** für ϕ_1 :
 - y_t **nicht Granger-kausal** für z_{2t} und
 - z_{2t} **ist schwach exogen** für ϕ_1
- Ist z_{2t} stark exogen (für ϕ_1), sind auch **verzögerte Rückkoppelungen von y_t ausgeschlossen**. Damit kann z_{2t}
 - in Mehr-Schritt-Prognosen als gegeben behandelt werden.
 - wie die nicht-stochastischen Variablen \mathbf{d}_t behandelt werden.
- Vgl. Davidson (2000, Section 4.5.4) oder Hendry (1995, Section 5.8).

Beispiel: SVAR-Modell

Frage: Unter welchen Parameterrestriktionen ist zur Erklärung von y_t durch Gleichung (13.2a) z_t schwach exogen?

Voraussetzung für schwache Exogenität ist, dass folgender 'sequential cut' der Parameter

$$\boldsymbol{\psi}_1 = (\alpha_{12} \quad \gamma_1 \quad \beta_{11} \quad \beta_{12} \quad \sigma_{11}), \quad \boldsymbol{\psi}_2 = (\alpha_{21} \quad \gamma_2 \quad \beta_{21} \quad \beta_{22} \quad \sigma_{12} \quad \sigma_{21} \quad \sigma_{22}) \quad (13.11)$$

existiert, wobei der Parametervektor $\boldsymbol{\psi}_1$ gerade die Parameter der strukturellen Gleichung (13.2a) und deren Fehlervarianz enthält.

Zum Überprüfen der Bedingung (13.8) wird erst die i) bedingte Dichte abgeleitet, ii) anschließend die Faktorisierung durchgeführt und iii) letztlich die Bedingung überprüft.

i) ist bereits in (13.5) erfolgt.

ii) **Faktorisieren der Dichte f :** Im vorliegenden Fall gibt es kein \mathbf{w}_t . Ist man an der Erklärung von y_t interessiert, benötigt man die Faktorisierung (13.6) $f_{y,z} = f_{y|z} f_z$.

Vorgehen:

- 1. Schritt: Ableitung der Faktorisierung von $f_{\varepsilon_{1t}, \varepsilon_{2t}} = f_{\varepsilon_{1t} | \varepsilon_{2t}} f_{\varepsilon_{2t}}$.
- 2. Schritt: Ersetzen von ε_{1t} und ε_{2t} mit Gleichungen in reduzierter Form (13.5b).

1. Schritt: Aufgrund der Normalverteilungsannahme und $E(\boldsymbol{\varepsilon}_t) = 0$ kann man

schreiben:

$$\varepsilon_{1t} = \rho\varepsilon_{2t} + \eta_t, \quad \text{wobei } E[\eta_t|\varepsilon_{2t}] = 0, \quad (13.12)$$

so dass gilt

$$E[\varepsilon_{1t}|\varepsilon_{2t}] = \rho\varepsilon_{2t}. \quad (13.13)$$

Da die ε 's die Fehler der reduzierten Form (13.5b) sind, gilt, dass ε_{1t} nicht z_t enthält und ε_{2t} nicht y_t enthält (im Gegensatz zu den Fehlern der strukturellen Form (13.2)). Damit lassen sich im 2. Schritt ε_{1t} und ε_{2t} ersetzen durch die jeweiligen Gleichungen in $\boldsymbol{\varepsilon}_t = \mathbf{x}_t - \mathbf{A}_0 - \mathbf{A}_1\mathbf{x}_{t-1}$ und anschließend die bedingte Dichte des konditionalen Modells bestimmen und deren Erwartungswert $E[y_t|z_t, \mathcal{X}_{t-1}]$ berechnen.

Zuvor muss noch der Parameter ρ in (13.12) bestimmt werden. Dies geschieht, indem für Gleichung (13.12) die Kovarianz und Varianz bestimmt werden:

$$\begin{aligned} \text{Cov}(\varepsilon_{1t}, \varepsilon_{2t}) &= a\text{Var}(\varepsilon_{2t}) = \omega_{12} = \rho\omega_{22} \quad \implies \rho = \frac{\omega_{12}}{\omega_{22}} \\ \text{Var}(\varepsilon_{1t}) &= \omega_{11} = \rho^2\omega_{22} + \text{Var}(\eta_t) \quad \implies \text{Var}(\eta_t) = \omega_{11} - \frac{\omega_{12}^2}{\omega_{22}}. \end{aligned}$$

Man erhält

$$\underbrace{\varepsilon_{1t}}_{\text{ist ohne } z_t} = \frac{\omega_{12}}{\omega_{22}} \underbrace{\varepsilon_{2t}}_{\text{ist ohne } y_t} + \eta_t. \quad (13.14)$$

2. Schritt: Jetzt werden ε_{1t} und ε_{2t} in (13.14) durch die jeweiligen Gleichungen in $\boldsymbol{\varepsilon}_t = \mathbf{x}_t - \mathbf{A}_0 - \mathbf{A}_1\mathbf{x}_{t-1}$ ersetzt. Nach einigen Umformungen (Details am Ende des Abschnitts) erhält man für y_t

$$y_t = \frac{\omega_{12}}{\omega_{22}}z_t + (P\gamma_1 - Q\gamma_2) + (P\beta_{11} - Q\beta_{21})y_{t-1} + (P\beta_{12} - Q\beta_{22})z_{t-1} + \eta_t \quad (13.15a)$$

mit

$$\frac{\omega_{12}}{\omega_{22}} = \frac{(1 + \alpha_{12}\alpha_{21})\sigma_{21} - \alpha_{12}\sigma_{22} - \alpha_{21}\sigma_{11}}{\sigma_{22} - 2\alpha_{21}\sigma_{21} + \alpha_{21}^2\sigma_{11}}, \quad (13.15b)$$

$$P = \frac{1 + \alpha_{21}\omega_{12}/\omega_{22}}{1 - \alpha_{21}\alpha_{12}}, \quad Q = \frac{\alpha_{12} + \omega_{12}/\omega_{22}}{1 - \alpha_{21}\alpha_{12}}. \quad (13.15c)$$

Die **bedingte Dichte** $f_{y_t|z_t, \mathbf{x}_{t-1}}$ **des konditionalen Modells für** y_t **gegeben** z_t ist demnach

$$y_t|z_t, \mathbf{x}_{t-1} \sim \quad (13.16)$$

$$N \left(\frac{\omega_{12}}{\omega_{22}} z_t + (P\gamma_1 - Q\gamma_2) + (P\beta_{11} - Q\beta_{21}) y_{t-1} + (P\beta_{12} - Q\beta_{22}) z_{t-1}, \omega_{11} - \frac{\omega_{12}^2}{\omega_{22}} \right).$$

Der **bedingte Erwartungswert** für y_t **gegeben** z_t **und Lags** ist

$$E(y_t | z_t, y_{t-1}, z_{t-1}) = \frac{\omega_{12}}{\omega_{22}} z_t + (P\gamma_1 - Q\gamma_2) + (P\beta_{11} - Q\beta_{21}) y_{t-1} + (P\beta_{12} - Q\beta_{22}) z_{t-1} \quad (13.17)$$

Analog lässt sich auch der Erwartungswert für z_t berechnen.

iii) Überprüfen der Bedingungen (13.8) für die schwache Exogenität

- In die bedingte Dichte (13.16) für das konditionale Modell für y_t gehen *alle* Parameter des Parametervektors ψ ein.
- Die Bedingungen (13.8) für schwache Exogenität können nur erfüllt sein, wenn ein sequential cut (13.11) existiert, so dass die Parameter der Gleichung für z_t nicht das bedingte Modell für y_t beeinflussen. Letzteres ist nur möglich, wenn gilt

$$\frac{\omega_{12}}{\omega_{22}} = -\alpha_{12} \implies P = 1, \quad Q = 0. \quad (13.18)$$

- Damit (13.18) gilt, muss
 - bei $\alpha_{21} \neq 0$ und/oder $\sigma_{21} \neq 0$ die Gleichung $\alpha_{12} = \alpha_{21}^{-1}$ erfüllt sein, wodurch \mathbf{B} nicht invertierbar ist und keine reduzierte Form existiert, oder
 - $\alpha_{21} = 0$ und $\sigma_{12} = 0$ gelten.
- Der **sequential cut** (13.11) ist also nur möglich, **wenn** $\alpha_{21} = \sigma_{21} = 0$.
 - Dann ist z_t **schwach exogen** für ψ_1 bzw. $\phi_1 = g(\psi_1)$ und
 - (13.2) ein **rekursives Modell**.
- Überprüfung von $\alpha_{21} = \sigma_{21} = 0$ nicht mit Regression möglich. Warum?

Bemerkung: Ist z_t **bzgl. des Parametervektors** ψ_1 **nicht schwach exogen**, dann schätzt der KQ-Schätzer nicht die Parameter der strukturellen Gleichung (13.2a), sondern die Parameter des bedingten Erwartungswertes (13.17).

Starke Exogenität z_t ist **stark exogen** für ψ_1 bzw. $\phi_1 = g(\psi_1)$, wenn in der reduzierten Form **zusätzlich** zur schwachen Exogenität ($\alpha_{12} = \sigma_{21} = 0$) gilt

$$\mathbf{B}^{-1} \mathbf{C} \mathbf{x}_{t-1} = \frac{1}{1 - \alpha_{12} \alpha_{21}} \begin{pmatrix} \beta_{11} - \alpha_{12} \beta_{21} & \beta_{12} - \alpha_{12} \beta_{22} \\ -\alpha_{21} \beta_{11} + \beta_{21} & -\alpha_{21} \beta_{12} + \beta_{22} \end{pmatrix} \mathbf{x}_{t-1} = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \mathbf{x}_{t-1},$$

also $\beta_{21} = 0$ (dann hat y_{t-1} keinen Einfluss auf z_t). Es gilt dann

$$E(z_t | \mathcal{X}_{t-1}) = E(z_t | z_{t-1}).$$

Vergleich von Exogenitätskonzepten

- **Schwache Exogenität versus partielle Unabhängigkeit**

- Partielle Unabhängigkeit (9.4) eines Regressors kann immer nur bzgl. eines Fehlerterms beurteilt werden.
- So ist per Konstruktion z_t partiell unabhängig bzgl. η_t , da aus (13.17) folgt: $E[\eta_t | z_t, y_{t-1}, z_{t-1}] = 0$.
- Und so ist im Allgemeinen z_t *nicht* partiell unabhängig bzgl. u_{1t} in (13.2a), weil man zeigen kann, dass im Allgemeinen $E[u_{1t} | z_t, y_{t-1}, z_{t-1}] \neq 0$.
- Ist jedoch z_t schwach exogen bzgl. ψ_1 , dann ist z_t partiell unabhängig von u_{1t} , da dann u_{1t} mit η_t identisch sind, da $\alpha_{12} = 0$.
- Ob ein Regressorvektor partiell unabhängig ist, hängt i) immer von dem zugrundeliegenden 'Fehler-definierenden' Modell ab, ii) und ggf. von den wahren Parametern des „größeren“ Modells. ab, dass den Fehler erzeugt - so wie bei schwacher Exogenität auch.
- Der Vorteil des Konzepts schwacher Exogenität gegenüber dem Konzept der Vorherbestimmtheit ist, dass es explizit anhand Parameterrestriktionen deutlich macht, wann es erfüllt ist.

- **Starke Exogenität versus strenge Exogenität**

- Die Regressoren \mathbf{X}_t werden als **streng exogen (strictly exogenous)** bezeichnet, wenn (9.1) erfüllt ist.
- Entsprechend den Bemerkungen zu schwacher Exogenität gilt: Liegt starke Exogenität einer Variablen bzgl. eines Parametervektors vor, ist diese Variable auch streng exogen.

Aktueller Stand in der Literatur: Eine Ausgabe des *Journal of Econometrics* (2006) ist Kausalität und Exogenität gewidmet, siehe [Bauwens et al. \(2006\)](#).

‡ Super-Exogenität (super-exogeneity)

- liegt vor, wenn die bedingte Verteilung $f_{\mathbf{y}|\mathbf{z}}$ in Abhängigkeit von ϕ_1 invariant gegenüber Veränderungen in der marginalen/gemeinsamen Verteilung $f_{\mathbf{z}}$ ist.
- **Formal:** Der Vektor nicht-stochastischer Variablen \mathbf{d}_t lässt sich in \mathbf{d}_{1t} und \mathbf{d}_{2t} zerlegen, z.B. $\mathbf{d}_t = (1, d_t)'$.

Mindestens ein Element von \mathbf{z}_t ist **super-exogen** für $\phi_1 = g(\psi_1)$:

- \mathbf{d}_{2t} variiert über Beobachtungsperiode und ist nicht triviales Argument der marginalen/gemeinsamen Dichte $f_{\mathbf{z}}$
- Die bedingte Dichte $f_{\mathbf{y}|\mathbf{z}}$ hängt nicht von \mathbf{d}_{2t} ab.

– \mathbf{z}_t ist schwach exogen für ϕ_1 .

- Damit \mathbf{z}_t super-exogen für ϕ_1 ist, muss (13.8) mit $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_1(\psi_1)$ und $\boldsymbol{\theta}_2 = \boldsymbol{\theta}_2(\psi_2)$ weiter restringiert werden:

$$f_{\mathbf{w}|\mathbf{y},\mathbf{z}} = f(\mathbf{w}_t | \mathbf{y}_t, \mathbf{z}_t, \mathcal{W}_{t-1}, \mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}, \mathbf{d}_{1t}, \mathbf{d}_{2t}, \boldsymbol{\theta}_2) \quad (13.19a)$$

$$f_{\mathbf{y}|\mathbf{z}} = f(\mathbf{y}_t | \mathbf{z}_t, \mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}, \mathbf{d}_{1t}, \boldsymbol{\theta}_1) \quad (13.19b)$$

$$f_{\mathbf{z}} = f(\mathbf{z}_t | \mathcal{W}_{t-1}, \mathcal{Y}_{t-1}, \mathcal{Z}_{t-1}, \mathbf{d}_{2t}, \boldsymbol{\theta}_2) \quad (13.19c)$$

Damit \mathbf{z}_t stark exogen für $\phi_1 = g(\psi_1)$ ist, darf \mathcal{Y}_{t-1} in (13.19c) nicht vorkommen.

- Super-Exogenität beseitigt Symmetrie zwischen \mathbf{y}_t und \mathbf{z}_t , so dass Kausalität trotz kontemporärer Korrelation bestimmt werden kann (Identifikation). Im folgenden Beispiel hängt nur eine der beiden bedingten Erwartungswerte von \mathbf{d}_{2t} ab.
- erlaubt Identifikation von Zusammenhängen,
 - die höchstwahrscheinlich strukturelle Interpretation erlauben,
 - die (wirtschafts-)politikinvariant sind und
 - eine Voraussetzung für Immunität gegenüber der Lucas-Kritik gegenüber ökonometrischen Modellen erfüllen.
- Im Kern lautet die **Lucas-Kritik**:

'Given that the structure of an econometric model consists of optimal decision rules for economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models.' (Lucas 1976, p. 41)

(Zitiert nach Hendry (1995, Section 14.14). Siehe ebendort für Tests bezüglich der Lucas-Kritik.)

Die Lucas-Kritik trifft z.B. auch zu, wenn Regressionsparameter sich aus Parametern aus Verhaltensgleichungen und Erwartungen zusammensetzen, wie z.B. bei rationalen Erwartungsmodellen, siehe z.B. Davidson (2000, Section 5.5).

Beispiel: SVAR-Modell

In der zweiten Strukturgleichung (13.2b) ist noch der Term $\delta_2 d_{2t}$ enthalten, die erste Strukturgleichung bleibt unverändert. Dann ist \mathbf{z}_t super-exogen für \mathbf{y}_t .

13.4. Dynamische lineare Regressionsmodelle

Dynamisches lineares Regressionsmodell

- Ein **dynamisches lineares Regressionsmodell** ist ein dynamisches ökonometrisches Modell (13.1), dessen zu erklärende Variable y_t durch eine Linearkombination von erklärenden Variablen und einem Fehlerterm bestimmt wird. Dabei können die erklärenden Variablen verzögerte endogene Variablen y_{t-j} , $j > 0$, enthalten.
- Dynamische lineare Regressionsmodelle modellieren im Allgemeinen den bedingten Erwartungswert der bedingten Dichte (13.6)

$$f_{y|\mathbf{z}} = f(y_t | \mathbf{z}_t, \mathcal{S}_{t-1}, \mathbf{d}_t; \boldsymbol{\theta}) \quad (13.20)$$

des konditionalen Modells, wobei hier y_t skalar ist.

- **Notation:** Im Folgenden nehmen wir an, dass
 - \mathbf{w}_t bereits als nicht relevant klassifiziert wurde.
 - Auch betrachten wir nur **eine** zu erklärende Variable, nämlich y_{1t} , die im weiteren als y_t notiert wird. Es gibt keine weiteren endogenen y_{jt} , $j \geq 2$. Die z_k erklärenden Variablen \mathbf{z}_t werden als

$$\mathbf{Z}_t = (z_{1t} \quad \cdots \quad z_{k_z,t}) = \mathbf{z}_t^T$$

zusammengefasst, so dass sich eine Stichprobenbeobachtung als

$$\begin{pmatrix} y_t \\ \mathbf{Z}_t^T \end{pmatrix} \quad (13.21)$$

schreiben lässt.

Damit lässt sich die bedingte Dichte (13.20) des konditionalen Modells angeben als

$$f_{Y_t | \mathbf{z}_t, \mathbf{z}_{t-1}, \dots, \mathbf{z}_1, Y_{t-1}, \dots, Y_1}(y_t | \mathbf{Z}_t, \mathbf{Z}_{t-1}, \dots, \mathbf{Z}_1, y_{t-1}, \dots, y_1, \mathbf{d}_t).$$

Der zu modellierende bedingte Erwartungswert lautet dann

$$E[y_t | \mathbf{Z}_t, \mathbf{Z}_{t-1}, \dots, \mathbf{Z}_1, y_{t-1}, \dots, y_1, \mathbf{d}_t].$$

Unterstellt man jetzt, dass der bedingte Erwartungswert linear in den Parametern ist, erhält man das dynamische *lineare* Regressionsmodell, das im Folgenden genauer besprochen wird.

- Die **inhaltlich (ökonomisch) relevanten Parameter** von dynamischen linearen Regressionsmodelle können mit dem KQ-Schätzer konsistent geschätzt werden, wenn bestimmte Voraussetzungen erfüllt sind, siehe Abschnitt 13.5. Dazu gehört die schwache Exogenität der Regressoren (bzw. deren Vorherbestimmtheit). Siehe hierzu die umfangreiche Diskussion in Abschnitt 13.3. Diese Voraussetzung wird im Folgenden bei der Definition zulässiger erklärender Variablen bereits berücksichtigt.

Ist die Voraussetzung schwacher Exogenität nicht gegeben, kann trotzdem der bedingte Erwartungswert der reduzierten Form konsistent geschätzt werden, wenn letzterer linear ist. Vgl. hierzu (13.17) im Beispiel des vorangegangenen Abschnitts. Allerdings sind dann die Parameter nicht interpretierbar. Für Prognosezwecke mag dies aber unerheblich sein.

Dynamische lineare Regressionsmodelle

- Alle Regressorvariablen, die zur Spezifikation eines dynamischen linearen Regressionsmodells für die endogene Variable y_t verwendet werden können, bilden die Informationsmenge Ω_t aller *potentiell* erklärenden Variablen. Die Informationsmenge der dann tatsächlich in einem Modell verwendeten Regressorvariablen wird mit $\mathcal{I}_t \subset \Omega_t$ bezeichnet. Siehe hierzu Abschnitt 5.2.
- Mögliche Regressorvariablen in \mathcal{I}_t sind:
 - deterministische Variablen, zusammengefasst im Zeilenvektor \mathbf{d}_t : Konstante, Zeitrend, Saisondummies, etc.,
 - verzögerte abhängige Variablen y_{t-j} , $j > 0$,
 - bezüglich des Fehlerterms u_t vorherbestimmte (kontemporäre) Variablen \mathbf{Z}_t , d. h. $\mathbf{Z}_t \in \Omega_t$, wobei $E[u_t|\Omega_t] = 0$ gilt,
 - verzögerte \mathbf{Z}_t , also \mathbf{Z}_{t-j} , $j > 0$,
 - (fast) jede Funktion der genannten Variablen.
- Ein **dynamisches lineares Regressionsmodell** mit Informationsmenge $\mathcal{I}_t = \{\mathbf{d}_t, \mathbf{Z}_t, \dots, \mathbf{Z}_{t-m}, y_{t-1}, \dots, y_{t-p}\}$ ist gegeben durch

$$y_t = \mathbf{d}_t \boldsymbol{\nu} + \mathbf{Z}_t \boldsymbol{\delta}_0 + \mathbf{Z}_{t-1} \boldsymbol{\delta}_1 + \dots + \mathbf{Z}_{t-m} \boldsymbol{\delta}_m + y_{t-1} \alpha_1 + \dots + y_{t-p} \alpha_p + u_t, \quad t \in \mathbb{T}. \quad (13.22)$$

- Mit

$$\mathbf{X}_t = \left(\mathbf{d}_t \quad \mathbf{Z}_t \quad \mathbf{Z}_{t-1} \quad \dots \quad \mathbf{Z}_{t-m} \quad y_{t-1} \quad \dots \quad y_{t-p} \right), \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\nu} \\ \boldsymbol{\delta}_0 \\ \boldsymbol{\delta}_1 \\ \vdots \\ \boldsymbol{\delta}_m \\ \alpha_1 \\ \vdots \\ \alpha_p \end{pmatrix} \quad (13.23)$$

lässt sich das dynamische lineare Regressionsmodell (13.22) wieder in der bekannten kompakten Form

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t \quad (13.24)$$

schreiben.

Dynamisch korrekt spezifiziertes Modell

- Ein dynamisches lineares Regressionsmodell ist **dynamisch korrekt spezifiziert**, wenn für $\beta = \beta_0$ und $\mathbf{X}_t \in \mathcal{I}_t$ gilt

$$E[y_t|\Omega_t] = E[y_t|\mathbf{d}_t, \mathbf{Z}_t, \mathbf{Z}_{t-1}, \dots, \mathbf{Z}_{t-m}, y_{t-1}, \dots, y_{t-p}] = E[y_t|\mathbf{X}_t, \beta_0] = \mathbf{X}_t\beta_0 \quad (13.25)$$

wobei β_0 der wahre Parametervektor ist.

13.5. KQ-Schätzung von dynamischen linearen Regressionsmodellen

Da AR(p)-Modelle ein Spezialfall von dynamischen linearen Regressionsmodellen sind, ist es ausreichend, die Schätzeigenschaften für letztere zu untersuchen.

- **Annahmen** für asymptotische Schätzeigenschaften des KQ-Schätzers von (13.24):

– (C1) \iff Annahme (B1): Der DGP ist für $\beta = \beta_0$ in (13.24) enthalten.

$$- \text{(C2): } u_t|\Omega_t \sim (0, \sigma^2) \iff \left\{ \begin{array}{l} \text{(C2a) Regressoren vorherbestimmt} \\ E(u_t|\Omega_t) = 0, \\ \text{(C2b) Bedingte Homoskedastie der Fehler} \\ E(u_t^2|\Omega_t) = \sigma^2 := E(u_t^2), \\ \text{wobei für die Fehlervarianz des DGP } \sigma^2 = \sigma_0^2 \text{ gilt.} \end{array} \right.$$

– (C3) \iff Annahme (A1)

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n E(\mathbf{X}_t^T \mathbf{X}_t) = \mathbf{S}_{\mathbf{X}^T \mathbf{X}} < \infty, \quad \mathbf{S}_{\mathbf{X}^T \mathbf{X}} \text{ invertierbar.}$$

– (C4a) **Strenge Stationarität** von $\{\mathbf{s}_t\} = (y_t \quad \mathbf{Z}_t)^T$,

– (C4b) $E|\boldsymbol{\lambda}^T \mathbf{X}_t u_t|^{2+\delta} \leq B < \infty$, $\delta > 0$, für alle feste $\boldsymbol{\lambda}$ mit $\boldsymbol{\lambda}^T \boldsymbol{\lambda} = 1$.

- **Asymptotische Schätzeigenschaften des KQ-Schätzers**

- **Konsistenz:** Unter den Annahmen **(C1)**, **(C2)**, **(C3)** ist der KQ-Schätzer konsistent, d. h.

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_n = \beta_0 \quad (13.26)$$

- **Asymptotische Normalverteilung:** Unter Annahmen **(C1)**, **(C2)**, **(C3)** und **(C4a)** oder **(C4b)** ist der KQ-Schätzer asymptotisch normalverteilt,

$$\sqrt{n} (\hat{\beta}_n - \beta_0) \xrightarrow{d} N(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}}^{-1}). \quad (13.27)$$

- Hier ohne Beweise. Die (aufwändigen) Beweise finden sich in den Folien zur MA-Veranstaltung **Fortgeschrittene Dynamische Ökonometrie** oder in [Davidson \(2000\)](#).

- **Anmerkungen zu den Annahmen:**

- Die Annahme **(C2a)** setzt voraus, dass alle Regressoren vorherbestimmt sind, also (9.4) gilt, **und** dass das **Modell dynamisch korrekt spezifiziert** ist, d. h.

$$E[y_t | \mathbf{X}_t, y_{t-1}, \mathbf{X}_{t-1}, y_{t-2}, \mathbf{X}_{t-2}, \dots] = E[y_t | \mathbf{X}_t] \quad (13.28)$$

Dann sind die Fehler $\{u_t\}$ unkorreliert.

- Die Annahme **(C2a)** ist schwächer als strenge Exogenität **(B2a)**, deshalb ist der KQ-Schätzer im dynamischen linearen Regressionsmodell im Allgemeinen verzerrt.
- Damit die Annahme **(C3)** gilt, muss beispielsweise im Fall eines AR(1)-Prozesses (12.16) gelten, dass

- * $|\alpha| < 1$ (Stabilitätsbedingung) gilt und

- * $E|u_t|^{2+\delta} \leq B < \infty, \delta > 0, t = 1, \dots, n$, d. h. für die Fehlerverteilung über die Varianz hinaus Momente existieren.

Für AR(p)-Prozesse muss die entsprechende Stabilitätsbedingung erfüllt sein (siehe z. B. BA-Veranstaltung **Zeitreihenökonomie** oder unten genannte MA-Veranstaltungen).

Wenn alle Regressoren **schwach stationär** sind, d. h.

- * $E[\mathbf{X}_t] = E[\mathbf{X}_s]$ und

- * $Cov(\mathbf{X}_s, \mathbf{X}_t) = Cov(\mathbf{X}_{s+k}, \mathbf{X}_{t+k})$ unabhängig von $s, t = 1, \dots$ und k gelten,

dann ist auch Annahme **(C3)** erfüllt (ohne Beweis).

- Annahme **(C4b)** erfordert, dass für die bedingte Fehlerverteilung über die Varianz hinaus Momente existieren. (Beispiel: bedingte Normalverteilung, t -Verteilung mit mindestens 4 Freiheitsgraden)
- Die Annahmen entsprechen den Voraussetzungen in [Davidson \(2000\)](#): Vgl. zu **(C2a)** ([Davidson 2000](#), Assumption 7.1.1), zu **(C2b)** ([Davidson 2000](#), Assumption 7.1.2), zu **(C3)** [Davidson \(2000, 7.1.3\)](#), zu **(C4b)** ([Davidson 2000](#), Eq. (7.1.12)).

Beispiel: Stationärer AR(1)-Prozess

R-Code

```
# ===== 13_5_KQ_AR1.R =====
# Programm zum Generieren und KQ-Schätzen eines AR(1)-Modells
# erstellt von : RT, 2011_01_19

graphics.off()      # Schließe alle Graphikfenster

# Setze Parameter des Modells und der Monte-Carlo-Simulation
set.seed(42)       # Randomseed
N      <- 50       # Stichprobengröße

beta  <- c(2,0.1)  # Parametervektor
sigma <- 2         # Standardabweichung des Fehlers
y0    <- 0         # Startwert des AR(1)-Prozesses

# Generieren einer Realisation eines AR(1)-Prozesses
u <- rnorm(N,mean=0,sd=sigma) # Ziehen von u
y <- rep(1,N)*y0
for (t in (2:N))
{
  y[t] <- beta[1] + y[t-1] * beta[2] + u[t] # Berechnen von y_t
}

# Plot der Zeitreihe
plot(y,xlab="Zeit",ylab="y",type="l")

# Scatterplot
plot(y[1:(N-1)],y[2:N])

# Berechnen des KQ-Schätzers
ols <- lm(y[2:N]~1+y[1:(N-1)]) # Beachte x=y_{t-1}. Deshalb y_t von t=2,...,N
summary(ols)
# ===== Ende =====
```

Listing 13.1: ./R_code/13_5_KQ_AR1.R

- **Beispiel:** Monte-Carlo-Simulation der KQ-Schätzung eines AR(1)-Prozesses mit folgendem

R-Code

```
# ===== 13_5_MC_KQ_AR1.R =====
# Programm für Monte-Carlo-Simulation
# zum Bestimmen der Verzerrung des KQ-Schätzers im AR(1)-Modell
# erstellt von : RT, 2010_11_25

graphics.off()      # Schließe alle Graphikfenster

# Setze Parameter des Modells und der Monte-Carlo-Simulation

set.seed(42)       # Randomseed
N      <- 50       # Stichprobengröße
R      <- 1000     # Zahl der Replikationen

beta  <- c(1,0.9)  # Parametervektor
sigma <- 2         # Standardabweichung des Fehlers
y0    <- 1         # Startwert des AR(1)-Prozesses

# Bilden einer Schleife
beta_hat_store <- matrix(0,nrow=R,ncol=length(beta))
# Initialisiere Matrix zum Abspeichern der KQ-Schätzungen
# für jede Realisation
```

```
for (r in (1:R))
{
  # Generieren einer Realisation eines AR(1)-Prozesses
  u <- rnorm(N,mean=0,sd=sigma)      # Ziehen von u
  y <- rep(1,N)*y0
  for (t in (2:N))
  {
    y[t] <- beta[1] + y[t-1] * beta[2] + u[t] # Berechnen von y_t
  }
  # Berechnen des KQ-Schätzers
  ols <- lm(y[2:N]~y[1:(N-1)])      # Beachte x=y_{t-1}. Deshalb y_t von t=2,...,N

  # Speichern der Parameterschätzung
  beta_hat_store[r,] <- coef(ols)
}

# Berechnen der Mittelwerte der Parameterschätzungen

colMeans(beta_hat_store)

# Erstellen von Histogrammen
par(mfrow=c(1,2))      # Zeichne zwei Plots in ein Graphikfenster

hist(beta_hat_store[,1],breaks=sqrt(R))
hist(beta_hat_store[,2],breaks=sqrt(R))

# ===== Ende =====
```

Listing 13.2: ./R_code/13_5_MC_KQ_AR1.R

14. Verallgemeinerter Kleinst-Quadrate-Schätzer und seine Anwendungen

- Das einfache lineare Modell lautet:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad E(\mathbf{u}|\mathbf{X}) = \mathbf{0}, \quad Var(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I}.$$

In der Praxis ist die Annahme homoskedastischer und unkorrelierter Fehler häufig verletzt.

- Verallgemeinertes lineares Modell mit (streng) exogenen Regressoren:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad E(\mathbf{u}|\mathbf{X}) = \mathbf{0}, \quad Var(\mathbf{u}|\mathbf{X}) = E(\mathbf{u}\mathbf{u}^T|\mathbf{X}) = \boldsymbol{\Omega} \quad (14.1)$$

wobei angenommen wird, dass die Kovarianzmatrix $\boldsymbol{\Omega}$ positiv definit ist:

$$\begin{aligned} \boldsymbol{\Omega} &= Var(\mathbf{u}|\mathbf{X}) = E[(\mathbf{u} - E[\mathbf{u}|\mathbf{X}])(\mathbf{u} - E[\mathbf{u}|\mathbf{X}])^T] \\ &= \begin{pmatrix} Var(u_1|\mathbf{X}) & Cov(u_1, u_2|\mathbf{X}) & \cdots & Cov(u_1, u_n|\mathbf{X}) \\ Cov(u_2, u_1|\mathbf{X}) & Var(u_2|\mathbf{X}) & \cdots & Cov(u_2, u_n|\mathbf{X}) \\ \vdots & \vdots & \ddots & \vdots \\ Cov(u_n, u_1|\mathbf{X}) & Cov(u_n, u_2|\mathbf{X}) & \cdots & Var(u_n|\mathbf{X}) \end{pmatrix}. \end{aligned} \quad (14.2)$$

Spezialfälle:

- Das einfache lineare Modell ist ein Spezialfall: $\boldsymbol{\Omega} = \sigma^2\mathbf{I}$.
- Ist $\boldsymbol{\Omega}$ eine Diagonalmatrix mit $\omega_t^2 = Var(u_t|\mathbf{X}) \neq \omega_s^2$ für einige $s, t, s \neq t$,

$$\boldsymbol{\Omega} = \begin{pmatrix} \omega_1^2 & 0 & \cdots & 0 \\ 0 & \omega_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_n^2 \end{pmatrix}, \quad (14.3)$$

sind die Fehler nicht korreliert, aber (bedingt) heteroskedastisch.

- **(Bedingte) Heteroskedastie** liegt vor, wenn die Fehlervarianz und damit die bedingte Varianz der abhängigen Variable gegeben die Informationsmenge Ω_t oder auch Teilen davon nicht konstant ist, also gilt:

$$\text{Var}(u_t|\Omega_t) = \omega_t^2 \neq \sigma^2, \quad (14.4a)$$

$$\text{Var}(y_t|\Omega_t) = E[(y_t - E[y_t|\Omega_t])^2|\Omega_t] = E[u_t^2|\Omega_t] = \omega_t^2. \quad (14.4b)$$

- **Heteroskedastie:** ω_t^2 ist eine Funktion von deterministischen Regressoren, z. B. Zeit.
- **Bedingte Heteroskedastie:** ω_t^2 ist eine Funktion von Regressoren, die Zufallsvariablen sind.

Beispiele:

- Die Varianz der Exporte hängt vom BIP des Exportlandes ab.
- Die Varianz der Konsumausgaben hängt von der Höhe des Einkommens ab.

Ein ziemlich allgemeines Modell für $E[u_t^2|\Omega_t]$ lautet:

$$\text{Var}(u_t|\Omega_t) = E[u_t^2|\Omega_t] = h(\delta + \mathbf{Z}_t\boldsymbol{\gamma}), \quad \mathbf{Z}_t \in \Omega_t. \quad (14.5)$$

Drei Fälle sind zu unterscheiden:

- Die Funktion $h(\cdot)$ ist inklusive aller Parameterwerte für $\delta, \boldsymbol{\gamma}$ bekannt, dann Verwendung des GLS-Schätzers (14.7), siehe Abschnitt 14.1.
- Die Funktion $h(\cdot)$ ist parametrisch, aber die Parameter $\delta, \boldsymbol{\gamma}$ sind unbekannt, dann Verwendung des FGLS-Schätzers (14.17), siehe Abschnitt 14.2.1.
- Die Funktion $h(\cdot)$ ist vollständig unbekannt, dann Verwendung von heteroskedastie-robusten Standardfehlern, siehe Abschnitt 14.3.

14.1. Verallgemeinerter Kleinst-Quadrate-Schätzer

- **Verallgemeinerter Linearer Kleinst-Quadrateschätzer (generalized least squares estimators (GLS estimator)):**

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y}.$$

- **Ableitung:**

- **Cholesky-Zerlegung:** Für jede symmetrische positiv definite Matrix \mathbf{A} existiert eine Zerlegung $\mathbf{B}\mathbf{B}^T$, wobei \mathbf{B} eine eindeutige untere Dreiecksmatrix ist mit positiven Elementen auf der Diagonale (Gentle (2007, Section 5.9.2), Lütkepohl (1996, Section 6.2.3 (2))).

- Da Ω symmetrisch positiv definit ist, existiert eine eindeutige untere Dreiecksmatrix Ψ , so dass

$$\Omega^{-1} = \Psi\Psi^T.$$

- Multiplizieren des verallgemeinerten linearen Modells (14.1) von links mit Ψ^T liefert

$$\underbrace{\Psi^T \mathbf{y}}_{\mathbf{y}^*} = \underbrace{\Psi^T \mathbf{X}}_{\mathbf{X}^*} \beta + \underbrace{\Psi^T \mathbf{u}}_{\mathbf{u}^*}$$

$$\mathbf{y}^* = \mathbf{X}^* \beta + \mathbf{u}^*, \quad (14.6)$$

wobei Ψ^T genau so gewählt wurde, dass $E[\mathbf{u}^* (\mathbf{u}^*)^T | \mathbf{X}] = \mathbf{I}$ (verifizieren!).

- Damit erfüllt das Modell mit den transformierten Variablen die Annahmen des einfachen linearen Modells an die Kovarianzmatrix des Fehlervektors, sodass sich der KQ-Schätzer anwenden lässt und daraus der GLS-Schätzer folgt:

$$\hat{\beta}_{GLS} = ((\mathbf{X}^*)^T \mathbf{X}^*)^{-1} (\mathbf{X}^*)^T \mathbf{y}^* \quad (14.7a)$$

$$= (\mathbf{X}^T \Psi \Psi^T \mathbf{X})^{-1} \mathbf{X}^T \Psi \Psi^T \mathbf{y} \quad (14.7b)$$

$$= (\mathbf{X}^T \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Omega^{-1} \mathbf{y}. \quad (14.7c)$$

- Der GLS-Schätzer lässt sich auch direkt aus den (theoretischen) Momentenbedingungen

$$\begin{aligned} \mathbf{X}^T \Psi (\Psi^T \mathbf{y} - \Psi^T \mathbf{X} \beta) &= \mathbf{0} \\ \mathbf{X}^T \Omega^{-1} (\mathbf{y} - \mathbf{X} \beta) &= \mathbf{0} \end{aligned} \quad (14.8)$$

ableiten, bzw. auch aus der Minimierung der SSR des Modells (14.6).

• Annahmen zur Bestimmung der Schätzeigenschaften

(vgl. Abschnitt 11.1 zu KQ-Annahmen)

- **(B1)** Das Modell ist korrekt, d. h. der DGP ist im Modell (14.1) enthalten.
- **(B2')** $\mathbf{u} | \mathbf{X} \sim (\mathbf{0}, \Omega)$.
- **(B3)** Keine perfekte Kollinearität in der Regressormatrix \mathbf{X} .
- **(B4')** $\mathbf{u} | \mathbf{X} \sim N(\mathbf{0}, \Omega)$.

Beachte, dass die Annahmen **(B2')** bzw. **(B4')** schwächer als die ursprünglichen Annahmen **(B2)** bzw. **(B4)** sind. Bedingt auf \mathbf{X} kann sowohl Heteroskedastie als auch bei Zeitreihendaten Autokorrelation in den Fehlern vorkommen.

• Schätzeigenschaften des GLS-Schätzers in endlichen Stichproben:

- Unter **(B1)**, **(B2a)** und **(B3)** ist der GLS-Schätzer **erwartungstreu**

$$E(\hat{\beta}_{GLS}) = \beta.$$

- Unter **(B1)**, **(B2')** und **(B3)** hat der GLS-Schätzer Kovarianzmatrix

$$\begin{aligned} \text{Var}(\hat{\beta}_{GLS}|\mathbf{X}) &= (\mathbf{X}^*{}^T \mathbf{X}^*)^{-1} \\ &= (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \end{aligned} \quad (14.9)$$

und ist **BLUE**, d. h. **effizient**. Nachweis siehe unten bei allgemeinem Momentenschätzer.

- **Allgemeiner Momentenschätzer:** Für eine gegebene Stichprobe seien die $(1 \times k)$ -Vektoren von Variablen $\mathbf{W}_t = (W_{t1} \ W_{t2} \ \dots \ W_{tk})$, $t = 1, \dots, n$, in der Matrix $\mathbf{W}^T = (\mathbf{W}_1^T \ \mathbf{W}_2^T \ \dots \ \mathbf{W}_n^T)$ zusammengefasst. Unter der Annahme/Eigenschaft

$$E(\mathbf{u}|\mathbf{X}, \mathbf{W}) = \mathbf{0}$$

ergibt sich ein Momentenschätzer durch Schätzung der theoretischen Momente $E[\mathbf{W}_t^T u_t] = \mathbf{0}$ auf Basis der daraus resultierenden Momentenbedingungen für eine gegebene Stichprobe mit

$$\mathbf{W}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}.$$

Man erhält:

$$\tilde{\boldsymbol{\beta}}_{\mathbf{W}} = (\mathbf{W}^T \mathbf{X})^{-1} \mathbf{W}^T \mathbf{y}.$$

Damit ergibt sich die Kovarianzmatrix

$$\text{Var}(\tilde{\boldsymbol{\beta}}_{\mathbf{W}}|\mathbf{X}, \mathbf{W}) = (\mathbf{W}^T \mathbf{X})^{-1} \mathbf{W}^T \boldsymbol{\Omega} \mathbf{W} (\mathbf{X}^T \mathbf{W})^{-1}.$$

GLS ist ein spezieller Momentenschätzer (vgl. (14.8)) mit

$$\mathbf{W} = \boldsymbol{\Omega}^{-1} \mathbf{X}.$$

Die Differenz der Präzision eines allgemeinen Momentenschätzers und der Präzision des GLS-Schätzers ist positiv semidefinit.

Da jeder lineare unverzerrte Schätzer $\tilde{\boldsymbol{\beta}} = \mathbf{A}\mathbf{y}$ mit $\mathbf{A}\mathbf{X} = \mathbf{I}$, vgl. Abschnitt 9.4, als Momentenschätzer dargestellt werden kann (wegen $\mathbf{y} = \mathbf{X}\tilde{\boldsymbol{\beta}} + \tilde{\mathbf{u}}$ folgt $\mathbf{A}\tilde{\mathbf{u}} = \mathbf{0}$), ist damit der GLS-Schätzer effizient.

- **Berechnen von GLS-Schätzern**

- Ist n groß, benötigt das Speichern und Invertieren von $\boldsymbol{\Omega}$ viel Speicherplatz ($n = 10000$ benötigt bspw. 1600 MB.) Deshalb besser: Vorheriges Anwenden von $\boldsymbol{\Psi}$ ohne Abspeichern von $\boldsymbol{\Psi}$ (sofern möglich).
- **Gewichteter Kleinst-Quadrate Schätzer (weighted least squares (WLS))** u_t heteroskedastisch und unkorreliert (d. h. $\boldsymbol{\Omega}$ diagonal). Damit ist $\boldsymbol{\Omega}$ diagonal (14.3) und der Ansatz (14.6) lautet

$$\frac{y_t}{\omega_t} = \frac{1}{\omega_t} \mathbf{X}_t \boldsymbol{\beta} + \frac{u_t}{\omega_t}$$

mit $\text{Var}(u_t/\omega_t|\mathbf{X}) = 1$. **Interpretation, Berechnung und Hinweise:**

- * Beobachtungen mit großer Fehlervarianz erhalten weniger Gewicht.
- * Wie Auswahl der Gewichte? Abhängig von Datenstruktur, z. B. durch eine Linearkombination erklärender Variable (Beispiel: Einkommenshöhe) oder Durchschnitte in verschiedenen Gruppen.
- * R^2 für gewichtete KQ-Schätzung mit Modell (14.6) berechnen, da die geschätzten Residuen auf $\Psi^T \mathbf{X}$ orthogonal stehen, jedoch nicht auf \mathbf{X} .
- * Für die nichtgewichtete Schätzung verwendet man am besten (7.33).

- **Asymptotische Schätzeigenschaften des GLS-Schätzers**

$$\hat{\beta}_{GLS} = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y} = \beta_0 + (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{u}.$$

Die Annahmen (A1), (A2) bzw. (A3) müssen entsprechend modifiziert werden, damit analog ein LLN und ein CLT gelten:

- (A1') $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X} = \mathbf{S}_{\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}}$, $\mathbf{S}_{\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}}$ hat vollen Rang.
- (A2') Es gilt ein LLN für $\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{u}/n$.
- (A3') $\frac{\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{u}}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \mathbf{S}_{\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}})$.

Dann lässt sich mit der bereits bekannten Vorgehensweise zeigen, dass der GLS-Schätzer

- **konsistent** ((B1), (B3) (A1'), (A2')) und
- **asymptotisch normalverteilt** ((B1), (B3) (A1'), (A3')) ist:

$$\sqrt{n} (\hat{\beta}_{GLS} - \beta_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{S}_{\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}}^{-1}).$$

- Oft ist Ψ bzw. $\boldsymbol{\Omega}$ unbekannt und muss geschätzt werden. Dann ist der GLS-Schätzer nicht anwendbar und muss durch folgenden Schätzer ersetzt werden.
- R-Befehl: `lm(,weights=)`, wobei `weights` die Gewichte $1/\omega_t^2$ übergeben werden muss.

14.2. Feasible GLS

- Ist die Fehlerkovarianzmatrix $\boldsymbol{\Omega}$ unbekannt, muss diese modelliert werden.
- **Asymptotische Eigenschaften von FGLS**
 - Kurz gesagt, der FGLS-Schätzer

$$\hat{\beta}_{FGLS} = (\mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y} \quad (14.10)$$

ist konsistent und asymptotisch normalverteilt, wenn die Fehlerkovarianzmatrix $\boldsymbol{\Omega}$ korrekt spezifiziert ist und konsistent geschätzt werden kann.

- Der Beweis ist aufwändiger, die Idee jedoch recht einfach. Sie ergibt sich aus den asymptotischen Eigenschaften des GLS-Schätzers. Diese bleiben für den FGLS-Schätzer erhalten, wenn gilt

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{X} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}, \quad (14.11a)$$

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \widehat{\boldsymbol{\Omega}}^{-1} \mathbf{u} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{u}. \quad (14.11b)$$

- * I. Allg. erfordert dies, dass in der ersten Stufe der $\boldsymbol{\beta}$ -Vektor konsistent geschätzt wird, so dass aus den durch $\hat{\mathbf{u}}$ konsistent geschätzten Fehlern z. B. $\boldsymbol{\gamma}$ in (14.5) und damit $\boldsymbol{\Omega}$ konsistent geschätzt werden kann.
- * **Achtung:** Ist $\boldsymbol{\Omega}$ nicht diagonal, ist der KQ-Schätzer für die erste Stufe i. Allg. inkonsistent! In diesem Fall sind andere Schätzverfahren notwendig.

14.2.1. Modellierung heteroskedastischer Fehler

- Ein häufig geeignetes Modell zur Modellierung von $\text{Var}(u_t | \Omega_t) = \omega_t^2$ ist

$$E[u_t^2 | \Omega_t] = e^{\delta + \mathbf{Z}_t \boldsymbol{\gamma}} = e^\delta e^{\mathbf{Z}_t \boldsymbol{\gamma}}. \quad (14.12)$$

Die Gleichung (14.12) spezifiziert die Funktion $h(\cdot)$ in (14.5) als $h(\cdot) = \exp(\cdot)$.

Wird für eine Zufallsvariable v_t festgelegt, dass

$$E[v_t | \Omega_t] = 0 \quad \text{und} \quad \text{Var}(v_t | \Omega_t) = 1,$$

lässt sich u_t^2 schreiben als

$$u_t^2 = e^{\delta + \mathbf{Z}_t \boldsymbol{\gamma}} v_t^2, \quad (14.13)$$

so dass (14.12) gilt. Um δ und $\boldsymbol{\gamma}$ mit einer linearen Regression schätzen zu können, wird (14.13) logarithmiert:

$$\ln u_t^2 = \delta + \mathbf{Z}_t \boldsymbol{\gamma} + \ln v_t^2. \quad (14.14)$$

Da $E[\ln v_t^2] \neq \ln E[v_t^2] = 0$ (Jensen-Ungleichung, siehe Abschnitt 2.7), verwendet man folgenden Trick:

$$\begin{aligned} \ln u_t^2 &= \underbrace{\delta + E[\ln v_t^2]}_{\delta'} + \mathbf{Z}_t \boldsymbol{\gamma} + \underbrace{(\ln v_t^2 - E[\ln v_t^2])}_{\eta_t \text{ wobei } E[\eta_t | \Omega_t] = 0} \\ \ln u_t^2 &= \delta' + \mathbf{Z}_t \boldsymbol{\gamma} + \eta_t. \end{aligned} \quad (14.15)$$

- **2-stufiger Schätzer:**

1. Schritt: Schätze das Modell (14.1) mit KQ und speichere die Residuen

$$\hat{\mathbf{u}} = \mathbf{M}_{\mathbf{X}} \mathbf{y}.$$

Einsetzen der Residuen in (14.15) zur KQ-Schätzung von δ' und $\boldsymbol{\gamma}$

$$\ln \hat{u}_t^2 = \delta' + \mathbf{Z}_t \boldsymbol{\gamma} + \text{Fehler}. \quad (14.16)$$

2. Schritt: Schätze (14.6)-Ansatz mit $\hat{\omega}_t^2 = \exp(\mathbf{Z}_t \hat{\boldsymbol{\gamma}})$

$$\frac{y_t}{\hat{\omega}_t} = \frac{1}{\hat{\omega}_t} \mathbf{X}_t \boldsymbol{\beta} + \frac{u_t}{\hat{\omega}_t}. \quad (14.17)$$

Der Faktor $e^{\hat{\boldsymbol{\gamma}}}$ kann weggelassen werden, da er für alle Beobachtungen konstant ist.

- **FGLS oder KQ mit heteroskedastierobusten Standardfehlern?** Die Frage ist wie gut $\boldsymbol{\Omega}$ geschätzt werden kann. Je unpräziser, desto eher wird man den KQ-Schätzer mit heteroskedastie-robuster Varianz-Kovarianzmatrix, siehe Abschnitt 14.3 nehmen.
- Es ist möglich, den FGLS-Schätzer zu iterieren. Dies hat *keinen* Einfluss auf die asymptotischen Eigenschaften, jedoch auf die Schätzeigenschaften in endlichen Stichproben.

14.2.2. Modelle mit autokorrelierten Fehlern

Siehe Davidson & MacKinnon (2004, Sections 7.6-7.9).

14.3. Heteroskedastie-robuste Standardfehler bei KQ-Schätzung

- **Ableitung heteroskedastie-robuster Standardfehler**

Liegen heteroskedastische Fehler vor, dann ist die Varianz-Kovarianzmatrix des KQ-Schätzers durch (9.7) gegeben

$$\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{u}|\mathbf{X}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \quad (9.7)$$

$$= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}. \quad (14.18)$$

Diese Varianz-Kovarianzmatrix wird oft als **sandwich covariance matrix** bezeichnet, wobei $(\mathbf{X}^T \mathbf{X})^{-1}$ die „Brotcheiben“ darstellen. Die Varianz-Kovarianzmatrix von ineffizienten Schätzern haben oft diese Form.

- Eine alternative Darstellung des „Belags“ ist:

$$\mathbf{X}^T \boldsymbol{\Omega} \mathbf{X} = \sum_{t=1}^n \omega_t^2 \mathbf{X}_t^T \mathbf{X}_t.$$

Da $E[u_t^2|\mathbf{X}] = \omega_t^2$, kann man ω_t^2 durch den „Durchschnitt auf Basis von einer Beobachtung“ u_t^2 schätzen. Dies ist natürlich kein besonders guter Schätzer aber für unseren Zweck tut er's. Da u_t unbekannt ist, nimmt man das Residuum \hat{u}_t .

Demnach kann man die Kovarianzmatrix (14.18) des KQ-Schätzers bei Heteroskedastie mittels

$$\widehat{\text{Var}}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{t=1}^n \hat{u}_t^2 \mathbf{X}_t^T \mathbf{X}_t \right) (\mathbf{X}'\mathbf{X})^{-1} \quad (14.19)$$

schätzen.

- Anmerkungen:
 - Die Standardfehler, die man aus (14.19) 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 unter recht allgemeinen Bedingungen **asymptotisch normalverteilt**. Es bleiben also die kritischen Werte und die p -values approximativ gültig, falls man (14.19) verwendet.
 - In Davidson & MacKinnon (2004, Section 5.5) wird erklärt, warum (14.19) ein konsistenter Schätzer von (14.18) ist.
 - Der KQ-Schätzer ist unabhängig von der Wahl der Standardfehler (White oder nicht-White) **unverzerrt** und **konsistent**, da die Annahmen (B1), (B2a), (B3) von Heteroskedastie unberührt bleiben.
 - Der KQ-Schätzer ist aber im Fall heteroskedastischer Fehler (asymptotisch) **nicht effizient**, da gezeigt werden kann, dass die Differenz der (asymptotischen) Präzision des KQ-Schätzers und des (F)GLS-Schätzers positiv semidefinit ist. Ist also etwas über die funktionale Form der Heteroskedastie bekannt und liegen genügend Stichprobenbeobachtungen vor, sollte der FGLS-Schätzer verwendet werden.
- **Alternative Schätzer** von (14.18) und deren Bezeichnungen in Davidson & MacKinnon (2004, Section 5.5) und den R-Paketen `car` oder `sandwich`.
 - „HC0“: White-Standardfehler (14.19).
 - „HC1“: Multipliziert White-Standardfehler (14.19) mit $n/(n - k)$. (Default in EViews.)
 - „HC2“: Ersetzt \hat{u}_t^2 White-Standardfehler (14.19) durch $\hat{u}_t^2/(1 - h_t)$, wobei h_t das t -te Diagonalelement von \mathbf{P}_X ist.
 - „HC3“: Ersetzt \hat{u}_t^2 White-Standardfehler (14.19) durch $\hat{u}_t^2/(1 - h_t)^2$, wobei h_t das t -te Diagonalelement von \mathbf{P}_X ist.

Alle Korrekturen haben zum Ziel, die Unterschätzung der Fehlervarianz durch Verwendung der Residuen anstelle der Fehler, siehe Abschnitt 9.5, zu korrigieren. Zur genaueren Begründung der jeweiligen Korrekturen siehe Davidson & MacKinnon (2004, Section 5.5).

 - R-Befehle zur Berechnung heteroskedastie-robuster Varianz-Kovarianzmatrizen:
 - Paket `car`: `hcmm(model, type="hc1")`

- Paket `sandwich`: `vcovHC(model, type="HC1")`
- R-Befehle zur Berechnung heteroskedastie-robuster Teststatistiken mit Paket `car`:
 - `coefTest(model, vcov=hccm(model, type="hc1"))` liefert üblichen Regressionsoutput mit heteroskedastie-robusten Standardfehlern.
 - `linearHypothesis(, vcov=hccm(model, type="hc1"))` liefert F -Test mit heteroskedastie-robuster Varianz-Kovarianzmatrix.

14.4. Empirische Analyse von Handelsströmen: Teil 4

Fortsetzung der Analyse von Modell (11.49):

$$\ln(\text{Importe}_i) = \beta_1 + \beta_2 \ln(\text{BIP}_i) + \beta_3 \ln(\text{Entfernung}_i) + \beta_4 \text{Offenheit}_i + \beta_5 \ln(\text{Flaeche}) + u_i. \quad (11.49)$$

- Eliminieren von **missing values** bzw. **not a number (NaN)** bzw. **not available/not applicable (NA)** (in R): Im verwendeten Datensatz kommt für die abhängige Variable `Importe` eine NA vor, die im weiteren Verlauf des R-Programms dazu führt, dass der Residuenvektor weniger Zeilen hat als die Regressormatrix. Deshalb ist es sinnvoll, vor Beginn der Schätzungen diese Beobachtung aus dem data frame zu eliminieren. Wenn der ursprüngliche data frame mit `daten_all` bezeichnet wird, geht dies mit dem Befehl

```
daten <- daten_all[!is.na(daten$trade_0_d_o),]
```

Erst danach den Befehl `attach(daten)` verwenden, damit R im richtigen data frame sucht!

- **FGLS-Schätzung**

R-Code (Ausschnitt aus R-Programm in Abschnitt A.4)

```
#### FGLS-Schätzung für Modell 4
mod_4_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o + log(cepii_area_o)

# 1. Schritt
resids <- residuals(mod_4_kq)
fits <- fitted(mod_4_kq)
mod_formula_ln_u_squared <- log(resids^2) ~ log(wdi_gdpusdcr_o) + log(cepii_dist)
+ ebrd_tfes_o + log(cepii_area_o)

# 2. Schritt
omega <- exp(fitted(lm(mod_formula_ln_u_squared)))
model_gls <- lm(mod_4_formula, weights=1/omega)
(summary(model_gls))
```

Listing 14.1: `./R_code/4_ff_Beispiel_Handelsstroeme.R`

Man erhält:

```
Call:
```

```
lm(formula = mod_4_formula, weights = 1/omega)

Weighted Residuals:
Min      1Q  Median      3Q      Max
-4.799 -1.227  0.544  1.174  3.006

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.20213    1.26254   1.744 0.088106 .
log(wdi_gdpusdcr_o)  1.07977    0.05715  18.893 < 2e-16 ***
log(cepii_dist)    -0.90934    0.11505  -7.904 5.54e-10 ***
ebrd_tfes_o        0.25397    0.17561   1.446 0.155201
log(cepii_area_o)  -0.20138    0.05343  -3.769 0.000485 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.897 on 44 degrees of freedom
Multiple R-squared:  0.9345,    Adjusted R-squared:  0.9286
F-statistic: 157.1 on 4 and 44 DF,  p-value: < 2.2e-16
```

- Heteroskedastie-robuste KQ-Schätzer

R-Code (Ausschnitt aus R-Programm in Abschnitt [A.4](#))

```
library(lmtest)
# Zu Wahlmöglichkeiten für die Schätzung der heteroskedastischen
# Varianz-Kovarianzmatrix siehe Abschnitt 14.3
(coeftest(mod_4_kq,vcov=hccm(mod_4_kq,type="hc1")))
```

Listing 14.2: ./R_code/4_ff_Beispiel_Handelsstroeme.R

liefert

```
t test of coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)      2.427777    1.337400   1.8153  0.076298 .
log(wdi_gdpusdcr_o)  1.025023    0.070679  14.5024 < 2.2e-16 ***
log(cepii_dist)    -0.888646    0.120775  -7.3579 3.428e-09 ***
ebrd_tfes_o        0.353154    0.180896   1.9522  0.057290 .
log(cepii_area_o)  -0.151031    0.050657  -2.9814  0.004662 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Zusammenfassen der Ergebnisse in einer Tabelle: Outputtable für Modell (11.49) für unterschiedliche Schätzer

Ergebnis: Sowohl die Parameterschätzungen selbst, als auch die Standardfehler unterscheiden sich nicht grundlegend. Mögliche Ursache: Es liegt keine Heteroskedastie in den Fehlervarianzen vor.

Abhängige Variable: $\ln(\text{Importe nach Deutschland})$		
Unabhängige Variablen/Modell	OLS	FGLS
Konstante	2.427 (2.132) [1.337]	2.024 (1.236)
$\ln(BIP)$	1.025 (0.076) [0.070]	1.080 (0.057)
$\ln(\text{Entfernung})$	-0.888 (0.156) [0.120]	-0.888 (0.110)
<i>Offenheit</i>	0.353 (0.206) [0.180]	0.263 (0.179)
$\ln(\text{Flaeche})$	-0.151 (0.085) [0.050]	-0.203 (0.048)
Stichprobengröße	49	49
R^2	0.906	0.9055
Standardfehler der Regression	0.853	
Residuenquadratsumme	32.017	
AIC	2.6164	
HQ	2.6896	
SC	2.8094	

Anmerkungen: KQ- bzw. FGLS-Standardfehler in runden, White-Standardfehler in eckigen Klammern

- Fortsetzung in Abschnitt [15.7](#).

15. Modellüberprüfung

Übersicht zum Modellierungsprozess, siehe Abschnitt 4.3.

Warum ist Modellüberprüfung notwendig?

Eigenschaften von Schätz- und Testverfahren gelten nur unter den getroffenen Annahmen!
⇒ Überprüfen dieser Annahmen essentiell, indem statistische Tests durchgeführt werden!

Wiederholung (Siehe Kapitel 11.3, S. 221)

Anwendungen von exakten Tests:

- **Spezifikation des normalen linearen Regressionsmodells und Überprüfen der Annahmen**, vgl. Abschnitt 11.1
 - **(B1)** und $E[\mathbf{u}|\mathbf{X}] = \mathbf{0}$ (**(B2a)**): $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ enthält DGP
 - * t -Tests, siehe Abschnitt 11.3.1; F -Tests, siehe Abschnitt 11.3.2.
 - * Testen der korrekten funktionalen Form, z. B. mit **RESET-Test**, siehe Abschnitt 15.3.
 - * Testen auf Parameterstabilität, z. B. mit **Chow-Test**, siehe (11.34) in Abschnitte 11.3.2.
 - **(B3)**: $\mathbf{X}^T\mathbf{X}$ hat Rang k : Verletzung führt zu Fehlermeldung „singuläre Matrix“.
 - **(B4)**: $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$:
 - * Setzt $E[\mathbf{u}|\mathbf{X}] = \mathbf{0}$ voraus, siehe oben.
 - * Setzt voraus: $Var(\mathbf{u}|\mathbf{X}_t) = \sigma^2$ (**Homoskedastie**): **Tests auf Heteroskedastie**, siehe Abschnitt 15.2.
 - * Erfordert normalverteilte Fehler: **Lomnicki-Jarque-Bera-Test**, siehe Abschnitt 15.4.
- **Überprüfen ökonomischer Hypothesen**

15.1. Tests auf Autokorrelation in den Fehlern

15.2. Tests auf heteroskedastische Fehler

- Wie bereits erwähnt, ist es nicht sinnvoll “automatisch” den FGLS-Schätzer (14.17) 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 .
- In R ist der Breusch-Pagan-Test im Paket `lmtest` enthalten.

Es wird angenommen, dass für das multiple lineare Regressionsmodell

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

die Annahmen für Erwartungstreue oder Konsistenz des KQ-Schätzers erfüllt sind. Der Hypothesentest bezieht sich auf die Gültigkeit von (B2b) bzw. (C2b), also auf das Vorliegen von Homoskedastizität.

Das zu testende **Hypothesenpaar** lautet

$$\begin{aligned} H_0 &: \text{Var}(u_t|\mathbf{X}_t) = \sigma^2 \quad (\text{Homoskedastie}), \\ H_1 &: \text{Var}(u_t|\mathbf{X}_t) = \omega_t^2 \neq \sigma^2 \quad (\text{Heteroskedastie}). \end{aligned}$$

Die Grundidee der Heteroskedastie-Tests ist, dass unter der Nullhypothese kein Regressor Erklärungsgehalt für $\text{Var}(u_t|\mathbf{X}_t)$ haben sollte. Gilt die Nullhypothese nicht, kann die bedingte Varianz $\text{Var}(u_t|\mathbf{X}_t)$ durch (beinahe) jede beliebige Funktion der Regressoren x_{tj} , ($1 \leq j \leq k$) oder anderer Regressoren bestimmt sein.

Beachte: Der Breusch-Pagan- und der White-Test unterscheiden sich bezüglich ihrer Alternativhypothese.

15.2.1. Breusch-Pagan-Test

- Idee: Betrachten wir die Regression

$$u_t^2 = \delta_0 + \delta_1 x_{t1} + \cdots + \delta_k x_{tk} + v_t, \quad t = 1, \dots, n. \quad (15.1)$$

Unter den Annahmen (B1), (B2a), (B3) ist der KQ-Schätzer für die δ_j 's unverzerrt.

Das Hypothesenpaar lautet somit:

$$\begin{aligned} H_0 &: \delta_1 = \delta_2 = \cdots = \delta_k = 0 \quad \text{versus} \\ H_1 &: \delta_1 \neq 0 \text{ und/oder } \delta_2 \neq 0 \text{ und/oder } \dots, \end{aligned}$$

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.

Damit muss ein asymptotischer F -Test verwendet werden. Mit den Ergebnissen aus Abschnitt 11.4 und entsprechenden Regularitätsannahmen folgt, dass k mal die F -Statistik asymptotisch $\chi^2(k)$ verteilt ist.

- 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 Grund hierfür ist, dass die Fehler konsistent durch die Residuen geschätzt werden, wenn die Parameter konsistent geschätzt werden. (Der formale Beweis ist formal recht aufwändig und unterbleibt hier.)
- Man kann auch die R^2 -Version der Teststatistik verwenden. Beachte, dass das R^2 wegen $SSR = SST$ 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 (15.1) mit $R_{\hat{u}^2}^2$ und erhalten

$$F = \frac{R_{\hat{u}^2}^2/k}{(1 - R_{\hat{u}^2}^2)/(n - k)}.$$

Die Teststatistik des Overall- F -Tests, der auf die gemeinsame Signifikanz aller Regressoren getestet, wird von den meisten Softwareprogrammen standardmäßig ausgegeben.

- H_0 wird dann abgelehnt, wenn F bzw. kF den kritischen Wert für ein gewähltes Signifikanzniveau auf Basis der $F_{k, n-k}$ bzw. $\chi^2(k)$ -Verteilung übersteigt (oder wenn der p -value kleiner ist als das Signifikanzniveau).
- In R wird der Test mit dem Befehl `bptest()` aus dem Paket `lmtest` durchgeführt und gibt die kF -Statistik aus, die asymptotisch $\chi^2(k)$ -verteilt ist.

- **Beachte:**

- Vermutet man, dass die Heteroskedastie von speziellen Variablen verursacht wird, die zuvor nicht in der Regression berücksichtigt wurden, können diese in die Regression (15.1) eingefügt werden.
- Falls H_0 nicht abgelehnt wird, bedeutet dies nicht automatisch, dass die u_i 's homoskedastisch sind. Sollte die Spezifikation (15.1) 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 $\sigma_t^2 = h(\delta + \mathbf{X}_t\boldsymbol{\beta})$. Wird etwa der Fall $h(\cdot) = \exp(\cdot)$

angenommen, erhält man die Testgleichung

$$\ln(\hat{u}_t^2) = \delta' + \mathbf{X}_t\boldsymbol{\beta} + \text{Fehler}. \quad (14.16)$$

15.2.2. White-Test

- **Hintergrund:** Um die asymptotische Verteilung des KQ-Schätzers abzuleiten, wird die Annahme homoskedastischer Fehler ((B2b) bzw. (C2b)) *nicht* benötigt.

Es reicht bereits aus, dass die quadrierten Fehler u_t^2 mit allen Regressoren, deren Quadraten und deren Kreuzprodukten unkorreliert sind.

Dies lässt sich recht einfach mit folgender Regression testen, wobei die unbekanntes Fehler bereits durch die Residuen ersetzt wurden:

$$\begin{aligned} \hat{u}_t^2 = & \delta_0 + \delta_1 x_{t1} + \cdots + \delta_k x_{tk} \\ & + \delta_{k+1} x_{t1}^2 + \cdots + \delta_{J_1} x_{tk}^2 \\ & + \delta_{J_1+1} x_{t1} x_{t2} + \cdots + \delta_{J_2} x_{tk-1} x_{tk} \\ & + v_t, \quad t = 1, \dots, n. \end{aligned} \quad (15.2)$$

- Das Hypothesenpaar lautet:

$$\begin{aligned} H_0 : & \delta_j = 0 \text{ für } j = 1, 2, \dots, J_2, \\ H_1 : & \delta_j \neq 0 \text{ für mindestens ein } j. \end{aligned}$$

Es kann wieder ein F -Test verwendet werden, dessen Verteilung approximativ die F -Verteilung ist (asymptotische Verteilung).

- Hat man viele Regressoren, ist es mühsam den F -Test für (15.2) 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_{tj}^2 in die Regression auf und vernachlässigt alle Kreuzprodukte.
- **Beachte:** Sollte die Nullhypothese abgelehnt werden, kann dies daran liegen, dass
 - die Fehler heteroskedastisch sind und/oder
 - das Modell nicht korrekt spezifiziert ist.
- In R ist der White-Test nicht automatisch verfügbar. Ein eigenes Programm `whitetest()` findet sich hierzu in Abschnitt B.2.

15.3. Test auf korrekte Spezifikation der funktionalen Form: RESET-Test

RESET Test (REgression Specification Error Test)

Idee und Durchführung:

- Der RESET-Test dient zur Überprüfung, ob das vorliegende Regressionsmodell

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t$$

korrekt spezifiziert ist, d. h. ob die Annahmen **(B1)** und **(B2a)** bzw. alternativ die Annahmen **(C1)** und **(C2a)** gelten und somit

$$E[y_t|\Omega_t] = \mathbf{X}_t\boldsymbol{\beta} \quad (10.1)$$

gilt. Vgl. Kapitel 10.

- 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 $\hat{y}_t = \mathbf{X}_t\hat{\boldsymbol{\beta}}$, die dem Modell hinzugefügt wurden, getestet werden kann. Beachte, dass die gefitteten Werte eine nichtlineare Funktion der Regressoren des Ausgangsmodells darstellen.
- In der Praxis erwiesen sich die zweite und dritte Potenz der \hat{y}_t als hinreichend, um den RESET-Test durchführen zu können:

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + \alpha\hat{y}_t^2 + \gamma\hat{y}_t^3 + Fehler. \quad (15.3)$$

Das Hypothesenpaar lautet

$$H_0 : \alpha = 0, \gamma = 0 \quad (\text{lineares Modell ist korrekt spezifiziert})$$

$$H_1 : \alpha \neq 0 \text{ und/oder } \gamma \neq 0.$$

Getestet wird diese Nullhypothese mittels eines F -Tests mit 2 Freiheitsgraden im Zähler und $n - k - 2$ im Nenner, wobei der daraus resultierenden kritische Wert nur asymptotisch korrekt ist.

- **Beachte:** Wird die Nullhypothese abgelehnt, dass das Ausgangsmodell korrekt spezifiziert ist, kann dies eine Reihe von Ursachen haben:
 - Die funktionale Form ist nichtlinear.
 - Es fehlen relevante Regressoren.
 - Es liegt Heteroskedastie vor.

- R-Befehl: `resettest()`, wobei ohne weitere Angaben die zweite und dritte Potenz berücksichtigt werden (erfordert R-Paket `lmtest`).
- Siehe [Davidson & MacKinnon \(2004, Section 15.2\)](#) für weitere Details.

15.4. Normalitätstest: Lomnicki-Jarque-Bera-Test

- Siehe [Davidson & MacKinnon \(2004, Section 15.2\)](#) für eine detaillierte Erklärung.
- In R wird der Lomnicki-Jarque-Bera-Test mit dem Befehl `jarque.test()` aus dem Paket `moments` durchgeführt.

15.5. Stabilitätstests

Chow-Test

siehe [\(11.34\)](#) in Abschnitt [11.3.2](#).

15.6. Zusammenfassung eines ökonometrischen Modellierungsprozesses

15.7. Empirische Analyse von Handelsströmen: Teil 5

Fortsetzung von Abschnitt [14.4](#).

- **RESET-Test** von Modell 4:

R-Code (Ausschnitt aus R-Programm in Abschnitt [A.4](#))

```
#### Durchführen des RESET-Tests f-ur Modell 4 mit
resettest(mod_4_kq)
```

Listing 15.1: `./R_code/4_ff_Beispiel_Handelsstroeme.R`

liefert:

```
RESET test
data:  mod_4_kq
RESET = 7.14, df1 = 2, df2 = 42, p-value = 0.002142
```

- **Breusch-Pagan-Test auf Heteroskedastie** von Modell 4:

R-Code (Ausschnitt aus R-Programm in Abschnitt [A.4](#))

```
#### Durchführen des Breusch-Pagan-Tests f-ur Modell 4
bptest(mod_4_kq)
```

Listing 15.2: ./R_code/4_ff_Beispiel_Handelsstroeme.R

liefert:

```
studentized Breusch-Pagan test

data:  mod_4_kq
BP = 4.2779, df = 4, p-value = 0.3697
```

- **White-Test mit Kreuzprodukte auf Heteroskedastie** mit den KQ-Residuen:

R-Code (Ausschnitt aus R-Programm in Abschnitt [A.4](#))

```
#####
#                               Beginn Funktion whitetest
#####
# White-Test auf homoskedastische Fehler mit Kreuzprodukten
# RW, 2011_01_26

whitetest <- function(model){

  # Daten aus model extrahieren
  dat <- model$model
  dat$resid_sq <- model$resid^2

  # Formel für die Hilfsregression erstellen
  regr <- attr(model$terms, "term.labels")
  form <- as.formula(paste("resid_sq~(",paste(regr,collapse="+"),")^2+",paste("I(",regr,"^2)",collapse="+"))
  )

  # Hilfsregression schätzen
  test_eq <- lm(form,data=dat)

  # Overall F-Test
  fstat <- summary(test_eq)$fstatistic

  # Ergebnis berechnen und ausgeben
  result1 <- c(fstat[1],fstat[2],fstat[3],pf(fstat[1],fstat[2],fstat[3],lower.tail=FALSE))
  names(result1) <- c("F-Statistic","df1","df2","P-Value")
  result <- list(result1,test_eq)
  return(result)
}

#####
#                               Ende Funktion whitetest
#####

#### Durchführen des White-Test f-ur Modell 4
whitetest(mod_4_kq)
```

Listing 15.3: ./R_code/4_ff_Beispiel_Handelsstroeme.R

liefert:

```
[[1]]
F-Statistic      df1      df2      P-Value
 0.5239004 14.0000000 34.0000000 0.9016863

[[2]]

Call:
```

```
lm(formula = form, data = dat)
Coefficients:
                (Intercept)                log(wdi_gdpusdcr_o)                log(
    cepii_dist)                -77.25103                3.91926
    4.08526
    I(log(wdi_gdpusdcr_o)^2)                I(log(cepai_dist)^2)                I(ebrd_
    tfes_o^2)                -0.02898                0.18986
    0.20200
    log(wdi_gdpusdcr_o):ebrd_tfes_o  log(wdi_gdpusdcr_o):log(cepai_area_o)                log(cepai_dist):ebrd
    _tfes_o                -0.71004                0.08044
    -0.98879
                ebrd_tfes_o                log(cepai_area_o)
                15.48882                -2.91347
    I(log(cepai_area_o)^2)  log(wdi_gdpusdcr_o):log(cepai_dist)
                -0.04813                -0.06623
    log(cepai_dist):log(cepai_area_o)                ebrd_tfes_o:log(cepai_area_o)
                -0.07797                0.61849
```

- **Ergebnisse:**

- Der RESET-Test liefert eine Ablehnung der Nullhypothese korrekter Spezifikation auf dem 1%-Signifikanzniveau. D. h. quadratische Terme spielen möglicherweise eine erklärende Rolle, z. B. $I((\log(\text{wdi_gdpusdcr_o}))$). Berücksichtigung dieses zusätzlichen Regressors führt jedoch zu keiner insignifikanten RESET-Test-Statistik. Möglicherweise ist dies auf Ausreißer zurückzuführen.
- 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 14.4 nicht effizient.
- Breusch-Pagan- und White-Test lehnen die Nullhypothese homoskedastischer standardisierter FGLS-Fehler ebenfalls nicht ab. Die p -values steigen nochmal deutlich auf über 50%.
- Ein endgültiges Modell ist aufgrund der starken Ablehnung des RESET-Tests auch mit quadratischem Regressor noch nicht gefunden.

A. R-Programme

A.1. R-Programme für Graphiken in Abschnitt 2.5 zu Verteilungs- und Dichtefunktionen

CDF und PDF der Standardnormalverteilung, siehe Abbildung 2.1

```
# Verteilungs- und Dichtefunktion der Standardnormalverteilung
# KK, 21.10.2010, RT, 29.09.2015 (pdf- statt eps-Graphiken)

#####
# Dichtefunktion der Standardnormalverteilung
#####

# Öffnen des Graphik-Outputs und Angabe, dass Speichern als .eps-Datei
# (Dateiname, Größe der Graphik)
#postscript("pdf_std_normal.eps", height=4, width=6, horizontal=FALSE)
# oder Speichern als .pdf
pdf("pdf_std_normal.pdf", height=4, width=6)

# Parameter für Graphiken: (optional)
# las=1: Achsenskalierung waagrecht
# mai: Breite der Ränder (unten, links, oben, rechts)
# mgp: Lage von Achsen, Achsenskalierung und Achsenbeschriftung
par(las=1, mai=c(0.6,0.1,0.1,0.1), mgp=c(1.5,1,0))

# Plotten der beiden Punkte (-3.5,0) und (3.5,0.48) (-> Ausmaße der Graphik)
# type="n": leerer Plot
# bty="n": keine Box um die Graphik
# xaxt="n", yaxt="n": keine x- und y-Achse
# xlab="x", ylab="": x-Achsenbeschriftung ist x, y-Achse unbeschriftet
plot(c(-3.5,3.5), c(0,0.48), type="n", bty="n", xaxt="n", yaxt="n",
      xlab="x", ylab="")

# Ergänzen der Achsen (1 x-Achse, 2 y-Achse)
# pos=0: Achse geht durch 0
# labels: Achsenskalierung
# at: Positionen der Achsenskalierung
axis(1, pos=0, labels=-3:3, at=-3:3)
axis(2, pos=0, labels=1:4/10, at=1:4/10)
# x-Achse zu kurz -> mit Linie bei y=0 verlängern (h horizontale Linie)
# y-Achse zu kurz -> Linie von 0 bis 0.44 einzeichnen
abline(h=0)
lines(c(0,0), c(0,0.44))
# y-Achsenbeschriftung
text(0, 0.472, expression(phi(x)))

# Plotten der Funktion
# dnorm: Dichte der Normalverteilung (analog pnorm: Verteilungsfunktion)
# from, to: Bereich, in dem die Funktion geplottet wird
# add=TRUE: plottet in bestehendes Graphikfenster
plot(function(x) dnorm(x), from=-3.5, to=3.5, add=TRUE)
```

```

# Schließen des aktuellen Graphikfensters (z.-B. .eps- oder .pdf-Datei)
dev.off()

#####
# Verteilungsfunktion der Standardnormalverteilung
#####

# postscript("cdf_std_normal.eps", height=4, width=6, horizontal=FALSE)
pdf("cdf_std_normal.pdf", height=4, width=6)

par(las=1, mai=c(0.6,0.1,0.1,0.1), mgp=c(1.5,1,0))

plot(c(-3.5,3.5), c(0,1.2), type="n", bty="n", xaxt="n", yaxt="n",
      xlab="x", ylab="")

axis(1, pos=0, labels=-3:3, at=-3:3)
axis(2, pos=0, labels=c("",0.5,1), at=c(0,0.5,1))
abline(h=0); lines(c(0,0), c(0,1.1))
text(0, 1.18, expression(Phi(x)))

plot(function(x) pnorm(x), from=-3.5, to=3.5, add=TRUE)

dev.off()

```

Listing A.1: ./R_code/2_5_Plot_PDF_CDF_StNormal.R

0.85-Quantil der Standardnormalverteilung, siehe Abbildung 2.2

```
# R-Programm zu Quantilen, Abschnitt Konfidenzintervalle
# RT, KK, 01.02.2011

alpha <- 0.85          # Lege Wahrscheinlichkeit für Quantil fest
dev.off()              # Schließe alle Graphikfenster
split.screen(c(1,2))  # teilt ein Graphikfenster
# Plote Dichtefunktion
screen(1)
plot(function(x) dnorm(x), from=-4, to=4, lwd=2, ylab="Dichte",
      main="Standardnormalverteilung")
abline(h=0)
abline(v=qnorm(alpha), col="red")

# Plan: Polygon zeichnen (das kann man dann einfärben)
# Polygon braucht zuerst alle x-Werte, dann alle y-Werte
# die werden dann verbunden
# x-Werte: von -4 bis qnorm(alpha) (-> x_tmp)
# y-Werte: Dichtewerte für die x-Werte
x_tmp <- seq(from=-4, to=qnorm(alpha), length.out=1000)
polygon(c(x_tmp, x_tmp[length(x_tmp)]),      # letzter doppelt (Punkt q_alpha,0)
        c(dnorm(x_tmp), 0),                  # letzter 0 (Punkt y_alpha,0)
        border=NA,                           # keine Umrandung
        density=10,                          # gestrichelt, 10%
        col="blue")                          # Farbe
# Plote Wahrscheinlichkeitsfunktion
screen(2)
plot(function(x) pnorm(x), from=-4, to=4, lwd=2, ylab="Wahrscheinlichkeitsfunktion",
      main="Standardnormalverteilung")
abline(h=alpha)
abline(v=qnorm(alpha), col="red")
```

Listing A.2: ./R_code/2_5_Plot_Quant_StNormal.R

PDF der bivariaten Normalverteilung, siehe Abbildung 2.3

```

# Dichtefunktion der bivariaten Normalverteilung
# RT, 21.10.2010, 25.10.2010

rm(list = ls()) # cleans workspace

# install package mnormt if not installed yet
if(!require(mnormt)){
  install.packages("mnormt")
}

library(mnormt) # load package mnormt

# parameters of bivariate normal distribution
mu_1 <- 0
mu_2 <- 0

sigma_1 <- 1
sigma_2 <- 1
rho <- 0.0

# determine mean vector
Mean <- c(mu_1,mu_2)

# compute variance-covariance matrix

sigma2_1 <- sigma_1^2
sigma2_2 <- sigma_2^2
sigma_12 <- sigma_1 * sigma_2 * rho

Sigma <- matrix(c(sigma2_1,sigma_12,sigma_12,sigma2_2),2)

# determine grid on which density is computed

x1_limit <- mu_1 + 3*sigma_1
x2_limit <- mu_2 + 3*sigma_2
ngridpoints <- 50

x1 <- seq(-x1_limit,x1_limit,2*x1_limit/(ngridpoints-1))
x2 <- seq(-x2_limit,x2_limit,2*x2_limit/(ngridpoints-1))
X <- expand.grid(x1=x1,x2=x2)

# compute density
Density <- apply(X,1,dmnorm,mean=Mean,varcov=Sigma)

Density <- matrix(Density,length(x1),length(x2),byrow=FALSE)

# Colors for surface = estimates
n_col <- 80
nrDensity <- nrow(Density)
ncDensity <- ncol(Density)
Densitylim <- c(min(Density),max(Density))
#couleurs <- tail(heat.colors(trunc(1 * n_col)),n_col)
couleurs <- topo.colors(trunc(1 * n_col))
Densitycol <- couleurs[trunc((Density-Densitylim[1])/
  (Densitylim[2]-Densitylim[1])*(n_col-1))+1]
dim(Densitycol) <- c(nrDensity,ncDensity)
Densitycol <- Densitycol[-nrDensity,-ncDensity]

# plot surface and contour lines
pdf("Biv_Normal_Surface_col.pdf", height=6, width=6)
#par(mfrow=c(1,1))
#split.screen(c(2,1))
#screen(1)
par(mai=c(0.5,0.5,0.3,0.1))
persp(y=x1, x=x2, z=Density, col=Densitycol,
  main="Density of Bivariate Normal Distribution for (x1,x2)" ,
  theta=35, phi=20 , r=10, shade=0.1, ticktype="detailed")

```

```
dev.off()
# ?contour
# screen(2)
pdf("Biv_Normal_Surface_con.pdf", height=6, width=6)
contour(x1,x2,Density,nlevels=50,main="Density of Bivariate Normal Distribution for (x1,x2)" )
dev.off()
# close.screen(all=TRUE)
```

Listing A.3: ./R_code/2_5_Plot_PDF_biv_Normal.R

A.2. R-Programm für Monte-Carlo-Simulation im Abschnitt 5.5.1 zum Gesetz der großen Zahlen

```

# ===== 5_4_MC_bar_y_LLN_CLT.R =====
#
# Programm für Monte-Carlo-Simulation
# zum Illustrieren des LLM und CLTs des arithmetischen Mittels
# Berechnet Mittelwert und Standardabweichung über alle Replikationen
# sowie Histogramme
# DGP: Mittelwert + chi-quadratverteilter Fehler
# Hinweis: Programm ist der Lesbarkeit halber mit for-Schleifen geschrieben
# Stand: RT, 2015_10_02

graphics.off()          # Schließe alle Graphikfenster

# Setze Parameter des Modells und der Monte-Carlo-Simulation

set.seed(42)           # Randomseed
N      <- c(10,50,100,500) # Stichprobengrößen
R      <- 10000          # Zahl der Replikationen

mu     <- 1             # Mittelwert
deg_freedom <- 1        # Freiheitsgrade der qui-quadrat-Verteilung
sigma  <- 2             # Standardabweichung des Fehlers

save.pdf <- 1          # 1=Erstelle PDFs von Graphiken, 0=sonst
# Bilden von zwei Schleifen:
# Äußere Schleife über die Zahl der Replikationen
# Innere Schleife über die Stichprobengröße

n_max <- N[length(N)] # maximale Stichprobengröße
# Initialisieren der Outputmatrizen
mu_hat_store <- matrix(0,nrow=R,ncol=length(N))
mu_tilde_store <- matrix(0,nrow=R,ncol=length(N))

for (r in (1:R))
{
  # Generieren einer Realisation eines einfachen linearen Regressionsmodells
  # für die maximale Stichprobengröße
  u <- rchisq(n_max,df=deg_freedom,) # Ziehen u
  u <- (u-deg_freedom)/sqrt(2*deg_freedom) # Standardisieren
  y <- mu+u

  for (i in (1:length(N)))
  {
    # Speichern der Schätzungen
    mu_hat_store[r,i] <- mean(y[1:N[i]]) # arithm. Mittel
    mu_tilde_store[r,i] <- (y[1]+y[N[i]])/2 # alternativer Schätzer
  }
}

# Berechnen der arithmetischen Mittelwerte der Parameterschätzungen
mu_hat_mean <- colMeans(mu_hat_store)
mu_tilde_mean <- colMeans(mu_tilde_store)

# Berechnen der Varianzen der Parameterschätzungen
mu_hat_sd <- sqrt(diag(var(mu_hat_store)))
mu_tilde_sd <- sqrt(diag(var(mu_tilde_store)))

# Darstellung am Bildschirm
(cbind(N,mu_hat_mean,mu_hat_sd,mu_tilde_mean,mu_tilde_sd))

# Erstellen von Histogrammen
if (save.pdf) pdf("plot_MC_mu_hat_Konsistenz.pdf", height=6, width=6)
par(mfrow=c(2,2)) # Zeichne vier Plots in ein Graphikfenster
for (i in (1:4))

```

```

{
# Stichprobengröße N[i]
hist(mu_hat_store[,i], breaks=sqrt(R),
      xlab=expression(hat(mu)), main=paste("Histogramm für n= ",N[i],sep=""))
}
if (save.pdf) dev.off()

if (save.pdf) pdf("plot_MC_mu_tilde_Konsistenz.pdf", height=6, width=6)
par(mfrow=c(2,2)) # Zeichne vier Plots in ein Graphikfenster
for (i in (1:4))
{
# Stichprobengröße N[i]
hist(mu_tilde_store[,i], breaks=sqrt(R),
      xlab=expression(tilde(mu)), main=paste("Histogramm für n= ",N[i],sep=""))
}
if (save.pdf) dev.off()
# ===== Ende =====

```

Listing A.4: ./R_code/5_4_MC_bar_y_LLN_CLT.R

A.3. R-Programme für Graphiken im Abschnitt 5.6 zu Grundlagen von Tests

Test auf Mittelwert des DAX, siehe Seite 129

```

# ===== 5_5_Test_Mean_DAX.R =====#
# Programm testet Erwartungswert der DAX-Renditen.
# Daten stammen von Yahoo-Finance
# Stand: 2015_10_01, 2018_11_28 RT

# Setze Arbeitsverzeichnis (working directory) auf das Verzeichnis der R-Datei
WD <- getwd() # Bestimme Verzeichnis der R-Datei und
setwd(WD) # setze es als Working Directory

# Einlesen der .xlsx-Datei
# 1) Verwende package "xlsx"
# erfordert, dass ältere Java-Version vorhanden ist,
# funktioniert deshalb nicht auf jedem MAC
if (!require(xlsx)){
install.packages("xlsx")
}
library(xlsx) # Library zum Einlesen von Dateien im Format xls oder xlsx

excel_daten <- read.xlsx("DAX_19930401_20171128.xlsx",
                        sheetIndex = 1, colIndex = 6, startRow=2,
                        colClasses = c("numeric"))

# 2) Verwende package "readxl", das Java nicht benötigt.
# Es gibt allerdings einen "tibble" anstelle eines "dataframes" zurück, so dass
# zusätzlich der Befehl "data.frame" benötigt wird, um einen tibble in einen dataframe
# umzuwandeln
if (!require(readxl)) install.packages("readxl")
library(readxl) # RT, 2017_05_15

excel_daten <- data.frame(read_excel(path="DAX_19930401_20171128.xlsx",
                                   range="Tabelle1!F2:F6355", col_names=FALSE, na="null"))

price <- excel_daten[,1]
price <- na.omit(price)
# price <- rev(price) # nur für Excel-Datei für Daten bis 2015
# sortiere Preise so, dass ältester Wert
# am Anfang des Vektors price, damit

```

```
                                # Renditen korrekt berechnet werden
n                                <- length(price)
r                                <- (log(price[2:n])-log(price[1:(n-1)])) / log(price[1:(n-1)])
mean(r)
sd(r)
(t                                <- mean(r)/(sd(r)/sqrt(n))) # Berechne Teststatistik
# ===== Ende =====
```

Listing A.5: ./R_code/5_5_Test_Mean_DAX.R

Gütefunktion des Tests bzgl. des Mittelwertes, siehe Abbildung 5.3

```

# ===== 5_5_Plot_Power_Function.R =====
# Programm zum Erstellen der Graphik zum Darstellen der
# Powerfunktion in Folien Methoden, Abschnitt 4.1
# erstellt von: RT, 2012_12_20, kor. 2015_12_01, 2021_10_21 (sigma_0 aus X-Unterschrift entfernt)

# ===== Definiere Funktion zur Berechnung der Power
z_power <- function(c, mu_v, sigma_mu, mu_H0)
{
  power_left   <- pnorm(-c, mean = (mu_v - mu_H0)/sigma_mu, sd=1)
  power_right  <- 1 - pnorm(c, mean = (mu_v - mu_H0)/sigma_mu, sd=1)
  return(power_left + power_right)
}
# ===== Ende Function =====

# Parameter für Plot

graphics.off()          # Schließe alle Graphikfenster

alpha    <- 0.05      # Signifikanzniveau
n        <- 50        # Anzahl Beobachtungen
mu_H0    <- 0         # Mittelwert unter H_0
sigma    <- 1         # Standardabweichung

save.pdf <- 1        # 1=Erstelle PDFs von Graphiken, 0=sonst

c        <- qnorm(1 - alpha/2) # Berechnen kritischer Wert
mu_v     <- seq(mu_H0 - 2, mu_H0 + 2, 0.1) # Gitter für Dichte unter H_1

# Erstelle Plot

if (save.pdf) pdf("plot_power_function_2021.pdf", height=4, width=7)
plot(mu_v, z_power(c, mu_v, sigma/sqrt(n), mu_H0), type="l",
      xlab = expression(mu[0]-mu[H[0]]#/(sigma[0]/sqrt(n)))
      ), ylab="Gütefunktion")
abline(h = 0.05, col = "red")
axis(2, at = 0.05, labels = expression(alpha), tick=FALSE)
lines(mu_v, z_power(c, mu_v, 2 * sigma/sqrt(n), mu_H0), type="l", col="blue")
if (save.pdf) dev.off()
# ===== Ende =====

```

Listing A.6: ./R_code/5_5_Plot_Power_Function.R

Darstellung der Gütefunktion auf einem Gitter, siehe Abbildung 5.4

```

# ===== power_function_persp =====
# Programm zum Erstellen der perspektischen Graphik zum Darstellen der
# Powerfunktion in Folien Methoden, Abschnitt 4.1
# erstellt von: RT, 2012_12_20
#
# Beachte: Falls library aplpack nicht installiert ist, vorher installieren!
# Laden der Library aplpack, die die Funktionen für slider enthält

graphics.off() # Schließe alle Graphikfenster
library(aplpack)# Laden der Library aplpack

# ===== Definiere Funktionen =====
# ----- z_power_grid-----
z_power_grid <- function(mu_d_sigma_mu,mu_0,c)
{
  mu_v      <- mu_d_sigma_mu[1]
  sigma_mu  <- mu_d_sigma_mu[2]
  power_left <- pnorm(-c,mean=(mu_v-mu_0)/sigma_mu,sd=1)
  power_right <- 1-pnorm(c,mean=(mu_v-mu_0)/sigma_mu,sd=1)
  return(power_left+power_right)
}
# ----- Ende z_power_grid -----

# ----- col_persp -----
# Funktion zum Kolorieren der Oberfläche
col_persp <- function(Z)
{
  # Colors for surface = estimates
  n_col <- dim(Z)
  nrZ <- nrow(Z)
  ncZ <- ncol(Z)
  Zlim <- c(min(Z),max(Z))
  couleurs <- heat.colors(trunc(1 * n_col))
  # couleurs <- topo.colors(trunc(1 * n_col))
  Zcol <- couleurs[trunc((Z-Zlim[1])/(Zlim[2]-Zlim[1])*(n_col-1))+1]
  dim(Zcol) <- c(nrZ,ncZ)
  return(Zcol      <- Zcol[-nrZ,-ncZ])
}
# ----- Ende col_persp-----

# ----- beweglicher plot -----
beweglicher_plot <- function(...)
# Erzeuge perspektivische Graphik
{
  persp(x=mu_v,y=sigma_mu_v,z=power_grid_mat, ticktype="detailed", col=power_grid_col,
        r=slider(no=3), #5,
        xlab = expression(mu[0]-mu[H[0]]), ylab = expression(sigma/sqrt(n)),
        zlab = "Gütefunktion",
        theta=slider(no=1), #35,
        phi=slider(no=2), #20,
        expand=1) -> res #phi = 30
}
# ----- Ende beweglicher plot -----
# ===== Ende Funktionen =====

# ===== Hauptprogramm =====
# Definieren von Parametern

alpha      <- 0.05      # Signifikanzniveau
mu_0       <- 0         # Mittelwert unter H_0
mu_diff    <- 1

sigma      <- 1         # Standardabweichung
sigma_min  <- sigma
sigma_max  <- sigma
n_min      <- 20
n_max      <- 1000

```

```

c          <- qnorm(1-alpha/2)      # kritischer Wert

# Grid
# grid fuer mu
mu_v      <- seq(mu_0-mu_diff,mu_0+mu_diff,0.05)
# grid fuer sigma_hat_mu
sigma_mu_step <- (sigma/sqrt(n_min)-sigma/sqrt(n_max))/(length(mu_v)-1)
sigma_mu_v   <- seq(sigma/sqrt(n_max),sigma/sqrt(n_min),by=sigma_mu_step)

grid      <- expand.grid(mu_v,sigma_mu_v)
power_grid <- apply(grid,1,z_power_grid,mu_0,c)
power_grid_mat <- matrix(power_grid,length(mu_v),length(sigma_mu_v),byrow=FALSE)

# Koloriere Hyperebene des Unterraums mit Funktion "col_persp", siehe oben
power_grid_col <- col_persp(power_grid_mat)

# Erstellen der 3D-Graphik
# rufe slider mit Funktion beweglicher_plot zum Erstellen
# und möglichem Drehen der 3D-Graphik auf
windows() # öffnet ein neues Graphikfenster
slider(beweglicher_plot,
  sl.names      = c("drehen", "kippen", "Entfernung"),
  sl.mins       = c(0, 0, 1),          # Minimumwerte für Schieber
  sl.maxs       = c(360, 360, 100),    # Maximumwerte für Schieber
  sl.deltas     = c(1, 1, 1),         # Schrittweite für Schieber
  sl.defaults   = c(35, 20, 5)        # Defaultwerte für Parameter
  , prompt = TRUE                    # sorgt dafür, dass man den Effekt einer Sliderbewegung
                                     # sofort auf dem Bildschirm sieht und nicht erst nach dem
                                     # Loslassen des Mousebuttons
)
# Ende slider

```

Listing A.7: ./R_code/5_5_Plot_Power_Function_Persp.R

A.4. R-Programm für empirisches Beispiel zu Handelsströmen, beginnend in Abschnitt 6.3

```

##### 4_ff_Beispiel_Handelsstroeme.R #####
#
#####
#####
# Handelsströme-Beispiel im Skript Methoden der Ökonometrie,
# Universität Regensburg
# Kommentierter R Code
# Stand: 01.10.2015
# Vorläufer:
# - aussenhandel_beispiel_hk.r WS 2014/15 für Teil bis
# ENDE PFLICHTKURS PO 2011 - STOFF
# - aussenhandel_beispiel_pflichtkurs.r
#####
#####
# Um das Skript ausführen zu können, werden die Daten für das
# Handelsströme-Beispiel "importe_ger_2004_ebrd.txt" benötigt.
#
# Hinweis: Zunächst werden die Funktionen stats und SelectCritEviews definiert.
# Anschließend beginnt das Hauptprogramm in Zeile ??
#####

```

```

#                               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 #####

#####
#                               Ende Definition Funktionen
#####

#####
#                               Beginn Hauptprogramm
#####
save.pdf    <- 1          # 1=Erstelle PDFs von Graphiken, 0=sonst

# 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")
}

# Festlegung des Arbeitsverzeichnisses (working directory)
# in welchem sich das R-Programm und die Daten befinden
WD          <- getwd() # Bestimme Verzeichnis der R-Datei und
setwd(WD)   # setze es als Working Directory

```

```

# Einlesen der Daten als data frame
daten_all <-read.table("importe_ger_2004_ebrd.txt", header = TRUE)
# Zuweisung der Variablenamen 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)

#####
# Abschnitt 6.3
#####

##### Scatterplot mit (linearer) Regressionsgerade #####
# I.1 Ziel/Wissenschaftliche Fragestellung: erster empirischer Versuch

# Für Ausgabe im PDF Format Dateiname definieren
if (save.pdf) pdf("plot_wdi_vs_trade.pdf", height=6, width=6)

# KQ-Schätzung eines einfachen linearen Regressionsmodells, abgespeichert in ols
ols <- 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, col = "red")
# Hinzufügen einer Legende
legend("bottomright", "Lineare Regression", col = "red", lty = 1, bty = "n")

# Device schließen
if (save.pdf) dev.off()

##### Schätzen von zwei multiplen linearen Regressionsmodellen #####
# II.3 Spezifizieren, Schätzen und Auswählen eines ökonomischen Modells
# Hinweis:
# Die Nummerierung der Regressionsmodelle orientiert sich an
# den Modellen im Skript, Abschnitt 10.3

# Ausführen einer linearen Regression und Speichern der Ergebnisse als Objekt
mod_2_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist))

# Anzeige der Regressionsergebnisse
summary(mod_2_kq)

# II.4 Überprüfen des geschätzten Modells
# Ausführen der linearen Regression mit zusätzlichem Regressor und
# Verwenden des formula-Befehls
mod_3a_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o
mod_3a_kq <- lm(mod_3a_formula)
# Anzeige der Regressionsergebnisse des zweiten linearen Regressionsmodells
summary(mod_3a_kq)

#####
# Abschnitt 8.2
#####
# Funktionale Form: level-level, ... , log-log

summary(lm(trade_0_d_o ~ wdi_gdpusdcr_o)) #level - level Modell
summary(lm(trade_0_d_o ~ log(wdi_gdpusdcr_o))) #level - log Modell
summary(lm(log(trade_0_d_o) ~ wdi_gdpusdcr_o)) #log - level Modell
summary(lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o))) #log - log Modelle
summary(lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o)+log(cepii_dist)))

#####
# Abschnitt 8.5
#####
# Besteht nichtlinearer Zusammenhang zwischen Importen und BIP?
# Einfache Modellierungsmöglichkeit: Regressor BIP geht auch quadratisch
# in das Modell ein

```

```

# Modell 5: Verwende auch log(BIP)^2 als Regressor
mod_5_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)

mod_5_kq <- lm(mod_5_formula)
summary(mod_5_kq)

# Generiere Plot der Elastizitäten für verschiedene BIPs
elast_gdp <- mod_5_kq$coef[2] + 2* mod_5_kq$coef[3]*log(wdi_gdpusdcr_o)
# Erstelle Scatterplot
if (save.pdf) pdf("plot_modell5_elast.pdf.pdf", height=6, width=6)
plot(wdi_gdpusdcr_o, elast_gdp, pch = 16, col = "blue", main = "GDP-Elasticity")
if (save.pdf) dev.off()

#####
# Abschnitt 9.5
#####
# Schätze Varianz-Kovarianzmatrix der KQ-Schätzer für Modell 3a
summary(mod_3a_kq)$cov

# Schätze Korrelationsmatrix der KQ-Schätzer für Modell 3a
cov2cor(summary(mod_3a_kq)$cov)

# Schätze Kovarianzmatrix der Stichprobenbeobachtungen für Modell 3a
cor(data.frame(log_wdi_gdpusdcr_o = log(wdi_gdpusdcr_o),
  log_cepii_dist=log(cepii_dist), ebrd_tfes_o))

#####
# Abschnitt 10.3 Informationskriterien
#####

# Berechnen der Werte der Tabelle.
# Anwenden der Funktion "SelectCritEviews" auf vier verschiedene Modelle

mod_1_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o))
summary(mod_1_kq)
deviance(mod_1_kq) # Berechnet SSR
SelectCritEviews(mod_1_kq) # Berechnet AIC, HQ, SC

mod_2_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist))
summary(mod_2_kq)
deviance(mod_2_kq) # Berechnet SSR
SelectCritEviews(mod_2_kq) # Berechnet AIC, HQ, SC

mod_3a_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o)
summary(mod_3a_kq)
deviance(mod_3a_kq) # Berechnet SSR
SelectCritEviews(mod_3a_kq) # Berechnet AIC, HQ, SC

mod_3b_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  log(cepii_area_o))
summary(mod_3b_kq)
deviance(mod_3b_kq) # Berechnet SSR
SelectCritEviews(mod_3b_kq) # Berechnet AIC, HQ, SC

mod_4_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o + log(cepii_area_o))
summary(mod_4_kq)
deviance(mod_4_kq) # Berechnet SSR
SelectCritEviews(mod_4_kq) # Berechnet AIC, HQ, SC

#####
# Abschnitt 11.3 Exakte Tests
#####

alpha <- 0.05 # Signifikanzniveau

```

```

# Schätzen von Modell 4
mod_4_kq <- lm(log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) +
              log(cepii_dist) + ebrd_tfes_o + log(cepii_area_o))
summary(mod_4_kq)
qf(1-alpha,2,44) # Kritischer Wert
library(car) # Library car laden für F-Test
# F-Test
F_stat <- linearHypothesis(mod_4_kq,
                          c("ebrd_tfes_o=0","log(cepii_area_o)=0"),
                          test=c("F"))
F_stat
# Chi^2-Test
F_stat <- linearHypothesis(mod_4_kq,
                          c("ebrd_tfes_o=0","log(cepii_area_o)=0"),
                          test=c("Chisq"))
F_stat

#####
# Abschnitt 11.7 Empirische Analyse von Handelsströmen
#####

# Modell 4 wurde in Abschnitt 10.3 berechnet
resid_mod_4_kq <- mod_4_kq$resid # Residuen von Modell 4
trade_0_d_o_fit <- mod_4_kq$fitted # Gefittete Werte von Modell 4

# Plot der Residuen vs. der gefitteten Werte
if (save.pdf) pdf("plot_fits_vs_resids_mod_4.pdf", 6, 6)
plot(trade_0_d_o_fit, resid_mod_4_kq, col = "blue", pch = 16, main = "Scatterplot")
if (save.pdf) dev.off()

# Plot des Histogramms der Residuen
if (save.pdf) pdf("plot_hist_resids_mod_4.pdf", 6, 6)
hist(resid_mod_4_kq, breaks = 20, col = "lightblue", prob = T, main = "Histogram")
# Geschätzte Dichte der Residuen
lines(density(resid_mod_4_kq), col = "black", prob = T, add=T)
# Einzeichnen der zugehörigen theoretischen Normalverteilung
curve(dnorm(x, mean = mean(resid_mod_4_kq), sd = sd(resid_mod_4_kq)),
      from = -3, to = 3, add = T, col = "red", lty = 2, lwd = 2)
legend("topleft", c("est. density", "theoretical\nnormal distribution"),
      col = c("black", "red"), lwd = 2, lty = c(1,2), bty = "n")
if (save.pdf) dev.off()

# statistische Auswertung der Residuen
stats(resid_mod_4_kq)

#### Konfidenzintervalle
confint(mod_4_kq)

#### t-Tests, zweiseitig und einseitig

# Zweiseitiger Test
# Bestimmen der kritischen Werte
alpha <- 0.05
qt(alpha/2, mod_4_kq$df)
qt(1-alpha/2, mod_4_kq$df)

# t-Statistik
(t <- (coefficients(mod_4_kq)[ "log(wdi_gdpusdcr_o)"] - 1) /
sqrt(vcov(mod_4_kq)[ "log(wdi_gdpusdcr_o)", "log(wdi_gdpusdcr_o)"])))

# p-value
2*pt(-abs(t), mod_4_kq$df)

# Alternative via F
# dazu benötigt man das car-paket
# download.packages("car", destdir="C:/Program Files/R/R-2.15.1/library")
# install.packages("car")
library("car")
(F_stat <- linearHypothesis(mod_4_kq, c("log(wdi_gdpusdcr_o)=1")))

```

```

# Einseitiger Test mit linksseitiger Alternative
# Kritische Werte
alpha      <- 0.05
(qt(alpha,mod_4_kq$df))

# t-Teststatistik
(t         <- summary(mod_4_kq)$coefficients["log(cepii_dist)",3])

# p-Value
(pt(t,mod_4_kq$df))

#### F-Test, Korrelationsmatrix und Konfidenzellipsen

# F-Test mit 2 Restriktionen
# kritischer Wert für F-Statistik
(qf(1-alpha,2,mod_4_kq$df))

# F-Statistik
(F2_stat   <- linearHypothesis(mod_4_kq,c("ebrd_tfes_o=0","log(cepii_area_o)=0"),
                                test=c("F")))

# chi^2-Test
# kritischer Wert für chi^2-Statistik
(qchisq(1-alpha,2))

# chi^2-Statistik
(Chisq_stat <- linearHypothesis(mod_4_kq,c("ebrd_tfes_o=0","log(cepii_area_o)=0"),
                                test=c("Chisq")))

#### Kovarianz- und Korrelationsmatrix

# Kovarianzmatrix
(cov_par   <- vcov(mod_4_kq))
# Korrelationsmatrix
(corr_par  <- cov2cor(cov_par))

#### Konfidenzellipsoide

# Konfidenzellipse
if (save.pdf) pdf("plot_conf_ellipse.pdf", 6, 6)
confidenceEllipse(mod_4_kq, which.coef = c(4, 5), levels = 0.95,
                  main = "confidence ellipse", col = "red")

# Konfidenzintervall
abline(v = confint(mod_4_kq, "ebrd_tfes_o", level = 0.95), lty = 2,
        col = "blue", lwd = 2)
abline(h = confint(mod_4_kq, "log(cepii_area_o)", level = 0.95), lty = 2,
        col = "blue", lwd = 2)
if (save.pdf) dev.off()

#####
# Abschnitt 14.4 FGLS und heteroskedastierobuste KQ-Schätzung
#####

#### FGLS-Schätzung für Modell 4
mod_4_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) +
  ebrd_tfes_o + log(cepii_area_o)

# 1. Schritt
resids      <- residuals(mod_4_kq)
fits       <- fitted(mod_4_kq)
mod_formula_ln_u_squared <- log(resids^2) ~ log(wdi_gdpusdcr_o) + log(cepii_dist)
+ ebrd_tfes_o + log(cepii_area_o)

# 2. Schritt
omega      <- exp(fitted(lm(mod_formula_ln_u_squared)))
model_gls <- lm(mod_4_formula, weights=1/omega)
(summary(model_gls))

```

```

#### Regressionsoutput mit heteroskedastie-robusten Standardfehlern

library(lmtest)
# Zu Wahlmöglichkeiten für die Schätzung der heteroskedastischen
# Varianz-Kovarianzmatrix siehe Abschnitt 14.3
(coeftest(mod_4_kq,vcov=hccm(mod_4_kq,type="hc1")))

#####
# Abschnitt 15.7
#####

#### Durchführen des RESET-Tests f-ur Modell 4 mit
resettest(mod_4_kq)

#### Durchführen des Breusch-Pagan-Tests f-ur Modell 4
bptest(mod_4_kq)

#####
#
#           Beginn Funktion whitetest
#####
# White-Test auf homoskedastische Fehler mit Kreuzprodukten
# RW, 2011_01_26

whitetest <- function(model){

  # Daten aus model extrahieren
  dat <- model$model
  dat$resid_sq <- model$resid^2

  # Formel für die Hilfsregression erstellen
  regr <- attr(model$terms, "term.labels")
  form <- as.formula(paste("resid_sq~(",paste(regr,collapse="+"),")^2+",paste("I(",regr,"^2)",collapse="+")))

  # Hilfsregression schätzen
  test_eq <- lm(form,data=dat)

  # Overall F-Test
  fstat <- summary(test_eq)$fstatistic

  # Ergebnis berechnen und ausgeben
  result1 <- c(fstat[1],fstat[2],fstat[3],pf(fstat[1],fstat[2],fstat[3],lower.tail=FALSE))
  names(result1) <- c("F-Statistic","df1","df2","P-Value")
  result <- list(result1,test_eq)
  return(result)
}

#####
#
#           Ende Funktion whitetest
#####

#### Durchführen des White-Test f-ur Modell 4
whitetest(mod_4_kq)

##### Ende Hauptprogramm #####

```

Listing A.8: ./R_code/4_ff_Beispiel_Handelsstroeme.R

A.5. R-Programm für Graphiken in Abschnitt 7.1 Die Geometrie des KQ-Schätzers

Geometrie des KQ-Schätzers, siehe Abbildungen 7.1 und 7.2 † **Ableitung der Funktion comp_d3** im folgenden R-Programm zur Berechnung der 3. Koordinate der Hyperebene, die durch $\delta(\mathbf{X})$ aufgespannt wird im Fall von $k = 2$ und $n = 3$:

Die Achsen der 3D-Graphik stehen orthogonal zueinander. Entsprechend der Richtungen der Achsen werden die drei Einheitsbasisvektoren \mathbf{e}_i , $i = 1, 2, 3$ gewählt (siehe Leverage-Effekt in Abschnitt 7.2). Für diese gilt deshalb $\mathbf{e}_i^T \mathbf{e}_j = 0$, $i \neq j$. Für alle Vektoren im Unterraum $\delta(\mathbf{X})$ gilt

$$\mathbf{X}\mathbf{a} = d_1\mathbf{e}_1 + d_2\mathbf{e}_2 + d_3\mathbf{e}_3 = \begin{pmatrix} d_1 \\ d_2 \\ d_3 \end{pmatrix} \quad (\text{A.1})$$

Zur Berechnung der Hyperebene des Unterraums in E^3 lassen sich d_1 und d_2 , jeweils auf einem Gitter vorgeben $d_1, d_2 = 0, 0.25, 0.5, \dots, 10$. Problem ist nun d_3 so zu bestimmen, dass (A.1) erfüllt ist:

1. Hierzu bestimmt man zunächst den (2×1) -Vektor \mathbf{a} in Abhängigkeit von d_1, d_2 :

$$\underbrace{\begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}}_{:=\mathbf{X}_I} \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}.$$

Multiplikation mit \mathbf{X}_I^{-1} (\mathbf{X}_I ist hier quadratisch) ergibt

$$\mathbf{a} = \mathbf{X}_I^{-1} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}.$$

2. Berechnen von d_3 mittels der 3. Zeile von (A.1) liefert:

$$d_3 = \mathbf{X}_3\mathbf{a} = \mathbf{X}_3\mathbf{X}_I^{-1} \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}.$$

```
# ===== 7_1_Projektion_KQ_n3.R =====
# Programm zum Erstellen der Graphiken im Skript Methoden der Ökonometrie,
# Abschnitt 7.1 Die Geometrie des KQ-Schätzers
# erstellt von: RT, KK, JS, 2010_11_24, 2015_09_30
#
# Falls library aplpack nicht installiert ist, wird dies jetzt installiert.
if (!require(aplpack)){
  install.packages("aplpack")
}
graphics.off() # SchlieÙe alle Graphikfenster
# Laden der Library aplpack, die die Funktionen für slider enthält
library(aplpack)# Laden der Library aplpack
```

```

# ===== Definiere Funktionen =====
# ----- comp_d3 -----
# Funktion "comp_d3", die d_3 gegeben d_1 und d_2 und X berechnet,
# vgl. Skript Methoden der Ökonometrie, Abschnitt A.3
comp_d3 <- function(d,X) X[3,] %*% solve(X[1:2,]) %*% d
# ----- Ende comp_d3 -----

# ----- col_persp -----
# Funktion zum Kolorieren der Oberfläche
col_persp <- function(Z)
{
  # Colors for surface = estimates
  n_col <- dim(Z)
  nrZ <- nrow(Z)
  ncZ <- ncol(Z)
  Zlim <- c(min(Z),max(Z))
  couleurs <- heat.colors(trunc(1 * n_col))
  # couleurs <- topo.colors(trunc(1 * n_col))
  Zcol <- couleurs[trunc((Z-Zlim[1])/(Zlim[2]-Zlim[1])*(n_col-1))+1]
  dim(Zcol) <- c(nrZ,ncZ)
  return(Zcol <- Zcol[-nrZ,-ncZ])
}
# ----- Ende col_persp -----

# ----- beweglicher_plot -----
# Definiere Funktion "beweglicher_plot", die vom Programm "slider"
# aufgerufen wird und die Oberflächen/Perspektivengraphik erzeugt
# Beachte: die Funktion verwendet Variable aus dem Hauptprogramm ohne
# dass diese explizit übergeben werden
beweglicher_plot <- function(...) {
  # Erzeuge perspektivische Graphik
  persp(x=d1,y=d2,z=d3_mat, ticktype="detailed", col=d3_col,
        r=slider(no=3), #5,
        xlab = "e1", ylab = "e2", zlab = "e3",
        theta=slider(no=1), #35,
        phi=slider(no=2), #20,
        expand=1) -> res #phi = 30
  # x_1 Vektor
  lines(trans3d(x=c(0,x1[1]), y=c(0,x1[1]), z=c(0,x1[3]), pmat=res),
        col="black", lwd=2)
  text(trans3d(x1[1], x1[2]+1, x1[3]), pmat=res), expression(x[1]),cex=sym_gr)

  # x_2 Vektor
  lines(trans3d(x=c(0,x2[1]), y=c(0,x2[2]), z=c(0,x2[3]), pmat=res),
        col="black", lwd=2)
  text(trans3d(x2[1]-1, x2[2]-0.5, x2[3]), pmat=res), expression(x[2]),cex=sym_gr)

  # X beta Vektor
  lines(trans3d(x=c(0,Xbeta[1]), y=c(0,Xbeta[2]), z=c(0,Xbeta[3]), pmat=res),
        col="grey2", lwd=2)
  text(trans3d(Xbeta[1]+1, Xbeta[2], Xbeta[3]), pmat=res), expression(X*beta),
        cex=sym_gr)

  # verschobener u Vektor
  lines(trans3d(x=c(Xbeta[1],y[1]), y=c(Xbeta[2],y[2]), z=c(Xbeta[3],y[3]),
        pmat=res), col="brown", lwd=2)

  # y Vektor
  lines(trans3d(x=c(0,y[1]), y=c(0,y[2]), z=c(0,y[3]), pmat=res),
        col="red", lwd=2)
  text(trans3d(y[1], y[2], y[3]+1, pmat=res), expression(y),cex=sym_gr)

  # X hat beta Vektor
  lines(trans3d(x=c(0,y_hat[1]), y=c(0,y_hat[2]), z=c(0,y_hat[3]), pmat=res),
        col="purple", lwd=2)
  text(trans3d(y_hat[1]+1, y_hat[2], y_hat[3]), pmat=res), expression(X*hat(beta)),
        cex=sym_gr )
}

```

```

# hat u Vektor
lines(trans3d(x=c(0,y[1]-y_hat[1]), y=c(0,y[2]-y_hat[2]), z=c(0,y[3]-y_hat[3]),
            pmat=res), col="green", lwd=2)
text(trans3d(y[1]-y_hat[1], y[2]-y_hat[2], y[3]-y_hat[3]+1, pmat=res),
     expression(hat(u)),cex=sym_gr)

# verschobener hat u Vektor
lines(trans3d(x=c(y[1],y_hat[1]), y=c(y[2],y_hat[2]), z=c(y[3],y_hat[3]),
            pmat=res), col="green", lwd=2)
}
# ----- Ende beweglicher_plot -----
# ===== Ende Funktionen =====

# ===== Hauptprogramm =====
# Definieren von Parametern
# Festlegen der Parameter für die Graphik
sym_gr <- 1.5 # Symbolgröße
step <- .25 # Schrittgröße für das Gitter, über dem X beta berechnet
# und gezeichnet werden soll

# Festlegen der drei Beobachtungen und des Parametervektors beta
x1 <- c(1,1,1)
x2 <- 2*c(0.5,2,1.3)
beta <- c(5,-1)
u <- c(-3,4,5)

X <- cbind(x1,x2) # X = { x_1 x_2 }
Xbeta <- X%*%beta # X*beta
y <- Xbeta + u

# Berechnen des KQ-Schätzers und der gefitteten Werte
beta_hat<- solve(t(X) %*% X) %*% t(X) %*% y
y_hat <- X %*% beta_hat

# Berechnen der 3D-Graphik
d1_min <- min(Xbeta,0)
d1_max <- max(Xbeta,10)
d1 <- seq(d1_min,d1_max, by=step) # Gitterpunkte in erster Richtung
d2 <- seq(d1_min,d1_max, by=step) # Gitterpunkte in zweiter Richtung
d_grid <- expand.grid(d1,d2) # Erstellen des Gitters, über dem Unterraum
# delta(X) geplottet werden soll
# Wende Funktion "comp_d3" auf Gitter von d_1 und d_2 an
# beachte: Argumente, die in apply der Funktion übergeben werden, werden
# ohne = übergeben
d3_grid <- apply(d_grid,1,comp_d3,X)
# apply gibt einen Vektor aus, der im Folgenden in eine Matrix umgewandelt
# wird, so dass d_3 zu den korrekten d_1 und d_2 passt.
d3_mat <- matrix(d3_grid,length(d1),length(d2),byrow=FALSE)

# Koloriere Hyperebene des Unterraums mit Funktion "col_persp", siehe oben
d3_col <- col_persp(d3_mat)

# Erstellen eines Scatterplots mit Regressionsgerade des DGP und
# geschätzte Regressionsgerade,
# Fehler und Residuen für 2. Beobachtung
plot(x2,y,col="red",pch=16,xlab=expression(x[2]),ylab=expression(y)) # Scatterplot
abline(a=beta[1],b=beta[2],col="black") # Regressionsgerade des DGP
points(x2,Xbeta,col="black",pch=16) # X beta auf der Regressionsgerade
abline(a=beta_hat[1],b=beta_hat[2],col="blue")# geschätzte Regressionsgerade
points(x2,y_hat,col="blue",pch=16) # hat y auf der gesch. Reg.ger.
# Einzeichnen des Fehlervektors und Residuenvektors
# für 2. Beobachtung
t <- 2
lines(cbind(x2[t],x2[t]),cbind(Xbeta[t],y[t]),col="brown") # Fehlervektor
text(x2[t]-.2,(y[t]-Xbeta[t])*0.75, expression(u[2]),cex=sym_gr)

lines(cbind(x2[t],x2[t]),cbind(y_hat[t],y[t]),col="green") # Residuenvektor
text(x2[t]-.2,y[t]-y_hat[t])*0.5, expression(hat(u)[2]),cex=sym_gr)
text(x2[t]-.4,y[t],expression((list(x[22],y[2]))),cex=sym_gr)

```

```
# Erstellen der 3D-Graphik
# rufe slider mit Funktion beweglicher_plot zum Erstellen
# und möglichem Drehen der 3D-Graphik auf
windows() # öffnet ein neues Graphikfenster
slider(beweglicher_plot,
       sl.names      = c("drehen", "kippen", "Entfernung"),
       sl.mins       = c(0, 0, 1),           # Minimumwerte für Schieber
       sl.maxs       = c(360, 360, 100),    # Maximumwerte für Schieber
       sl.deltas     = c(1, 1, 1),         # Schrittweite für Schieber
       sl.defaults   = c(35, 20, 5)        # Defaultwerte für Parameter
       , prompt      = TRUE                # sorgt dafür, dass man den Effekt einer Sliderbewegung
                                           # sofort auf dem Bildschirm sieht und nicht erst nach dem
                                           # Loslassen des Mousebuttons
)
# Ende slider
# ===== Ende Hauptprogramm =====
```

Listing A.9: ./R_code/7_1_Projection_KQ_n3.R

A.6. R-Programm für Regressionsergebnisse in Abschnitt 8.3 zu qualitative Daten als Regressoren

```

# ===== 8_4_Interpretationen_Wage.R =====
# Programm für Lohnregressionen mit Dummies und Interaktionstermen,
# siehe Abschnitt 8.4 in Skript Methoden der Ökonometrie
# Stand: 2015_10_02
# Vorläufer: app_interpretationen_wage.r aus WS 2013/14

# Festlegung des Arbeitsverzeichnisses (working directory)
# in welchem sich das R-Program und die Daten befinden

WD          <- getwd() # Bestimme Verzeichnis der R-Datei und
setwd(WD)    # setze es als Working Directory

# Einlesen der Daten
# Die Datendatei "wage1.txt" muss in demselben Verzeichnis wie die
# R-Datei liegen
wage_data   <- read.table("wage1.txt", header = TRUE)
attach(wage_data)

# Lohnregression mit Dummyvariable, siehe Abschnitt 8.4.1
wage_mod_1_kq <- lm(log(wage) ~ female +
                    educ + exper + I(exper^2) + tenure + I(tenure^2))
summary(wage_mod_1_kq)

# Relative Differenz der unbedingten Lohnmittelwerte von Frauen und Männern
(mean(wage[female==1]) - mean(wage[female==0])) / mean(wage[female==0])
# alternative Berechnungsmöglichkeit
wage_mean <- lm(wage ~ 0 + female + I(1 - female))
(wage_mean$coef[1] - wage_mean$coef[2]) / wage_mean$coef[2]

# Lohnregression mit mehreren Dummyvariablen: Interaktion von Dummies,
# siehe Abschnitt 8.4.2

# Definiere Dummvariablen für Untergruppen
femmarr     <- female * married
malesing    <- (1 - female) * (1 - married)
malemarr    <- (1 - female) * married

wage_mod_2_kq <- lm(log(wage) ~ femmarr + malesing + malemarr +
                    educ + exper + I(exper^2) + tenure + I(tenure^2))
summary(wage_mod_2_kq)

# Lohnregression mit einer Dummy und Dummyinteraktionsterm
wage_mod_3_kq <- lm(log(wage) ~ female +
                    educ + exper + I(exper^2) + tenure + I(tenure^2) +
                    I(female*educ))
summary(wage_mod_3_kq)

```

Listing A.10: ./R_code/8_4_Interpretationen_Wage.R

A.7. R-Programm für Graphiken in Abschnitt 9.1 zu Erwartungstreue des KQ-Schätzers

Monte-Carlo-Simulation zur Erwartungstreue, siehe Abbildung 9.1

```
# ===== 9_1_MC_KQ_einf_lin_Reg.R =====
#
# Programm für Monte-Carlo-Simulation
# zum Illustrieren der Erwartungstreue des KQ-Schätzers
# im einfachen linearen Regressionsmodell.
# Außerdem wird mit einem Scatterplot die Kovarianz zwischen den geschätzten
# KQ-Parametern illustriert.
# erstellt von : RT, 2010_11_25

graphics.off()      # Schließe alle Graphikfenster

# Setze Parameter des Modells und der Monte-Carlo-Simulation

set.seed(42)        # Randomseed
n                   <- 50      # Stichprobengröße
R                   <- 1000    # Zahl der Replikationen

beta_0              <- c(1,0.9) # Parametervektor
sigma_0             <- 2       # Standardabweichung des Fehlers

save.pdf            <- 1       # 1=Erstelle PDFs von Graphiken, 0=sonst

# Bilden einer Schleife
beta_hat_store <- matrix(0,nrow=R,ncol=length(beta))
                # Initialisiere Matrix zum Abspeichern der KQ-Schätzungen
                # für jede Realisation
for (r in (1:R))
{
  # Generieren einer Realisation eines einfachen linearen Regressionsmodells
  u           <- rnorm(n,mean=0,sd=sigma_0) # Ziehen von u
  x           <- sample(1:20, n, replace=TRUE) # Ziehen von x
  y           <- beta_0[1] + x * beta_0[2] + u # Berechnen von y

  # Berechnen des KQ-Schätzers
  ols         <- lm(y~x)

  # Speichern der Parameterschätzung
  beta_hat_store[r,] <- coef(ols)
}

# Berechnen der Mittelwerte der Parameterschätzungen
colMeans(beta_hat_store)

# Erstellen von Histogrammen
if (save.pdf) pdf("plot_MC_KQ_einf_lin_Reg_hist.pdf", height=6, width=6)
par(mfrow=c(1,2)) # Zeichne zwei Plots in ein Graphikfenster
hist(beta_hat_store[,1],breaks=sqrt(R))
hist(beta_hat_store[,2],breaks=sqrt(R))
if (save.pdf) dev.off()

# Varianz-Kovarianzmatrix der Schätzer aus den R Realisationen
(var(beta_hat_store))

# Asymptotische Varianz-Kovarianzmatrix
S_XX      <- matrix(c(1,10.5,10.5,143.5),2,2)
cov_asymp <- sigma_0^2 * solve(S_XX)
          # Anpassung an Stichprobengröße
          (cov_asymp / n)

# Scatterplot der R KQ-Schätzungen
par(mfrow=c(1,1))
plot(beta_hat_store[,1],beta_hat_store[,2])
```

```
# ===== Ende =====
```

Listing A.11: ./R_code/9_1_MC_KQ_einf_lin_Reg.R

A.8. R-Programm für Monte-Carlo-Simulation im Abschnitt 9.2 zur Konsistenz des KQ-Schätzers

Monte-Carlo-Simulation zur Konsistenz und zum zentralen Grenzwertsatz, siehe Abbildungen 9.2 und 9.3

```
# ===== 9_2_MC_KQ_Konsistenz_einf_lin_Reg.R =====
# Programm für Monte-Carlo-Simulation
# zum Illustrieren der Konsistenz und der asymptotischen Normalverteilung
# des KQ-Schätzers im einfachen linearen Regressionsmodell.
# Berechnet Mittelwert und Standardabweichung über alle Replikationen
# sowie Histogramme.
# Hinweis: Programm ist der Lesbarkeit halber mit for-Schleifen geschrieben
# Stand: RT, 2015_10_04

graphics.off()          # Schließe alle Graphikfenster

# Setze Parameter des Modells und der Monte-Carlo-Simulation

set.seed(42)           # Randomseed
N                      <- c(50,100,500,1000,10000,100000) # Stichprobengrößen
R                      <- 10000 # Zahl der Replikationen

beta                   <- c(1,0.9) # Parametervektor
sigma                  <- 2 # Standardabweichung des Fehlers

save.pdf               <- 1 # 1=Erstelle PDFs von Graphiken, 0=sonst
# Bilden von zwei Schleifen:
# Äußere Schleife über die Zahl der Replikationen
# Innere Schleife über die Stichprobengröße

n_max                  <- N[length(N)] # maximale Stichprobengröße
# Initialisieren der Outputmatrizen
beta_1_hat_store <- matrix(0,nrow=R,ncol=length(N))
# Initialisiere Matrix zum Abspeichern der KQ-Schätzungen
# für beta_1 jede Realisation und jede Stichprobengröße
beta_2_hat_store <- matrix(0,nrow=R,ncol=length(N))
# Initialisiere Matrix zum Abspeichern der KQ-Schätzungen
# für beta_2 jede Realisation und jede Stichprobengröße

for (r in (1:R))
{
  # Generieren einer Realisation eines einfachen linearen Regressionsmodells
  # für die maximale Stichprobengröße
  u <- rnorm(n_max,mean=0,sd=sigma) # Ziehen von u
  x <- sample(1:20, n_max, replace=TRUE) # Ziehen von x
  y <- beta[1] + x * beta[2] + u # Berechnen von y

  for (i in (1:length(N)))
  {
    # Berechnen des KQ-Schätzers für alle Stichprobengrößen

    # ols <- lm(y[1:N[i]]~x[1:N[i]]) # Standard-Befehl für KQ-Schätzung
    # schneller lm-Befehl, um in der Simulation Zeit zu sparen
    ols <- lm.fit(cbind(rep(1,N[i]),x[1:N[i]]),y[1:N[i]])

    # Speichern der Parameterschätzung
    beta_1_hat_store[r,i] <- coef(ols)[1]
    beta_2_hat_store[r,i] <- coef(ols)[2]
  }
}

# Berechnen der Mittelwerte der Parameterschätzungen
beta_1_hat_mean <- colMeans(beta_1_hat_store)
beta_2_hat_mean <- colMeans(beta_2_hat_store)
```

```

# Berechnen der Standardabweichungen der Parameterschätzungen
beta_1_hat_sd <- sqrt(diag(var(beta_1_hat_store)))
beta_2_hat_sd <- sqrt(diag(var(beta_2_hat_store)))

# Darstellung am Bildschirm
(cbind(N,beta_1_hat_mean,beta_1_hat_sd,beta_2_hat_mean,beta_2_hat_sd))

# Erstellen von Histogrammen
if (save.pdf) pdf("plot_MC_KQ_Konsistenz_einf_lin_Reg1.pdf", height=6, width=6)
par(mfrow=c(2,2)) # Zeichne vier Plots in ein Graphikfenster
for (i in (1:2))
{
  # Stichprobengröße N[i]
  hist(beta_1_hat_store[,i], breaks=sqrt(R),
       xlab=expression(hat(beta)[1]), main=paste("Histogramm für n= ",N[i],sep=""))
  hist(beta_2_hat_store[,i], breaks=sqrt(R),
       xlab=expression(hat(beta)[2]), main=paste("Histogramm für n= ",N[i],sep=""))
}
if (save.pdf) dev.off()

if (save.pdf) pdf("plot_MC_KQ_Konsistenz_einf_lin_Reg2.pdf", height=6, width=6)
par(mfrow=c(2,2)) # Zeichne vier alle Plots in ein Graphikfenster
for (i in (3:4))
{
  # Stichprobengröße N[i]
  hist(beta_1_hat_store[,i], breaks=sqrt(R),
       xlab=expression(hat(beta)[1]), main=paste("Histogramm für n= ",N[i],sep=""))
  hist(beta_2_hat_store[,i], breaks=sqrt(R),
       xlab=expression(hat(beta)[2]), main=paste("Histogramm für n= ",N[i],sep=""))
}
if (save.pdf) dev.off()

if (save.pdf) pdf("plot_MC_KQ_Konsistenz_einf_lin_Reg3.pdf", height=6, width=6)
par(mfrow=c(2,2)) # Zeichne vier alle Plots in ein Graphikfenster
for (i in (5:6))
{
  # Stichprobengröße N[i]
  hist(beta_1_hat_store[,i], breaks=sqrt(R),
       xlab=expression(hat(beta)[1]), main=paste("Histogramm für n= ",N[i],sep=""))
  hist(beta_2_hat_store[,i], breaks=sqrt(R),
       xlab=expression(hat(beta)[2]), main=paste("Histogramm für n= ",N[i],sep=""))
}
if (save.pdf) dev.off()
# ===== Ende =====

```

Listing A.12: ./R_code/9_2_MC_KQ_Konsistenz_einf_lin_Reg.R

A.9. R-Programm zur Darstellung von ifo Geschäftsklimazeitreihen im Abschnitt 12 zu univariaten Zeitreihenmodellen

```

# ===== 12_0_ifo_Geschaeftsklima.R =====
#
# erzeugt Graphik der Zeitreihen zu ifo Geschäftserwartungen,
# der ifo Geschäftsbeurteilung und des ifo Geschäftsklimas für die
# gewerbliche Wirtschaft
# letzte Änderung: 2018_11_29, CR

save.pdf      <- 1          # 1=Erstelle PDFs von Graphiken, 0=sonst

# Falls diese nicht installiert sind, werden diese zunächst installiert:
if (!require(xlsx)){
  install.packages("xlsx")
}
if (!require(dynlm)){
  install.packages("dynlm")
}
if (!require(readxl)) {
  install.packages("readxl")
}

# Festlegung des Arbeitsverzeichnisses (working directory)
# in welchem sich das R-Programm und die Daten befinden
WD      <- getwd() # Bestimme Verzeichnis der R-Datei und
setwd(WD)        # setze es als Working Directory

# Einlesen der Daten
# 1) Verwende package "xlsx"
#     erfordert, dass ältere Java-Version vorhanden ist,
#     funktioniert deshalb nicht auf jedem MAC

library(xlsx) # Einlesen von Dateien im Format xls oder xlsx
excel_daten <- read.xlsx("ifo-geschaeftsklima_1991_01_2015_09.xls",
                        sheetIndex = 1, colIndex = c(2,3,4),startRow=20,
                        colClasses = c("numeric","numeric","numeric"))

# 2) Verwende package "readxl", das Java nicht benötigt.
#     Es gibt allerdings einen "tibble" anstelle eines "dataframes" zurück, so dass
#     zusätzlich der Befehl "data.frame" benötigt wird, um einen tibble in einen dataframe
#     umzuwandeln
library(readxl)
excel_daten <- data.frame(read_excel(path="ifo-geschaeftsklima_1991_01_2015_09.xls",
                                    sheet = 1, range = "B20:D313", col_names=FALSE))

# Erstellen eines Zeitreihenobjekts mit dataframe-Eigenschaften
daten <- ts((excel_daten[-1,1:3]),
           start = c(1991, 1), end = c(2015,9), frequency = 12,
           names = c("Geschaeftsklima", "Geschaeftsbeurteilung",
                    "Geschaeftserwartungen"))

# Zeitreihen plotten
if (save.pdf) pdf("ifo_geschaeftsklima_1991_01_2015_09.pdf", height=6,width=6)
plot(daten,xlab="Zeit",main="ifo Geschäftsdaten")
if (save.pdf) dev.off()

# Scatterplot für Geschäftserwartungen erstellen
n <- nrow(daten)
if (save.pdf) pdf("ifo_geschaeftsklima_scatter_1991_01_2015_09.pdf",
                height=6, width=6)
plot(Geschaefterwartungen[2:n] ~ Geschaeftserwartungen[1:(n-1)], data=daten)
if (save.pdf) dev.off()

# AR(1)-Modell für Geschaeftsklima und Geschäftsklima schätzen

```

```
gk_ols <- lm(Geschaeftsklima[2:n] ~ Geschaeftsklima[1:(n-1)],data=daten)
summary(gk_ols)

ge_ols <- lm(Geschaefterwartungen[2:n] ~ Geschaefterwartungen[1:(n-1)],
            data=daten)
summary(ge_ols)

# alternativ mit dynlm-Paket (lässt Lag-Schreibweise wie in EViews zu)
library(dynlm) # erleichterte Regression mit Zeitreihen
gk_dynlm <- dynlm(Geschaeftsklima ~ L(Geschaeftsklima),data=daten)
summary(gk_dynlm)
# ===== Ende =====
```

Listing A.13: ./R_code/12_0_ifo_Geschaeftsklima.R

A.10. R-Programm zur Darstellung verschiedener Realisierungen von Zeitreihen im Abschnitt 12.1 zu stochastischen Prozessen

```
# ===== 12_1_Traj_RW.R =====
#
# erzeugt Graphik zehn Realisationen eines Random Walks
# und plottet die Zeitreihen
# letzte Änderung: 2015_10_10, RT

save.pdf <- 1 # 1=Erstelle PDFs von Graphiken, 0=sonst

# Parameter des DGPs und der MC
n <- 20 # Länge der Zeitreihen

alpha <- 1 # AR-Parameter von AR(1)-Prozess mit Mittelwert 0
          # 0: Gaussches Weißes Rauschen
          # 1: Random Walk
          # 0 < |alpha| < 1: stationärer Prozess

R <- 10 # Zahl der Trajektorien

var_z <- 0 # Varianz von z zur Illustration von Ergodizität:
          # 0 => ergodisch
          # >0 => nicht ergodisch

# Parameter für Plots
lwd <- 3
cexmu <- 2

set.seed(42) # seed value

# Initialisierung der Outputmatrizen
y <- matrix(rnorm(n*R),n) # Initialisierung der Zeitreihenvektoren mit
                          # standardnormalverteilten Fehlerprozess (Gaussches
                          # Weißes Rauschen)
z <- rnorm(R) * var_z # Ziehen einer Zufallszahl, die für alle t gleich ist

# Erzeugen aller R Trajektorien des AR(1)-Prozesses

for (i in 1:R) y[,i] <- filter(y[,i],alpha,method="recursive")

# Plotten der Zeitreihen - Darstellung Ensemble
if (save.pdf) pdf("Traj_RW_points.pdf")
# Erste Trajektorie
plot(y[,1]+z[1], cex.lab=cexmu, cex.axis=cexmu, lwd=lwd, ylim=c(min(y+min(z)),
                        max(y)+max(z)), ylab=expression(y[t]), xlab="t")
# 2te bis R-te Trajektorie
for (i in 2:R) points(y[,i]+z[i], col=i, lwd=lwd)
dev.off()
```

```
# Plotten der Zeitreihen - Darstellung Trajektorien
if (save.pdf) pdf("Traj_RW_lines.pdf")
# Erste Trajektorie
plot(y[,1], cex.lab=cexmu, cex.axis=cexmu, lwd=lwd, type="l", ylim=c(min(y),
  max(y)), ylab=expression(x[t]), xlab="t")
# 2te bis R-te Trajektorie
for (i in 2:R) lines(y[,i], col=i, lwd=lwd)
dev.off()
# ===== Ende =====
```

Listing A.14: ./R_code/12_1_Traj_RW.R

A.11. R-Programm für Monte-Carlo-Simulation im Abschnitt 12.2 zu linearen stochastischen Prozessen und MA-Prozessen

```
# =====12_2_MA2_Realisation =====
# Programm zum Erstellen einer Realisation eines MA(2)-Prozesses
# erstellt von: RT, 2015_09_11

# Festlegen der Stichprobengröße und der Parameter eines MA(2)-Prozesses
n <- 100 # Stichprobengröße
sigma <- 2 # Standardabweichung des Weißen Rauschens
psi <- c(1,0.8,0.6) # MA-Parameter für  $y_t = u_t + 0.8 u_{t-1} + 0.6 u_{t-2}$ 
set.seed(1) # Setze Seed Value für Zufallsgenerator

save.pdf <- 0 # 1=Erstelle PDFs von Graphiken, 0=sonst

# Generieren einer Realisation
u <- rnorm(n+length(psi)-1,sd=sigma) # Generieren von Weißem Rauschen
# Generieren einer Realisation eines MA(2)-Prozesses
y <- filter(u,filter=psi,sides=1,method="convolution")

if (save.pdf) pdf("MA_Realisation.pdf", height=6, width=6)
plot(y,xlab="Zeit",ylab=expression(y[t])) # Plotten einer MA(2)-Zeitreihe
if (save.pdf) dev.off()

# Berechnen der theoretischen Autokorrelationsfunktion für k=0,1,...,10
ARMAacf(ma=psi[2:3],lag.max=10)
# ===== Ende =====
```

Listing A.15: ./R_code/12_2_MA2_Realisation.R

A.12. R-Programm für Monte-Carlo-Simulation im Abschnitt 12.3.1 zu AR(1)-Prozessen

```
# ===== 12_3_AR1_Realisierung.R =====
# Programm zum Erstellen einer Realisation eines AR(1)-Prozesses
# erstellt von: RT, 2015_10_10

# AR(1)-Parameter des DGP
nu <- 1
alpha_1 <- 0.8
sigma2 <- 4
y_0 <- 0

# Länge der Zeitreihe
n <- 500
```

```

set.seed(15) # Lege seed value fest
u           <- rnorm(n,sd=sqrt(sigma2)) # Generiere Gaussian White Noise
y           <- rep(y_0,n)               # Initialisiere Outputvektor

save.pdf    <- 1                       # 1=Erstelle PDFs von Graphiken, 0=sonst
# Festlegung des Arbeitsverzeichnisses (working directory)
# in welchem sich das R-Program und die Daten befinden
WD          <- getwd() # Bestimme Verzeichnis der R-Datei und
setwd(WD)   # setze es als Working Directory

# Generiere AR(1)-Realisation
for (i in (2:n))
{
  y[i]      <- nu + alpha_1 * y[i-1] + u[i]
}
# Plot Realisierung
if (save.pdf) pdf("AR1_Realisierung.pdf", height=6, width=6)
plot(seq(1:n),y,xlab="Zeit",ylab=expression(y[t]),type="l")
if (save.pdf) dev.off()

```

Listing A.16: ./R_code/12_3_AR1_Realisierung.R

A.13. R-Programm für Monte-Carlo-Simulation im Abschnitt 12.3.3 zu $AR(p)$ -Prozessen und mehr

Realisierung, ACF, MA-Parameter, PACF eines $AR(2)$ -Prozesses, siehe Abbildung 12.7

```

# ===== 12_3_AR2_Realisierung.R =====
# Programm zum Erstellen einer Realisation eines AR(2)-Prozesses
# sowie zum Berechnen der ACF, der MA-Darstellung und der Wurzeln
# erstellt von: RT, 2015_29_09

# AR(2)-Parameter des DGP

alpha_0    <- 1
alpha      <- c(-0.5,-0.8)
sigma2     <- 4

# Startwerte
y_start    <- c(0,0)

# Länge der Zeitreihe
n          <- 500

save.pdf   <- 1                       # 1=Erstelle PDFs von Graphiken, 0=sonst
# Festlegung des Arbeitsverzeichnisses (working directory)
# in welchem sich das R-Program und die Daten befinden
WD         <- getwd() # Bestimme Verzeichnis der R-Datei und
setwd(WD)  # setze es als Working Directory

# Check Stabilität des AR(2)-Polynoms
AR2_wurzeln <- polyroot(c(1,-alpha))
abs(AR2_wurzeln)

set.seed(15) # Lege seed value fest
u           <- rnorm(n,sd=sqrt(sigma2)) # Generiere Gaussian White Noise
y           <- rep(NA,n)               # Initialisiere Outputvektor
y[1:length(y_start)] <- y_start       # Setze Startwerte ein

# Generiere AR(2)-Realisation

```

```

for (i in ((length(alpha)+1):n))
{
  y[i]      <- alpha_0 + alpha %*% y[(i-1):(i-2)] + u[i]
}

# Generate plots
if (save.pdf) pdf("AR2_Realisierung.pdf", height=6, width=6)
split.screen(c(2,2))

# Plot Realisation
screen(1)
plot(seq(1:n),y,xlab="Zeit",ylab=expression(y[t]),type="l")

# Plot theoretische ACF
screen(2)
plot(ARMAacf(ar=alpha,lag.max=20),type="h",ylab="ACF",xlab="Lags")

# Plot theoretische PACF
screen(3)
plot(ARMAacf(ar=alpha,lag.max=20,pacf=TRUE),type="h",ylab="PACF",xlab="Lags")

# Plot MA-Parameter des invertierten AR(2)-Prozesses
screen(4)
plot((1:20),ARMAtoMA(ar=alpha,lag.max=20),type="h",ylab="MA-Parameter",xlab="Lags")
if (save.pdf) dev.off()

if (save.pdf) pdf("AR2_Realisierung_ACF.pdf", height=6, width=6)
acf(y,lag.max=20,type="correlation")
if (save.pdf) dev.off()
# ===== Ende =====

```

Listing A.17: ./R_code/12_3_AR2_Realisierung.R

A.14. R-Programm für Schätzung der Autokorrelationsfunktion im Abschnitt 12.4 zur Schätzung erster und zweiter Momente im Fall stationärer Prozesse

Geschätzte Autokorrelationsfunktion einer Realisation von Weißem Rauschen, siehe Abbildung 12.9

```

# ===== 12_4_WN_ACF_Est =====
# Programm zum Schätzen der Autokorrelationsfunktion einer Realisation eines
# Gausschen White Noise Prozesses mit n=100 Beobachtungen
# erstellt von: RT, 2015_18_10

# Varianz
sigma2 <- 4
# Länge der Zeitreihe
n <- 100
save.pdf <- 1 # 1=Erstelle PDFs von Graphiken, 0=sonst

# Setze seed value
set.seed(15)
# Generieren von Gausschem Weißem Rauschen
y <- rnorm(n,sd=sqrt(sigma2))

# Plotten der geschätzten Autokorrelationsfunktion
# mit 95%-Konfidenzintervallen
if (save.pdf) pdf("ACF_WN_Est.pdf", height=6, width=6)
acf(y,lag.max=20,type="correlation")

```

```
if (save.pdf) dev.off()
```

Listing A.18: ./R_code/12_4_WN_ACF_Est.R

A.15. R-Programm für die Simulation und Schätzung von AR(1)-Prozessen im Abschnitt 13.5 zur KQ-Schätzung von dynamischen linearen Regressionsmodellen

Generierung und Schätzung eines Prozesses

```
# ===== 13_5_KQ_AR1.R =====
# Programm zum Generieren und KQ-Schätzen eines AR(1)-Modells
# erstellt von : RT, 2011_01_19

graphics.off()      # Schließe alle Graphikfenster

# Setze Parameter des Modells und der Monte-Carlo-Simulation
set.seed(42)        # Randomseed
N      <- 50         # Stichprobengröße

beta   <- c(2,0.1)  # Parametervektor
sigma  <- 2         # Standardabweichung des Fehlers
y0     <- 0         # Startwert des AR(1)-Prozesses

# Generieren einer Realisation eines AR(1)-Prozesses
u <- rnorm(N,mean=0,sd=sigma) # Ziehen von u
y <- rep(1,N)*y0
for (t in (2:N))
{
  y[t] <- beta[1] + y[t-1] * beta[2] + u[t] # Berechnen von y_t
}

# Plot der Zeitreihe
plot(y,xlab="Zeit",ylab="y",type="l")

# Scatterplot
plot(y[1:(N-1)],y[2:N])

# Berechnen des KQ-Schätzers
ols <- lm(y[2:N]~1+y[1:(N-1)]) # Beachte x=y_{t-1}. Deshalb y_t von t=2,...,N
summary(ols)
# ===== Ende =====
```

Listing A.19: ./R_code/13_5_KQ_AR1.R

Monte-Carlo-Simulation

```
# ===== 13_5_MC_KQ_AR1.R =====
# Programm für Monte-Carlo-Simulation
# zum Bestimmen der Verzerrung des KQ-Schätzers im AR(1)-Modell
# erstellt von : RT, 2010_11_25

graphics.off()      # Schließe alle Graphikfenster

# Setze Parameter des Modells und der Monte-Carlo-Simulation
set.seed(42)        # Randomseed
N      <- 50         # Stichprobengröße
R      <- 1000       # Zahl der Replikationen
```

```
beta <- c(1,0.9) # Parametervektor
sigma <- 2      # Standardabweichung des Fehlers
y0    <- 1      # Startwert des AR(1)-Prozesses

# Bilden einer Schleife
beta_hat_store <- matrix(0,nrow=R,ncol=length(beta))
                # Initialisiere Matrix zum Abspeichern der KQ-Schätzungen
                # für jede Realisation
for (r in (1:R))
{
  # Generieren einer Realisation eines AR(1)-Prozesses
  u <- rnorm(N,mean=0,sd=sigma) # Ziehen von u
  y <- rep(1,N)*y0
  for (t in (2:N))
  {
    y[t] <- beta[1] + y[t-1] * beta[2] + u[t] # Berechnen von y_t
  }
  # Berechnen des KQ-Schätzers
  ols <- lm(y[2:N]~y[1:(N-1)]) # Beachte x=y_{t-1}. Deshalb y_t von t=2,...,N

  # Speichern der Parameterschätzung
  beta_hat_store[r,] <- coef(ols)
}

# Berechnen der Mittelwerte der Parameterschätzungen
colMeans(beta_hat_store)

# Erstellen von Histogrammen
par(mfrow=c(1,2)) # Zeichne zwei Plots in ein Graphikfenster
hist(beta_hat_store[,1],breaks=sqrt(R))
hist(beta_hat_store[,2],breaks=sqrt(R))

# ===== Ende =====
```

Listing A.20: ./R_code/13_5_MC_KQ_AR1.R

B. R Befehle für die Regressionsanalyse

B.1. Übersicht über verfügbare Befehle

Benötigte R-Pakete: `stats` normalerweise geladen, `car`, `lmtest`, `moments`, `sandwich`.

Durchführen einer linearen Regression:

```
model_kq <- lm()
```

erstellt Regressionsobjekt, das Grundlage für folgende Befehle ist:

R-Befehl	Funktionsbeschreibung	R-Paket
Schätz- und Prognoseoutput		
<code>print()</code>	einfaches gedrucktes Display	
<code>summary()</code>	Standard Regressionsoutput	
<code>coef()</code>	(oder <code>coefficients()</code>) extrahiert geschätzte Regressionsparameter	
<code>residuals()</code>	(oder <code>resid()</code>) extrahiert Residuen	
<code>fitted()</code>	(oder <code>fitted.values()</code>) extrahiert angepasste/gefttete Werte	
<code>anova()</code>	Vergleich von geschachtelten Modellen (nested models)	
<code>predict()</code>	Vorhersagen für neue Regressionswerte	
<code>confint()</code>	Konfidenzintervalle für Regressionskoeffizienten	
<code>confidenceEllipse()</code>	Konfidenzintervalle für Regressionskoeffizienten	
<code>deviance()</code>	Residuenquadratsumme (SSR)	
<code>vcov()</code>	(geschätzte) Varianz-Kovarianzmatrix der Parameterschätzer	
<code>logLik()</code>	Log-Likelihood (unter der Annahme normalverteilter Fehler)	<code>car</code>
Testen		
<code>hccm()</code>	Heteroskedastie-korrigierte Varianz-Kovarianzmatrix der Parameterschätzer; mit <code>type="hc0"</code> White-Varianz-Kovarianzmatrix (14.19)	<code>car</code>
<code>coefTest()</code>	Standard-Regressionoutput, ggf. mit heteroskedastie-robusten Standardfehlern	<code>lmtest</code>
<code>linearHypothesis()</code>	F -Test <code>test=c("F")</code> oder (asymptotischer) χ^2 -Test <code>test=c("Chisq")</code> ; mit <code>white.adjust=c(FALSE, TRUE, "hc0")</code>	<code>car</code>
<code>lrtest()</code>	White-heteroskedastierobuste-Varianz-kovarianzmatrix Likelihoodratio-Test, siehe Weiterführende Fragen der Ökonometrie oder Fortgeschrittene Ökonometrie	<code>lmtest</code>
<code>waldtest()</code>	Wald-Test, siehe Weiterführende Fragen der Ökonometrie oder Fortgeschrittene Ökonometrie	<code>lmtest</code>

Fortsetzung auf der nächsten Seite.

R-Befehl	Funktionsbeschreibung	R-Paket
Modellspezifikation		
AIC()	Informationskriterien einschließlich AIC, BIC/SC (unter der Annahme normalverteilter Fehler) - Beachte: Im Gegensatz zu EViews wird die geschätzte Parametervarianz als Parameter mitgezählt und nicht durch die Zahl der Beobachtungen dividiert, siehe Abschnitt 10.1	eigenes R-Programm, siehe Abschnitt B.2
SelectCritEViews()	Informationskriterien a la EViews, siehe Abschnitt 10.1	lmtest
encomptest()	Encompassing-Test zum Testen nicht geschachtelter Regressionsmodelle, siehe Abschnitt 10.2 oder Davidson & MacKinnon (2004, Section 15.3)	lmtest
jtest()	J-Test zum Testen nicht geschachtelter Regressionsmodelle, siehe Abschnitt 10.2 oder Davidson & MacKinnon (2004, Section 15.3)	lmtest
Modelldiagnose		
plot()	Graphiken zur Modellüberprüfung	lmtest
resettetest()	RESET-Test zum Testen der funktionalen Form, siehe Abschnitt 15.3	moments
jarque.test()	Lomnicki-Jarque-Bera-Test zum Überprüfen normalverteilter Fehler, siehe Abschnitt 15.4	lmtest
bptest()	Breusch-Pagan-Test zum Testen auf Vorliegen von heteroskedastischen Fehlern, siehe Abschnitt 15.2.1	eigenes R-Programm, siehe Abschnitt B.2
whitetest()	White-Test zum Testen auf Vorliegen von heteroskedastischen Fehlern, siehe Abschnitt 15.2.2	

Befehle für Graphiken/Plots:

```
# INFOS FÜR GRAPHIKEN

# Speichern als .eps-Datei (Dateiname, Größe der Graphik)
# postscript("pdf_std_normal.eps", height=4, width=6, horizontal=FALSE)
# oder Speichern als .pdf
# pdf("pdf_std_normal.pdf", height=4, width=6)
# windows()          # öffnet ein neues Graphikfenster

# split.screen(c(2,1)) # teilt ein Graphikfenster
# screen(1)            # Ansteuern von Fenster 1
# dev.off()           # Schließt geöffnetes Graphikfenster
# close.screen(all=TRUE)# Schließt alle Fenster

# Parameter für Graphiken: (optional)
#   las=1: Achsenskalierung waagrecht
#   mai: Breite der Ränder (unten, links, oben, rechts)
#   mgp: Lage von Achsen, Achsenskalierung und Achsenbeschriftung
```

B.2. Eigene R-Programme

```
# ----- SelectCritEViews -----
# function to compute model selection criteria for linear regressions as 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(sellist)
}
# -----
```

```
# ----- whitetest -----
# function to conduct White test including cross terms
# RW, 2011_01_26

whitetest <- function(model){

# Daten aus model extrahieren
dat <- model$model
dat$resid_sq <- model$resid^2

# Formel für die Hilfsregression erstellen
regr <- attr(model$terms, "term.labels")
form <- as.formula(paste("resid_sq~(",paste(regr,collapse="+"),")^2+",
  paste("I(",regr,"^2)",collapse="+")))

# Hilfsregression schätzen
test_eq <- lm(form,data=dat)

# Overall F-Test
fstat <- summary(test_eq)$fstatistic

# Ergebnis berechnen und ausgeben
result1 <- c(fstat[1],fstat[2],fstat[3],pf(fstat[1],fstat[2],
```

```
fstat[3,lower.tail=FALSE))
names(result1) <- c("F-Statistic","df1","df2","P-Value")
result <- list(result1,test_eq)
return(result)
}
# -----
```

Mehr:

- Kurs **Programmieren in R**
- Kleiber & Zeileis (2008)
- Übersicht über verfügbare Pakete in R

<http://cran.r-project.org/web/views/>

C. Daten für die Schätzung der Gravitationsgleichung

Abschnitt entspricht Abschnitt 10.4 in Kursmaterial zu [Einführung in die Ökonometrie](#).

Legende für die Daten in `importe_ger_2004_ebrd.txt`

• Länder und Ländercodes

1	ALB	Albanien	26	ISL	Island
2	ARM	Armenien	27	ITA	Italien
3	AUT	Österreich	28	JPN	Japan
4	AZE	Aserbaidshan	29	KAZ	Kasachstan
5	BEL	Belgien und Luxemburg	30	KGZ	Kirgisistan
6	BGR	Bulgarien	31	LTU	Litauen
7	BIH	Bosnien und Herzegowina	32	LVA	Lettland
8	BLR	Weißrussland	33	MDA	Republik Moldau
9	CAN	Kanada	34	MKD	EYR Mazedonien
10	CHE	Schweiz	35	MLT	Malta
11	CYP	Zypern	36	NLD	Niederlande
12	CZE	Tschechische Republik	37	NOR	Norwegen
13	DNK	Dänemark	38	POL	Polen
14	ESP	Spanien	39	PRT	Portugal
15	EST	Estland	40	ROM	Rumänien
16	FIN	Finnland	41	RUS	Russland
17	FRA	Frankreich	42	SVK	Slowakei
18	GBR	Vereinigtes Königreich	43	SVN	Slowenien
19	GEO	Georgien	44	SWE	Schweden
20	GER	Deutschland	45	TJK	Tadschikistan
21	GRC	Griechenland	46	TKM	Turkmenistan
22	HKG	Hong Kong	47	TUR	Türkei
23	HRV	Kroatien	48	UKR	Ukraine
24	HUN	Ungarn	49	USA	Vereinigte Staaten
25	IRL	Irland	50	UZB	Usbekistan

Länder, die nur als Herkunftsländer auftauchen:

BIH	Bosnien und Herzegowina	CHN	China	KOR	Südkorea
TJK	Tadschikistan	HKG	Hong Kong	TWN	Taiwan
UZB	Usbekistan	JPN	Japan	THA	Thailand

R code:

```
setwd('d:/..') # setze working directory ein
daten <- read.table("importe_ger_2004_ebrd.txt", header=TRUE, sep="\t")
```

```
attach(daten)
cbind(matrix(iso_o,50,1),matrix(d_o,50,1)) # Abkürzungen der Ländernamen
                                           # und Handelsrichtungen
```

- **Endogene Variable:**

- TRADE_0_D_O: 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:

WDI_GDPUSDCR_O				Weltbank - World Development Indicators
WDI_GDPPCUSD CR_O				Weltbank - World Development Indicators
WEO_GDPFCR_O				IWF - World Economic Outlook database
WEO_GDPPCCR_O				IWF - World Economic Outlook database
WEO_POP_O				IWF - World Economic Outlook database
CEPII_AREA_O				CEPII
CEPII_COL45				CEPII
CEPII_COL45_REV				CEPII
CEPII_COLONY				CEPII
CEPII_COMCOL				CEPII
CEPII_COMCOL_REV				CEPII
CEPII_COMLANG_ETHNO				CEPII
CEPII_COMLANG_ETHNO_REV				CEPII
CEPII_COMLANG_OFF				CEPII
CEPII_CONTIG				CEPII
CEPII_DISINT_O				CEPII
CEPII_DIST				CEPII
CEPII_DISTCAP				CEPII
CEPII_DISTW				CEPII
CEPII_DISTWCES				CEPII
CEPII_LAT_O				CEPII
CEPII_LON_O				CEPII
CEPII_SMCTRY_REV				CEPII, revidiert
ISO_O				CEPII
EBRD_TFES_O				EBRD
Herkunftsland (O-Land)				
	Herkunftsland BIP-Daten; in laufenden US-Dollars			
	Herkunftsland BIP-pro Kopf-Daten; in laufenden US-Dollars			
	Ziel- und Herkunftsland BIP-Daten; in laufenden US-Dollars			
	Ziel- und Herkunftsland BIP-pro-Kopf-Daten; in laufenden US-Dollars			
	Herkunftsland Bevölkerungsdaten			
	Fläche des Herkunftslandes in km ²			
	dummy; d- und o-Land besaßen eine Kolonialbeziehung nach 1945			
	dummy; revidiert durch "Expertenwissen"			
	dummy; d- und o-Land hatten jemals eine Kolonialbeziehung			
	dummy; d- und o-Land teilen einen gemeinsamen Kolonialherren nach 1945			
	dummy; revidiert durch "Expertenwissen"			
	dummy; d- und o-Land teilen eine gemeinsame Sprache			
	wird von mindestens 9% der Bevölkerung gesprochen			
	dummy; d- und o-Land teilen gemeinsame Amtssprache			
	dummy; d- und o-Land sind Nachbarstaaten			
	inländische Entfernung in Herkunftsland			
	geodätische Entfernung zwischen d- und o-Land			
	Entfernung zwischen d- und o-Land basierend auf deren Hauptstädten			
	$0.67 \sqrt{\text{Fläche}/\pi}$			
	gewichtete Entfernungen, für Details siehe CEPII			
	gewichtete Entfernungen, für Details siehe CEPII			
	Breitengrad der Stadt			
	Längengrad der Stadt			
	dummy; d- und o-Land waren/sind dasselbe Land			
	ISO-Code mit drei Buchstaben für Herkunftsland			
	EBRD-Maß für den Grad der Liberalisierung der Handels- und Zahlungsströme des o-Landes			
Zielland (D-Land)				
	Zielland BIP-Daten; in laufenden US-Dollars			
	Zielland BIP-pro Kopf-Daten; in laufenden US-Dollars			
	Ziel- und Herkunftsland BIP-Daten; in laufenden US-Dollars			
	Ziel- und Herkunftsland BIP-pro-Kopf-Daten; in laufenden US-Dollars			
	Zielland Bevölkerungsdaten			
WDI_GDPUSDCR_D				Weltbank - World Development Indicators
WDI_GDPPCUSD CR_D				Weltbank - World Development Indicators
WEO_GDPFCR_D				IWF - World Economic Outlook database
WEO_GDPPCCR_D				IWF - World Economic Outlook database
WEO_POP_D				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 fortgeschrittener Industriestaaten erreicht hat, d.h., von OECD Staaten. Diese Variable ist per Konstruktion qualitativer und nicht kardinaler Art.

- **Dank:** an Richard Frensch, [Osteuropa-Institut, Regensburg](#), der die Daten zur Verfügung gestellt hat.
- **Internetseiten** [CEPII](#)

D. Grundbegriffe Mengen

Definition

Fischer (vgl. 2014, S. 43) Eine Menge G zusammen mit einer Verknüpfung $*$ heißt **Gruppe**, wenn folgende Axiome erfüllt sind:

1. Assoziativgesetz: $(a * b) * c = a * (b * c)$ für alle $a, b, c \in G$
2. Existenz eines neutralen Elements $e \in G$ mit den folgenden Eigenschaften:
 - a) $e * a = a$ für alle $a \in G$.
 - b) Existenz eines inversen Elements: Zu jedem $a \in G$ gibt es ein $a' \in G$ mit $a' * a = e$.

Die Gruppe heißt **abelsch** oder **kommutativ**, falls außerdem $a * b = b * a$ für alle $a, b \in G$.

Definition

Fischer (vgl. 2014, S. 54) Eine Menge R zusammen mit zwei Verknüpfungen

$$\begin{aligned} + : R \times R &\rightarrow R, & (a, b) &\rightarrow a + b, & \text{und} \\ \cdot : R \times R &\rightarrow R, & (a, b) &\rightarrow a \cdot b, \end{aligned}$$

heißt **Ring**, wenn folgendes gilt:

1. R zusammen mit der Addition $+$ ist eine abelsche Gruppe (= kommutative Gruppe).
2. Die Multiplikation \cdot ist assoziativ.
3. Es gelten die Distributivitätsgesetze, d.h. für alle $a, b, c \in R$ gilt

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{und} \quad (a + b) \cdot c = a \cdot c + b \cdot c.$$

Definition

Fischer (vgl. 2014, S. 56) Eine Menge K zusammen mit zwei Verknüpfungen

$$\begin{aligned} + : R \times R &\rightarrow R, & (a, b) &\rightarrow a + b, & \text{und} \\ \cdot : R \times R &\rightarrow R, & (a, b) &\rightarrow a \cdot b, \end{aligned}$$

heißt **Körper**, wenn folgendes gilt:

1. K zusammen mit der Addition $+$ ist eine abelsche Gruppe.

2. Existenz einer Untermenge für die die Multiplikation auch eine abelsche Gruppe ist:
Bezeichnet $K^* := K \setminus \{0\}$, so gilt für $a, b \in K^*$ auch $a \cdot b \in K^*$, und K^* zusammen mit der so erhaltenen Multiplikation ist eine abelsche Gruppe.
3. Es gelten die Distributivitätsgesetze, d.h. für alle $a, b, c \in K$ ist

$$a \cdot (b + c) = a \cdot b + a \cdot c \quad \text{und} \quad (a + b) \cdot c = a \cdot c + b \cdot c.$$

Literaturverzeichnis

- Anderson, J. E. & Wincoop, E. v. (2003), 'Gravity with gravitas: A solution to the border puzzle', *The American Economic Review* **93**, 170–192. [149](#)
- Angrist, J. & Pischke, J. (2009), *Mostly harmless econometrics. An Empiricist's Companion*, Princeton University Press.
- Bauwens, L., Boswijk, H. P. & Urbain, J.-P. (2006), 'Causality and exogeneity in econometrics', *Journal of Econometrics* **132**, 305 – 309. [301](#)
- Brockwell, P. J. & Davis, R. A. (1991), *Time Series: Theory and Methods*, 2. edn, Springer, New York, NY. [265](#), [266](#), [283](#), [284](#), [285](#), [287](#), [288](#)
- Cameron, A. & Trivedi, P. (2005), *Microeconometrics*, Cambridge University Press.
- Casella, G. & Berger, R. L. (2002), *Statistical Inference*, 2nd edn, Duxbury - Thomson. [38](#), [71](#), [84](#)
- Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press. [121](#), [124](#)
- Davidson, J. (2000), *Econometric Theory*, Blackwell Publishers. [72](#), [83](#), [102](#), [121](#), [217](#), [225](#), [257](#), [286](#), [291](#), [293](#), [294](#), [297](#), [298](#), [302](#), [306](#)
- Davidson, R. & MacKinnon, J. (1993), *Estimation and Inference in Econometrics.*, Oxford University Press.
URL: <http://www.oup.com/uk/catalogue/?ci=9780195060119>
- Davidson, R. & MacKinnon, J. G. (2004), *Econometric Theory and Methods*, Oxford University Press, Oxford. [26](#), [38](#), [55](#), [100](#), [103](#), [105](#), [109](#), [110](#), [116](#), [124](#), [139](#), [140](#), [145](#), [150](#), [162](#), [164](#), [165](#), [167](#), [168](#), [169](#), [193](#), [209](#), [213](#), [214](#), [239](#), [241](#), [242](#), [243](#), [245](#), [315](#), [316](#), [325](#), [363](#)
- Engle, R., Hendry, D. & Richard, J.-F. (1983), 'Exogeneity', *Econometrica* **51**, 277–304. [294](#), [296](#)
- Fahrmeier, L., Künstler, R., Pigeot, I. & Tutz, G. (2004), *Statistik*, Springer. [39](#)
- Fahrmeir, L., Heumann, C., Künstler, R., Pigeot, I. & Tutz, G. (2016), *Statistik: Der Weg zur Datenanalyse*, 8 edn, Springer. [38](#)
- Fischer, G. (2010), *Lineare Algebra*, 17 edn, Vieweg + Teubner. [6](#), [26](#)
- Fischer, G. (2014), *Lineare Algebra*, 18 edn, Springer Spektrum.
URL: <http://dx.doi.org/10.1007/978-3-658-03945-5> [3](#), [370](#)

- Fratianni, M. (2007), The gravity equation in international trade, Technical report, Dipartimento di Economia, Universita Politecnica delle Marche. [149](#)
- Gentle, J. E. (2007), *Matrix Algebra. Theory, Computations, and Applications in Statistics*, Springer Texts in Statistics, Springer.
URL: <http://www.springerlink.com/content/x4rj03/> [3](#), [9](#), [10](#), [11](#), [26](#), [310](#)
- Granger, C. (1969), ‘Investigating causal relations by econometric models and cross-spectral methods’, *Econometrica* **37**, 424 – 438. [297](#)
- Greene, W. (2008), *Econometric Analysis*, 6 edn, Pearson.
URL: <http://www.pearsonhighered.com/educator/academic/product/0,3110,0135132452,00.html>
[168](#)
- Greene, W. (2012), *Econometric Analysis*, 7 edn, Pearson.
- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press. [265](#), [274](#), [278](#)
- Hansen, B. E. (2015), *Econometrics*.
- Hassler, U. (2007), *Stochastische Integration und Zeitreihenmodellierung*, Springer, Berlin, Heidelberg. [257](#), [258](#)
- Hayashi, F. (2000), *Econometrics*, Princeton University Press, Princeton, NJ [u.a.]. [263](#), [265](#)
- Hendry, D. F. (1995), *Dynamic Econometrics*, Oxford University Press. [124](#), [297](#), [298](#), [302](#)
- Horowitz, J. (2001), The bootstrap, in J. Heckman & E. Leamer, eds, ‘Handbook of Econometrics’, Vol. 5, North-Holland. [242](#)
- Horowitz, J. (2003), ‘The bootstrap in econometrics’, *Statistical Science* **18**, 211–218. [242](#)
- Kirchgässner, G. & Wolters, J. (2008), *Introduction To Modern Time Series Analysis*, Springer, Berlin, [u.a.]. [278](#), [279](#)
- Kirchgässner, G., Wolters, J. & Hassler, U. (2013), *Introduction To Modern Time Series Analysis*, 2nd. ed. edn, Springer, Berlin, [u.a.]. [265](#)
- Kleiber, C. & Zeileis, A. (2008), *Applied Econometrics with R*, Springer. [365](#)
- Li, Q. & Racine, J. (2007), *Nonparametric Econometrics*, Princeton University Press. [105](#)
- Lucas, R. (1976), Econometric policy evaluation: A critique, in K. Brunner & A. Meltzer, eds, ‘The Phillips Curve and Labor Markets’, Vol. Vol. 1 of *Carnegie-Rochester Conferences on Public Policy*, North-Holland, Amsterdam, pp. 19 – 46. [302](#)
- Lütkepohl, H. (1996), *Handbook of Matrices*, John Wiley & Sons, Chichester. [3](#), [287](#), [310](#)
- Lütkepohl, H. (2004), Vector autoregressive and vector error correction models, in H. Lütkepohl & M. Krätzig, eds, ‘Applied Time Series Econometrics’, Cambridge University Press, Cambridge, chapter 3, pp. 86–158. [297](#)

- Lütkepohl, H. & Kraetzig, M. (2008), *Applied Time Series Econometrics*, Cambridge University Press. 265
- Mikosch, T. (1998), *Elementary Stochastic Calculus*, World Scientific Publishing, Singapore. 257
- Neusser, K. (2006), *Zeitreihenanalyse in den Wirtschaftswissenschaften*, Teubner, Wiesbaden. 265, 266
- Neusser, K. (2009), *Zeitreihenanalyse in den Wirtschaftswissenschaften*, 2. edn, Teubner, Wiesbaden. 265, 274, 275, 276
- Peracchi, F. (2001), *Econometrics*, John Wiley and Sons.
URL: <http://www.wiley-vch.de/publish/dt/books/bySubjectEC00/ISBN0-471-98764-6/?sID=he2l84vhvc6o6e4f1mc7i17k05>
- Robinson, P. M., ed. (2003), *Time Series with Long Memory*, Oxford University Press. 281
- Ruud, P. (2000), *An Introduction to Classical Econometric Theory*, Oxford University Press.
URL: <http://www.oup.com/uk/catalogue/?ci=9780195111644> 169
- Schmidt, K. & Trenkler, G. (2006), *Einführung in die Moderne Matrix-Algebra. Mit Anwendungen in der Statistik*, Springer. 24, 25, 26, 29, 228
- Schmidt, K. & Trenkler, G. (2015), *Einführung in die Moderne Matrix-Algebra. Mit Anwendungen in der Statistik*, 3 edn, Springer. 3
- Steland, A. (2010), *Basiswissen Statistik : Kompaktkurs für Anwender aus Wirtschaft, Informatik und Technik*, 2., komplett überarb. und erw. Aufl. edn, Springer, Berlin ; Heidelberg : Springer. 46
- Steland, A. (2013), *Basiswissen Statistik : Kompaktkurs für Anwender aus Wirtschaft, Informatik und Technik*, 3., überarb. und erw. Aufl. edn, Springer, Berlin ; Heidelberg : Springer.
URL: <http://link.springer.com/book/10.1007/978-3-642-37201-8> 38
- Steland, A. (2016), *Basiswissen Statistik : Kompaktkurs für Anwender aus Wirtschaft, Informatik und Technik*, 4 edn, Springer, Berlin ; Heidelberg : Springer.
URL: <http://link.springer.com/book/10.1007/978-3-642-37201-8>
- Stock, J. H. & Watson, M. W. (2007), *Introduction to Econometrics*, 2nd. edn, Pearson, Boston, Mass. 86, 87, 96
- Stock, J. H. & Watson, M. W. (2012), *Introduction to Econometrics*, 3rd. edn, Pearson, Boston, Mass.
- Tschernig, R. (1994), *Wechselkurse, Unsicherheit und Long Memory*, Physica-Verlag, Heidelberg.
URL: <http://epub.uni-regensburg.de/6928/> 281
- Vaart, A. v. d. (1998), *Asymptotic Statistics*, Cambridge series in statistical and probabilistic mathematics, Cambridge University Press. 82

Verbeek, M. (2012), *A guide to modern econometrics*, Wiley, Chichester.

Wooldridge, J. M. (2009), *Introductory Econometrics. A Modern Approach*, 4th edn, Thomson South-Western, Mason. [61](#), [64](#), [65](#), [109](#), [133](#), [136](#), [139](#), [140](#), [161](#), [168](#), [176](#), [177](#), [186](#), [190](#), [191](#), [193](#), [205](#), [214](#)

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press. [121](#)