

Fortgeschrittene Ökonometrie

Folien

Rolf Tschernig

Universität Regensburg

Stand: 30. April 2022¹

¹Diese Version entspricht bis auf einige Korrekturen und dem Aktualisieren von Links und Verweisen auf das Handout für Methoden der Ökonometrie den Folien für Fortgeschrittene Ökonometrie vom 26. Juli 2011 bzw. im Seitenkopf angegebenem Datum 12.04.2012. Außerdem wurde das Kapitel zur Maximum-Likelihood-Schätzung nach vorne gerückt. Ich danke sehr herzlich Roland Jucknewitz für seine Zuarbeit, wichtigen Korrekturen und substantiellen Verbesserungsvorschläge.

© Rolf Tschernig. 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, Folien für Fortgeschrittene Ökonometrie, Universität Regensburg, 30. April 2022. Downloaded am [Tag Monat Jahr].

Inhaltsverzeichnis

1. Wiederholung und Motivation	1
1.1. Wiederholung	2
1.2. Beispiele	12
1.2.1. Beispiel IM — Analyse von Importen nach Deutschland	12
1.2.2. Beispiel CO ₂ : Co ₂ -Ausstoß und BIP	13
1.3. Allgemeinere Modelle in diesem Kurs	16
2. Nichtlineare Regression	18
2.1. Momentenschätzer für nichtlineare Modelle	28
2.2. Nichtlinearer Kleinst-Quadrate-Schätzer	42

2.3.	Numerische Optimierung	49
2.3.1.	Newton's Methode/Newton-Raphson-Algorithmus	49
2.3.2.	Quasi-Newton-Methoden/Gradientenverfahren	53
2.3.3.	Gauss-Newton Methode	56
2.3.4.	Wichtige Anmerkungen	60
2.3.5.	Beispiel IM - Fortsetzung von Abschnitt 1.2.1	63
2.3.6.	Beispiel CO ₂ - Fortsetzung von Abschnitt 1.2.2	70
2.4.	Hypothesentests	74
2.5.	Smooth Transition (STR) Modelle	83
2.6.	Nichtlinearitätstests für STR-Modelle	88
3.	Maximum-Likelihood-Schätzung	94
3.1.	Einführendes Beispiel	94
3.2.	ML-Schätzung im Falle stetiger Zufallsvariablen	98
3.3.	ML-Schätzung des normalen linearen Regressionsmodells	101
3.4.	Asymptotische Verteilung des ML-Schätzers	104
3.4.1.	Identifikation bei ML-Schätzung	104
3.4.2.	Konsistenz	106

3.4.3.	Asymptotische Normalverteilung	112
3.5.	Numerische Optimierung	130
3.6.	Hypothesentests	133
3.6.1.	Likelihood-Quotienten-Test (LR-Test)	136
3.6.2.	Wald-Test	139
3.6.3.	LM oder Score-Test	142
4.	Verallgemeinerter KQ-Schätzer und Anwendungen	152
4.1.	Verallgemeinerter Kleinst-Quadrate-Schätzer	156
4.2.	Feasible GLS	164
4.2.1.	Modellierung heteroskedastischer Fehler	166
4.2.2.	Modelle mit autokorrelierten Fehlern	168
4.3.	Heteroskedastie-robuste Standardfehler	169
4.4.	Empirische Analyse von Handelsströmen: Teil 3	174
4.5.	Paneldaten	179
4.6.	Multivariate Regression	199

5. Instrumentvariablenschätzung	200
5.1. Anwendungsbeispiele für den IV-Schätzer	202
5.1.1. Modelle mit Fehlern in den Variablen (error in variables)	203
5.1.2. Simultane Gleichungsmodelle	206
5.2. Der einfache IV-Schätzer	210
5.3. Der verallgemeinerte IV-Schätzer	228
5.4. Asymptotische Tests	237
5.5. Testen von überidentifizierenden Restriktionen	238
6. Generalized Method of Moments (GMM)	240
6.1. Übersicht	240
6.2. GMM-Schätzer für lineare Modelle	247
6.3. Effizienter GMM-Schätzer für lineare Modelle	254
6.4. Vollständig effizienter GMM-Schätzer	258
6.5. HAC-Kovarianzschätzer	269
6.6. Tests auf Basis der GMM-Zielfunktion	283

A. Mehr zu Wahrscheinlichkeitstheorie	296
A.1. L^p -Raum und L^p -Norm	296
A.2. Konvergenz von Folgen von Zufallsvariablen	298
A.3. Wahrscheinlichkeitsungleichungen	302

Organisation

Kontakt

Prof. Dr. Rolf Tschernig

Gebäude RW(L), 5. Stock, Raum 514

Universitätsstr. 31, 93040 Regensburg

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

Email: rolf.tschernig@ur.de

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

Zeiten, Räume und Kursleiter

siehe Kurshomepage:

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

Voraussetzungen

Inhalt des Pflichtmoduls **Methoden der Ökonometrie**.

Prüfungsmodalitäten und Notengebung

Der Kurs **Fortgeschrittene Ökonometrie** ist eine Pflichtveranstaltung der MA-Schwerpunktmodulgruppe **Empirische Wirtschaftsforschung** und ist der Inhalt des Moduls **Empirische Wirtschaftsforschung 1**. Weitere Informationen enthält

der Modulkatalog:

<http://wiwi-service.uni-regensburg.de/studyprogram/public.php>

Aktuelle Regelungen finden Sie auf der Kurshomepage: <https://www.uni-regensburg.de/wirtschaftswissenschaften/vwl-tschernig/lehre/master/fortgeschrittene-oekonometrie/index.html>

Software

Im Kurs wird die freie Software **R** (<http://www.r-project.org/>) verwendet. Eine Übersicht zu Software für ökonometrische Analysen findet sich am Anfang des Handouts zum Kurs **Methoden der Ökonometrie**

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

Pflichtliteratur

- Davidson und MacKinnon (2004). *Econometric Theory and Methods*, Oxford University Press.
Stoff: Chapters 4.6 - 4.7, 5.3, 5.5-5.6, 6, 7.6-7.10, 8, 9, 10, 15.
Voraussetzung: Chapters 1, 2.1-2.5, 3, 4-4.5, 5-5.2, 5.4, 7 - 7.5.

Ergänzungsliteratur

- Davidson (2000). *Econometric Theory*, Blackwell Publishers.
- Greene (2018). *Econometric Analysis*, Pearson.
- Hayashi (2000). *Econometrics*, Princeton University Press.
- Kleiber und Zeileis (2008). *Applied Econometrics with R*, Springer.
- Wooldridge (2010). *Econometric Analysis of Cross Section and Panel Data*, The MIT Press.

Ergänzungsliteratur, speziell für Zeitreihen

- Davidson (2000)
- Hassler (2007) (Online-Zugang im UR-Netz)
- Lütkepohl (2005)
- Lütkepohl und Krätzig (2004); Lütkepohl (2008) (Online-Zugang im UR-Netz; Version 2008 is Reprint)
- Kirchgässner und Wolters (2008) (Online-Zugang im UR-Netz)
- Kirchgässner, Wolters, und Hassler (2013)

weitere Literatur wird bei Bedarf bekanntgegeben.

1. Wiederholung und Motivation

Ökonometrische Modelle, die in **Methoden der Ökonometrie** betrachtet wurden:

- **Multiples lineares Regressionsmodell**
- **Dynamisches Regressionsmodell**

Betrachtete Schätzmethoden:

- KQ-Schätzer
- GLS-Schätzer
- FGLS-Schätzer
- KQ-Schätzer mit heteroskedastierobusten Standardfehlern

1.1. Wiederholung

- **Multiple lineares Regressionsmodell** mit **streng exogenen Regressoren**

$$y = \mathbf{X}\beta + \mathbf{u} \quad (1.1)$$

– Annahmen für **Unverzerrtheit** und **Schätzvarianz** des KQ-/GLS-Schätzers

- * **(B1) Korrekt spezifiziertes Modell**

Der DGP ist für $\beta = \beta_0$ im multiplen linearen Regressionsmodell (1.1) enthalten.

- * **(B2a):** $E(\mathbf{u}|\mathbf{X}) = 0$ **Regressoren streng exogen** und

- * **(B2b):** $Var(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{I}$ **Fehler homoskedastisch** oder

- * **(B2b')** $Var(\mathbf{u}|\mathbf{X}) = \Omega$, Ω Diagonalmatrix **Fehler heteroskedastisch.**

- * **(B3) Keine perfekte Kollinearität**

\mathbf{X} (bzw. $\mathbf{X}^T\mathbf{X}$) hat vollen Rang.

Siehe Annahmen **(B1)**, **(B2)**, **(B3)** in **Methoden der Ökonometrie**, Abschnitt 9.4 *Die Effizienz unverzerrter Schätzer* bzw. Annahme **(B2')** in Abschnitt 14.1 *Verallgemeinerter Kleinst-Quadrate-Schätzer*.

– Zusätzliche **Annahme** für **exakte Inferenz**

- * **(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. Vgl. zur Schreibweise Davidson (2000, Section 2.4.1). Oder

- * **(B4')** $\mathbf{u}|\mathbf{X} \sim N(\mathbf{0}, \Omega)$.

– Zusätzliche **Annahme** für **asymptotische Inferenz**

- * **(A1)** $\text{plim}_{n \rightarrow \infty} \left(\frac{\mathbf{X}^T \mathbf{X}}{n} \right) = \mathbf{S}_{\mathbf{X}^T \mathbf{X}}$ und $\mathbf{S}_{\mathbf{X}^T \mathbf{X}}$ hat vollen Rang und (entspricht Davidson und MacKinnon, 2004, Gleichungen (3.17) bzw. (4.49)) oder

- * **(A1')** $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \Omega^{-1} \mathbf{X} = \mathbf{S}_{\mathbf{X}^T \Omega^{-1} \mathbf{X}}$, $\mathbf{S}_{\mathbf{X}^T \Omega^{-1} \mathbf{X}}$ hat vollen Rang.

- * **(A3)** $\frac{1}{\sqrt{n}} \mathbf{X}^T \mathbf{u} \xrightarrow{d} \mathbf{w}_\infty \sim N(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}})$ oder

- * **(A3')** $\frac{\mathbf{X}^T \Omega^{-1} \mathbf{u}}{\sqrt{n}} \xrightarrow{d} N(\mathbf{0}, \mathbf{S}_{\mathbf{X}^T \Omega^{-1} \mathbf{X}})$. oder

- * **(B5)** Es liegt eine Zufallsstichprobe vor.

Siehe Annahme **(B4)** in **Methoden der Ökonometrie**, Abschnitt *11.1 Exakte Verteilung des KQ-Schätzers*, Annahme **(A1)** in Abschnitt *9.2 Konsistenz des KQ-Schätzers*, bzw. Abschnitt *11.2 Asymptotische Verteilung des KQ-Schätzers* für Annahme **(A3)** bzw. **(B2)** plus Annahme einer Zufallsstichprobe. Siehe Abschnitt *14.1 Verallgemeinerter Kleinst-Quadrate-Schätzer* für **(B4')**, **(A1')** und **(A3')**.

- **Dynamisches Regressionsmodell** : Vgl. **Methoden der Ökonometrie**, Abschnitt *13.4 Dynamische lineare Regressionsmodelle*

$$\begin{aligned}
 y_t &= \mathbf{d}_t \boldsymbol{\nu} + \mathbf{Z}_t \boldsymbol{\delta}_0 + \mathbf{Z}_{t-1} \boldsymbol{\delta}_1 + \cdots + \mathbf{Z}_{t-m} \boldsymbol{\delta}_m + y_{t-1} \alpha_1 + \cdots + y_{t-p} \alpha_p + u_t, \\
 y_t &= \mathbf{X}_t \boldsymbol{\beta} + u_t
 \end{aligned}
 \tag{1.2a}$$

Beachte: In einem dynamischen Regressionsmodell bezeichnet der Index t die Zeit und ist damit geordnet. Letzteres ist in einem Modell für Querschnittsdaten nicht der Fall.

• **Informationsmengen** für dynamische Regressionsmodelle:

- Ω_t : Die Menge Ω_t bezeichnet die Menge aller **potentiell** erklärenden Variablen, die zur Spezifikation eines Modells für die endogene Variable y_t als **kausale** Variablen in Frage kommen.

Mögliche Variablen in Ω_t :

- * deterministische Variablen, zusammengefasst im Zeilenvektor \mathbf{d}_t : Konstante, Zeittrend, Saisondummies, etc.,
 - * verzögerte abhängige Variablen y_{t-j} , $j > 0$,
 - * kontemporäre Variablen $\mathbf{Z}_t \in \Omega_t$, so dass $E(u_t | \Omega_t) = 0$ gilt, also die Variablen \mathbf{Z}_t bezüglich des Fehlers u_t vorherbestimmt sind.
(‡ Dies entspricht der Voraussetzung, dass die kontemporären Variablen \mathbf{Z}_t schwach exogen bezüglich β, σ^2 sind, siehe Davidson (2000, Sections 5.3.1, 7.1.1).)
 - * verzögerte \mathbf{Z}_t , also \mathbf{Z}_{t-j} , $j > 0$,
 - * (fast) jede Funktion der genannten Variablen.
- \mathcal{I}_t : Die Menge aller erklärenden Variablen, die zur Spezifikation eines Modells

für die endogene Variable y_t **tatsächlich** verwendet werden, wird mit $\mathcal{I}_t \in \Omega_t$ bezeichnet.

Siehe **Methoden der Ökonometrie**, Abschnitte 5.2 *Spezifikation ökonometrischer Modelle* und 13. *Modelle für multivariate Zeitreihen*.

– **Annahmen** für asymptotische Schätzeigenschaften des KQ-Schätzers für (1.2)

* **(C1)** \iff Annahme **(B1)**: Der DGP ist für $\beta = \beta_0$ in (1.2a) enthalten.

(C2a) Regressoren vorherbestimmt

$$E(u_t|\Omega_t) = 0,$$

* **(C2)**: $u_t|\Omega_t \sim (0, \sigma^2)$ \iff **(C2b) Bedingte Homoskedastie der Fehler**

$$E(u_t^2|\Omega_t) = \sigma^2 \equiv E(u_t^2),$$

wobei für die Fehlervarianz des DGP $\sigma^2 = \sigma_0^2$ gilt.

* **(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 $\{\mathbf{w}_t\} = \begin{pmatrix} y_t & \mathbf{Z}_t \end{pmatrix}^T$,

* **(C4b)** $E|\boldsymbol{\lambda}^T \mathbf{X}_t u_t|^{2+\delta} \leq B < \infty$, $\delta > 0$, für alle feste $\boldsymbol{\lambda}$ mit $\boldsymbol{\lambda}^T \boldsymbol{\lambda} = 1$.

Siehe **Methoden der Ökonometrie**, Abschnitt 13.5 KQ-Schätzung von dynamischen linearen Regressions-

modellen.

- Bedingung **(C2a)** impliziert für (1.2), dass gilt:

$$E(y_t|\Omega_t) = E(u_t|\mathbf{X}_t) = \mathbf{X}_t\boldsymbol{\beta}. \quad (1.3)$$

Deshalb wird ein dynamisches lineares Regressionsmodell, das (1.3) erfüllt, als **dynamisch vollständig und korrekt spezifiziert** bezeichnet.

Vgl. **Methoden der Ökonometrie**, Abschnitt 13.4 *Dynamische lineare Regressionsmodelle* oder (vgl. [Wooldridge, 2009](#), Section 11.4, Equation (11.38)).

- Regressoren \mathbf{X}_t , die Bedingung **(C2a)** erfüllen, werden als **vorherbestimmt** bezüglich des Fehlerterms u_t bezeichnet. Die Fehler u_t werden auch als **Innovationen** oder als **Schocks** bezeichnet, da sie gegeben \mathbf{X}_t nicht prognostizierbar sind.

Vgl. in **Methoden der Ökonometrie**, Abschnitt 9.1.2 *Vorherbestimmte Regressoren* oder [Davidson und MacKinnon \(2004\)](#), Sections 3.2, 4.5).

• Notation

- Zur Erinnerung: Aus **(C2a)** folgt für das dynamische Regressionsmodell (1.2), dass die Fehler unkorreliert sind, da

$$E(u_t u_{t-j} | \Omega_t) = E[u_{t-j} E(u_t | \mathbf{X}_t, y_{t-1}, \mathbf{X}_{t-1}, y_{t-2}, \dots)] = E[u_{t-j} 0] = 0 \quad \text{für alle } j > 0$$

Siehe **Methoden der Ökonometrie**, Abschnitt 13.4 *Dynamische lineare Regressionsmodelle*.

Aus **(C2)** folgt jedoch nicht, dass die bedingten Verteilungen $f(u_t | \Omega_t)$ und $f(u_s | \Omega_s)$ unabhängig sind, da bspw.

$$E(u_t^2 u_s | \Omega_t) \neq 0$$

möglich ist.

- Die stärkere Annahme der Unabhängigkeit der bedingten Verteilungen wird folgendermaßen notiert:

* **(C5) IID der Fehler bedingt auf Informationsmenge**

$$u_t | \Omega_t \sim IID(0, \sigma^2)$$

Die Annahme bedeutet, dass

- * die Dichten $f(u_t|\Omega_t)$ von u_t **gegeben** die Informationsmenge Ω_t für alle t
 - **gleich** und
 - **unabhängig** sind, d. h. die Bedingung auf alle anderen Fehlerterme keinen Effekt hat, also

$$f(u_t|\Omega_t, u_1, \dots, u_{t-1}, u_{t+1}, \dots, u_n) = f(u_t|\Omega_t)$$

gilt und dass

- * $f(u_t|\Omega_t) = f(u_t)$ gilt.

Dies bedeutet auch, dass $E(u_t|\Omega_t) = 0$ und $Var(u_t|\Omega_t) = \sigma^2$.

Beachte: Davidson und MacKinnon (2004, S. 86 in Section 3.1, S. 213 in Section 6.1) schreiben lediglich $u_t \sim IID(0, \sigma^2)$ und weisen auf die Bedingtheit hinsichtlich Ω_t nur im Text hin.

– Eine noch strengere Annahme ist

* **(C6) Bedingte Normalverteilung der Fehler**

$$u_t | \Omega_t \sim NID(0, \sigma^2).$$

Bisher betrachtete Modelle

1. Dynamisches Regressionsmodell mit homoskedastischen Fehlern

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t, \quad E(y_t | \Omega_t) = \mathbf{X}_t \boldsymbol{\beta}, \quad \text{Var}(u_t | \Omega_t) = \sigma^2$$

$$\implies \text{OLS-Schätzer } \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

2. Dynamisches Regressionsmodell mit heteroskedastischen Fehlern

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t, \quad E(y_t | \Omega_t) = \mathbf{X}_t \boldsymbol{\beta}, \quad \text{Var}(u_t | \Omega_t) = \sigma_t^2, \quad \boldsymbol{\Omega} = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

\implies OLS-Schätzer mit heteroskedastierobusten Standardfehlern

$$\text{GLS-Schätzer } \hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y}$$

$$\text{FGLS-Schätzer } \hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}$$

1.2. Beispiele

1.2.1. Beispiel IM — Analyse von Importen nach Deutschland

Beispiel einer Querschnittsanalyse.

Siehe **Methoden der Ökonometrie**, Abschnitte

- *6.3 Empirische Analyse von Handelsströmen: Teil 1 — ein kurzer Überblick,*
- *10.3 Empirische Analyse von Handelsströmen: Teil 2,*
- *11.7 Empirische Analyse von Handelsströmen: Teil 3,*
- *14.4 Empirische Analyse von Handelsströmen: Teil 4,*
- *15.7 Empirische Analyse von Handelsströmen: Teil 5.*

Die Analyse wird im Verlauf dieses Kurses fortgesetzt.

1.2.2. Beispiel CO₂: Analyse des Zusammenhangs zwischen dem CO₂-Ausstoß eines Landes und dessen BIP, sowie weiteren möglichen Einflussfaktoren

Die Analyse erfolgt im Laufe des Kurses mit unterschiedlichen Modellen.

Folgende Daten werden jetzt betrachtet:

Daten für 151 Länder für das Jahr 2005 (später auch Paneldaten):

- CO₂-Ausstoß pro Kopf (in Tonnen),
Quelle: [Weltbank, World Development Indicators](#) CO2 emissions (metric tons per capita)
- GDP, real, chained US\$, pro Kopf,
Quelle: [Penn World Tables, Real GDP chained US-Dollar](#)
- Services, Anteil an GDP, Quelle: Weltbank, WDI
- Industry, Anteil an GDP, Quelle: Weltbank, WDI
- Agriculture, Anteil an GDP, Quelle: Weltbank, WDI

Multiple Regression mit Interaktionsterm, $t = 1, \dots, 151$:

$$\ln(CO2_t) = \beta_1 + \beta_2 \ln(GDP_t) + \beta_3 \ln(GDP_t) * Services_t + \beta_4 Services_t + \varepsilon_t. \quad (1.4)$$

wobei

$$\Omega_t = \{GDP_t, Services_t, Industry_t, Agriculture_t, Inflation_t, Weather_t, Population_t, \dots, GDP_{t-1}, \dots\}$$

$$\mathcal{I}_t = \{GDP_t, Services_t\}$$

R-Programm und CO₂-Daten zu folgendem Output

Call:

```
lm(formula = log(CO2) ~ log(GDP) + I(log(GDP) * Services) + Services,
    data = data2005)
```

Residuals:

Min	1Q	Median	3Q	Max
-3.530299	-0.471693	0.001403	0.435345	2.509990

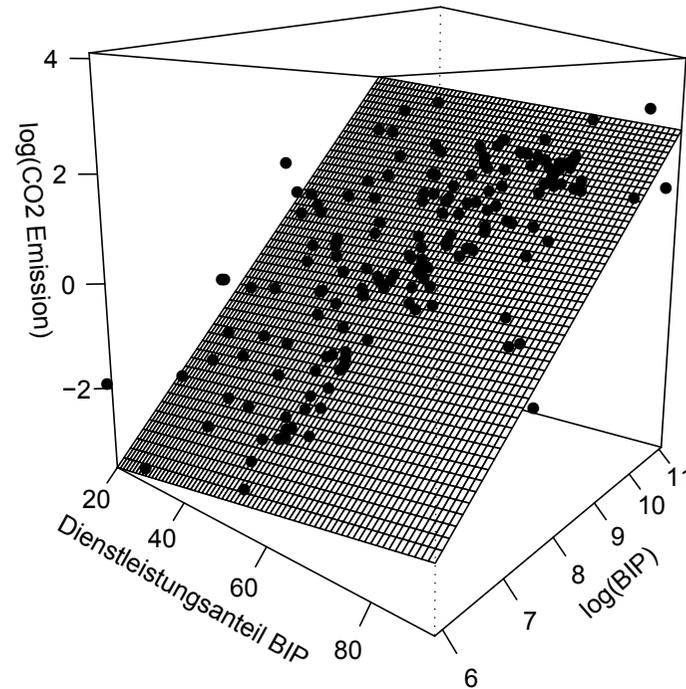
Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	-13.316995	1.510081	-8.819	2.43e-15	***
log(GDP)	1.604615	0.174719	9.184	2.76e-16	***
I(log(GDP) * Services)	-0.006412	0.003186	-2.012	0.0459	*
Services	0.052214	0.028952	1.803	0.0733	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7857 on 153 degrees of freedom

Multiple R-squared: 0.7749, Adjusted R-squared: 0.7705
F-statistic: 175.5 on 3 and 153 DF, p-value: < 2.2e-16



Geschätzte Regressionsfunktion

Ist diese Regression korrekt spezifiziert?

- Sind alle relevanten Variablen enthalten?
- Wird eine mögliche Nichtlinearität korrekt erfasst?
- Sind die Regressoren vorherbestimmt?

1.3. Allgemeinere Modelle in diesem Kurs

3. Nichtlineares Modell mit vorherbestimmten Variablen und homoskedastischen Fehlern

$$y_t = x_t(\boldsymbol{\beta}) + u_t, \quad E(y_t|\Omega_t) = x_t(\boldsymbol{\beta}), \quad Var(u_t|\Omega_t) = \sigma^2$$

$$\implies \text{Momentenschätzer } \mathbf{Z}^T(\mathbf{y} - \mathbf{x}(\hat{\boldsymbol{\beta}})) = \mathbf{0}$$

$$\text{NLS-Schätzer } \mathbf{X}(\hat{\boldsymbol{\beta}})^T(\mathbf{y} - \mathbf{x}(\hat{\boldsymbol{\beta}})) = \mathbf{0}$$

\implies Kapitel **2 Nichtlineare Regression**.

4. Regressionsmodelle mit *vollständig spezifizierter Fehlerverteilung*

\implies Kapitel **3 Maximum-Likelihood-Schätzung**.

5. Regressionsmodell mit *streng exogenen Regressoren* und heteroskedastischen und autokorrelierten Fehlern

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \quad E(y_t|\Omega_t) = \mathbf{X}_t\boldsymbol{\beta}, \quad \mathbf{u}|\mathbf{X} = (\mathbf{0}, \boldsymbol{\Omega})$$

$$\implies \text{GLS-Schätzer } \hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\Omega}^{-1}\mathbf{y}$$

$$\text{FGLS-Schätzer } \hat{\boldsymbol{\beta}}_{FGLS} = (\mathbf{X}^T\hat{\boldsymbol{\Omega}}^{-1}\mathbf{X})^{-1}\mathbf{X}^T\hat{\boldsymbol{\Omega}}^{-1}\mathbf{y}$$

\implies Kapitel **4. Verallgemeinerter KQ-Schätzer**

6. Lineares Modell mit *endogenen Regressoren* und homoskedastischen Fehlern

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \quad E(y_t|\Omega_t) = E(\mathbf{X}_t|\Omega_t)\boldsymbol{\beta} \neq \mathbf{X}_t\boldsymbol{\beta}, \quad Var(u_t|\Omega_t) = \sigma^2$$

$$\begin{aligned} &\implies \text{IV-Schätzer } \hat{\beta}_{IV} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y} \\ &\text{verallgemeinerter IV-Schätzer } \hat{\beta}_{IV} = (\mathbf{X}^T \mathbf{P}_W \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_W \mathbf{y} \end{aligned}$$

mit $(n \times k)$ -Instrumentenmatrix \mathbf{Z} und $(n \times l)$ -Instrumentenmatrix \mathbf{W}

\implies Kapitel **5 Instrumentvariablenschätzung**.

7. Lineares Modell mit *endogenen* Regressoren und heteroskedastischen und korrelierten Fehlern

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t, \quad E(y_t | \Omega_t) = E(\mathbf{X}_t | \Omega_t) \boldsymbol{\beta} \neq \mathbf{X}_t \boldsymbol{\beta}, \quad E(u_t u_s | \mathbf{W}_t, \mathbf{W}_s) = \omega_{ts}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \cdots & \omega_{nn} \end{pmatrix}$$

$$\begin{aligned} &\implies \text{IV-Schätzer } \hat{\beta}_{IV} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y} \\ &\text{verallgemeinerter IV-Schätzer } \hat{\beta}_{IV} = (\mathbf{X}^T \mathbf{P}_W \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_W \mathbf{y} \end{aligned}$$

mit $(n \times k)$ -Instrumentenmatrix \mathbf{Z} und $(n \times l)$ -Instrumentenmatrix \mathbf{W}

\implies Kapitel **6 Generalized Method of Moments (GMM)**.

2. Nichtlineare Regression

- Das (dynamische) Regressionsmodell

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t \quad (1.2)$$

wird als **linear** bezeichnet, da die Regressionsfunktion $\mathbf{X}_t\boldsymbol{\beta}$ eine lineare Funktion in den Parametern $\boldsymbol{\beta}$ ist.

Zu den linearen Regressionsmodellen gehören auch Modelle, die durch Logarithmieren linearisierbar sind wie bspw. eine Cobb-Douglas-Funktion mit multiplikativem positivem Fehler

$$Y_t = AK_t^\alpha L_t^\beta e^{\varepsilon_t}$$
$$\ln Y_t = \ln A + \alpha \ln K_t + \beta \ln L_t + \varepsilon_t$$

- Ein **nichtlineares Regressionsmodell** liegt vor, wenn die Regressionsfunktion $m(\mathbf{X}_t, \boldsymbol{\beta})$ in

$$y_t = m(\mathbf{X}_t, \boldsymbol{\beta}) + u_t \quad (2.1)$$

eine nichtlineare Funktion in $\boldsymbol{\beta}$ ist. Die Regressionsfunktion $m(\cdot, \boldsymbol{\beta})$ kann nicht-linear in den k Parametern $\boldsymbol{\beta}$ sein, wobei aber die Anzahl der Regressoren *nicht* gleich k sein muss.

- **Beispiele** für nichtlineare Regressionsmodelle inkl. nichtlineare Zeitreihenmodelle:

- **Nichtlinearisierbare ökonomische Modelle** - Beispiele:

- * **CES-Produktionsfunktion**

$$Y_t = y[(1 - \delta)K_t^\xi + \delta L_t^\xi]^{1/\xi} + \varepsilon_t.$$

- * **(Hedonische) Preistheorie und Preisabsatzfunktionen** (z.B. [Haupt, Schnurbus, und Tschernig, 2010](#))

- * **Konjunktur-/Endogene Wachstumsmodelle**

* **Zinsstrukturkurve** (Nelson und Siegel, 1987)

$$r_t = \beta_0 + (\beta_1 + \beta_2) \frac{1 - \exp(-m_t/\tau)}{m_t/\tau} - \beta_2 \exp(-m_t/\tau) + u_t \quad (2.2)$$

r_t : Rendite einer Nullcouponanleihe t ; m_t : Restlaufzeit.

* **Nachfrage nach Gesundheitsleistungen** (Chakroun, 2009): logistisches 2-Regime Smooth Transition Regression Modell

$$lhe_{it} = \beta_1 + \beta_2 lgd_{pit} + \beta_3 lgd_{pit} G(s_{it}; c, \nu) + \beta_4 phe_{sh_{it}} + u_{it}$$

$$G(s_{it}; c, \nu) = \frac{1}{1 + \exp(-\nu(s_{it} - c))}, \quad \nu > 0.$$

lhe_{it} : logarithmierte Gesundheitsausgaben von Land i zum Zeitpunkt t ;
 lgd_{pit} : logarithmierte BIP; s_{it} : Proxy für die Produktivität im Gesundheitssektor (Transitionsvariable; etwa Lebenserwartung o.ä.).

- * **Asymmetrische Zinseffekte auf die reale Ökonomie** (Sensier, Osborn, und Öcal, 2002): Smooth Transition Modell

$$y_t = \phi_{00} + \sum_{i=1}^8 \phi_{0i} y_{t-i} + \sum_{i=1}^8 \delta_{0i} z_{t-i} + G(r_{t-1}; \gamma, c) \left(\phi_{10} + \sum_{i=1}^8 \phi_{1i} y_{t-i} + \sum_{i=1}^8 \delta_{1i} z_{t-i} \right) + u_t,$$

$$G(r_{t-1}; c, \nu) = \frac{1}{1 + \exp(-\nu(r_{t-1} - c))}, \quad \nu > 0.$$

y_t : Wachstumsrate des BIP zum Zeitpunkt (Quartal) t ; z_t : Änderung des kurzfristigen Zinses und Zinssatz r_t .

- * **Prognose von Aktienrenditen** (McMillan und Wohar, 2009): exponentielles Smooth Transition Model (STR) Modell

$$r_t = \beta_0 + \{1 - \exp[-\gamma(dp_{t-1} - c)^2]\} \beta_1 dp_{t-1} + u_t$$

r_t : (Über)rendite eines Aktienportfolios zum Zeitpunkt t ; dp_t : Dividenden-Kursrelation dieses Portfolios.

- * **Nichtlineare Geldnachfrage** (Teräsvirta und Eliasson, 2001): Smooth Transition Modell (STR)

$$\begin{aligned}\Delta(m - p)_t &= \beta_0 + \beta_1 \Delta(m - p)_{t-1} + \beta_2 \Delta p_t + \beta_3 r n_t^a + \beta_4 e c_t \\ &+ G(\Delta i_t; \gamma, c) \times [\gamma_0 + \gamma_1 \Delta(m - p)_{t-1} + \gamma_2 \Delta p_t + \gamma_3 r n_t^a + \gamma_4 e c_t] \\ &+ u_t,\end{aligned}$$

$$G(\Delta i_t; \gamma, c_1, c_2) = \left\{ 1 + \exp[-\gamma(\Delta i_t - c_1)(\Delta i_t - c_2)/\sigma_{\Delta i_t}^2] \right\}^{-1}$$

m_t : logarithmierte Geldmenge; p_t : aggregiertes Preisniveau; $r n_t^a$: Kurzfristzins; $e c_t$: Abweichungen von der langfristig gleichgewichtigen Geldnachfrage.

- * **Schwelleneffekte der Inflation** (Tsiionas und Christopoulos, 2003): Smooth Transition Modell (STR)

$$\begin{aligned}y_t &= \alpha_0 + \alpha_1 \pi_t + \alpha_2 s_t + \alpha_3 q_t + G(\pi_t; \gamma, \pi^*) \times (\beta_0 + \beta_1 \pi_t + \beta_2 s_t + \beta_3 q_t) \\ &+ u_t,\end{aligned}$$

$$G(\pi_t; \gamma, c) = \left\{ 1 + \exp[-\gamma(\pi_{t-1} - \pi^*)] \right\}^{-1}$$

y_t : BIP-Wachstum; π_t : Inflationsrate; s_t : Investitionsanteil des BIP; q_t : Terms-of-Trade Wachstum; π^* : unbekannter Parameter (Schwellenwert).

- * Prognose von **realen Wechselkursen** (Donauer, Heinen, und Sibbertsen, 2010): Smooth Transition Autoregression (T-STAR)

$$y_t = y_{t-1} - \phi[1 - (1 + (y_{t-d} - c)^2)^{-\gamma}] + u_t,$$

y_t : realer Wechselkurs zwischen zwei Währungen zum Zeitpunkt t .

– Komplexere ökonometrische Modelle

- * **Mikroökonometrie**: Modelle mit zensierten oder gestutzten oder diskreten abhängigen Variablen
- * **Lineares Regressionsmodell mit autokorrelierten Fehlern**

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \quad u_t = \rho u_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim IID(0, \sigma^2), \quad |\rho| < 1$$

- Auflösen der ersten Gleichung nach u_t (bzw. u_{t-1}) und Einsetzen in die zweite Gleichung ergibt den Ansatz

$$\begin{aligned} y_t &= \mathbf{X}_t\boldsymbol{\beta} + \rho(y_{t-1} - \mathbf{X}_{t-1}\boldsymbol{\beta}) + \varepsilon_t \\ &= \rho y_{t-1} + \mathbf{X}_t\boldsymbol{\beta} - \rho\mathbf{X}_{t-1}\boldsymbol{\beta} + \varepsilon_t, \end{aligned}$$

der dynamisch und nichtlinear in den Parametern ρ und $\boldsymbol{\beta}$ ist.

- Unrestringiert könnte man mit OLS

$$y_t = \rho y_{t-1} + \mathbf{X}_t \boldsymbol{\beta} + \mathbf{X}_{t-1} \boldsymbol{\gamma} + v_t$$

schätzen. Allerdings müssen dann $2k+1$ anstatt $k+1$ Parameter geschätzt werden, was ineffizient (und ein anderes Modell) ist.

– Nichtlineare Modelle für andere Fachgebiete

- * **Medizinischer Dosis-Wirkung-Zusammenhang**, etwa Michaelis-Menten Modell ([Raaijmakers, 1987](#))

$$y_t = \beta_1 + \frac{\beta_2 x_t}{x_t + \beta_3} + u_t \quad (2.3)$$

oder EMAX-Modell

$$y_t = \beta_1 + \frac{\beta_2 x_t^\gamma}{x_t^\gamma + \beta_3^\gamma} + u_t.$$

Hier ist x_t die Dosis, y_t die Wirkung eines bestimmten Medikamentes bei Patient $t = 1, 2, \dots, n$.

- **Dynamisch vollständige und korrekte Spezifikation** der nichtlinearen Regressionsfunktion:

- Das (nichtlineare) Regressionsmodell (2.1) ist **dynamisch vollständig und korrekt spezifiziert**, wenn

$$E(y_t|\Omega_t) = E(y_t|\mathbf{X}_t) = m(\mathbf{X}_t, \boldsymbol{\beta}) \quad (2.4)$$

gilt. Dabei ergibt das erste Gleichheitszeichen, dass das Modell *dynamisch vollständig* ist, und das zweite Gleichheitszeichen, dass das Modell *korrekt spezifiziert* ist. Das bedeutet, dass der DGP für $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ in (2.1) enthalten ist.

- Die Bedingung (2.4) gilt, wenn und nur wenn in (2.1) die Annahme

$$E(u_t|\Omega_t) = 0. \quad (\mathbf{C2a})$$

erfüllt ist.

- **Beachte:**

- Es gilt für beliebige $\mathcal{I}_t \in \Omega_t$ immer, dass

$$y = E(y_t|\mathcal{I}_t) + \underbrace{y_t - E(y_t|\mathcal{I}_t)}_{v_t} \Leftrightarrow E(v_t|\mathcal{I}_t) = 0, \quad (2.5)$$

wobei **jedoch**

$$\begin{aligned} E(v_t|\Omega_t) &= E(y_t|\Omega_t) - E[E(y_t|\mathcal{I}_t)|\Omega_t] \\ &= E(y_t|\Omega_t) - E(y_t|\mathcal{I}_t) \begin{cases} = 0 & \text{falls alle relevanten Regressoren in } \mathcal{I}_t, \\ \neq 0 & \text{sonst.} \end{cases} \end{aligned}$$

$E[E(y_t|\mathcal{I}_t)|\Omega_t] = E(y_t|\mathcal{I}_t)$, da die Informationsmenge \mathcal{I}_t in Ω_t enthalten ist.

- **Dynamisch vollständige und korrekte Spezifikation** erfordert also, dass das nichtlineare Regressionsmodell (2.1)
 1. alle relevanten Variablen enthält und
 2. die die Klasse der betrachteten nichtlinearen Regressionsfunktionen $m(\cdot, \beta)$ die wahre Regressionsfunktion enthält.

- **Notation** in Davidson und MacKinnon (2004, Chapter 6) für nichtlineare (homoskedastische) Regressionsmodelle:

$$y_t = \underbrace{m(\mathbf{X}_t, \boldsymbol{\beta})}_{x_t(\boldsymbol{\beta})} + u_t, \quad u_t | \Omega_t \sim IID(0, \sigma^2), \quad \text{für } t = 1, \dots, n \quad (2.6a)$$

$$y_t = x_t(\boldsymbol{\beta}) + u_t. \quad (2.6b)$$

Die skalare Funktion $x_t(\boldsymbol{\beta})$ hängt wegen \mathbf{X}_t vom Index t ab.

Matrixschreibweise:

$$\mathbf{y} = \mathbf{m}(\mathbf{X}, \boldsymbol{\beta}) + \mathbf{u}, \quad \text{bzw.} \quad (2.6c)$$

$$\mathbf{y} = \mathbf{x}(\boldsymbol{\beta}) + \mathbf{u}, \quad \mathbf{u} \sim IID(\mathbf{0}, \sigma^2 \mathbf{I}), \quad (2.6d)$$

wobei $\mathbf{x}(\boldsymbol{\beta})$ eine Vektorfunktion ist und $u \sim IID(\mathbf{0}, \sigma^2 \mathbf{I})$ bedeutet, dass $u_t | \Omega_t \sim IID(0, \sigma^2)$.

- **Schätzmethoden:**
 - Momentenschätzer für nichtlineare Modelle,
 - Nichtlineare Kleinst-Quadrate-Schätzung,
 - Maximum Likelihood-Schätzung, siehe Kapitel 3.

2.1. Momentenschätzer für nichtlineare Modelle

- **Zu lesen:** Davidson und MacKinnon (2004, Chapter 6).

- **Lineare Modelle**

Ist die Regressionsfunktion in (1.2) dynamisch vollständig und korrekt spezifiziert, d. h. gilt **(C2a)**, und damit die Regressoren **vorherbestimmt**, $\mathbf{X}_t \in \Omega_t$ folgt

$$E(u_t | \mathbf{X}_t) = 0.$$

Daraus ergibt sich die Momentenbedingung

$$E(\mathbf{X}_t u_t) = E[E(\mathbf{X}_t u_t | \mathbf{X}_t)] = E[\mathbf{X}_t E(u_t | \mathbf{X}_t)] = \mathbf{0}.$$

- **Nichtlineare Modelle**

Aus der Annahme **(C2a)**

$$E(u_t | \Omega_t) = 0$$

ergibt sich für jeden Vektor $\mathbf{W}_t \in \Omega_t$ wieder eine Momentenbedingung

$$E(\mathbf{W}_t u_t) = E[\mathbf{W}_t E(u_t | \Omega_t)] = \mathbf{0}.$$

Einsetzen von $u_t = y_t - x_t(\beta_0)$ ergibt

$$E[\mathbf{W}_t(y_t - x_t(\beta_0))] = \mathbf{0}.$$

Hingegen ist i.A. $E[\mathbf{W}_t(y_t - x_t(\beta))] \neq \mathbf{0}$, sofern $\beta \neq \beta_0$.

- D. h. könnte man diesen Vektor der Erwartungswerte (der gerade dem Kovarianzvektor (2. Moment) zwischen \mathbf{W}_t und u_t entspricht) berechnen, ließe sich β_0 finden, sofern ausschließlich β_0 Kovarianzen von Null ergibt (d. h. β ist *identifizierbar*, siehe später).
- Der Erwartungswert ist unbekannt und muss geschätzt werden, wie üblich, durch Mittelwertbildung über die n Stichprobenbeobachtungen.

Für die vorliegende Stichprobe resultieren die **Momentenbedingungen**

$$\mathbf{W}^T(\mathbf{y} - \mathbf{x}(\beta)) \stackrel{!}{=} \mathbf{0}.$$

Eine Lösung der Momentenbedingungen ergibt den Momentenschätzer $\hat{\beta}$ auf Basis von \mathbf{W} , der abhängig ist von der **Wahl der Matrix \mathbf{W}** :

- \mathbf{W} muss Rang k haben.
- \mathbf{W} beeinflusst asymptotische Kovarianzmatrix.

- **Identifikation eines Parametervektors**

- Ein Parametervektor ist in einem ökonometrischen Modell bezüglich eines DGPs und eines bestimmten Schätzverfahrens **identifiziert**, wenn das Schätzverfahren zu einer **eindeutigen Lösung** führt.
- Im vorliegenden Fall muss das Gleichungssystem der Momentenbedingungen eine eindeutige Lösung aufweisen.

- **Asymptotische Identifikation**

- Ein Parametervektor ist in einem ökonometrischen Modell bezüglich eines DGPs und eines bestimmten Schätzverfahrens **asymptotisch identifiziert**, wenn er für $n \rightarrow \infty$ identifiziert ist.
- Es gibt Modelle und DGPs, in denen der Parametervektor asymptotisch identifiziert, aber nicht identifiziert ist und umgekehrt.
(Siehe Beispiele in [Davidson und MacKinnon \(2004, S. 217-218\)](#)).

- Vorliegender Fall: Betrachte plim der Momentenbedingungen für Parametervektor β_0 des DGP

$$\frac{1}{n} \mathbf{W}^T (\mathbf{y} - \mathbf{x}(\beta_0)) = \frac{1}{n} \sum_{t=1}^n \mathbf{W}_t^T u_t.$$

Aufgrund **(C2a)** gilt $E(u_t W_t) = 0$. Gilt darüber hinaus, dass $E(|W_t u_t|^{1+\delta}) \leq B < \infty$, $\delta > 0$, dann lässt sich ein Gesetz der Großen Zahlen anwenden und man erhält

$$\alpha(\beta_0) \equiv \text{plim} \frac{1}{n} \sum_{t=1}^n \mathbf{W}_t^T u_t = \mathbf{0}.$$

‡ Genauer gesagt, ein Gesetz der Großen Zahlen für Martingaldifferenzen, siehe z.B. [Davidson \(2000, Section 11.3.3\)](#) oder **Fortgeschrittene Dynamische Ökonometrie**.

- Ist $\beta \neq \beta_0$, dann ist im Allgemeinen

$$\alpha(\beta) \equiv \text{plim} \frac{1}{n} \mathbf{W}^T (\mathbf{y} - \mathbf{x}(\beta)) \neq \mathbf{0}.$$

Ist $\alpha(\beta) \neq \mathbf{0}$ für alle $\beta \neq \beta_0$, dann ist der Parametervektor **asymptotisch identifizierbar**, da $\min_{\beta} \|\alpha(\beta)\|$ eindeutig für β_0 gegeben ist.

– **Notwendige Voraussetzung für asymptotische Identifizierbarkeit:**

Es gibt kein $\beta_1 \neq \beta_0$, so dass $\mathbf{x}(\beta_1) = \mathbf{x}(\beta_0)$.

– Beachte: Asymptotische Identifizierbarkeit ist von der gewählten Schätzmethode (hier auch Auswahl von \mathbf{W}) abhängig, da

$$\boldsymbol{\alpha}(\boldsymbol{\beta}) = \underbrace{\text{plim} \frac{1}{n} \mathbf{W}^T \mathbf{u}}_{=\boldsymbol{\alpha}(\boldsymbol{\beta}_0)=\mathbf{0}} + \text{plim} \frac{1}{n} \mathbf{W}^T (\mathbf{x}(\boldsymbol{\beta}_0) - \mathbf{x}(\boldsymbol{\beta}))$$

• Konsistenz

– Grundidee des Beweises: Damit $\hat{\beta}_n$ konsistent sein kann, muss $\text{plim } \hat{\beta}_n$ existieren.

(Beweis technisch anspruchsvoll und wird hier vorausgesetzt).

Notation: $\beta_\infty \equiv \text{plim } \hat{\beta}_n$.

Nun gilt für $\hat{\beta}_n$

$$\frac{1}{n} \mathbf{W}^T (\mathbf{y} - \mathbf{x}(\hat{\beta}_n)) = \mathbf{0}.$$

Möchte man nun den plim auf beiden Seiten dieses Ausdrucks berechnen, hat man die Schwierigkeit, dass $\hat{\beta}_n$ mit variierender Stichprobengröße nicht konstant ist. Man löst dieses Problem, indem man

$$\alpha(\beta_\infty) = \text{plim } \frac{1}{n} \mathbf{W}^T (\mathbf{y} - \mathbf{x}(\text{plim } \hat{\beta}_n)) = \text{plim } \mathbf{0}$$

betrachtet. Man nehme nun an, dass $\beta_\infty \neq \beta_0$ gilt. Gilt nun asymptotische Identifizierbarkeit von β , dann folgt

$$\alpha(\beta_\infty) \neq \alpha(\beta_0) = \mathbf{0},$$

was im Widerspruch zu obiger Gleichung steht. Also muss $\hat{\beta}_n$ konsistent, d. h. $\beta_\infty = \beta_0$ sein, wenn asymptotische Identifizierbarkeit vorliegt.

– **Annahmen für Konsistenz:**

- * **(NL1)** Der DGP ist für $\beta = \beta_0$ im nichtlinearen Regressionsmodell (2.1) enthalten.

(NL2a) Die Regressoren \mathbf{X}_t sind vorherbestimmt **(C2a)**:

$$E(u_t | \Omega_t) = 0,$$

- * **(NL2)** $u_t | \Omega_t \sim IID(0, \sigma^2) \implies$

(NL2b) Bedingte Homoskedastie der Fehler

$$E(u_t^2 | \Omega_t) = \sigma^2 \equiv E(u_t^2),$$

wobei für die Fehlervarianz des DGP $\sigma^2 = \sigma_0^2$ gilt.

- * **(NL3)** Der Parametervektor β ist asymptotisch identifizierbar, d. h. $\alpha(\beta) = \text{plim} \frac{1}{n} \mathbf{W}^T (\mathbf{y} - \mathbf{x}(\beta))$ existiert für alle zulässigen β und $\mathbf{W}_t \in \Omega_t$ und es gilt $\alpha(\beta) \neq \alpha(\beta_0) = \mathbf{0}$, wenn $\|\beta - \beta_0\| > 0$.
- * **(NL4)** $\text{plim} \hat{\beta}$ existiert.

- **Konsistenz**: Unter **(NL1)** bis **(NL4)** und weiteren technischen Regularitätsbedingungen, ist der Momentenschätzer konsistent.
- Konsistenz erfordert, dass

$$\alpha(\beta_0) = \text{plim} \frac{1}{n} \mathbf{W}^T \mathbf{u} = \mathbf{0}.$$

Dies gilt unter allgemeineren Annahmen, z.B. bei

* heteroskedastischen Fehlern

jedoch nicht, wenn sowohl verzögerte endogene Regressoren als auch korrelierte Fehler vorliegen.

• Asymptotische Normalität

– **Notation:** Wir definieren die partiellen Ableitungen

$$X_{ti}(\bar{\boldsymbol{\beta}}) = \left. \frac{\partial x_t(\boldsymbol{\beta})}{\partial \beta_i} \right|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}} \quad (1 \times 1)$$

$$\mathbf{X}_t(\bar{\boldsymbol{\beta}}) = \left. \frac{\partial x_t(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \right|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}} \quad (1 \times k)$$

$$\mathbf{X}(\bar{\boldsymbol{\beta}}) = \left. \frac{\partial \mathbf{x}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^T} \right|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}} \quad (n \times k)$$

– Insbesondere sei

$$\mathbf{X}_0 = \mathbf{X}(\boldsymbol{\beta}_0)$$

die $(n \times k)$ -Matrix aller partiellen Ableitungen der n nichtlinearen Regressionsfunktionen $x_t(\boldsymbol{\beta}_0)$, $t = 1, \dots, n$, nach den k Parametern an der Stelle $\boldsymbol{\beta} = \boldsymbol{\beta}_0$.

– Beachte: für das lineare Regressionsmodell gilt

$$\mathbf{X}_t(\boldsymbol{\beta}) = \mathbf{X}_t \text{ bzw. } \mathbf{X}(\boldsymbol{\beta}) = \mathbf{X}.$$

Weitere Annahmen für die asymptotische Normalverteilung:

* **(NL5)** Die nichtlineare Regressionsfunktion ist zweimal partiell stetig differenzierbar. (Die zweiten partiellen Ableitungen erscheinen im folgenden nicht, werden aber bei technischen Details benötigt.)

* **(NL6)** LLN: Es gilt ein Gesetz der Großen Zahlen für $\frac{1}{n}\mathbf{W}^T\mathbf{X}_0$

$$\text{plim} \frac{1}{n}\mathbf{W}^T\mathbf{X}_0 = \mathbf{S}_{\mathbf{W}^T\mathbf{X}}$$

und $\mathbf{S}_{\mathbf{W}^T\mathbf{X}}$ hat vollen Rang k . Im speziellen Fall $\mathbf{W} = \mathbf{X}_0 = \mathbf{X}$ erhält man Annahme (A1) bzw. Annahme **(C3)**.

* **(NL7)** CLT: Es gilt für die asymptotische Kovarianzmatrix

$$\lim_{n \rightarrow \infty} \text{Var} \left(\frac{1}{\sqrt{n}}\mathbf{W}^T\mathbf{u} \right) = \sigma_0^2 \text{plim} \frac{1}{n}\mathbf{W}^T\mathbf{W} = \sigma_0^2\mathbf{S}_{\mathbf{W}^T\mathbf{W}}$$

und es gilt ein Zentraler Grenzwertsatz für Vektoren

$$\frac{1}{\sqrt{n}}\mathbf{W}^T\mathbf{u} \xrightarrow{d} N(\mathbf{0}, \sigma_0^2\mathbf{S}_{\mathbf{W}^T\mathbf{W}}).$$

Diese Annahme entspricht für $\mathbf{W} = \mathbf{X}$ Annahme **(A3)**.

Sie können diese Annahmen analog zum Vorgehen im genannten Abschnitt unter der zusätzlichen Annahme einer Zufallsstichprobe zeigen. Allgemeinere Beweise für abhängige Daten sind wesentlich schwieriger.

– Asymptotische Normalverteilung

Unter den Annahmen **(NL1)** bis **(NL7)** und weiteren technischen Annahmen ist der Momentenschätzer **asymptotisch normalverteilt**

$$\sqrt{n} \left(\hat{\beta} - \beta_0 \right) \xrightarrow{d} N \left(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{W}^T \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{W}^T \mathbf{W}} \left(\mathbf{S}_{\mathbf{W}^T \mathbf{X}}^T \right)^{-1} \right).$$

– Ableitung

* Grundidee: Multiplizieren der Momentenbedingungen mit $\frac{1}{\sqrt{n}}$ und Einsetzen des DGPs $\mathbf{y} = \mathbf{x}(\beta_0) + \mathbf{u}$ (Verwendung von **(NL1)**) ergibt

$$\frac{1}{\sqrt{n}} \mathbf{W}^T \left(\mathbf{x}(\beta_0) + \mathbf{u} - \mathbf{x}(\hat{\beta}) \right) = \mathbf{0}.$$

Umstellen liefert

$$\frac{1}{\sqrt{n}} \mathbf{W}^T \left(\mathbf{x}(\hat{\beta}) - \mathbf{x}(\beta_0) \right) = \frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u}$$

- * Taylorentwicklung (Annahme **(NL5)**) von $\mathbf{x}(\boldsymbol{\beta})$ an der Stelle des wahren Parametervektors $\boldsymbol{\beta}_0$ ergibt mit $\|\bar{\boldsymbol{\beta}}_t - \boldsymbol{\beta}_0\| \leq \|\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|$:

$$\begin{aligned} x_t(\hat{\boldsymbol{\beta}}) &= x_t(\boldsymbol{\beta}_0) + \sum_{i=1}^k \frac{\partial x_t(\boldsymbol{\beta})}{\partial \beta_i} \Big|_{\boldsymbol{\beta}=\bar{\boldsymbol{\beta}}_t} (\hat{\beta}_i - \beta_{0i}) \\ &= x_t(\boldsymbol{\beta}_0) + \mathbf{X}_t(\bar{\boldsymbol{\beta}}_t) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \\ \mathbf{x}(\hat{\boldsymbol{\beta}}) &= \mathbf{x}(\boldsymbol{\beta}_0) + \mathbf{X}(\bar{\boldsymbol{\beta}}_1, \dots, \bar{\boldsymbol{\beta}}_n) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0). \end{aligned}$$

Man beachte, dass $\mathbf{X}(\bar{\boldsymbol{\beta}}_1, \dots, \bar{\boldsymbol{\beta}}_n)$ bedeutet, dass $\bar{\boldsymbol{\beta}}_t$ für jede Beobachtung verschieden sein kann.

- * Einsetzen der Taylorentwicklung ergibt

$$\begin{aligned} \frac{1}{\sqrt{n}} \mathbf{W}^T \left(\mathbf{x}(\boldsymbol{\beta}_0) - \mathbf{x}(\boldsymbol{\beta}_0) + \mathbf{X}(\bar{\boldsymbol{\beta}}_1, \dots, \bar{\boldsymbol{\beta}}_n) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \right) &= \frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u} \\ \frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{X}(\bar{\boldsymbol{\beta}}_1, \dots, \bar{\boldsymbol{\beta}}_n) (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) &= \frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u}. \end{aligned}$$

- * Jetzt ist man fast fertig. Ist der Schätzer konsistent (**(NL1)** bis **(NL4)**),

so ist $\text{plim } \bar{\beta}_t = \beta_0$ und somit erhält man die asymptotische Äquivalenz

$$\frac{1}{n} \mathbf{W}^T \mathbf{X}(\beta_0) \sqrt{n} (\hat{\beta} - \beta_0) \stackrel{a}{=} \frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u},$$

wobei $\stackrel{a}{=}$ 'asymptotisch äquivalent' bedeutet.

Verwendet man nun **(NL6)** $\text{plim } \frac{1}{n} \mathbf{W}^T \mathbf{X}(\beta_0) = \mathbf{S}_{\mathbf{W}^T \mathbf{X}}$, ergibt sich als weiterer Approximationsschritt

$$\sqrt{n} (\hat{\beta} - \beta_0) \stackrel{a}{=} \mathbf{S}_{\mathbf{W}^T \mathbf{X}}^{-1} \frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u}.$$

Unter Anwendung des zentralen Grenzwertsatzes (Annahme **(NL7)**) erhält man schließlich die **asymptotische Normalverteilung**

$$\sqrt{n} (\hat{\beta} - \beta_0) \xrightarrow{d} N \left(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{W}^T \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{W}^T \mathbf{W}} (\mathbf{S}_{\mathbf{W}^T \mathbf{X}}^T)^{-1} \right).$$

- Es kann gezeigt werden, dass die asymptotische Kovarianzmatrix am 'kleinsten' ist, wenn $\mathbf{W} = \mathbf{X}_0$ gewählt wird. Allerdings ist \mathbf{X}_0 nicht beobachtbar und damit ist dieser Momentenschätzer **'infeasible'**.
- Man kann allerdings einen anwendbaren Schätzer erhalten, indem man β_0 mit

einem konsistenten Schätzer $\tilde{\beta}$ konsistent schätzt und $\mathbf{X}_0 = \mathbf{X}(\beta_0)$ mit $\mathbf{X}(\tilde{\beta})$ konsistent schätzt:

- * Zweistufiger Schätzer
- * Nichtlinearer Kleinstquadrateschätzer

2.2. Nichtlinearer Kleinst-Quadrate-Schätzer

Um einen **effizienten** Schätzer zu erhalten, ist es notwendig, dass für große Stichproben \mathbf{W} gegen \mathbf{X}_0 konvergiert.

- Einfaches **zweistufiges** Verfahren:

1. Schätze mit einer zulässigen Matrix \mathbf{W} mit dem Momentenschätzer den Parametervektor β , so dass β konsistent geschätzt wird. Bezeichne den Parameterschätzer mit $\hat{\beta}$.
2. Schätze β abermals mit dem Momentenschätzer, diesmal jedoch mit $\mathbf{W} = \mathbf{X}(\hat{\beta})$.

Da $\text{plim } \hat{\beta} = \beta_0$ und somit $\mathbf{X}(\hat{\beta})$ gegen $\mathbf{X}(\beta_0)$ mit wachsendem Stichprobenumfang tendiert, ergibt sich ein effizienter Schätzer.

- In endlichen Stichproben wird die verfügbare Information in dem zweistufigen Verfahren nicht optimal ausgenutzt, da man ja β in der Momentenbedingung gleichzeitig schätzen könnte, also

$$\mathbf{X}^T(\beta) (\mathbf{y} - \mathbf{x}(\beta)) = \mathbf{0} \quad (2.7)$$

betrachten könnte. Falls der Schätzer aus der Momentenbedingung (2.7) konsistent ist, sollte zumindest für große Stichproben $\hat{\beta}$ nahe am wahren Parametervektor β_0 liegen.

- Die Lösung der in β nichtlinearen Momentenbedingungen (2.7) ist identisch mit der **Minimierung der Fehlerquadratsumme**

$$SSR(\beta) = \sum_{t=1}^n (y_t - x_t(\beta))^2 = (\mathbf{y} - \mathbf{x}(\beta))^T (\mathbf{y} - \mathbf{x}(\beta)). \quad (2.8)$$

Da aufgrund der in β nichtlinearen Funktionen $x_t(\beta)$ keine geschlossene Lösung existiert, wird der Schätzer $\hat{\beta}$ auf Basis von (2.7) bzw. (2.8) als **nichtlinearer Kleinst-Quadrate-Schätzer (nonlinear least squares (NLS) estimator)** bezeichnet.

Eigenschaften des nichtlinearer KQ-Schätzers

- Die endlichen Stichprobeneigenschaften des NLS-Schätzers

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} SSR(\boldsymbol{\beta}) = \arg \min_{\boldsymbol{\beta}} (\mathbf{y} - \mathbf{x}(\boldsymbol{\beta}))^T (\mathbf{y} - \mathbf{x}(\boldsymbol{\beta})) \quad (2.9)$$

lassen sich nicht ohne Kenntnis des DGP bestimmen. Die asymptotischen Eigenschaften entsprechen denen des effizienten Momentenschätzers.

- **Konsistenz**

- Modifikation von Annahme **(NL3)** indem $\mathbf{W}_t = \mathbf{X}_t(\boldsymbol{\beta})$ gewählt wird:

- * **(NL3')** Der Parametervektor $\boldsymbol{\beta}$ ist asymptotisch identifizierbar, d. h.

$$\boldsymbol{\alpha}_{NLS}(\boldsymbol{\beta}) \equiv \text{plim} \frac{1}{n} \mathbf{X}^T(\boldsymbol{\beta})(\mathbf{y} - \mathbf{x}(\boldsymbol{\beta})) \quad (2.10)$$

existiert für alle zulässigen $\boldsymbol{\beta}$ und $\mathbf{X}(\boldsymbol{\beta}) \in \Omega_t$ und es gilt $\boldsymbol{\alpha}_{NLS}(\boldsymbol{\beta}) \neq \boldsymbol{\alpha}_{NLS}(\boldsymbol{\beta}_0) = \mathbf{0}$, wenn $\|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| > 0$.

Das Gleichheitszeichen gilt, da $\mathbf{X}_t(\boldsymbol{\beta})$ für jedes beliebige $\boldsymbol{\beta}$ in der Informationsmenge Ω_t zum Zeitpunkt t liegt, so dass **(NL2a)** und damit $E(\mathbf{X}_t(\boldsymbol{\beta})u_t) = 0$ erfüllt ist.

- **Konsistenz:** Unter **(NL1)**, **(NL2)**, **(NL3')** und **(NL4)** und weiteren technischen Regularitätsbedingungen, ist der nichtlineare Kleinst-Quadrate-Schätzer (2.9) konsistent.

- **Asymptotische Normalität**

- Gelten

- * die Annahmen **(NL1)**, **(NL2)**, **(NL3')** und **(NL4)** ($\hat{\beta}$ konsistent) und

- * Annahmen **(NL5)** bis **(NL7)** mit $\mathbf{W} = \mathbf{X}(\beta_0) = \mathbf{X}_0$,

dann ist der nichtlineare KQ-Schätzer (2.9) asymptotisch normalverteilt

$$\sqrt{n} \left(\hat{\beta} - \beta_0 \right) \xrightarrow{d} N \left(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{X}_0^T \mathbf{X}_0}^{-1} \right), \quad (2.11)$$

mit

$$\mathbf{S}_{\mathbf{X}_0^T \mathbf{X}_0} \equiv \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}_0^T \mathbf{X}_0.$$

- Eine Beweisskizze ist komplizierter als für den Momentenschätzer, da $\mathbf{W} = \mathbf{X}(\beta)$ nicht mehr fix ist, sondern sich mit β ändert. Man kann dies berücksichtigen, indem man eine Taylorapproximation von $\mathbf{X}(\beta)$ an der Stelle β_0 macht

und zeigt, dass alle Terme mit Ordnung zwei und höher asymptotisch keine Rolle spielen, siehe Davidson und MacKinnon (2004, Section 6.3). Ein ausführlicher Beweis findet sich in Davidson und MacKinnon (1993, Chapter 5).

- **Zu den Annahmen:**

- Insbesondere die Annahmen **(NL3)**, **(NL4)**, **(NL6)**, **(NL7)** sind nicht leicht überprüfbar, sollten aber für ein gewähltes nichtlineares Regressionsmodell überprüft werden. Dies übersteigt jedoch das Niveau dieses Kurses. Allerdings machen sie deutlich, dass z.B. ein Zeittrend als Regressor ausgeschlossen ist. Warum?
- Liegen Zeitreihendaten vor, wird dies noch komplizierter, da nicht nur das Abklingen der Autokorrelation mit zunehmenden Lags begrenzt werden muss, sondern das Abklingen der gesamten stochastischen Abhängigkeit zwischen den Beobachtungen. Dies übersteigt ebenfalls das Niveau dieses Kurses.

- **Effizienz**

Der nichtlineare KQ-Schätzer ist asymptotisch effizient, da die asymptotische Kovarianzmatrix des Momentenschätzers

$$\sigma_0^2 \mathbf{S}_{\mathbf{W}^T \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{W}^T \mathbf{W}} (\mathbf{S}_{\mathbf{W}^T \mathbf{X}}^T)^{-1}$$

sich wesentlich vereinfacht, denn nun lässt sich unter den genannten (Regularitäts)bedingungen zeigen, dass

$$\mathbf{S}_{\mathbf{W}^T \mathbf{X}} = \mathbf{S}_{\mathbf{X}_0^T \mathbf{X}_0} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T(\hat{\boldsymbol{\beta}}) \mathbf{X}(\hat{\boldsymbol{\beta}}). \quad (2.12)$$

Damit ergibt sich die asymptotische Kovarianzmatrix $\sigma_0^2 \mathbf{S}_{\mathbf{X}_0^T \mathbf{X}_0}^{-1}$.

- **In der Praxis** wird die approximative Verteilung

$$\hat{\beta} \stackrel{a}{\sim} N \left(\beta, s^2 \left(\mathbf{X}(\hat{\beta})^T \mathbf{X}(\hat{\beta}) \right)^{-1} \right) \quad (2.13)$$

angewendet.

- **Schätzung der Varianz des Störterms:**

Es ist vorteilhaft,

$$s^2 = \frac{1}{n - k} \sum_{t=1}^n \hat{u}_t^2$$

anstelle des ML-Schätzers $\hat{\sigma}^2 = n^{-1} \sum_{t=1}^n \hat{u}_t^2$ zu verwenden. Der Grund ist prinzipiell derselbe wie im linearen Modell. Der ML -Schätzer $\hat{\sigma}^2$ unterschätzt die Fehlervarianz systematisch, da $E[SSR(\hat{\beta})] \leq n\sigma_0^2 = E[SSR(\beta_0)]$ aufgrund der Minimierung von SSR gelten muss. Dieser Nachteil tritt bei s^2 nicht auf.

Unter den getroffenen (Regularitäts)bedingungen kann die Konsistenz von s^2 gezeigt werden:

$$s^2 = \frac{1}{n - k} \sum_{t=1}^n \hat{u}_t^2 \xrightarrow{P} \sigma_0^2 \quad (2.14)$$

2.3. Numerische Optimierung

Es bezeichne $Q(\boldsymbol{\beta})$ eine **Zielfunktion**, die bezüglich des $(k \times 1)$ -Vektors $\boldsymbol{\beta}$ optimiert werden soll. Im Folgenden wird eine Minimierung betrachtet.

2.3.1. Newton's Methode/Newton-Raphson-Algorithmus

- **Idee:** Approximiere $Q(\boldsymbol{\beta})$ durch Taylorapproximation an der Stelle $\boldsymbol{\beta}_{(0)}$

$$\begin{aligned}
 Q(\boldsymbol{\beta}) &= Q(\boldsymbol{\beta}_{(0)}) + \mathbf{g}^T(\boldsymbol{\beta}_{(0)}) (\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)}) + \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)})^T \mathbf{H}(\bar{\boldsymbol{\beta}}) (\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)}) \\
 &\approx \underbrace{Q(\boldsymbol{\beta}_{(0)}) + \mathbf{g}^T(\boldsymbol{\beta}_{(0)}) (\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)}) + \frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)})^T \mathbf{H}(\boldsymbol{\beta}_{(0)}) (\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)})}_{\equiv Q^*(\boldsymbol{\beta})}
 \end{aligned}$$

wobei $\|\bar{\boldsymbol{\beta}} - \boldsymbol{\beta}_{(0)}\| \leq \|\boldsymbol{\beta} - \boldsymbol{\beta}_{(0)}\|$.

Notation:

- **Gradient:** Der $(k \times 1)$ -Vektor

$$\mathbf{g}(\boldsymbol{\beta}') \equiv \left. \frac{\partial Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}'}$$

fasst alle k partiellen Ableitungen erster Ordnung $\partial Q(\boldsymbol{\beta})/\partial \beta_j$, $j = 1, \dots, k$ an der Stelle $\boldsymbol{\beta} = \boldsymbol{\beta}'$ zusammen und wird als der **Gradient** von $Q(\cdot)$ bezeichnet.

- **Hessematrix:** Die $(k \times k)$ -Matrix

$$\mathbf{H}(\boldsymbol{\beta}') \equiv \left. \frac{\partial}{\partial \boldsymbol{\beta}^T} \mathbf{g}(\boldsymbol{\beta}) \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}'} = \left. \frac{\partial^2 Q(\boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}'}$$

ist die **Hessematrix** von $Q(\cdot)$. Sie enthält alle partiellen Ableitungen zweiter Ordnung bzw. alle partiellen Ableitungen erster Ordnung des Gradienten.

- Die Minimierung von $Q^*(\beta)$ ist aufgrund ihrer quadratischen Form sehr einfach. Nachdem man die partiellen Ableitungen erster Ordnung von $Q^*(\cdot)$ gleich Null gesetzt hat

$$\mathbf{g}(\beta_{(0)}) + \mathbf{H}(\beta_{(0)}) (\beta - \beta_{(0)}) \stackrel{!}{=} \mathbf{0},$$

kann man nach $\beta_{(0)}$ auflösen, sofern $\mathbf{H}(\cdot)$ nicht singulär (d. h. invertierbar) ist:

$$\beta_{(1)} = \beta_{(0)} - \mathbf{H}^{-1}(\beta_{(0)})\mathbf{g}(\beta_{(0)}), \quad (2.15)$$

wobei $\beta_{(1)}$ die Lösung bezeichnet. Der Parametervektor $\beta_{(1)}$ optimiert $Q^*(\beta)$.

- Ist $Q(\beta)$ nahe $Q^*(\beta)$, dann ist $\beta_{(1)}$ beinahe bereits die Lösung für $Q(\beta)$. Da im Optimum β_0 gilt, dass $Q(\beta_0) = Q^*(\beta_0)$, hat es Sinn, den Schritt (2.15) zu wiederholen bis

$$|Q(\beta_{(j)}) - Q(\beta_{(j-1)})| \leq \text{vorgegebene Toleranz}$$

erfüllt ist. Der Vektor $\hat{\beta} = \beta_{(j)}$ bezeichnet dann z.B. den nichtlineare KQ-Schätzer.

(EViews: Newton-Raphson, R-Funktion: `nlm`)

- **Problem:**

- Ist $Q(\beta)$ nicht global konvex und liegt der Startwert $\beta_{(0)}$ in einem konkaven Bereich, dann ist $\mathbf{H}(\cdot)$ negativ definit (im univariaten Fall würde dies bedeuten, dass die zweite Ableitung das falsche Vorzeichen hat), so dass der nächste Schritt in die falsche Richtung (weg vom Minimalwert) geht. Eine graphische Veranschaulichung finden Sie in [Davidson und MacKinnon \(2004, Figure 6.2\)](#).

Lösung: Quasi-Newton-Methoden und Gradientenverfahren.

2.3.2. Quasi-Newton-Methoden/Gradientenverfahren

- Quasi-Newton-Methoden verallgemeinern die Newton-Methode in zweierlei Hinsicht:
 - **Mögliche Wahl einer Schrittlänge** $\alpha_{(j)}$ Die Schrittlänge $\alpha_{(j)}$ wird im Allgemeinen mit einer univariaten Optimierung bestimmt. Im einfachsten Fall wird sie in jedem Schritt so gewählt, dass keine Verschlechterung eintritt, also bei Minimierung von $Q(\cdot)$ der neue Wert der Zielfunktion keinesfalls größer ist als im Schritt vorher.
 - **Ersetzen der Hessematrix** durch eine andere geeignete Matrix $\mathbf{D}(\cdot)$:

$$\boldsymbol{\beta}_{(j+1)} = \boldsymbol{\beta}_{(j)} - \alpha_{(j)} \mathbf{D}^{-1}(\boldsymbol{\beta}_{(j)}) \mathbf{g}(\boldsymbol{\beta}_{(j)}) \quad (2.16)$$

Die Matrix $\mathbf{D}(\boldsymbol{\beta}_{(j)})$ sollte die Hessematrix in der Umgebung des Extremums gut approximieren

- Alternative Algorithmen:

- **Gradientenverfahren** erfordern ausschließlich die Berechnung der partiellen Ableitungen erster Ordnung:

- * **Steilster Abstieg (steepest descent):** $\mathbf{D}(\boldsymbol{\beta}_{(j)}) = \mathbf{I}$.

- * **Gauss-Newton-Methode:** speziell für den Nichtlinearen-Kleinst-Quadrate-Schätzer konzipiert, siehe Abschnitt 2.3.3 $\mathbf{D}(\boldsymbol{\beta}_{(j)}) = \frac{2}{n} \mathbf{X}^T(\boldsymbol{\beta}_{(j)}) \mathbf{X}(\boldsymbol{\beta}_{(j)})$
(EViews, R-Funktion nls)

- * **Berndt, Hall, Hall, Hausman Algorithmus:** speziell für die Maximum-Likelihood-Schätzung, d. h. die Maximierung der Log-Likelihoodfunktion $l(\boldsymbol{\beta})$, siehe Abschnitt 3.2, konzipiert: $\mathbf{D}(\boldsymbol{\beta}_{(j)}) = \sum_{t=1}^T \frac{\partial l}{\partial \boldsymbol{\beta}} \Big|_{\boldsymbol{\beta}_{(j)}} \frac{\partial l}{\partial \boldsymbol{\beta}'} \Big|_{\boldsymbol{\beta}_{(j)}}$
(EViews, R-package maxLik)

- * **Marquardt-Algorithmus:** $\mathbf{D}(\boldsymbol{\beta}_{(j)}) = \frac{2}{n} \mathbf{X}^T(\boldsymbol{\beta}_{(j)}) \mathbf{X}(\boldsymbol{\beta}_{(j)}) + \mu \mathbf{I}$
(EViews)

– **Verfahren mit partiellen Ableitungen zweiter Ordnung:**

* **Quadratic hill climbing:** $\mathbf{D}(\boldsymbol{\beta}_{(j)}) = \mathbf{H}(\boldsymbol{\beta}_{(j)}) - \mu \mathbf{I}$.

Beachte, dass, je größer μ ist, diese Methode des steilsten Abstiegs immer näher kommt.

(EViews)

* **Levenberg-Marquardt-Algorithmus:** $\mathbf{D}^{-1}(\boldsymbol{\beta}_{(j)}) = \mathbf{H}^{-1}(\boldsymbol{\beta}_{(j)}) + \mu \mathbf{I}$

(EViews, R-Funktion `nls.lm` in R-package `minpack.lm`)

* **Scoring-Algorithmus:** speziell für die Maximum-Likelihood-Schätzung, d. h. die Maximierung der Log-Likelihoodfunktion $l(\boldsymbol{\beta})$, siehe Abschnitt

3.2, konzipiert: $\mathbf{D}(\boldsymbol{\beta}_{(j)}) = E \left[\frac{\partial^2 l}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \Big|_{\boldsymbol{\beta}_j} \right]$.

* **Davidon-Fletcher-Powell-Algorithmus (DFP-Algorithmus)** und **Broyden-Fletcher-Goldfarb-Shanno-Algorithmus (BFGS-Algorithmus):** Ma-

trix $\mathbf{D}^{-1}(\boldsymbol{\beta}_{(j)})$ nicht komplett neu berechnet, sondern schrittweise berechnet. Siehe z.B. (Davidson, 2000, Section 9.2.3).

(BFGS: R-Funktion `optim`)

2.3.3. Gauss-Newton Methode

- Wird als Zielfunktion $SSR(\boldsymbol{\beta})$ gewählt, dann lassen sich Gradient und Hessesche der Zielfunktion durch Gradient und Hessesche der nichtlinearen Regressionsfunktion ausdrücken. Damit erhält man

$$\mathbf{g}(\boldsymbol{\beta}) = -\frac{2}{n} \mathbf{X}^T(\boldsymbol{\beta}) (\mathbf{y} - \mathbf{x}(\boldsymbol{\beta})) \quad (2.17a)$$

$$H_{ij}(\boldsymbol{\beta}) = -\frac{2}{n} \sum_{t=1}^n \left[(y_t - x_t(\boldsymbol{\beta})) \underbrace{\frac{\partial X_{ti}(\boldsymbol{\beta})}{\partial \beta_j}}_{=\frac{\partial^2 x_t(\boldsymbol{\beta})}{\partial \beta_i \partial \beta_j}} - X_{ti}(\boldsymbol{\beta}) X_{tj}(\boldsymbol{\beta}) \right], \quad i, j = 1, \dots, k. \quad (2.17b)$$

- Man beachte, dass gegebene Annahme **(NL2a)** für $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ der Erwartungswert

$$E[H_{ij}(\boldsymbol{\beta}_0)] = \frac{2}{n} \sum_{t=1}^n X_{ti}(\boldsymbol{\beta}_0) X_{tj}(\boldsymbol{\beta}_0)$$

beträgt und damit auch $E\left(u_t \frac{\partial X_{ti}(\boldsymbol{\beta}_0)}{\partial \beta_j}\right) = 0$ gilt.

- Damit ist die Hessesche Matrix asymptotisch äquivalent zu

$$\mathbf{D}(\boldsymbol{\beta}_0) = \frac{2}{n} \mathbf{X}^T(\boldsymbol{\beta}_0) \mathbf{X}(\boldsymbol{\beta}_0). \quad (2.18)$$

- Das Bemerkenswerte an $\mathbf{D}(\boldsymbol{\beta})$ ist, dass die Matrix **immer** positiv definit ist, **sofern $\mathbf{X}(\boldsymbol{\beta})$ vollen Rang hat!**

– Verwendet man $\mathbf{D}(\boldsymbol{\beta})$ anstelle von $\mathbf{H}(\boldsymbol{\beta})$, vermeidet man das Problem der 'falschen Richtung'.

– Allerdings funktioniert die Minimierung nicht gut, wenn eben $\mathbf{D}(\boldsymbol{\beta})$ beinahe singulär ist. Abhilfe: Marquardt-Algorithmus, siehe Abschnitt 2.3.2.

- Durch Einsetzen von (2.17) und (2.18) in den Newton-Algorithmus erhält man für den $j + 1$ Iterationsschritt des **Gauss-Newton-Verfahrens**

$$\boldsymbol{\beta}_{(j+1)} = \boldsymbol{\beta}_{(j)} + \underbrace{\left(\mathbf{X}^T(\boldsymbol{\beta}_{(j)}) \mathbf{X}(\boldsymbol{\beta}_{(j)}) \right)^{-1} \mathbf{X}^T(\boldsymbol{\beta}_{(j)}) (\mathbf{y} - \mathbf{x}(\boldsymbol{\beta}_{(j)}))}_{\equiv \hat{\mathbf{b}}_{(j)}} \quad (2.19)$$

- Das Besondere am Gauss-Newton Verfahren ist, dass sich jede Iteration mit einer **Hilfsregression (artificial regression)** berechnen lässt, wobei $\hat{\mathbf{b}}_{(j)}$ den OLS-

Schätzer von $\mathbf{b}_{(j)}$ in der **Hilfsregression**

$$\mathbf{y} - \mathbf{x}(\boldsymbol{\beta}_{(j)}) = \mathbf{X}(\boldsymbol{\beta}_{(j)})\mathbf{b}_{(j)} + \text{errors} \quad (2.20)$$

bezeichnet. Die Hilfsregression (2.20) wird von Davidson und MacKinnon (2004, Section 6.5) und deshalb auch im Folgenden als **Gauss-Newton-Regression (GNR)** bezeichnet.

- Ergänzt man (2.19) um eine variable Schrittlänge $\alpha_{(j)}$ lautet der allgemeine Gauss-Newton-Iterationsschritt

$$\boldsymbol{\beta}_{(j+1)} = \boldsymbol{\beta}_{(j)} + \alpha_{(j)}\hat{\mathbf{b}}_{(j)}. \quad (2.21)$$

- Man beachte, dass die Hilfsregression an der Stelle $\hat{\boldsymbol{\beta}}$,

$$\mathbf{y} - \mathbf{x}(\hat{\boldsymbol{\beta}}) = \mathbf{X}(\hat{\boldsymbol{\beta}})\mathbf{b} + \text{residuals}$$

– einen Parameterschätzer von $\mathbf{0}$ erzeugt, da

$$\hat{\mathbf{b}} = \left(\mathbf{X}^T(\hat{\boldsymbol{\beta}})\mathbf{X}(\hat{\boldsymbol{\beta}}) \right)^{-1} \underbrace{\mathbf{X}^T(\hat{\boldsymbol{\beta}}) \left(\mathbf{y} - \mathbf{x}(\hat{\boldsymbol{\beta}}) \right)}_{=0 \text{ (Momentenbedingung)}}$$

- eine konsistente Schätzung der Kovarianzmatrix $\sigma_0^2 \mathbf{S}_{\mathbf{X}^T \mathbf{X}}$ erzeugt, da die Kovarianzmatrix von $\hat{\mathbf{b}}$ gerade

$$\widehat{Var}(\hat{\mathbf{b}}) = s^2 \left(\mathbf{X}^T(\hat{\boldsymbol{\beta}}) \mathbf{X}(\hat{\boldsymbol{\beta}}) \right)^{-1}$$

ergibt, wobei s^2 aus der NLS-Schätzung stammt (warum?).

- Diese Eigenschaft ist sehr praktisch, da zur Minimierung verschiedenste $\mathbf{D}(\cdot)$ Matrizen eingesetzt werden können, diese aber bei Abbruch der Optimierung dann nicht die Kovarianzmatrix schätzen. So kann man nach Beenden der Optimierung die Kovarianzmatrix aus der oben dargestellten Hilfsregression gewinnen.
- Jeder heteroskedastie-konsistente Schätzer für $Var(\mathbf{b})$ der Hilfsregression mit $\mathbf{y} - \mathbf{x}(\hat{\boldsymbol{\beta}})$ und $\mathbf{X}(\hat{\boldsymbol{\beta}})$ ergibt auch einen heteroskedastie-konsistenten Schätzer für $Var(\hat{\boldsymbol{\beta}})$.
- Man kann die Hilfsregression verwenden, um in einem Schritt einen asymptotisch effizienten Schätzer zu erhalten, sofern der Startwert konsistent geschätzt wurde (wie auch im Fall einer zweifachen Anwendung des Momentenschätzers). Siehe [Davidson und MacKinnon \(2004, Section 6.6\)](#) für einen Beweis.

2.3.4. Wichtige Anmerkungen

Methoden ohne Ableitungen:

Zielfunktionen, die nicht (partiell) differenzierbar sind, lassen sich mit **Methoden ohne Ableitungen** optimieren. Dazu gehören:

- **Gittersuche** (grid search): im univariaten Fall: Berechne Funktion an J bestimmten Punkten innerhalb eines Intervalls $[\beta_{min}, \beta_{max}]$, z.B. $\beta_j = \beta_{min} + j * \frac{\beta_{max} - \beta_{min}}{J}$, $j = 0, 1, \dots, J$.
(R: optimize bei univariater Optimierung)
- Axial-Line-Suche (für multivariate Probleme)(siehe z.B. Davidson (2000, Section 9.2.4))
- Conjugate-Directions-Methode (siehe z.B. Davidson (2000, Section 9.2.4))

Vor Beginn einer numerischen Optimierung:

- Alle nichtlinearen Optimierungsverfahren brauchen Startwerte. Die **Wahl der Startwerte** entscheidet mit, ob man in einem **lokalen** oder **globalen** Extremum landet.

Auswahlmethoden:

- beliebig
 - Gittersuche
 - Hinweise durch ökonomische Theorie
 - Ziehen aus einer Wahrscheinlichkeitsverteilung. Dann zunächst pro Startwerte geringe Anzahl an Iterationen wählen und aus allen Optimierungen das Beste Ergebnis wählen. Jetzt Anzahl der Iterationen erhöhen und Konvergenzkriterium senken.
- Verfahren mit partiellen Ableitungen benötigen den Gradientenvektor der Zielfunktion. Dieser kann, falls möglich, analytisch berechnet werden, oder mit Hilfe geeigneter Verfahren numerisch bestimmt werden (Standard/default).

- **Konvergenzkriterien**

Mögliche Konvergenzkriterien/Abbruchkriterien (stopping rules) sind

$$\begin{aligned} |Q(\boldsymbol{\beta}_{(j)}) - Q(\boldsymbol{\beta}_{(j-1)})| &< \epsilon_1 \\ \|\boldsymbol{\beta}_{(j)} - \boldsymbol{\beta}_{(j-1)}\| &< \epsilon_2 \\ \mathbf{g}^T(\boldsymbol{\beta}_{(j)})\mathbf{D}^{-1}(\boldsymbol{\beta}_{(j)})\mathbf{g}(\boldsymbol{\beta}_{(j)}) &< \epsilon_3, \end{aligned}$$

wobei ϵ_i im allgemeinen kleiner als 10^{-4} gewählt wird. Wählt man ϵ_i zu klein, läuft der Algorithmus möglicherweise zu lange, wählt man ϵ_i zu groß, ist man möglicherweise nicht nahe genug am wahren Extremum.

2.3.5. Beispiel IM - Fortsetzung von Abschnitt 1.2.1

Logistic Smooth Transition Regression, siehe für Details Abschnitt 2.5, insbesondere (2.31)

- **Spezifikation A**

$$\begin{aligned} \ln(\text{Importe}_i) = & \beta_1 + \beta_2 \ln(\text{BIP}_i) \\ & + (\beta_3 + \beta_4 \ln(\text{BIP}_i)) \left[1 + \exp(-\beta_5 (\ln(\text{BIP}_i) - \ln(\bar{\text{BIP}}))) \right]^{-1} \\ & + \beta_6 \ln(\text{Entfernung}_i) + \beta_7 \text{Offenheit}_i + \beta_8 \ln(\text{Flaeche}_i) + u_i \end{aligned} \quad (2.22)$$

Startwerte werden mittels Gittersuche für β_5 im Intervall $[0.0001, 5]$ in Schritten von 0.01 und dann möglicher KQ-Schätzung gesucht. Anschließend wird die Gauss-Newton-Methode angewendet.

R-Programm zu folgendem Output (mit R-Funktion nls geschätzt):

```
Formula: log(trade_0_d_o) ~ beta1 + beta2 * log(wdi_gdpusdcr_o) + (beta3 +  
  beta4 * log(wdi_gdpusdcr_o)) * LSTR1(s_trans, beta5, s_trans_mean) +  
  beta6 * log(cepii_dist) + beta7 * ebrd_tfes_o + beta8 * log(cepii_area_o)
```

Parameters:

	Estimate	Std. Error	t value	Pr(> t)	
beta1	-8.15754	4.87590	-1.673	0.1019	
beta2	1.41638	0.19456	7.280	6.66e-09	***
beta3	15.86989	7.34922	2.159	0.0367	*
beta4	-0.64552	0.27994	-2.306	0.0262	*
beta5	2.80227	11.20264	0.250	0.8037	
beta6	-0.75349	0.16168	-4.660	3.32e-05	***
beta7	0.42750	0.19642	2.176	0.0353	*
beta8	-0.13303	0.08482	-1.568	0.1245	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

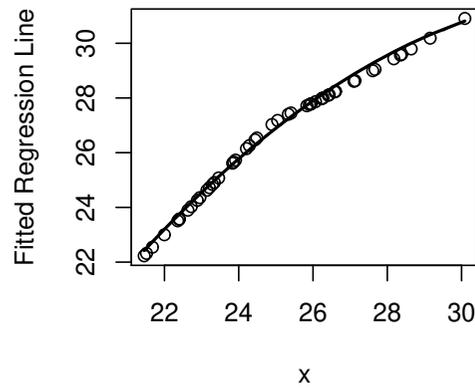
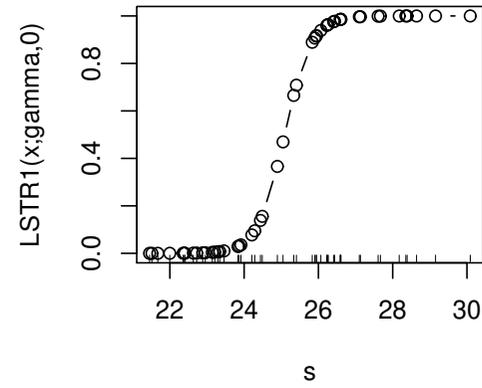
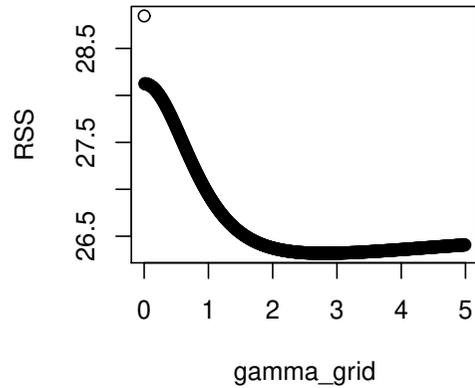
Residual standard error: 0.8012 on 41 degrees of freedom

Number of iterations to convergence: 41

Achieved convergence tolerance: 8.227e-06

(1 observation deleted due to missingness)

RSS für verschiedene β_5 , Geschätzte Übergangsfunktion, Geschätzte nichtlineare Regressionsfunktion (nur in Abhängigkeit von $\ln(BIP)$)



Wie häufig der Fall, ist der Übergangparameter β_5 (überhaupt) nicht signifikant. Eine Möglichkeit ist, $\beta_3 = 0$ zu setzen. Dies ergibt die folgende Spezifikation.

• **Spezifikation B**

$$\begin{aligned} \ln(\text{Importe}_i) = & \beta_1 + \beta_2 \ln(\text{BIP}_i) \\ & + \beta_3 \ln(\text{BIP}_i) \left[1 + \exp \left(-\beta_4 (\ln(\text{BIP}_i) - \ln(\bar{\text{BIP}})) \right) \right]^{-1} \\ & + \beta_5 \ln(\text{Entfernung}_i) + \beta_6 \text{Offenheit}_i + \beta_7 \ln(\text{Flaeche}_i) + u_i \end{aligned} \tag{2.23}$$

R-Programm zu folgendem Output (mit R-Funktion nls geschätzt):

```
Formula: log(trade_0_d_o) ~ beta1 + beta2 * log(wdi_gdpusdcr_o) + beta3 *
log(wdi_gdpusdcr_o) * LSTR1(s_trans, beta4, s_trans_mean) +
beta5 * log(cepii_dist) + beta6 * ebrd_tfes_o + beta7 * log(cepii_area_o)
```

Parameters:

	Estimate	Std. Error	t value	Pr(> t)	
beta1	-45.2492	19.4546	-2.326	0.02493	*
beta2	3.4519	1.0111	3.414	0.00143	**
beta3	-1.1458	0.6442	-1.779	0.08254	.
beta4	0.2718	0.1212	2.242	0.03027	*
beta5	-0.7821	0.1653	-4.731	2.54e-05	***
beta6	0.4479	0.2014	2.224	0.03160	*
beta7	-0.1244	0.0823	-1.512	0.13810	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

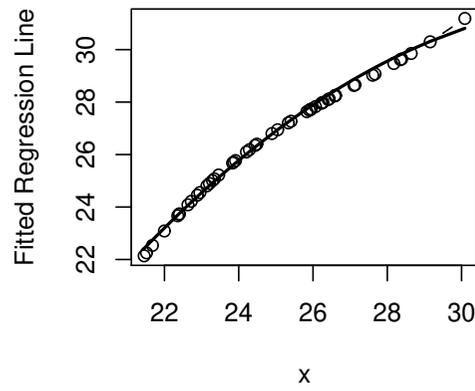
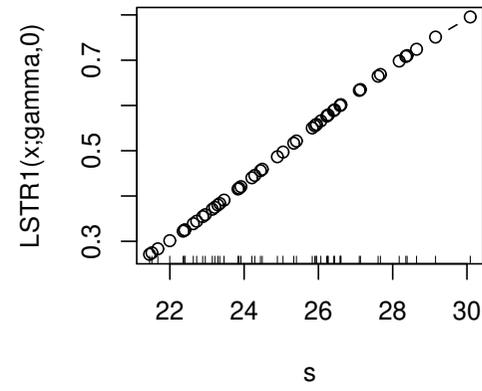
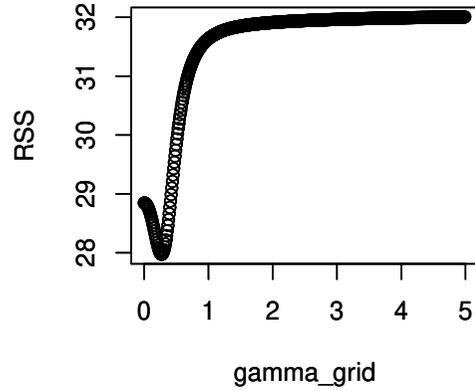
Residual standard error: 0.8161 on 42 degrees of freedom

Number of iterations to convergence: 3

Achieved convergence tolerance: 5.834e-07

(1 observation deleted due to missingness)

RSS für verschiedene β_5 , Geschätzte Übergangsfunktion, Geschätzte nichtlineare Regressionsfunktion (nur in Abhängigkeit von $\ln(BIP)$)



Beachte, dass der Verlauf der Übergangsfunktion nahe legt, dass $\ln(BIP_i)$ mit $\ln(BIP_i)$ multipliziert wird, man also folgendes lineares Regressionmodell schätzen könnte

$$\begin{aligned}\ln(Importe_i) = & \beta_1 + \beta_2 \ln(BIP_i) + \beta_3 \ln(BIP_i)^2 \\ & + \beta_4 \ln(Entfernung_i) + \beta_5 Offenheit_i + \beta_6 \ln(Flaeche_i) + u_i\end{aligned}\tag{2.24}$$

- Lineare Regressionen mit quadratischen oder kubischen Termen sind auch geeignet, um auf das Vorliegen von Nichtlinearität der STR-Form zu testen, siehe Abschnitt 2.6.

2.3.6. Beispiel CO₂ - Fortsetzung von Abschnitt 1.2.2

Analyse des Zusammenhangs zwischen dem CO₂-Ausstoß eines Landes und dessen BIP, sowie weiteren möglichen Einflussfaktoren:

Nichtlineare Regression (**Smooth Transition Regression**), $t = 1, \dots, 151$:

$$\ln(CO2_t) = b_1 + b_2 \ln(GDP_t) + (a_1 + a_2 \ln(GDP_t)) \frac{1}{1 + \exp(-\gamma(\text{Services}_t - c))} + \varepsilon_t. \quad (2.25)$$

mit Übergangsfunktion (**transition function**)

$$G(x; \gamma, c) = \frac{1}{1 + \exp(-\gamma(x - c))}, \quad 0 < G(\cdot) < 1.$$

Überlegen Sie sich, ob die Parameter in der nichtlinearen Regressionsfunktion in (2.25) für alle möglichen Parameterwerte (asymptotisch) identifiziert sind.

R-Programm zu folgendem Output (mit R-Funktion nls geschätzt):

Formula: $\log(\text{CO2}) \sim b1 + b2 * \log(\text{GDP}) + (a1 + a2 * \log(\text{GDP})) / (1 + \exp(-\text{gamma} * (\text{Services} - c)))$

Parameters:

	Estimate	Std. Error	t value	Pr(> t)	
b1	-12.12224	0.77787	-15.584	< 2e-16	***
b2	1.45266	0.09426	15.411	< 2e-16	***
a1	4.04036	1.39548	2.895	0.00435	**
a2	-0.46327	0.15387	-3.011	0.00305	**
gamma	0.46274	0.80529	0.575	0.56640	
c	55.47209	3.43542	16.147	< 2e-16	***

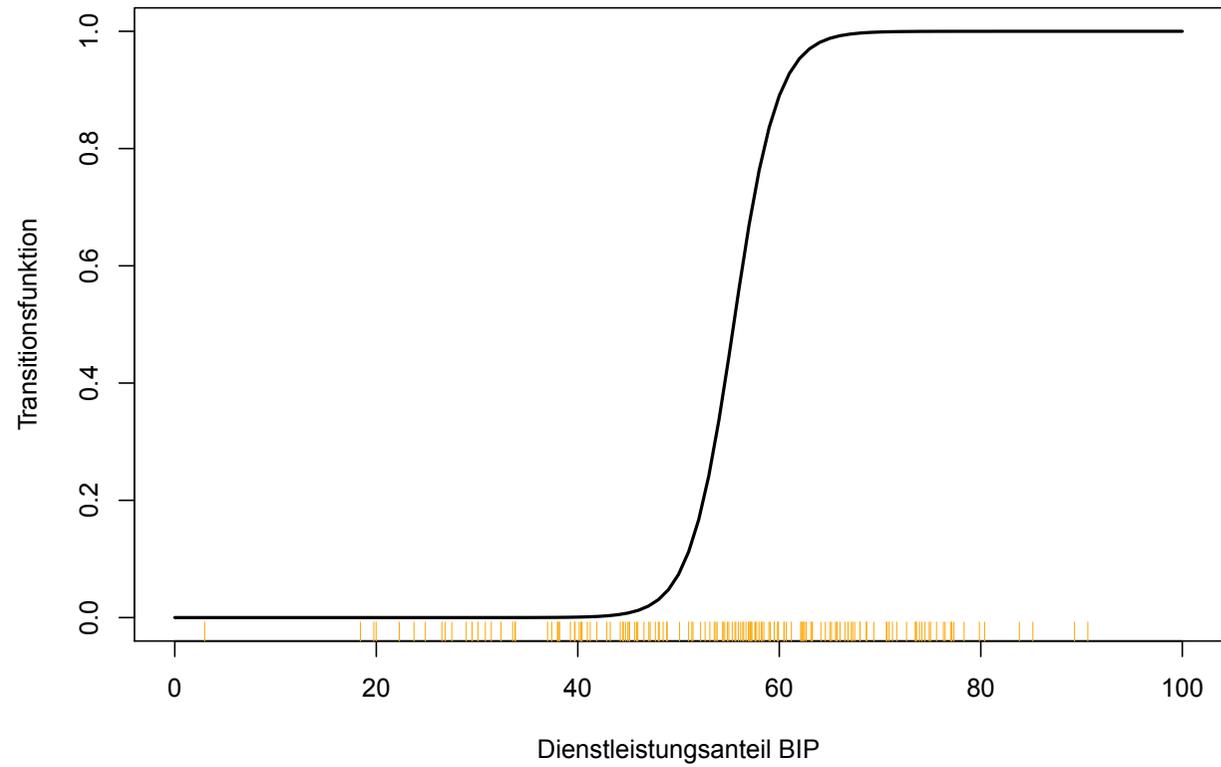
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7768 on 151 degrees of freedom

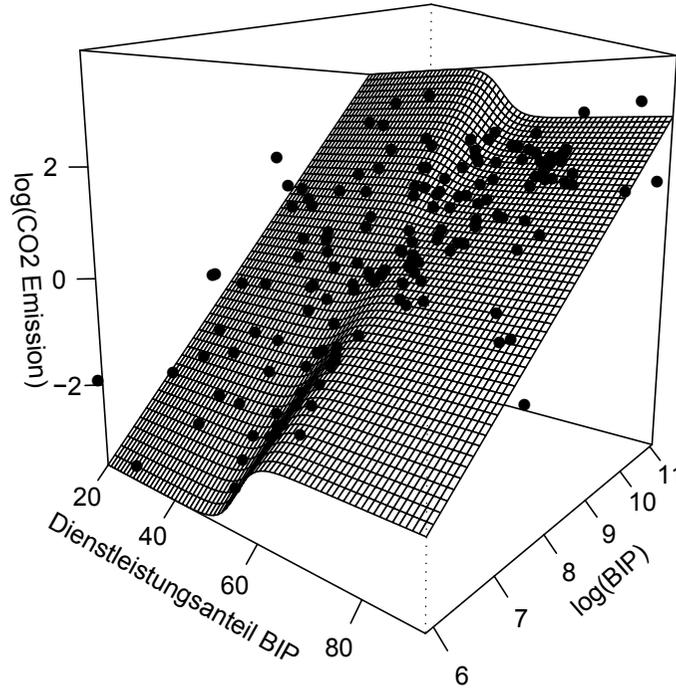
Number of iterations to convergence: 20

Achieved convergence tolerance: 5.882e-06

Geschätzte Switchingfunktion $G(\text{Services})$



Geschätzte nichtlineare Regressionsfunktion



- Ist das nichtlineare Regressionsmodell korrekt spezifiziert?
- Sind die Parameter signifikant?

2.4. Hypothesentests

Testen von **linearen Restriktionen** bezüglich des Parametervektors mit dem Hypothesenpaar:

$$H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r} \quad \text{versus} \quad H_1 : \mathbf{R}\boldsymbol{\beta} \neq \mathbf{r}$$

wobei \mathbf{R} eine $(q \times k)$ Matrix und \mathbf{r} ein $(q \times 1)$ Vektor sind.

- **Wald-Test:** Ableitung der Teststatistik und -verteilung

- Ist $\hat{\boldsymbol{\beta}}$ asymptotisch normalverteilt (Annahmen **(NL1)**, **(NL2)**, **(NL3')**, **(NL4)** bis **(NL7)**), gilt **(2.11)**:

$$\sqrt{n} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) \xrightarrow{d} N \left(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{X}_0^T \mathbf{X}_0}^{-1} \right). \quad (2.11)$$

- Unter H_0 , d. h. entspricht der DGP der Nullhypothese, gilt $\mathbf{r} = \mathbf{R}\boldsymbol{\beta}_0$. Somit erhält man $\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} = \mathbf{R} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right)$ und wegen **(2.11)**

$$\sqrt{n} \mathbf{R} \left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 \right) \xrightarrow{d} N \left(\mathbf{0}, \sigma_0^2 \mathbf{R} \mathbf{S}_{\mathbf{X}_0^T \mathbf{X}_0}^{-1} \mathbf{R}^T \right). \quad (2.26)$$

- Die Kovarianzmatrix von $\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)$ kann wegen (2.12), (2.14) unter den genannten (Regularitäts)bedingungen durch

$$\widehat{Var}(\mathbf{R}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)) = s^2 \mathbf{R} \left(\mathbf{X}^T(\hat{\boldsymbol{\beta}}) \mathbf{X}(\hat{\boldsymbol{\beta}}) \right)^{-1} \mathbf{R}^T \xrightarrow{P} \sigma_0^2 \mathbf{R} \mathbf{S}_{\mathbf{X}_0^T \mathbf{X}_0}^{-1} \mathbf{R}^T \quad (2.27)$$

konsistent geschätzt werden.

- Da die geschätzten Abweichungen von der Nullhypothese asymptotisch normalverteilt sind, folgt aus dem Theorem für stetige Abbildungen (siehe **Methoden der Ökonometrie**, Abschnitt 3.5 *Konvergenz in Verteilung*) und der konsistenten Varianzschätzung (2.27) analog zu dem Vorgehen bei der Ableitung des asymptotischen F-Tests (siehe **Methoden der Ökonometrie**, Abschnitt 11.4 *Asymptotische Tests*), dass die gewichteten quadrierten geschätzten Abweichungen

$$\lambda_W = \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} \right)^T \left(\widehat{Var}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \right)^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} \right) \xrightarrow{d} \chi^2(q). \quad (2.28)$$

asymptotisch $\chi^2(q)$ -verteilt sind, wobei $\hat{\boldsymbol{\beta}}$ unter der Alternative, d. h. dem unrestringierten Modell geschätzt wird. (Die Teststatistik λ_W wird i.A. in Referenz an den Ökonometriker Wald (1943) als **Wald-Statistik** bezeichnet.)

- Alternativ lässt sich eine F -**Statistik** der Form

$$F = \frac{(SSR_{H_0} - SSR_{H_1})/q}{SSR_{H_1}/(n - k)} \xrightarrow{d} F_{q,\infty} \quad (2.29)$$

verwenden. Der Nachweis für die asymptotische F -Verteilung erfolgt weiter unten für den einfacheren Fall von Nullrestriktionen.

- **Anmerkungen:**

- Im linearen Regressionsmodell existieren verschiedene äquivalente Darstellungen der F -Statistik, insbesondere

$$\begin{aligned} F &= \frac{(SSR_{H_0} - SSR_{H_1})/q}{SSR_{H_1}/(n - k)} \\ &= \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} \right)^T \left(\widehat{Var}(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r}) \right)^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}} - \mathbf{r} \right) / q. \end{aligned}$$

Im nichtlinearen Regressionsmodell sind die entsprechende Waldstatistik λ_W und die F -Statistik F nur noch asymptotisch — jedoch *nicht* in endlichen Stichproben — äquivalent.

- Es hat sich in Monte Carlo Simulationen gezeigt, dass in kleinen Stichproben das nominale und tatsächliche Signifikanzniveau besser übereinstimmen, wenn anstelle der asymptotisch gültigen $F_{q,\infty}$ Verteilung die $F_{q,n-k}$ -Verteilung genommen wird (diese wird bspw. auch in EViews zur Berechnung der p -Werte verwendet).
- Offensichtlich gilt:

$$qF \xrightarrow{d} \chi^2(q)$$

Die Verwendung dieser Teststatistik mit den kritischen Werten aus der $\chi^2(q)$ -Verteilung entspricht dann der Verwendung der kritischen Werte aus der $F_{q,\infty}$ -Verteilung (z.B. werden diese Teststatistiken inklusive p -Werte in EViews angegeben).

- **Warnung:** Die asymptotischen Testverteilungen existieren nur dann, wenn
 - * der Parametervektor β konsistent geschätzt werden kann, und
 - * die Kovarianzmatrix invertierbar ist.

Dies setzt voraus, dass die Parameter (**asymptotisch**) **identifizierbar** sind. Diese Voraussetzung ist häufig verletzt, z.B., wenn man testen möchte, ob ein DGP linear ist. Dann sind der Wald-Test und der F-Test mit den üblichen asymptotischen Verteilungen *nicht* anwendbar.

– **Beispiele:**

- * Der DGP sei

$$y_t = \mathbf{X}_t \beta_0 + u_t, \quad u_t \sim NID(0, \sigma_0^2)$$

Geschätztes Modell:

$$y_t = \mathbf{X}_t \beta + x_{t2}^\gamma \delta + u_t, \quad u_t \sim NID(0, \sigma_0^2)$$

H_0 : DGP ist linear, d. h. $\gamma = 0$ und/oder $\delta = 0$ versus

H_1 : $\delta \neq 0$ und $\gamma \neq 0$

Da im DGP $\delta = 0$, kann γ jeden beliebigen Wert annehmen. Damit ist γ

nicht identifiziert. Welche Parameter sind nicht identifiziert, wenn $\gamma = 0$?

- * **Beispiel CO₂**: Ist in der nichtlinearen Regressionsfunktion (2.25) $\gamma = 0$, dann ist nur $a_1 + b_1$ sowie $a_2 + b_2$ identifiziert, nicht jedoch die einzelnen Parameter. Warum?

– **Auswege:**

- * Einschränken des Parameterraums (falls sinnvoll)
- * Alternative Testverfahren, z.B. mit Hilfe von Taylorapproximationen oder mit kompliziertere Asymptotik. Dazu u.U. später mehr.
- * RESET-Test auf Basis eines linearen Modells. Siehe z.B. BA-Kurs Ökonometrie I, Abschnitt Modelldiagnose.

- Spezialfall: Testen von **Nullrestriktionen**:

- Ein häufiger Spezialfall von Tests ist das Testen von Nullrestriktionen.

Notation:

$$\boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{pmatrix}$$

wobei $\boldsymbol{\beta}_1$ ein $(k_1 \times 1)$ und $\boldsymbol{\beta}_2$ ein $(k_2 \times 1)$ Vektor ist. Dann lässt sich das nichtlineare Modell schreiben als

$$\mathbf{y} = \mathbf{x}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) + \mathbf{u}$$

- Hypothesenpaar:

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{versus} \quad H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$$

bzw.

$$H_0 : \mathbf{y} = \mathbf{x}(\boldsymbol{\beta}_1, \mathbf{0}) + \mathbf{u} \quad \text{versus} \quad H_1 : \mathbf{y} = \mathbf{x}(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2) + \mathbf{u}.$$

Beachte: im nichtlinearen Fall sind Nullrestriktionen nicht (zwangsläufig) Ausschlussrestriktionen!

- Hier bietet sich die Berechnung der F -**Statistik** an, so wie im linearen Regressionsmodell mit $r = k_2$

$$F = \frac{(SSR_{H_0} - SSR_{H_1})/r}{SSR_{H_1}/(n - k)} \xrightarrow{d} F_{r, \infty}$$

Die asymptotische F -Verteilung ergibt sich aus der Tatsache, dass der nicht-lineare KQ-Schätzer unter geeigneten Annahmen asymptotisch normalverteilt ist. Der Nachweis (siehe [Davidson und MacKinnon \(2004, S. 243-244\)](#)) basiert auf folgender Beobachtung: asymptotisch ergeben sich die **unrestringierten NLS-Residuen $\hat{\mathbf{u}}$** (unter H_1) aus

$$\hat{\mathbf{u}} \stackrel{a}{=} \mathbf{y} - \mathbf{X}(\boldsymbol{\beta}_0) (\mathbf{X}^T(\boldsymbol{\beta}_0)\mathbf{X}(\boldsymbol{\beta}_0))^{-1} \mathbf{X}^T(\boldsymbol{\beta}_0)\mathbf{y} = \mathbf{M}_{\mathbf{X}(\boldsymbol{\beta}_0)}\mathbf{y} = \mathbf{M}_{\mathbf{X}(\boldsymbol{\beta}_0)}\mathbf{u}.$$

Entsprechendes gilt für die **restringierten NLS-Residuen $\tilde{\mathbf{u}}$** (unter H_0)

$$\tilde{\mathbf{u}} \stackrel{a}{=} \mathbf{y} - \mathbf{X}_1(\boldsymbol{\beta}_0) (\mathbf{X}_1^T(\boldsymbol{\beta}_0)\mathbf{X}_1(\boldsymbol{\beta}_0))^{-1} \mathbf{X}_1^T(\boldsymbol{\beta}_0)\mathbf{y} = \mathbf{M}_{\mathbf{X}_1(\boldsymbol{\beta}_0)}\mathbf{y} = \mathbf{M}_{\mathbf{X}_1(\boldsymbol{\beta}_0)}\mathbf{u}$$

mit

$$\mathbf{X}_1(\boldsymbol{\beta}_0) = \left. \frac{\partial \mathbf{x}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1^T} \right|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}.$$

Siehe auch **Methoden der Ökonometrie**, Abschnitte *11.3.2 F -Tests: Testen mehrerer Restriktionen*, *11.4 Asymptotische Tests*.

- **Nichtlineare Restriktionen** können z.B. mit der allgemeineren Version des Wald-Tests getestet werden, siehe hierzu die Ableitung für den Maximum-Likelihood-Schätzer in Abschnitt 3.6.
- **Weitere Anmerkungen:**
 - Es lassen sich auch Tests direkt auf Basis der Gauss-Newton Hilfsregressionen durchführen (siehe Davidson und MacKinnon (2004, S. 244-249)).
 - **Heteroskedastierobuste Tests** lassen sich entsprechend dem Vorgehen bei linearen Modellen konstruieren. Der wichtigste Unterschied ist, dass \mathbf{X} durch $\mathbf{X}(\beta_0)$ ersetzt wird. Siehe z.B. Davidson und MacKinnon (2004, Section 6.8).
 - **Bootstrap Tests** funktionieren für nichtlineare Modelle genauso wie für lineare Modelle. Allerdings ist deren Durchführung häufig wesentlich zeitintensiver, da ja im Falle des NLS-Schätzers für jede Schätzung einer Bootstrapstichprobe eine nichtlineare Optimierung durchgeführt werden muss. Um Computerzeit zu sparen, werden deshalb zur Erstellung der Bootstrapschätzungen häufig zweistufige Verfahren verwendet (s.o.).

- Für eine weiterführende Beschäftigung sollte man auch die Geometrie der nichtlinearen Schätzung/Optimierung studieren (siehe hierzu [Davidson und MacKinnon \(1993\)](#)).
- Es gibt außer dem Wald-Test noch zwei weitere wichtige Testansätze, nämlich den Likelihood-Ratio-Test und den Lagrange-Multiplikator-Test. Diese werden bei der Maximum-Likelihood-Schätzung in Abschnitt [3.6](#) eingeführt.

Zu lesen: [Davidson und MacKinnon \(2004, Chapter 6\)](#).

2.5. Smooth Transition (STR) Modelle

- Ausführliche Darstellungen finden sich in
 - [Teräsvirta, Tjøstheim, und Granger \(2010\)](#)
 - [Teräsvirta \(1998\)](#)
 - [Granger und Teräsvirta \(1993, Chapter 7\)](#)
- **Übersicht über die Modelle**

– Allgemeines STR-Modell

Stetiger Übergang zwischen linearen Regimen durch stetig differenzierbare und beschränkte Funktion(en) G . Im Fall von 2 Regimen erhält man

$$\begin{aligned} y_t &= \mathbf{X}_t \boldsymbol{\beta}_1 + \mathbf{X}_t \boldsymbol{\beta}_2 G(\gamma, c, s_t) + u_t, & t = 1, \dots, n \\ &= \mathbf{X}_t (\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 G(\gamma, c, s_t)) + u_t \end{aligned} \quad (2.30)$$

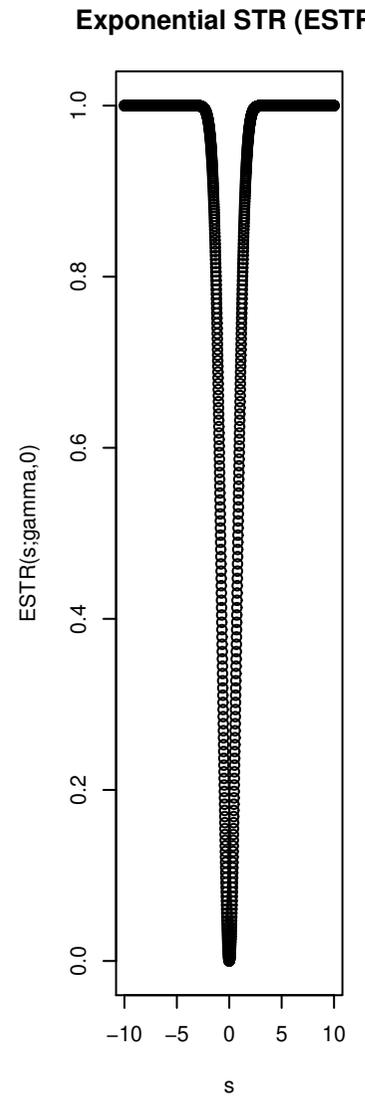
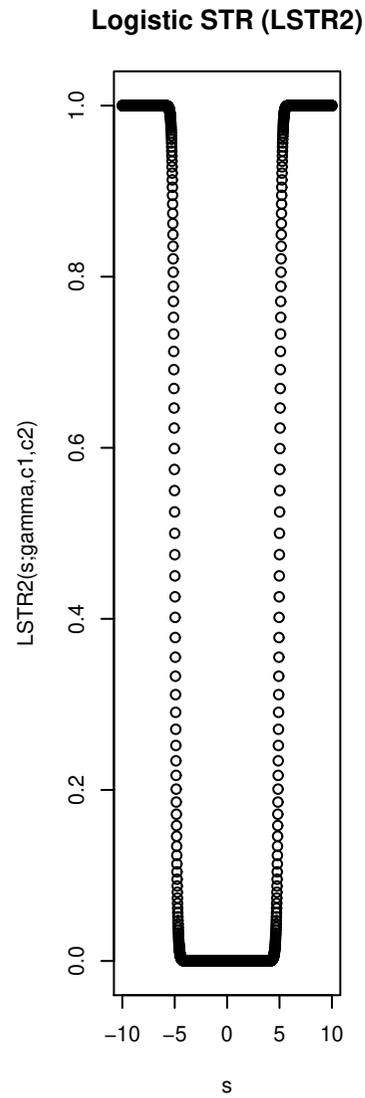
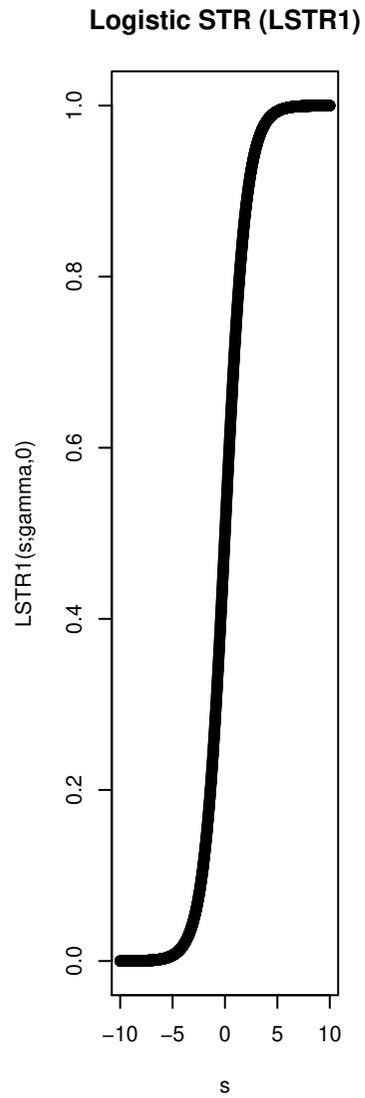
Im Allgemeinen:

- * $0 \leq G(\cdot) \leq 1$

- * s_t steuert den Übergang

Gebräuchliche Spezifikationen für die Übergangsfunktion (transition function) $G(\cdot)$ sind folgende:

Gebräuchliche Übergangsfunktionen (transition functions)



– Logistic STR (LSTR1) Model

$$G_1(\gamma, c, s_t) = \{1 + \exp[-\gamma(s_t - c)]\}^{-1}, \quad \gamma > 0 \quad (2.31)$$

Eigenschaften:

- * G_1 monoton steigend in s_t
- * γ bezeichnet die Übergangsgeschwindigkeit und muss positiv sein (Identifizierungsrestriktion)
- * c ist Übergangspunkt
- * $\gamma \rightarrow \infty$ führt zu einem 'switching regression model' (=diskreter Übergang zwischen zwei linearen Regimen) mit 2 Regimen.

– Logistic STR (LSTR2) Model

$$G_2(\gamma, c, s_t) = \{1 + \exp[-\gamma(s_t - c_1)(s_t - c_2)]\}^{-1}, \quad \gamma > 0, \quad c_1 \leq c_2 \quad (2.32)$$

Eigenschaften:

- * Spezialfall eines 3-Regime Modells.
- * G_2 symmetrisch an $\frac{c_1+c_2}{2}$.

$$* s_t \rightarrow \pm\infty \Rightarrow G_2 \rightarrow 1.$$

$$* 0 \leq \min_{s_t} G_2(\gamma, c, s_t) \leq \frac{1}{2} \text{ (oder } = \frac{1}{2}, \text{ falls } c_1 = c_2).$$

$$* \gamma \rightarrow \infty \Rightarrow G_2(\gamma, c, s_t) \rightarrow \begin{cases} 0 & \text{falls } c_1 \leq s_t \leq c_2, \\ 1 & \text{sonst.} \end{cases}$$

– Exponential (ESTR) Model

$$G(\gamma, c, s_t) = 1 - \exp[-\gamma(s_t - c)^2], \quad \gamma > 0 \quad (2.33)$$

Eigenschaften:

$$* G \text{ symmetrisch bei } c$$

$$* s_t \rightarrow \pm\infty \Rightarrow G \rightarrow 1$$

$$* s_t = c \Rightarrow G = 0$$

* $\gamma \rightarrow \infty \Rightarrow G \rightarrow 1$ außer im Fall $s_t = c \Rightarrow$ für große γ kann es schwierig sein, ESTR Model von einem linearen Modell zu unterscheiden.

2.6. Nichtlinearitätstests mit STR-Modellen in der Alternative

- Die Parameter der STR-Modelle sind nicht identifiziert, wenn der DGP linear ist. Deshalb kann man nicht einfach ein STR-Modell schätzen und dann Nullrestriktionstests durchführen, da die Teststatistiken keine asymptotische F -Verteilung aufweisen können!!
- **Ausweg:** Man approximiert die nichtlineare Regressionsfunktion durch eine nichtlineare Funktion, die auch unter einem linearen DGP identifiziert ist. Dies geschieht am einfachsten durch die Approximation von $f(\cdot)$ durch eine Taylorentwicklung (Granger und Teräsvirta (1993)).
- Zur Vereinfachung der weiteren Notation sei

$$G^*(\cdot) = G(\cdot) - 1/2$$

für alle genannten Übergangsfunktionen. Die STR-Modelle lassen sich dann schreiben als

$$y_t = \mathbf{X}_t \boldsymbol{\beta}_1^* + \mathbf{X}_t \boldsymbol{\beta}_2 G^*(\gamma, c, s_t) + u_t, \quad t = 1, \dots, n$$

wobei $\beta_1^* = \beta_1 + 1/2$.

- Das Hypothesenpaar lautet

$$H_0 : \gamma = 0 \quad \text{versus} \quad H_1 : \gamma \neq 0.$$

Man beachte, dass für $\gamma = 0$ die Funktionen $G(\cdot)$ und $G^*(\cdot)$ konstant sind und man deshalb nur ein lineares Regime hat!

- Die Taylorentwicklung erster Ordnung von $G(\cdot)^*$ nach s_t an der Nullhypothese $H_0 : \gamma = 0$ lautet

$$G_T^* = g_0 + g_1 s_t + R(\gamma, c; s_t)$$

wobei $g_0 = G_T(0, c, s_t) = \textit{konstant}$ und $g_1 = \partial G_T(\gamma, c, s_t) / \partial s_t |_{\gamma=0} = \textit{konstant}$ und $R(\gamma, c; s_t)$ Restterm bezeichnen. Einsetzen von $G_T^*(\cdot)$ in das STR-Modell liefert

$$\begin{aligned} y_t &= \mathbf{X}_t \beta_1^* + \mathbf{X}_t \beta_2 (g_0 + g_1 s_t + R(\gamma, c; s_t)) + u_t \\ &= \mathbf{X}_t \underbrace{(\beta_1^* + \beta_2 g_0)}_{\delta_1} + \mathbf{X}_t s_t \underbrace{\beta_2 g_1}_{\delta_2} + \underbrace{\mathbf{X}_t \beta_2 R(\gamma, c, s_t)}_{\nu_t} + u_t \\ y_t &= \mathbf{X}_t \delta_1 + \mathbf{X}_t s_t \delta_2 + \nu_t. \end{aligned} \tag{2.34}$$

- Man beachte, dass die erste partielle Ableitung der Übergangsfunktionen LSTR1, LSTR2 und ESTR für gegebenes s_t als $\partial G_T^* / \partial s_t = \gamma \omega$ geschrieben werden kann, wobei ω eine positive Konstante bezeichnet. Deshalb ergibt sich $\delta_2 = \gamma \beta_2 \omega$ und $H_0 : \gamma = 0$ impliziert

$$H'_0 : \delta_2 = \mathbf{0} \quad \text{versus} \quad H'_1 : \delta_2 \neq \mathbf{0}.$$

- Beachte: die Modellgleichung (2.34) ist
 - nur unter H'_1 eine Approximation während unter H'_0 die Gleichung den DGP enthält.
 - sowohl unter H'_0 als auch unter H'_1 identifiziert, sofern nicht \mathbf{X}_t und $\mathbf{X}_t s_t$ perfekt korreliert sind. Deshalb kann dann jeder Standardtest für lineare Modelle angewendet werden.
 - nicht identifiziert, falls \mathbf{X}_t eine Konstante enthält und s_t ein Element von \mathbf{X}_t ist, denn dann sind \mathbf{X}_t und $\mathbf{X}_t s_t$ perfekt korreliert. Warum? Die Lösung des Problems ist einfach: Man ersetzt $\mathbf{X}_t s_t$ durch $\tilde{\mathbf{X}}_t s_t$, wobei $\mathbf{X}_t = \begin{pmatrix} 1 & \tilde{\mathbf{X}}_t \end{pmatrix}$ und

schätzt

$$y_t = \mathbf{X}_t \boldsymbol{\delta}_1 + (\tilde{\mathbf{X}}_t s_t) \boldsymbol{\delta}_2 + \nu_t.$$

- **F-test**

Wenn \mathbf{X} und \mathbf{u} unabhängig sind und \mathbf{u} NID, dann ist unter $H_0 : \gamma = 0$ das wahre Modell ein normales lineares Regressionsmodell und somit auch (2.34). Deshalb lässt sich der F -Test anwenden mit *exakter* F -Verteilung

$$F = \frac{\tilde{\mathbf{v}}' \tilde{\mathbf{v}} - \hat{\mathbf{v}}' \hat{\mathbf{v}}}{\hat{\mathbf{v}}' \hat{\mathbf{v}}} \frac{T - L}{J} \stackrel{D}{\sim} F(J, n - L)$$

wobei

$$J = \begin{cases} k_1 & \text{falls } \mathbf{X}_t \text{ verwendet wird,} \\ k_1 - 1 & \text{falls } \tilde{\mathbf{X}}_t \text{ verwendet wird.} \end{cases}, \quad L = \begin{cases} 2k_1 & \text{falls } \mathbf{X}_t s_t \text{ verwendet wird} \\ 2k_1 - 1 & \text{falls } \tilde{\mathbf{X}}_t s_t \text{ verwendet wird.} \end{cases}$$

- Ist der DGP unter H_0 nicht normalverteilt und/oder \mathbf{X}_t nur partiell unabhängig, gelten die üblichen asymptotischen Ergebnisse für lineare Modelle und die F -Statistik ist asymptotisch $F_{J, \infty}$ verteilt.

- **Bemerkungen:**

- Dieser Linearitätstest scheint alle Probleme zu lösen. Man beachte jedoch, dass die Güte des Tests davon abhängt, wie gut die Taylorapproximation $G(\cdot)$ approximiert. Ist die Approximation ungenau, dann kann es leicht passieren, dass H_0' nur mit geringer Wahrscheinlichkeit abgelehnt wird, obwohl der DGP nichtlinear ist. Man hat dann ein Güteproblem!
- Es gibt noch völlig andere Testansätze, zu denen kommen wir vielleicht später.
- Wenn im wahren Modell $G(\cdot)$ nur auf die Konstante wirkt, dann ist im wahren Modell $\beta_2 = (\beta_{21}, 0, \dots, 0)'$. Da das Modell eine Konstante enthält man muss im Hilfsmodell (2.34) die Regressoren \tilde{X}_{tS_t} verwenden, wobei allerdings wegen β_2 man $\delta_2 = \mathbf{0}$ erhält — selbst unter H_1 . Um dieses Problem zu lösen, schlagen **Luukkonen, Saikkonen, und Teräsvirta (1988)** vor, eine Taylorentwicklung dritter Ordnung zu verwenden.
- Das approximierende Modell auf Basis einer Taylorentwicklung erster Ordnung ist für das ESTR und das LSTR2 identisch.
- Man beachte, dass die genannten Test nur gültig sind, wenn die Momente von

$\mathbf{X}_t s_t / n$ existieren.

- Das vorgestellte Verfahren kann immer als Linearitätstest verwendet werden. Allerdings ist die Güte am größten, wenn die Nichtlinearität das gewählte s_t betrifft. In der Praxis müssen deshalb die Tests für verschiedene s_t durchgeführt werden.

3. Maximum-Likelihood-Schätzung

Es gibt 2 allgemeine Schätzprinzipien:

- Momentenschätzer bzw. GMM, siehe Kapitel 6,
- Maximum-Likelihood-Schätzer (wobei dieser auch als GMM-Schätzer aufgefasst werden kann).

3.1. Einführendes Beispiel

- **Einführendes Beispiel:** Nach 10-maligem Münzwurf ergibt sich 9 mal 'Kopf' und 1 mal 'Zahl'. Glauben Sie, dass dies eine 'faire' Münze ist (eine Münze, für

die die Wahrscheinlichkeit, Kopf zu erhalten, gerade 0.5 ist)?

- Man beachte, dass die Wahrscheinlichkeit, von insgesamt n Würfeln k mal 'Kopf' zu erhalten, durch die Binomialverteilung

$$P('k \text{ mal 'Kopf' bei } n \text{ Würfeln}' | p) = \frac{n!}{(n-k)!k!} p^k (1-p)^{(n-k)} \quad (3.1)$$

gegeben ist, wobei p die Wahrscheinlichkeit angibt, in einem Wurf 'Kopf' zu erhalten.

- Deshalb ist die Wahrscheinlichkeit, das obengenannte Ergebnis zu erhalten, für unterschiedliche p :

$$p = 1/2 \quad \longrightarrow \quad P('k \text{ mal 'Kopf' bei } n \text{ Würfeln}' | p = 1/2) \approx 0.01$$

$$p = 3/4 \quad \longrightarrow \quad P('k \text{ mal 'Kopf' bei } n \text{ Würfeln}' | p = 3/4) \approx 0.19$$

$$p = 9/10 \quad \longrightarrow \quad P('k \text{ mal 'Kopf' bei } n \text{ Würfeln}' | p = 9/10) \approx 0.39$$

und man würde wohl sehr zögern, die benutzte Münze als 'fair' zu bezeichnen.

- Man kann nun die Verwendung der Wahrscheinlichkeitsfunktion (3.1) ändern und sie benutzen, um einem **gegebenen** Ereignis eine Wahrscheinlichkeit auf Basis eines gewählten Wertes für p zuzuweisen. Mit dieser Interpretation nennt

man (3.1) eine **Likelihood-Funktion**, um so die Verwendung von der als Wahrscheinlichkeitsfunktion (gegebenes p) unterscheiden zu können. Für den gegebenen Fall erhält man

$$L(p|k \text{ mal 'Kopf' bei } n \text{ Würfeln}') = \frac{n!}{(n-k)!k!} p^k (1-p)^{(n-k)}.$$

- Da man die Likelihood $L(p|k \text{ mal 'Kopf' bei } n \text{ Würfeln}')$ für ein gegebenes Ereignis, z.B. '9 mal 'Kopf' bei 10 Würfeln', für jedes beliebige p berechnen kann, kann man die Likelihood $L(p|k \text{ mal 'Kopf' bei } n \text{ Würfeln}')$ bezüglich p maximieren. Man erhält dann eine Schätzung \hat{p} für p , die *die Likelihood, das beobachtete Ereignis tatsächlich beobachten zu können, maximiert*. Deshalb wird dieser Schätzer **Maximum-Likelihood-Schätzer (ML-Schätzer)** genannt.
- Im gegebenen Fall kann man sehr einfach den ML-Schätzer \hat{p} ableiten, indem man die partielle Ableitung erster Ordnung von (3.1) bezüglich p gleich Null setzt und nach p auflöst.
- Sehr häufig ist es einfacher, die Likelihood zu maximieren, nachdem man logarithmiert hat. Dies hat keinen Einfluss auf die Schätzung, da Logarithmieren

eine strikt monotone Transformation ist. Es macht jedoch die analytische oder numerische Optimierung viel einfacher. Die **Log-Likelihood Funktion** lautet im gegebenen Fall

$$\begin{aligned} l(p|'k \text{ mal 'Kopf' bei } n \text{ Würfeln}') &= \ln L(p|'k \text{ mal 'Kopf' bei } n \text{ Würfeln}') \\ &= \ln \left(\frac{n!}{(n-k)!k!} \right) + k \ln p + (n-k) \log(1-p). \end{aligned}$$

Die erste Ableitung ist

$$\frac{\partial \ln L(p|\cdot)}{\partial p} = \frac{k}{p} - \frac{n-k}{1-p} \stackrel{!}{=} 0.$$

Die ML-Schätzung für p ist demnach $\hat{p} = k/n = 9/10$. (Der Vollständigkeit halber müsste man auch überprüfen, ob dieses Extremum ein Maximum ist. Hierfür muss die partielle Ableitung zweiter Ordnung in der Umgebung von \hat{p} negativ sein.)

3.2. Maximum-Likelihood-Schätzung im Falle stetiger Zufallsvariablen

- Für eine stetige Zufallsvariable Y gilt, dass die Wahrscheinlichkeit 'Y nimmt den Wert y an' gerade Null ist, d. h. $P(Y = y) = 0$.
- Stattdessen betrachtet man die Wahrscheinlichkeit, dass Y in einem (sehr) kleinen Intervall liegt

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

siehe "Eine kurze Einführung in die [Wahrscheinlichkeitstheorie Sommer 2009](#)",
Diese Wahrscheinlichkeit ist approximativ proportional zur Wahrscheinlichkeitsdichte $f(y)$.

- Entsprechend ergibt sich, falls die Wahrscheinlichkeit und die Dichte vom Parametervektor θ abhängen,

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

Die Maximierung der Likelihood, Y in einem winzigen Intervall um y herum zu beobachten, kann deshalb erfolgen, indem man die Dichte bezüglich θ maximiert.

Für stetige Zufallsvariable erhält man deshalb die Interpretation der **Likelihood-Funktion**

$$L(\boldsymbol{\theta}|y) = f(y|\boldsymbol{\theta}).$$

Der **ML-Schätzer $\hat{\boldsymbol{\theta}}$ für $\boldsymbol{\theta}$** ist deshalb gegeben durch

$$\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|y) \quad (= \max_{\boldsymbol{\theta}} f(y|\boldsymbol{\theta})).$$

- Um den ML-Schätzer für ein spezifisches Problem abzuleiten, muss man deshalb eine geeignet parametrisierte Dichtefunktion wählen.
- Für eine Stichprobe von n Beobachtungen $\mathbf{y} = \left(y_1 \ y_2 \ \dots \ y_n \right)^T$ ist die Likelihood-Funktion die gemeinsame Dichte bezüglich $\boldsymbol{\theta}$

$$L(\boldsymbol{\theta}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\theta}).$$

- Die gemeinsame Dichte für n IID Beobachtungen ist das Produkt der n marginalen Dichten. (Vgl. in **Methoden der Ökonometrie**, Abschnitt 2.6 *Bedingte Wahrscheinlichkeiten*.) Die Likelihood lautet deshalb

$$L(\boldsymbol{\theta}|\mathbf{y}) = f(\mathbf{y}|\boldsymbol{\theta}) = f(y_1|\boldsymbol{\theta}) \cdots f(y_n|\boldsymbol{\theta})$$

und die **Log-Likelihood ist die Summe der Log-Likelihood für jede einzelne Beobachtung**

$$l(\boldsymbol{\theta}|\mathbf{y}) = \ln f(\mathbf{y}|\boldsymbol{\theta}) = \sum_{t=1}^n \ln f(y_t|\boldsymbol{\theta}).$$

Diese Eigenschaft ist sehr praktisch zur Maximierung der (Log)-Likelihood!!

- Im Fall von nicht-IID Beobachtungen kann man folgende Zerlegung verwenden

$$\begin{aligned} L(\boldsymbol{\theta}|\mathbf{y}) &= f(\mathbf{y}|\boldsymbol{\theta}) \\ &= f(y_n|y_{n-1}, \dots, y_1; \boldsymbol{\theta}) f(y_{n-1}, y_{n-2}, \dots, y_1|\boldsymbol{\theta}) \\ &= f(y_n|y_{n-1}, \dots, y_1; \boldsymbol{\theta}) f(y_{n-1}|y_{n-2}, \dots, y_1; \boldsymbol{\theta}) f(y_{n-2}, \dots, y_1|\boldsymbol{\theta}) \\ &= f(y_n|y_{n-1}, \dots, y_1; \boldsymbol{\theta}) f(y_{n-1}|y_{n-2}, \dots, y_1; \boldsymbol{\theta}) \cdots f(y_2|y_1; \boldsymbol{\theta}) f(y_1|\boldsymbol{\theta}). \end{aligned}$$

Nach Logarithmieren erhält man die Summe

$$\begin{aligned} l(\boldsymbol{\theta}|\mathbf{y}) &= \ln f(\mathbf{y}|\boldsymbol{\theta}) \\ &= \sum_{t=2}^n \ln f(y_t|y_{t-1}, \dots, y_1; \boldsymbol{\theta}) + \ln f(y_1|\boldsymbol{\theta}). \end{aligned}$$

- Wenn der Term $\ln f(y_1|\boldsymbol{\theta})$ vernachlässigt wird, erhält man die **bedingte Likelihood-Funktion** bedingt auf y_1 . Ihre Maximierung ergibt den **bedingten Maximum-Likelihood-Schätzer**.
- Eine generelle Anmerkung:

Die Ableitung eines ML-Schätzers erfordert eine **vollständige Spezifikation** des Modells, d. h. die gemeinsame Dichte aller abhängigen Beobachtungen ist bis auf die spezifischen Parameterwerte bekannt. Vgl. **Methoden der Ökonometrie**, Abschnitt 5.2 *Ökonometrische Modelle*.

3.3. ML-Schätzung des normalen linearen Regressionsmodells

- Gegeben sei das normale multiple lineare Regressionsmodell

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \quad t = 1, 2, \dots, n, \quad \boldsymbol{\beta} \in \mathbb{B}, \quad (3.2)$$

so dass die Annahmen **(B1)**, **(B3)** und **(B4)** erfüllt sind. Durch die Normalverteilungsannahme

$$\begin{aligned} \mathbf{u}|\mathbf{X} &\sim N(\mathbf{0}, \sigma^2\mathbf{I}), \quad \text{bzw.} && ((\mathbf{B4})) \\ u_t|\mathbf{X} &\sim NID(0, \sigma^2), \quad t = 1, 2, \dots, n. \end{aligned}$$

erhält man die Dichte der Normalverteilung

$$\begin{aligned} f(u_t|\mathbf{X}; \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{u_t^2}{\sigma^2}\right), \quad t = 1, 2, \dots, n, \quad \text{bzw.} \\ f(\mathbf{u}|\mathbf{X}; \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{u}^T\mathbf{u}\right). \end{aligned}$$

Wegen **(B1)** und **(B4)** gilt $\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, $Var(\mathbf{y}|\mathbf{X}) = \sigma^2\mathbf{I}$ und damit

$$\begin{aligned} f(y_t|\mathbf{X}_t; \boldsymbol{\beta}, \sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{(y_t - \mathbf{X}_t\boldsymbol{\beta})^2}{\sigma^2}\right), \quad t = 1, 2, \dots, n, \quad \text{bzw.} \\ f(\mathbf{y}|\mathbf{X}, \boldsymbol{\beta}, \sigma^2) &= \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right). \end{aligned}$$

Die Likelihood-Funktion lautet demnach

$$L(\boldsymbol{\beta}, \sigma^2|\mathbf{y}, \mathbf{X}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right).$$

Logarithmieren ergibt die Log-Likelihood-Funktion

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma | \mathbf{y}, \mathbf{X}) &= \sum_{t=1}^n \left\{ -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{[y_t - \mathbf{X}_t\boldsymbol{\beta}]^2}{2\sigma^2} \right\} \\ &= \text{const} - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{t=1}^n [y_t - \mathbf{X}_t\boldsymbol{\beta}]^2 \\ &= \text{const} - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= \text{const} - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} SSR(\boldsymbol{\beta}). \end{aligned} \quad (3.3)$$

Man sieht, dass die **Maximierung der Log-Likelihood (3.3)** bezüglich $\boldsymbol{\beta}$ **im Fall des normalen linearen Regressionsmodells identisch** ist mit dem **KQ-Schätzer**. Im Fall streng exogener Regressoren sind deshalb bezüglich der Schätzung von $\boldsymbol{\beta}$ auch die **Schätzeigenschaften identisch**.

- Gilt die Annahme **(B2a)**, die in Annahme **(B4)** enthalten ist, nicht, d. h. sind die Regressoren nicht streng exogen, sondern lediglich vorherbestimmt bzgl. des Fehlerterms u_t (**(C2a)**), ist die asymptotische Verteilung des ML-Schätzers zu verwenden, die im Folgenden abgeleitet wird.

3.4. Asymptotische Eigenschaften des Maximum-Likelihood-Schätzers

3.4.1. Identifikation bei ML-Schätzung

- Im Folgenden werden die Definitionen für Identifikation aus Abschnitt 2.1 präzisiert und auf ML-Schätzer angepasst.
- Ein Parametervektor $\theta_1 \in \Theta$ heißt **für eine gegebene Stichprobe global/lokal identifiziert**, wenn für die Log-Likelihood gilt

$$l(\theta_1 | \mathbf{y}, \mathbf{X}) = \max_{\theta \in \Theta} l(\theta | \mathbf{y}, \mathbf{X}) \neq l(\theta_2 | \mathbf{y}, \mathbf{X}) \quad \text{falls } \theta_1 \neq \theta_2. \quad (3.4)$$

Globale Identifikation liegt vor, wenn $\theta_2 \in \Theta$, lokale Identifikation liegt vor, wenn θ_2 in einer offenen Umgebung um θ_1 liegt.

- Ein Parametervektor $\theta_1 \in \Theta$ heißt **global/lokal identifiziert**, wenn gilt

$$E[l(\theta_1 | \mathbf{y}, \mathbf{X})] = \max_{\theta \in \Theta} E[l(\theta | \mathbf{y}, \mathbf{X})] \neq E[l(\theta_2 | \mathbf{y}, \mathbf{X})] \quad \text{falls } \theta_1 \neq \theta_2. \quad (3.5)$$

Globale Identifikation liegt vor, wenn $\theta_2 \in \Theta$, lokale Identifikation liegt vor, wenn

θ_2 in einer offenen Umgebung um θ_1 liegt.

- Ein Parametervektor $\theta_1 \in \Theta$ heißt **asymptotisch global/lokal identifiziert**, wenn für die Log-Likelihood gilt

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\theta_1 | \mathbf{y}, \mathbf{X}) \neq \text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\theta_2 | \mathbf{y}, \mathbf{X}) \quad \text{falls } \theta_1 \neq \theta_2. \quad (3.6)$$

Globale Identifikation liegt vor, wenn $\theta_2 \in \Theta$, lokale Identifikation liegt vor, wenn θ_2 in einer offenen Umgebung um θ_1 liegt.

- Die Identifikation eines Parametervektors für andere Schätzverfahren, z.B. KQ erfordert weniger, da nicht die ganze bedingte Dichte, sondern die ersten und zweiten bedingten Momente betrachtet werden.
- Vgl. Davidson und MacKinnon (2004, Sections 6.2, 10.3). Beachte: die Definitionen sind in der Literatur nicht ganz einheitlich, siehe z.B. Davidson (2000, Sections 9.3.3, 11.3.2).

3.4.2. Konsistenz

- **Annahmen**

- **(ML.1)** Das Modell \mathbb{M} ist korrekt und vollständig spezifiziert. Es gilt für ein oder mehrere $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \sigma^2)^T \in \Theta$, dass der DGP im Modell enthalten ist, also

$$L(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X}) = f(\mathbf{y} | \mathbf{X}, \boldsymbol{\theta}_0) \subseteq \mathbb{M}$$

gilt, wobei darüber hinaus angenommen wird, dass der 'wahre' Parametervektor $\boldsymbol{\theta}_0$ im Inneren des Parameterraums Θ liegt, d. h. jeder Vektor $\boldsymbol{\theta}$ in einer beliebig kleinen lokalen Umgebung von $\boldsymbol{\theta}_0$ ebenfalls im Parameterraum Θ liegt.

- **(ML.2)**

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X}) = \bar{l}_\infty(\boldsymbol{\theta})$$

gleichmäßig in Θ , wobei $\bar{l}_\infty(\boldsymbol{\theta})$ eine nicht-stochastische Funktion in $\boldsymbol{\theta}$ ist.

- **(ML.3a)** Der 'wahre' Parametervektor $\boldsymbol{\theta}_0$ ist in endlichen Stichproben zumindest lokal **identifiziert**, d. h. für globale Identifikation gilt (3.4) für $\boldsymbol{\theta}_0 \in \Theta$.

- **(ML.3b)** Der Parametervektor θ_0 ist **asymptotisch global/lokal identifiziert**, d. h. (3.6) gilt für $\theta_0 \in \Theta$.

Annahme **(ML.2)** ist klarerweise Voraussetzung für **(ML.3b)**, aber nicht leicht zu zeigen, da dies weitergehende technische Konzepte erfordert, da die Log-Likelihood-Funktion im Allgemeinen nichtlinear ist. Siehe hierzu z.B. Davidson (2000, Section 9.3.2).

- Um zu verdeutlichen, dass der ML-Schätzer für θ mit der Stichprobengröße und der Stichprobe (\mathbf{y}, \mathbf{X}) selbst variiert, wird der Schätzer mit n indiziert.

- **Konsistenz:**

Unter den Annahmen **(ML.1)**, **(ML.2)**, **(ML.3a)** und **(ML.3b)** ist der Maximum-Likelihood-Schätzer

$$\hat{\theta}_n = \arg \max_{\theta} l(\theta_n | \mathbf{y}, \mathbf{X}) \quad \left(= \arg \max_{\theta} \sum_{t=1}^n \ln f(y_t | \mathbf{X}_t, \theta) \right) \quad (3.7)$$

konsistent, d. h.

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\hat{\theta}_n | \mathbf{y}, \mathbf{X}) = \text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\theta_0 | \mathbf{y}, \mathbf{X}) \quad (3.8)$$

- **Beweis:**

Um (3.8) zu zeigen, zeigt man, dass

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\hat{\boldsymbol{\theta}}_n | \mathbf{y}, \mathbf{X}) \geq \text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X}), \quad (3.9a)$$

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\hat{\boldsymbol{\theta}}_n | \mathbf{y}, \mathbf{X}) \leq \text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X}) \quad (3.9b)$$

zusammen gelten.

- Die Gleichung (3.9a) folgt unter den Annahmen (ML.1), (ML.2) und (ML.3b), sowie aus der Maximierung der Likelihood, denn $l(\hat{\boldsymbol{\theta}}_n | \mathbf{y}, \mathbf{X}) \geq l(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X})$ für jedes n und jede Stichprobe. Wenn der DGP nicht im Modell enthalten wäre, also (ML.1) nicht gelten würde, dann muss die Ungleichung nicht notwendigerweise gelten!
- Wie zeigt man (3.9b) ?

1. Zunächst ist festzuhalten, dass (3.9b) gilt, wenn

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X}) \leq \text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X})$$

für beliebige $\boldsymbol{\theta}^*$ gilt, denn $\hat{\boldsymbol{\theta}}_n$ ist ja immer irgendein beliebiges $\boldsymbol{\theta}^*$.

2. Da die Log-Likelihood die Summe

$$l(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X}) = \sum_{t=1}^n l_t(\boldsymbol{\theta}^* | y_t, \mathbf{X}_t)$$

ist, ergibt sich nach Division durch n ein Mittelwertschätzer. Wenn also ein Gesetz der großen Zahlen gilt, konvergiert die Log-Likelihood-Funktion dividiert durch die Zahl der Beobachtungen und gegeben den Parametervektor $\boldsymbol{\theta}^*$ gegen den Grenzwert des Mittelwerts, also gilt

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X}) = \lim_{n \rightarrow \infty} E_0 \left[\frac{1}{n} l(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X}) \right] \quad (3.10)$$

und entsprechend

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} l(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X}) = \lim_{n \rightarrow \infty} E_0 \left[\frac{1}{n} l(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X}) \right], \quad (3.11)$$

wobei $E_0[\cdot]$ den Erwartungswert auf der Basis der Dichte mit dem wahren Parametervektor $\boldsymbol{\theta}_0$ bezeichne. Aufgrund **(ML.2)** existieren **(3.10)** und **(3.11)**.

(‡ Anmerkung für Zeitreihenexperten: Bei vorherbestimmten Variablen ist $l_t(\boldsymbol{\theta}^* | y_t, \mathbf{X}_t)$ im Allgemeinen nur für $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0$ eine Martingaldifferenz.)

D. h. (3.9b) gilt, wenn

$$\lim_{n \rightarrow \infty} E_0 \left[\frac{1}{n} l(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X}) \right] \leq \lim_{n \rightarrow \infty} E_0 \left[\frac{1}{n} l(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X}) \right]$$

gilt. Dies ist gleichbedeutend mit der Bedingung

$$\lim_{n \rightarrow \infty} \left(E_0 \left[\frac{1}{n} l(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X}) \right] - E_0 \left[\frac{1}{n} l(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X}) \right] \right) \leq 0.$$

bzw. mit

$$\lim_{n \rightarrow \infty} \left(\frac{1}{n} E_0 \left[\log \frac{L(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X})}{L(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X})} \right] \right) \leq 0. \quad (3.12)$$

3. Im nächsten Schritt zeigt man, dass die Differenz für jedes beliebige n kleiner oder gleich 0 ist. Hierfür benötigt man als Werkzeug **Jensen's Ungleichung** (A.15).

Da $\log(\cdot)$ eine streng konkave Funktion ist, ergibt sich unter (ML.3a)

$$E_0 \left[\log \left(\frac{L(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X})}{L(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X})} \right) \right] < \log E_0 \left[\frac{L(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X})}{L(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X})} \right] \quad \text{für beliebige } \boldsymbol{\theta}^* \neq \boldsymbol{\theta}_0, \boldsymbol{\theta}^* \in \Theta$$

4. Unter **(ML.1)**

$$E_0 \left[\frac{L(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X})}{L(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X})} \right] = \int_{-\infty}^{\infty} \frac{L(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X})}{L(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X})} f(\mathbf{y} | \mathbf{X}, \boldsymbol{\theta}_0) d\mathbf{y} = \int_{-\infty}^{\infty} L(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X}) d\mathbf{y} = 1,$$

da das Integral jeder Dichte (bzw. Likelihood) über den Definitionsbereich der Zufallsvariable 1 ist.

Deshalb gilt

$$E_0 \left[\log \left(\frac{L(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X})}{L(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X})} \right) \right] = E_0[l(\boldsymbol{\theta}^* | \mathbf{y}, \mathbf{X})] - E_0[l(\boldsymbol{\theta}_0 | \mathbf{y}, \mathbf{X})] < \log 1 = 0.$$

Damit ist (3.12) gezeigt und damit wiederum (3.9b).

q.e.d.

- Lese Davidson und MacKinnon (2004, Section 10.1 bis 10.3). Davidson (2000, Section 11.3.2) zeigt die Konsistenz unter sehr allgemeinen Bedingungen.

3.4.3. Asymptotische Normalverteilung

- **Einige Definitionen** (k ist die Zahl der Parameter) (vgl. Abschnitt 2.3.1:
 - **Gradientenvektor oder Score-Vektor** der Log-Likelihood-Funktion

$$\frac{\partial l}{\partial \theta_i} \Big|_{\boldsymbol{\theta}^*} \equiv \frac{\partial l(\boldsymbol{\theta} | \mathbf{y}, \mathbf{X})}{\partial \theta_i} \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*} = \sum_{t=1}^n \frac{\partial l_t(\boldsymbol{\theta} | y_t, \mathbf{X}_t)}{\partial \theta_i} \Big|_{\boldsymbol{\theta} = \boldsymbol{\theta}^*} \quad (3.13)$$

$$(\equiv g_i(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta}^*)), \quad i = 1, \dots, k$$

$$\frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}^*} \equiv \begin{pmatrix} \frac{\partial l}{\partial \theta_1} \Big|_{\boldsymbol{\theta}^*} \\ \frac{\partial l}{\partial \theta_2} \Big|_{\boldsymbol{\theta}^*} \\ \vdots \\ \frac{\partial l}{\partial \theta_k} \Big|_{\boldsymbol{\theta}^*} \end{pmatrix} (\equiv \mathbf{g}(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta})). \quad (3.14)$$

- **Hessesche Matrix** (=Matrix der partiellen Ableitungen zweiter Ord-

nung) (Hessian matrix)

$$\mathbf{H}(\boldsymbol{\theta}^*) \equiv \left(\begin{array}{c|c} \frac{\partial^2 l}{\partial \theta_1 \partial \boldsymbol{\theta}^T} & \boldsymbol{\theta}^* \\ \vdots & \\ \frac{\partial^2 l}{\partial \theta_k \partial \boldsymbol{\theta}^T} & \boldsymbol{\theta}^* \end{array} \right) = \frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}^*}. \quad (3.15)$$

$$\mathbf{H}(\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*) \equiv \left(\begin{array}{c|c} \frac{\partial^2 l}{\partial \theta_1 \partial \boldsymbol{\theta}^T} & \boldsymbol{\theta}_1^* \\ \vdots & \\ \frac{\partial^2 l}{\partial \theta_k \partial \boldsymbol{\theta}^T} & \boldsymbol{\theta}_k^* \end{array} \right). \quad (3.16)$$

– Asymptotische Hessesche Matrix

$$\mathcal{H}(\boldsymbol{\theta}) = \text{plim}_{n \rightarrow \infty} \boldsymbol{\theta} \frac{1}{n} \mathbf{H}(\boldsymbol{\theta}) \quad (3.17)$$

– (Fisher) Informationsmatrix

$$\begin{aligned} \mathbf{I}(\boldsymbol{\theta}^*) &\equiv \text{Cov}_{\boldsymbol{\theta}^*} \left(\frac{\partial l}{\partial \boldsymbol{\theta}} \middle| \boldsymbol{\theta}^* \right) = E_{\boldsymbol{\theta}^*} \left[\frac{\partial l}{\partial \boldsymbol{\theta}} \middle| \boldsymbol{\theta}^* \frac{\partial l}{\partial \boldsymbol{\theta}^T} \middle| \boldsymbol{\theta}^* \right] \\ &= \sum_{t=1}^n E_{\boldsymbol{\theta}^*} \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \middle| \boldsymbol{\theta}^* \frac{\partial l_t}{\partial \boldsymbol{\theta}^T} \middle| \boldsymbol{\theta}^* \right]. \end{aligned} \quad (3.18)$$

– Asymptotische Informationsmatrix

$$\mathcal{I}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{I}(\boldsymbol{\theta}). \quad (3.19)$$

‡ Davidson und MacKinnon (2004, Gleichung (10.32)) verwenden anstelle des Limes den Wahrscheinlichkeitslimes, da in ihrer Notation die Informationsmatrix implizit auf die exogenen Variablen \mathbf{X} bedingt ist.

- **Zusätzliche Annahmen**

- **(ML.4a)** Die Log-Likelihood-Funktion $l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})$ ist bezüglich $\boldsymbol{\theta}$ im Inneren des Parameterraums Θ mit Wahrscheinlichkeit 1 zweimal partiell differenzierbar. (Beachte, dass die Log-Likelihood-Funktion eine Zufallsvariable in Abhängigkeit der Stichprobe (\mathbf{y}, \mathbf{X}) ist.)
- **(ML.4b)** (Davidson, 2000, Assumption 11.3.1)

$$\left\| \left\| \frac{\partial l_t}{\partial \theta_i} \right\| \right\|_{2+\delta} \leq B < \infty, \quad \delta > 0, \quad t = 1, 2, \dots, \quad i = 1, \dots, k.$$

Diese Annahme entspricht im KQ-Fall der Annahme **(C4b)**.

- **(ML.5):**

$$\frac{1}{n} \mathbf{H}(\boldsymbol{\theta}) \xrightarrow{P} \boldsymbol{\mathcal{H}}(\boldsymbol{\theta}).$$

gleichmäßig in einer Umgebung von $\boldsymbol{\theta}_0$, wobei $\boldsymbol{\mathcal{H}}(\cdot)$ eine nicht-stochastische und invertierbare Matrixfunktion ist.

• Ableitung

Um $l(\cdot)$ zu maximieren, müssen wir den Score-Vektor $\frac{\partial l}{\partial \boldsymbol{\theta}}$ gleich Null setzen. Schreibe die i -te partielle Ableitung erster Ordnung mit Hilfe des Mittelwertsatzes in Abhängigkeit vom wahren Parametervektor $\boldsymbol{\theta}_0$ als (dies ist wegen **(ML.4a)** möglich)

$$\frac{\partial l}{\partial \theta_i} \Big|_{\hat{\boldsymbol{\theta}}} = \frac{\partial l}{\partial \theta_i} \Big|_{\boldsymbol{\theta}_0} + \frac{\partial^2 l}{\partial \theta_i \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}_i^*} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \quad i = 1, \dots, k,$$

wobei $\boldsymbol{\theta}_i^*$ zwischen $\hat{\boldsymbol{\theta}}$ und $\boldsymbol{\theta}_0$ liegt. Beachte, dass sich $\boldsymbol{\theta}_i^*$ im Allgemeinen mit n und i ändert. Man kann nun diese Gleichungen untereinander in einen Spaltenvektor schreiben und mit obiger Notation umformen. Man erhält schließlich

$$\frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} = \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} + \mathbf{H}(\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

Da die FOC $\frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} \stackrel{!}{=} \mathbf{0}$ lautet, erhalten wir

$$\sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = - \underbrace{\left[\frac{1}{n} \mathbf{H}(\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*) \right]^{-1}}_{(i)} \underbrace{\frac{1}{\sqrt{n}} \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0}}_{(ii)}, \quad (3.20)$$

falls $\mathbf{H}(\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*)$ invertierbar ist.

Analyse des asymptotischen Verhaltens der Terme (i) und (ii):

- Term (i): Da der ML-Schätzer konsistent ist, gilt $\text{plim} \frac{1}{n} \mathbf{H}(\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*) = \text{plim} \frac{1}{n} \mathbf{H}(\boldsymbol{\theta}_0)$, wobei letzterer plim durch Annahme **(ML.5)** bestimmt ist, so dass Term (i) durch $\mathcal{H}(\boldsymbol{\theta}_0)$ ersetzt werden kann.
- Term (ii): Hier muss geprüft werden, ob für den skalierten Zufallsvektor $\frac{1}{\sqrt{n}} \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0}$ der Scorefunktion ein multivariater Grenzwertsatz gilt.

Falls $\{y_t, \mathbf{X}_t\}$, $t = 1, \dots, n$ **i.i.d. sind**, ist der Score-Vektor die Summe von n heterogenen, aber unabhängigen Zufallsvektoren

$$\frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} = \sum_{t=1}^n \frac{\partial l_t}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0}. \quad (3.21)$$

Damit entspricht der Score-Vektor n mal einem Mittelwertschätzer und es lässt sich unter bestimmten Regularitätsbedingungen ein Zentraler Grenzwertsatz

für heterogene, aber unabhängige Zufallsvariable auf

$$\sqrt{n} \left(\frac{1}{n} \sum_{t=1}^n \frac{\partial l_t}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} - E_0 \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} \right] \right).$$

anwenden, siehe **Methoden der Ökonometrie**, Abschnitt 5.5 *Werkzeuge für die asymptotische Analyse*. Hierfür benötigt man den Mittelwert und die Varianz von $\partial l_t / \partial \boldsymbol{\theta}$.

Bei **Zeitreihen** existiert unter bestimmten Voraussetzungen (u. a. Stationarität oder Fehler als Martingalprozess) ebenfalls ein Zentraler Grenzwertsatz, siehe z. B. Davidson (2000, Section 11.3.3). In Davidson und MacKinnon (2004, Section 10.3) ist eine vereinfachte Darstellung zu finden.

* Der Mittelwert des Score-Vektors ist ein Nullvektor. Nachweis:

$$\begin{aligned}
 E_0 \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} \right] &= \int \frac{\partial \ln L(\boldsymbol{\theta}_0 | y_t, \mathbf{X}_t)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} f(y_t | \mathbf{X}_t, \boldsymbol{\theta}_0) dy_t \\
 &= \int \frac{1}{L(\boldsymbol{\theta}_0 | y_t, \mathbf{X}_t)} \frac{\partial L(\boldsymbol{\theta} | y_t, \mathbf{X}_t)}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} f(y_t | \mathbf{X}_t, \boldsymbol{\theta}_0) dy_t \\
 &= \int \frac{f(y_t | \mathbf{X}_t, \boldsymbol{\theta}_0)}{f(y_t | \mathbf{X}_t, \boldsymbol{\theta}_0)} \frac{\partial f(y_t | \mathbf{X}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} dy_t \\
 &= \int \frac{\partial f(y_t | \mathbf{X}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} dy_t, \tag{3.22}
 \end{aligned}$$

wobei das letzte Gleichheitszeichen nur im Falle korrekter Spezifikation gilt, d. h. $L(\cdot) = f(\cdot)$ **(ML.1)**. Berechnung bei $\boldsymbol{\theta}_0$ führt zu

$$E_0 \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} \right] = \int \frac{\partial f(y_t | \mathbf{X}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} dy_t = \mathbf{0}.$$

Das zweite Gleichheitszeichen folgt aus $\int f(y_t | \mathbf{X}_t, \boldsymbol{\theta}) dy_t = 1$ für alle $\boldsymbol{\theta}$. Falls $\boldsymbol{\theta}$ den Integrationsbereich nicht beeinflusst (für genauere Bedingungen zur Vertauschbarkeit von Ableitung und Integration von Zufallszahlen siehe

z.B. Davidson (2000, Assumption 11.3.1, p. 274, Theorem B.10.1, p. 460))
 können Ableitung und Integral vertauscht werden, also

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\theta}} \int f(y_t | \mathbf{X}_t, \boldsymbol{\theta}) dy_t &= \frac{\partial}{\partial \boldsymbol{\theta}} 1 \\ \int \frac{\partial}{\partial \boldsymbol{\theta}} f(y_t | \mathbf{X}_t, \boldsymbol{\theta}) dy_t &= \mathbf{0}. \end{aligned} \quad (3.23)$$

* Da der Mittelwert von $\partial l_t / \partial \boldsymbol{\theta} |_{\boldsymbol{\theta}_0}$ bei $\boldsymbol{\theta}_0$ ein Nullvektor ist, ist die Varianz bei $\boldsymbol{\theta}_0$ gegeben durch

$$E_0 \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} \frac{\partial l_t}{\partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}_0} \right].$$

Da l_t und l_s , $s \neq t$ unabhängig und somit unkorreliert sind, ist die Varianz des Score-Vektors bei $\boldsymbol{\theta}_0$, d. h. die Summe aller $\partial l_t / \partial \boldsymbol{\theta} |_{\boldsymbol{\theta}_0}$ gegeben durch die (Fisher) Informationsmatrix (3.18) an der Stelle $\boldsymbol{\theta}_0$

$$\mathbf{I}(\boldsymbol{\theta}_0) = E_0 \left[\frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} \frac{\partial l}{\partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}_0} \right] = \sum_{t=1}^n E_0 \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} \frac{\partial l_t}{\partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}_0} \right]. \quad (3.24)$$

Jetzt lässt sich ein Zentraler Grenzwertsatz für heterogene, aber unabhängige

Zufallsvariable anwenden. Dabei wird die Informationsmatrix (3.18) am wahren Parametervektor $\mathbf{I}(\boldsymbol{\theta}_0)$ durch die asymptotische Informationsmatrix (3.19)

$$\mathcal{I}(\boldsymbol{\theta}_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{I}(\boldsymbol{\theta}_0)$$

ersetzt und man erhält

$$\sqrt{n} \left(\frac{1}{n} \frac{\partial l}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}_0} \right) = \frac{1}{\sqrt{n}} \frac{\partial l}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}_0} \xrightarrow{d} \mathbf{1}_\infty \sim N(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)). \quad (3.25)$$

(iii) Da (3.25) asymptotisch normalverteilt ist und Annahme (ML.5) gilt, folgt aus den Eigenschaften von plim's, dass

$$\begin{aligned} \sqrt{n} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &= \left[\frac{1}{n} \mathbf{H}(\boldsymbol{\theta}_1^*, \dots, \boldsymbol{\theta}_k^*) \right]^{-1} \frac{1}{\sqrt{n}} \frac{\partial l}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}_0} \\ &\xrightarrow{d} -\mathcal{H}(\boldsymbol{\theta}_0)^{-1} \mathbf{1}_\infty \sim N \left(\mathbf{0}, \underbrace{\mathcal{H}(\boldsymbol{\theta}_0)^{-1} \mathcal{I}(\boldsymbol{\theta}_0) \mathcal{H}(\boldsymbol{\theta}_0)^{-1}}_{\boldsymbol{\nu}(\boldsymbol{\theta}_0)} \right). \end{aligned} \quad (3.26)$$

Beachte: Man erhält eine asymptotische Kovarianzmatrix des ML-Schätzers,

die eine Sandwichform aufweist. Vgl. hierzu KQ- oder GMM-Schätzer mit Sandwich-Kovarianzmatrix.

- (iv) Zu guter Letzt muss man noch überprüfen, unter welchen Umständen sich die Sandwichform vereinfacht, d. h. $\mathcal{V}(\boldsymbol{\theta}_0) = \mathcal{I}(\boldsymbol{\theta}_0)^{-1}$ gilt. $\boldsymbol{\theta}_0$ wird in den folgenden Gleichungen zur Vereinfachung der Notation unterdrückt. Man betrachte

$$\begin{aligned}
 E_0 \left[\frac{\partial^2 l_t(\boldsymbol{\theta} | y_t, \mathbf{X}_t)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] &= \int \frac{\partial}{\partial \boldsymbol{\theta}^T} \left[\frac{1}{L_t(\boldsymbol{\theta} | y_t, \mathbf{X}_t)} \frac{\partial L_t(\boldsymbol{\theta} | y_t, \mathbf{X}_t)}{\partial \boldsymbol{\theta}} \right] f(y_t | \mathbf{X}_t, \boldsymbol{\theta}_0) dy_t \\
 &= \int -\frac{1}{L_t^2} \frac{\partial L_t}{\partial \boldsymbol{\theta}} \frac{\partial L_t}{\partial \boldsymbol{\theta}^T} f(y_t | \mathbf{X}_t, \boldsymbol{\theta}_0) dy_t + \int \frac{1}{L_t} \frac{\partial^2 L_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} f(y_t | \mathbf{X}_t, \boldsymbol{\theta}_0) dy_t \\
 &= -E_0 \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \frac{\partial l_t}{\partial \boldsymbol{\theta}^T} \right] + \int \frac{1}{L_t} \frac{\partial^2 L_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} f(y_t | \mathbf{X}_t, \boldsymbol{\theta}_0) dy_t \\
 &= -E_0 \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \frac{\partial l_t}{\partial \boldsymbol{\theta}^T} \right] + \int \frac{\partial^2 f(y_t | \mathbf{X}_t, \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} dy_t.
 \end{aligned}$$

Man beachte, dass bei korrekter Spezifikation $f(\cdot) = L(\cdot)$ gilt und dass das zweite Integral auf der rechten Gleichungsseite eine Nullmatrix ist. Dies lässt

sich zeigen, indem (3.23) ein weiteres Mal abgeleitet wird, was zu

$$\begin{aligned}\frac{\partial}{\partial \boldsymbol{\theta}^T} \int \frac{\partial f(y_t | \mathbf{X}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} dy_t &= \mathbf{0} \\ \int \frac{\partial^2 f(y_t | \mathbf{X}_t, \boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} dy_t &= \mathbf{0}\end{aligned}$$

führt. Im Falle einer korrekt spezifizierten Likelihood-Funktion gilt deshalb

$$E_0 \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \bigg|_{\boldsymbol{\theta}_0} \frac{\partial l_t}{\partial \boldsymbol{\theta}^T} \bigg|_{\boldsymbol{\theta}_0} \right] = -E_0 \left[\frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \bigg|_{\boldsymbol{\theta}_0} \right]$$

und da l_t und l_s unkorreliert sind, gilt auch

$$\mathbf{I}(\boldsymbol{\theta}_0) = -E_0 \left[\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \bigg|_{\boldsymbol{\theta}_0} \right] = -E_0 (\mathbf{H}(\boldsymbol{\theta}_0)). \quad (3.27)$$

Gelten also die Annahmen (ML.1) und (ML.5), ist das Modell korrekt spezifiziert und es gilt

$$\mathcal{I}(\boldsymbol{\theta}_0) = -\mathcal{H}(\boldsymbol{\theta}_0) \quad (3.28)$$

und damit

$$\boldsymbol{\mathcal{V}}(\boldsymbol{\theta}_0) = \mathcal{I}(\boldsymbol{\theta}_0)^{-1} \mathcal{I}(\boldsymbol{\theta}_0) \mathcal{I}(\boldsymbol{\theta}_0)^{-1} = \mathcal{I}(\boldsymbol{\theta}_0)^{-1}$$

und die **asymptotische Normalverteilung des ML-Schätzers** ist gegeben durch

$$\sqrt{n} \left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \right) \xrightarrow{D} N \left(\mathbf{0}, \mathcal{I}(\boldsymbol{\theta}_0)^{-1} \right). \quad (3.29)$$

- Einige Anmerkungen:

- Der Beweis der asymptotischen Normalität wurde unter der IID-Annahme durchgeführt. Er lässt sich mit etwas mehr Aufwand natürlich auch für abhängige Variable durchführen, siehe indem gezeigt wird, dass die $\partial l_t / \partial \boldsymbol{\theta}$ Martingaldifferenzen sind, siehe z.B. Davidson (2000, Section 11.3.3) oder **Fortgeschrittene Dynamische Ökonometrie**.
- (3.27) wird als **Informationsmatrixgleichheit** und (3.28) als **asymptotische Informationsmatrixgleichheit** bezeichnet.
- Die asymptotische Verteilung (3.29) des ML-Schätzers wird in der Praxis verwendet um

$$\hat{\boldsymbol{\theta}} \overset{\text{approx.}}{\sim} N \left(\boldsymbol{\theta}_0, \hat{\mathbf{I}}(\hat{\boldsymbol{\theta}})^{-1} \right)$$

zu rechtfertigen.

• **Zur Schätzung der Kovarianzmatrix $\mathbf{I}(\boldsymbol{\theta}_0)^{-1}$:**

Im Fall einer korrekt spezifizierten Likelihood-Funktion **(ML.1)** stehen zur Verfügung:

1. minus die Inverse der **empirischen Hessematrix**:

$$-\mathbf{H}(\hat{\boldsymbol{\theta}})^{-1} = - \left[\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\hat{\boldsymbol{\theta}}} \right]^{-1} = - \left[\sum_{t=1}^n \left[\frac{\partial^2 l_t}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\hat{\boldsymbol{\theta}}} \right] \right]^{-1}, \quad (3.30)$$

2. die Inverse der **Fisher Informationsmatrix** (3.18) an der Stelle des ML-Schätzers $\hat{\boldsymbol{\theta}}$

$$\mathbf{I}(\hat{\boldsymbol{\theta}})^{-1} = Cov_{\hat{\boldsymbol{\theta}}}^{-1} \left(\frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} \right) = \left[\sum_{t=1}^n E_{\hat{\boldsymbol{\theta}}} \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} \frac{\partial l_t}{\partial \boldsymbol{\theta}^T} \Big|_{\hat{\boldsymbol{\theta}}} \right] \right]^{-1} \quad (3.31)$$

oder des Schätzers von $\mathbf{I}(\boldsymbol{\theta}_0)$,

3. die Inverse des **äußeren Produkts des Score-Vektors**

$$\left[\sum_{t=1}^n \left[\frac{\partial l_t}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} \frac{\partial l_t}{\partial \boldsymbol{\theta}^T} \Big|_{\hat{\boldsymbol{\theta}}} \right] \right]^{-1}. \quad (3.32)$$

Beachte, dass die negative Hessesche Matrix der Log-Likelihood-Funktion in Variante (a) z.B. auch bei der numerischen Optimierung nach der Newton-Raphson-Methode verwendet wird (siehe Abschnitt 2.3). Variante (b) ist vorzuziehen, wenn sich die Erwartungswerte explizit berechnen und die sich ergebenden Terme approximieren lassen (z.B. bei (nicht-)linearen Regressionen mit exogenen Variablen). In komplizierteren Modellen ist dies jedoch häufig nicht möglich. In Monte Carlo-Studien hat sich gezeigt, dass die Variante (c) in endlichen Stichproben nicht so zuverlässig wie die beiden anderen Varianten funktioniert.

- **Quasi-Maximum-Likelihood-Schätzer (QMLE):**

Ist die Log-Likelihood fehlspezifiziert, so dass die (asymptotische) Informationsmatrixgleichheit nicht gilt, jedoch der ML-Schätzer konsistent ist, lässt sich die ML-Methode anwenden. Jedoch muss dann die allgemeine Kovarianzmatrix

$$\mathcal{V}(\boldsymbol{\theta}_0) = \mathcal{H}(\boldsymbol{\theta}_0)^{-1} \mathcal{I}_0 \mathcal{H}(\boldsymbol{\theta}_0)^{-1}$$

aus Gleichung (3.26) verwendet werden. Deshalb bezeichnet man den Schätzer als **Quasi-Maximum-Likelihood-Schätzer (quasi-maximum likelihood estimator (QMLE))**. Da deren Schätzung mehr Parameter erfordert und damit die Schätzung unsicherer wird, ist für die Fehlspezifikation in der Praxis der Preis ei-

nes größeren Schätzfehlers zu zahlen. QMLE sollte also, wenn möglich, vermieden werden.

- **Cramér-Rao-Untergrenze:**

- Cramér und Rao haben gezeigt, dass kein unverzerrter Schätzer existiert, dessen Kovarianzmatrix die Inverse der Fisher Informationsmatrix $\mathbf{I}(\boldsymbol{\theta}_0)$ "unterschreiten" kann. Man bezeichnet diese Grenze als **Cramér-Rao-Untergrenze**.
- Der ML-Schätzer ist **(asymptotisch) effizient**, d. h. die Kovarianzmatrix des ML-Schätzers erreicht die asymptotische Version der Cramér-Rao-Untergrenze.

– ‡ **Ableitung der Cramér-Rao Untergrenze:**

Sei $\tilde{\theta}(\mathbf{z})$ ein unverzerrter Schätzer für die Parameter θ auf Basis der Daten \mathbf{z} , welchem eine Dichte $f(\mathbf{z}|\theta_0)$ zugrunde liegt. Die Bedingung eines unverzerrten Schätzers lautet

$$E_0(\tilde{\theta}(\mathbf{z})) = \int \tilde{\theta}(\mathbf{z}) f(\mathbf{z}|\theta_0) d\mathbf{z} = \theta_0.$$

Ableiten dieser Bedingung nach θ ergibt (gegeben die Vertauschbarkeit von Integration und Ableitung)

$$\int \tilde{\theta}(\mathbf{z}) \frac{\partial f(\mathbf{z}|\theta)}{\partial \theta^T} \Big|_{\theta_0} d\mathbf{z} = \frac{\partial \theta}{\partial \theta^T} \Big|_{\theta_0}$$

bzw. $\int \tilde{\theta}(\mathbf{z}) \frac{\partial \ln f(\mathbf{z}|\theta)}{\partial \theta^T} \Big|_{\theta_0} f(\mathbf{z}|\theta_0) d\mathbf{z} = \mathbf{I}$

bzw. $E_0 \left[\tilde{\theta}(\mathbf{z}) \frac{\partial \ln f(\mathbf{z}, \theta)}{\partial \theta^T} \Big|_{\theta_0} \right] = \mathbf{I}$

bzw. $Cov_0 \left(\tilde{\theta}(\mathbf{z}), \frac{\partial \ln f(\mathbf{z}, \theta)}{\partial \theta} \Big|_{\theta_0} \right) = \mathbf{I}.$

Die letzte Zeile folgt wegen $E \left[\frac{\partial l}{\partial \theta} \Big|_{\theta_0} \right] = \mathbf{0}.$

Es gilt für skalare Zufallszahlen Z, W

$$Var(Z)Var(W) \geq Cov(Z, W)^2$$

und damit im Fall $k = 1$

$$Var(\tilde{\theta}(\mathbf{z})) \geq Var^{-1} \left(\frac{\partial \ln f(\mathbf{z}|\theta)}{\partial \theta} \Big|_{\theta_0} \right) = I(\theta_0)^{-1}.$$

Im multivariaten Fall lässt sich dies analog zeigen. Hierzu benötigt man die multivariate Cauchy-Schwartz-Ungleichung für Zufallsvektoren (siehe z.B. [Gourieroux und Monfort \(1995, Section B.2.5\)](#)), d. h.

$$(\text{Var}(\mathbf{Z}) - \text{Cov}(\mathbf{Z}, \mathbf{W})\text{Var}(\mathbf{W})^{-1}\text{Cov}(\mathbf{W}, \mathbf{Z})) \text{ ist positiv semidefinit.}$$

3.5. Numerische Optimierung

• Numerische Optimierungsverfahren

Siehe zur Einführung in die numerische Optimierung Abschnitt 2.3. Für die Quasi-Newton-Verfahren/Gradientenverfahren (2.16), vgl. Abschnitt 2.3.2, hier für Parametervektor $\boldsymbol{\theta}$

$$\boldsymbol{\theta}_{(j+1)} = \boldsymbol{\theta}_{(j)} - \alpha_{(j)} \mathbf{D}^{-1}(\boldsymbol{\theta}_{(j)}) \mathbf{g}(\boldsymbol{\theta}_{(j)})$$

gibt es speziell für die ML-Schätzung:

– Gradientenverfahren:

BHHH-Algorithmus Berndt, Hall, Hall, Hausman Algorithmus:

$$\mathbf{D}(\boldsymbol{\theta}_{(j)}) = \sum_{t=1}^n \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_{(j)}} \frac{\partial l}{\partial \boldsymbol{\beta}^T} \Big|_{\boldsymbol{\theta}_{(j)}}$$

basiert auf (3.32). (EViews, R-package maxLik)

– Verfahren mit partiellen Ableitungen zweiter Ordnung:

Scoring-Algorithmus: $\mathbf{D}(\boldsymbol{\theta}_{(j)}) = E \left[\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}_j} \right].$

(R-package maxLik, Übergeben der erwarteten Hessematrix)

- **Konzentrieren der Likelihood-Funktion**

Beispiel: konzentrierte Log-Likelihood-Funktion für das normale lineare Regressionsmodell, vgl. Abschnitt 3.3 mit Log-Likelihood-Funktion

$$l(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) = -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

Ableiten nach σ^2 und Nullsetzen ergibt

$$\frac{\partial l(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X})}{\partial \sigma} = -\frac{n}{\sigma} + \frac{1}{\sigma^3} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0.$$

Auflösen ergibt den ML-Schätzer für die Fehlervarianz für gegebene $\boldsymbol{\beta}$ (vgl. **Methoden der Ökonometrie**, Abschnitt 3.5 *Schätzen der Fehlervarianz*)

$$\hat{\sigma}^2 = \frac{1}{n} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (3.33)$$

Einsetzen in die Log-Likelihood-Funktion ergibt die **konzentrierte Log-Likelihood-Funktion**

$$\begin{aligned} l^c(\boldsymbol{\beta} | \mathbf{y}, \mathbf{X}) &= -\frac{n}{2} \ln 2\pi - \frac{n}{2} \ln \left(\frac{1}{n} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right) - \frac{n}{2} \\ &= -\frac{n}{2} (1 + \ln 2\pi - \ln n) - \frac{n}{2} \ln SSR(\boldsymbol{\beta}) \end{aligned} \quad (3.34)$$

Die **Log-Likelihood-Funktion** für die geschätzten Parameter lautet demnach

$$l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2 | \mathbf{y}, \mathbf{X}) = -\frac{n}{2} (1 + \ln 2\pi - \ln n) - \frac{n}{2} \ln SSR(\hat{\boldsymbol{\beta}}) \quad (3.35)$$

3.6. Hypothesentests

Wiederholung: Bestandteile und Charakteristika eines Tests

- **Null und Alternativhypothese**

$$H_0 : \mathbf{r}(\boldsymbol{\theta}) = \mathbf{0} \quad \text{versus} \quad H_1 : \mathbf{r}(\boldsymbol{\theta}) \neq \mathbf{0},$$

wobei

$$\mathbf{r}(\boldsymbol{\theta}) = \begin{pmatrix} r_1(\boldsymbol{\theta}) \\ r_2(\boldsymbol{\theta}) \\ \vdots \\ r_q(\boldsymbol{\theta}) \end{pmatrix}$$

ein Vektor von q (möglicherweise nichtlinearen) Restriktionen ist.

Beispiel: $r_1(\boldsymbol{\theta}) = \theta_1 - \frac{1}{\theta_2} = 0$, $r_2(\boldsymbol{\theta}) = \theta_1\theta_2 = 3$

- **Teststatistik:** $\lambda = \text{function}(\mathbf{y}, \mathbf{X})$
- **Entscheidungsregel:** Lehne H_0 ab, falls Teststatistik im kritischen Bereich C liegt.

- **Definitionen**

- **Tatsächliches Signifikanzniveau eines Tests = Niveau eines Tests = Fehler 1. Art**

$$\alpha(\boldsymbol{\theta}) = P(\lambda \in C | \boldsymbol{\theta} \text{ in } H_0)$$

- **Größe eines Tests**

$$\alpha = \max_{\boldsymbol{\theta} \text{ in } H_0} \alpha(\boldsymbol{\theta})$$

- **Fehler 2. Art**

$$\beta(\boldsymbol{\theta}) = 1 - P(\lambda \in C | \boldsymbol{\theta} \text{ in } H_1)$$

- **Macht:**

$$\pi(\boldsymbol{\theta}) = P(\lambda \in C | \boldsymbol{\theta} \text{ in } H_1) = 1 - \beta(\boldsymbol{\theta})$$

- **uniformly most powerful test:** Ein Test, der unter allen konkurrierenden Tests die größte Macht hat und zwar bei jedem möglichen Parametervektor in H_1 .
- **konsistenter Test:** $\lim_{n \rightarrow \infty} \pi(\boldsymbol{\theta}) = 1$ für $\boldsymbol{\theta}$ in H_1 .

Überblick:

Ausgehend vom ML-Schätzer lassen sich drei grundlegende Testprinzipien ableiten:

- der Likelihood-Quotienten-Test (likelihood ratio test, LR test)
- der Wald-Test
- der Lagrange-Multiplikator-Test (LM test)

Notation:

Um die Lesbarkeit zu erhöhen, wird im Folgenden in der Log-Likelihood-Funktion die Stichprobenbedingung weggelassen, also

$$l(\boldsymbol{\theta}) = l(\boldsymbol{\theta}|\mathbf{y}, \mathbf{X})$$

verwendet.

3.6.1. Likelihood-Quotienten-Test (LR-Test)

- **Idee:** Vergleiche die Log-Likelihood des unrestringierten und des restringierten Modells miteinander.
- **Teststatistik und asymptotische Verteilung:**

$$LR \equiv 2 \left(l(\hat{\boldsymbol{\theta}}) - l(\tilde{\boldsymbol{\theta}}) \right) \xrightarrow{d} \chi^2(q), \quad (3.36)$$

wobei $\hat{\boldsymbol{\theta}}$ und $\tilde{\boldsymbol{\theta}}$ die ML-Schätzer des unrestringierten und des restringierten Modells bezeichnen und letzteres q Restriktionen unterliegt.

- **Heuristische Ableitung:**

Eine Taylorapproximation zweiter Ordnung der restringierten Log-Likelihood an der Stelle des geschätzten Parametervektors $\hat{\boldsymbol{\theta}}$ des unrestringierten Modells ergibt

$$l(\tilde{\boldsymbol{\theta}}) \approx l(\hat{\boldsymbol{\theta}}) + \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) + \frac{1}{2} (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})^T \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\hat{\boldsymbol{\theta}}} (\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}).$$

Da unter H_0 gilt $\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}} = \mathbf{0}$ und $\text{plim } n^{-1} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\hat{\boldsymbol{\theta}}} = \boldsymbol{\mathcal{H}}(\boldsymbol{\theta}_0)$ (wegen Konsistenz

und **(ML.5)**) gilt, erhält man approximativ

$$LR = 2 \left(l(\hat{\boldsymbol{\theta}}) - l(\tilde{\boldsymbol{\theta}}) \right) \stackrel{asy}{\sim} -n \left(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}} \right)^T \boldsymbol{\mathcal{H}}(\boldsymbol{\theta}_0) \left(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}} \right).$$

Da die Parametervektoren $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ und $\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ asymptotisch normalverteilt sind, siehe **(3.29)**, (und damit auch deren Differenz), ist die Summe der Quadrate mit der inversen Kovarianzmatrix gewichtet asymptotisch χ^2 -verteilt (Theorem über stetige Abbildungen). Die Schwierigkeit ist, die korrekte Anzahl an Freiheitsgraden zu bestimmen.

Da im restringierten Modell q Restriktionen vorliegen (im Fall von Nullrestriktionen q Parameter weniger zu schätzen sind), ist zu erwarten, dass die Schätzer für die $k - q$ Parameter $\left(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}} \right)$, die in beiden Modellen vorliegen, sich ähnlich verhalten, dies jedoch für die q verbleibenden Parameter nicht der Fall ist. Deshalb liegt Unsicherheit nur bezüglich der q restringierten Parameter vor und man erhält eine asymptotische $\chi^2(q)$ -Verteilung

$$LR \xrightarrow{d} \chi^2(q).$$

Für den Fall von Nullrestriktionen findet sich in **Davidson und MacKinnon (2004, Section 10.6)** eine präzise Ableitung.

- **Einige Anmerkungen:**

- Der LR-Test funktioniert nur, wenn das restringierte Modell im unrestringierten Modell enthalten ist (**nested**).
- Der LR-Test ist immer dann leicht zu berechnen, wenn sowohl das restringierte als auch das unrestringierte Modell leicht geschätzt werden können.
- Für das normale multiple lineare Regressionsmodell ist die LR-Statistik der F -Statistik sehr ähnlich (Davidson und MacKinnon, 2004, p. 421)

$$F = \frac{(SSR(\tilde{\beta}) - SSR(\hat{\beta})) / q}{SSR(\hat{\beta}) / (n - k)}$$

$$LR = 2 \left(l(\hat{\beta}, \hat{\sigma}^2 | \mathbf{y}, \mathbf{X}) - l(\tilde{\beta}, \tilde{\sigma}^2 | \mathbf{y}, \mathbf{X}) \right)$$

$$= n \ln \left(\frac{SSR(\tilde{\beta})}{SSR(\hat{\beta})} \right) \approx qF,$$

wobei für das letzte Gleichheitszeichen (3.35) verwendet wurde.

3.6.2. Wald-Test

- **Teststatistik und asymptotische Verteilung:**

$$W = \mathbf{r}(\hat{\boldsymbol{\theta}})^T \left[\mathbf{R}(\hat{\boldsymbol{\theta}}) \widehat{\mathbf{Var}}(\hat{\boldsymbol{\theta}}) \mathbf{R}^T(\hat{\boldsymbol{\theta}}) \right]^{-1} \mathbf{r}(\hat{\boldsymbol{\theta}}) \xrightarrow{d} \chi^2(q). \quad (3.37)$$

wobei $\widehat{\mathbf{Var}}(\hat{\boldsymbol{\theta}}) = \left[n \widehat{\mathcal{I}}(\boldsymbol{\theta}_0) \right]^{-1}$

- **Ableitung:**

- Die Wald-Teststatistik erhält man, indem man die Restriktionsfunktion $\mathbf{r}(\boldsymbol{\theta})$ mittels einer Taylorapproximation am wahren Parametervektor $\boldsymbol{\theta}_0$ approximiert

$$\mathbf{r}(\boldsymbol{\theta}) \approx \mathbf{r}(\boldsymbol{\theta}_0) + \underbrace{\frac{\partial \mathbf{r}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}_0}}_{\mathbf{R}(\boldsymbol{\theta}_0)} (\boldsymbol{\theta} - \boldsymbol{\theta}_0). \quad (3.38)$$

- Man beachte, dass unter H_0 gilt: $\mathbf{r}(\boldsymbol{\theta}_0) = \mathbf{0}$.
- Man beachte auch, dass bei linearen Restriktionen $\mathbf{R}(\boldsymbol{\theta}_0) = \mathbf{R}$ gilt.
- Da unter Regularitätsannahmen $\sqrt{n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ asymptotisch normalverteilt ist,

siehe (3.29), gilt dies auch für die rechte Seite von (3.38), wenn $\boldsymbol{\theta}$ durch $\hat{\boldsymbol{\theta}}$ ersetzt wird, und, sofern die Annahmen „komische“ Approximationsfehler ausschließen, unter H_0 auch für $\sqrt{n} \mathbf{r}(\hat{\boldsymbol{\theta}})$ selbst. Unter der Annahme korrekter Spezifikation und H_0 ergibt sich deshalb

$$\sqrt{n} \mathbf{r}(\hat{\boldsymbol{\theta}}) \xrightarrow{d} N(\mathbf{0}, \mathbf{R}(\boldsymbol{\theta}_0) \mathcal{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{R}^T(\boldsymbol{\theta}_0)).$$

Man beachte, dass die Kovarianzmatrix Rang q hat, da $\mathbf{R}(\boldsymbol{\theta}_0)$ eine $(q \times k)$ Matrix ist.

- Nunmehr wendet man das Theorem für stetige Abbildungen und die Regeln für Summen quadrierter normalverteilter Zufallsvektoren an, um eine (asymptotische) χ^2 -Verteilung zu erhalten. Damit ergibt sich

$$n \mathbf{r}(\hat{\boldsymbol{\theta}})^T [\mathbf{R}(\boldsymbol{\theta}_0) \mathcal{I}(\boldsymbol{\theta}_0)^{-1} \mathbf{R}^T(\boldsymbol{\theta}_0)]^{-1} \mathbf{r}(\hat{\boldsymbol{\theta}}) \xrightarrow{d} \chi^2(q).$$

- Man erhält die **Wald-Teststatistik** (3.37), indem man alle unbekannt GröÙen unter dem unrestringierten Modell mit einem konsistenten Schätzer bestimmt.

- **Einige Bemerkungen:**

- Man kann jeden konsistenten Schätzer für die Kovarianzmatrix verwenden, vgl. (3.30), (3.31) und (3.32) in Abschnitt 3.4. Man beachte, dass die Wahl des Kovarianzschätzers im Allgemeinen das Testergebnis beeinflusst.
- In manchen Fällen sollte man den Wald-Test nicht wählen. Dies ist dann der Fall, wenn eine Umformulierung der Restriktionsfunktion zu einer anderen Teststatistik führt, denn dann können die Verteilungseigenschaften in endlichen Stichproben sehr unterschiedlich sein. Dies geschieht leicht bei nichtlinearen Restriktionsfunktionen, siehe Davidson und MacKinnon (2004, p. 423).
- Die F -Teststatistik im linearen Regressionsmodell mit linearen Restriktionen ist ein Spezialfall der Wald-Teststatistik.
- Ein zweiseitiger t -Test entspricht einem Wald-Test.
- Davidson und MacKinnon (2004, p. 424) warnen (bei Modellen, die nichtlinear in den Parametern sind) vor der Benutzung von gewöhnlichen t -Teststatistiken, da sie auf der Wald-Teststatistik basieren. Alternativ kann man den LR-Test verwenden.
- Es ist nur eine Schätzung unter H_1 notwendig.

3.6.3. Lagrange-Multiplikator-Test (LM) Test oder Score-Test

- **Teststatistik und asymptotische Verteilung:**

Es bezeichne $\tilde{\theta}$ den ML-Schätzer des restringierten Modells (unter H_0).

- **LM-Version** der Teststatistik

$$LM = \tilde{\lambda}^T \mathbf{R}(\tilde{\theta}) \widetilde{\text{Var}}(\tilde{\theta}) \mathbf{R}(\tilde{\theta})^T \tilde{\lambda} \xrightarrow{d} \chi^2(q), \quad (3.39)$$

wobei λ den Vektor der Lagrangemultiplikatoren bezeichnet.

- **Score-Statistik**

$$LM = \frac{\partial l}{\partial \theta^T} \Big|_{\tilde{\theta}} \widetilde{\text{Var}}(\tilde{\theta}) \frac{\partial l}{\partial \theta} \Big|_{\tilde{\theta}} \xrightarrow{d} \chi^2(q). \quad (3.40)$$

- **Ableitung:**

- Man schätzt das Modell nur unter der Nullhypothese.

- D. h. man muss die Log-Likelihood unter Nebenbedingungen $\mathbf{r}(\theta) = \mathbf{0}$ maximieren. Hierzu lässt sich der **Lagrange-Ansatz** verwenden und man maxi-

miert die Lagrangefunktion

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = l(\boldsymbol{\theta}) - \mathbf{r}(\boldsymbol{\theta})^T \boldsymbol{\lambda}, \quad (3.41)$$

wobei der Vektor $\boldsymbol{\lambda}$ alle q Lagrange-Parameter enthält. Die Bedingungen erster Ordnung für den restringierten ML-Schätzer $\tilde{\boldsymbol{\theta}}$ lauten:

$$\begin{aligned} \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\tilde{\boldsymbol{\theta}}} - \mathbf{R}(\tilde{\boldsymbol{\theta}})^T \tilde{\boldsymbol{\lambda}} &= \mathbf{0} \\ \mathbf{r}(\tilde{\boldsymbol{\theta}}) &= \mathbf{0}. \end{aligned}$$

- Mit Hilfe einer Taylorapproximation erster Ordnung ergibt sich für den Score-Vektor unter H_0

$$\frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\tilde{\boldsymbol{\theta}}} \approx \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} + \frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}_0} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0).$$

- Da für den restringierten Schätzer $\tilde{\boldsymbol{\theta}}$

$$\frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\tilde{\boldsymbol{\theta}}} = \mathbf{R}(\tilde{\boldsymbol{\theta}})^T \tilde{\boldsymbol{\lambda}}$$

und unter H_0 , siehe (3.38),

$$\mathbf{r}(\tilde{\boldsymbol{\theta}}) \approx \mathbf{R}(\boldsymbol{\theta}_0)(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$$

gilt, erhält man

$$\begin{aligned} \frac{1}{n} \mathbf{R}(\tilde{\boldsymbol{\theta}})^T \sqrt{n} \tilde{\boldsymbol{\lambda}} - \frac{1}{n} \frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big|_{\boldsymbol{\theta}_0} \sqrt{n} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\approx \frac{1}{\sqrt{n}} \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} \\ \frac{1}{n} \mathbf{R}(\boldsymbol{\theta}_0) \sqrt{n} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) &\approx \mathbf{0}, \end{aligned}$$

wobei noch jeweils die Gleichungen mit $\frac{1}{\sqrt{n}}$ multipliziert wurden.

- Als nächstes ersetzt man $\mathbf{R}(\tilde{\boldsymbol{\theta}})^T$ durch $\mathbf{R}(\boldsymbol{\theta}_0)^T$ (unter H_0 ist der Approximationsfehler asymptotisch Null). Mit der Definition der Hesseschen Matrix $\mathbf{H}(\boldsymbol{\theta})$ erhält man in Matrixschreibweise

$$\frac{1}{n} \begin{pmatrix} -\mathbf{H}(\boldsymbol{\theta}_0) & \mathbf{R}(\boldsymbol{\theta}_0)^T \\ \mathbf{R}(\boldsymbol{\theta}_0) & \mathbf{0} \end{pmatrix} \begin{pmatrix} \sqrt{n} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \\ \sqrt{n} \tilde{\boldsymbol{\lambda}} \end{pmatrix} \approx \begin{pmatrix} \frac{1}{\sqrt{n}} \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0} \\ \mathbf{0} \end{pmatrix}$$

- Nunmehr muss die partitionierte Matrix invertiert werden.

– **Regeln zur Invertierung partitionierter Matrizen**

(Lütkepohl, 1996, Section 3.5.3, (2) und (3))

\mathbf{A} und $(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})$ invertierbar:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} & -\mathbf{A}^{-1}\mathbf{B}(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \\ -(\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1}\mathbf{CA}^{-1} & (\mathbf{D} - \mathbf{CA}^{-1}\mathbf{B})^{-1} \end{pmatrix} \quad (3.42)$$

\mathbf{D} und $(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})$ invertierbar:

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{C} & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} (\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{BD}^{-1}\mathbf{C})^{-1}\mathbf{BD}^{-1} \end{pmatrix} \quad (3.43)$$

– Mit (3.42) ergibt sich für den Vektor der Lagrange-Multiplikatoren

$$\sqrt{n}\tilde{\boldsymbol{\lambda}} \approx n \left[\mathbf{R}(\boldsymbol{\theta}_0)\mathbf{H}(\boldsymbol{\theta}_0)^{-1}\mathbf{R}(\boldsymbol{\theta}_0)^T \right]^{-1} \mathbf{R}(\boldsymbol{\theta}_0)\mathbf{H}(\boldsymbol{\theta}_0)^{-1} \frac{1}{\sqrt{n}} \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0}$$

und mit (ML.5)

$$\frac{1}{\sqrt{n}}\tilde{\boldsymbol{\lambda}} \approx \underbrace{\left[\mathbf{R}(\boldsymbol{\theta}_0)\mathcal{H}(\boldsymbol{\theta}_0)^{-1}\mathbf{R}(\boldsymbol{\theta}_0)^T \right]^{-1} \mathbf{R}(\boldsymbol{\theta}_0)\mathcal{H}(\boldsymbol{\theta}_0)^{-1}}_{\mathbf{B}} \frac{1}{\sqrt{n}} \frac{\partial l}{\partial \boldsymbol{\theta}} \Big|_{\boldsymbol{\theta}_0}.$$

Damit erhält man unter Verwendung der asymptotischen Informationsmatrixgleichheit (3.28) und (3.25) die asymptotische Verteilung

$$\frac{1}{\sqrt{n}} \tilde{\boldsymbol{\lambda}} \xrightarrow{d} N(\mathbf{0}, \mathbf{B} \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0) \mathbf{B}^T)$$

bzw.

$$\frac{1}{\sqrt{n}} \tilde{\boldsymbol{\lambda}} \xrightarrow{d} N\left(\mathbf{0}, [\mathbf{R}(\boldsymbol{\theta}_0) \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)^{-1} \mathbf{R}(\boldsymbol{\theta}_0)^T]^{-1}\right).$$

Da die Kovarianzmatrix q Freiheitsgrade hat, ist die quadratische Form unter H_0

$$\frac{1}{n} \tilde{\boldsymbol{\lambda}}^T [\mathbf{R}(\boldsymbol{\theta}_0) \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)^{-1} \mathbf{R}(\boldsymbol{\theta}_0)^T] \tilde{\boldsymbol{\lambda}} \xrightarrow{d} \chi^2(q)$$

asymptotisch χ^2 -verteilt mit q Freiheitsgraden. Um eine Teststatistik zu erhalten, schätzt man die Informationsmatrix $n \boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)$ unter H_0 mit einem konsistenten Kovarianzschätzer $\widetilde{\text{Var}}(\tilde{\boldsymbol{\theta}}) = \left[n \widetilde{\boldsymbol{\mathcal{I}}(\boldsymbol{\theta}_0)} \right]^{-1}$ und man erhält die **LM-Version** (3.39)

$$LM = \tilde{\boldsymbol{\lambda}}^T \mathbf{R}(\tilde{\boldsymbol{\theta}}) \widetilde{\text{Var}}(\tilde{\boldsymbol{\theta}}) \mathbf{R}(\tilde{\boldsymbol{\theta}})^T \tilde{\boldsymbol{\lambda}} \xrightarrow{d} \chi^2(q).$$

- Die **Score-Statistik** (3.40) erhält man, indem man den Vektor der Lagrange-Multiplikatoren durch den Score ersetzt (aus Bedingungen erster Ordnung).

• Einige Anmerkungen

- Für das normale **lineare Regressionsmodell** mit streng exogenen Regressoren ergibt sich für die Score-Funktion und die Fisher Informationsmatrix

$$\mathbf{g}(\tilde{\boldsymbol{\beta}}, \sigma_0^2) = \left. \frac{\partial l}{\partial \boldsymbol{\beta}} \right|_{\tilde{\boldsymbol{\beta}}} = \mathbf{R}^T \tilde{\boldsymbol{\lambda}} = -\frac{1}{\sigma_0^2} \mathbf{X}^T \tilde{\mathbf{u}}$$

$$\mathbf{I}(\boldsymbol{\beta}_0, \sigma_0^2) = E \left[\left. \frac{\partial l}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}_0} \left. \frac{\partial l}{\partial \boldsymbol{\beta}^T} \right|_{\boldsymbol{\beta}_0} \right] = \frac{1}{\sigma_0^4} E \left[\mathbf{X}^T E(\tilde{\mathbf{u}}\tilde{\mathbf{u}}^T | \mathbf{X}) \mathbf{X} \right] = \frac{1}{\sigma_0^2} E(\mathbf{X}^T \mathbf{X}),$$

wobei das letzte Gleichheitszeichen nur bei homoskedastischen Fehlern gilt. Vernachlässigen des Erwartungswertes und Einsetzen in eine der beiden LM-Versionen ergibt

$$\frac{\tilde{\mathbf{u}}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \tilde{\mathbf{u}}}{\sigma_0^2} \sim \chi^2(q).$$

Indem σ_0^2 aus den Residuen des restringierten Modells geschätzt wird, erhält man die **Lagrange-Multiplikator-Statistik (LM) Statistik** im Falle des linearen Regressionsmodells

$$LM = n \frac{\tilde{\mathbf{u}}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \tilde{\mathbf{u}}}{\tilde{\mathbf{u}}^T \tilde{\mathbf{u}}} \xrightarrow{d} \chi^2(q). \quad (3.44)$$

Die LM -Statistik hat im Gegensatz zur F -Statistik keine *exakte* F -Verteilung, da der Zähler und der Nenner in endlichen Stichproben stochastisch nicht unabhängig sind. Deshalb lässt sich nur die asymptotische Verteilung berechnen. Dies erklärt, weshalb im normalen linearen Regressionsmodell mit linearen Restriktionen der LM-Test nicht verwendet wird.

- Die asymptotische Verteilung von (3.44) gilt auch, wenn die Fehler nicht normalverteilt sind. Wieso?
- Die **Berechnung der LM-Teststatistik** ist sehr **einfach**, wenn **Nullrestriktionen** getestet werden sollen. Dann erhält man $\tilde{\mathbf{u}}$ aus der Regression

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

und berechnet das Bestimmtheitsmaß R^2 für die Hilfsregression

$$\tilde{\mathbf{u}} = \mathbf{X}\boldsymbol{\gamma} + \mathbf{v}.$$

Dann kann gezeigt werden, dass

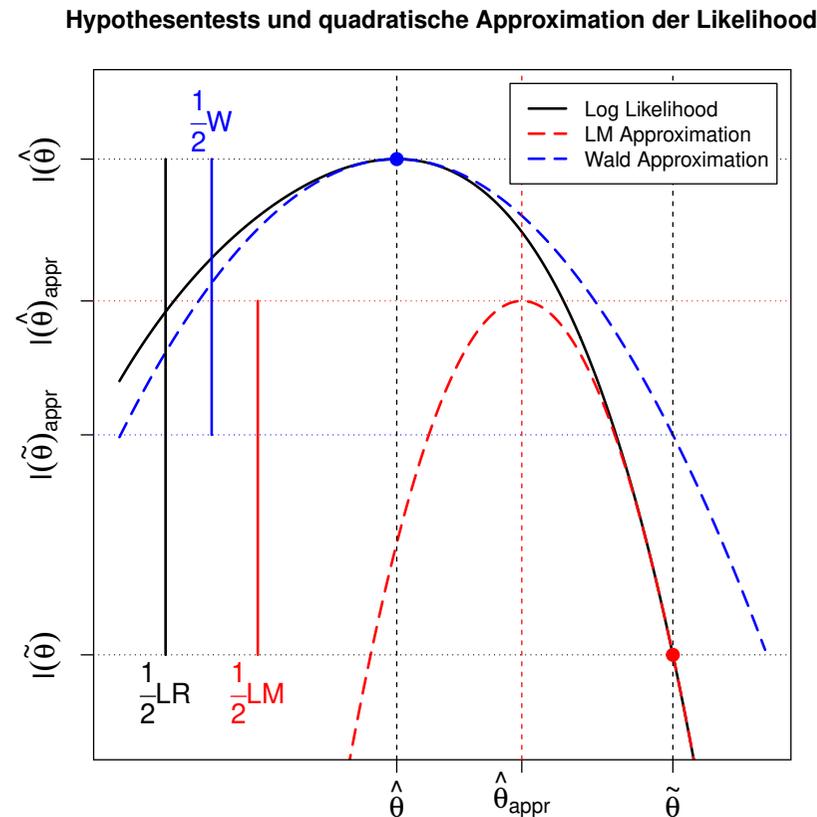
$$LM = n \frac{SSE}{SST} = nR^2 \longrightarrow \chi^2(q),$$

wobei $SSE = \tilde{\mathbf{u}}^T \underbrace{\mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T}_{\mathbf{P}_X} \tilde{\mathbf{u}}$ und $SST = \tilde{\mathbf{u}}^T \tilde{\mathbf{u}}$ ist.

- Die LM-Teststatistik ist **invariant** gegenüber Reparameterisierungen.

Noch einige allgemeine Bemerkungen:

- Die LM- und die Wald-Statistik sind zwei unterschiedliche Approximationen der LR-Statistik, da keine der beiden Ersteren die Schätzung des restringierten als auch des unrestringierten Modells erfordert. Ein Vergleich der drei Testprinzipien zeigt folgende Abbildung



- Für das **normale lineare Regressionsmodell** mit linearen Restriktionen und unbekannter Fehlervarianz σ_0 gilt, dass

$$LM < LR < W.$$

Wenn σ_0 bekannt ist, sind alle drei Tests numerisch identisch und exakt $\chi^2(q)$ -verteilt, siehe z.B. Davidson und MacKinnon (2004, S. 432).

- Davidson und MacKinnon (2004, S. 428-429) empfehlen zur Approximation der endlichen Stichprobenverteilung aller drei Testverfahren Bootstrap-Verfahren zu verwenden.
- Man beachte, dass die vorliegende Testtheorie viel allgemeiner ist als die für das KQ-Modell besprochene Testtheorie für den Fall linearer Restriktionen. Im vorliegenden Fall können auch nichtlineare Restriktionen getestet werden und die Art des Modells wird nur durch die Regularitätsbedingungen an die Likelihood-Funktion beschränkt!

4. 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:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad E(\mathbf{u}|\mathbf{X}) = \mathbf{0}, \quad Var(\mathbf{u}|\mathbf{X}) = E(\mathbf{u}\mathbf{u}^T|\mathbf{X}) = \boldsymbol{\Omega} \quad (4.1)$$

wobei angenommen wird, dass die Kovarianzmatrix $\boldsymbol{\Omega}$ positiv definit ist:

$$\begin{aligned} \mathbf{\Omega} &= \text{Var}(\mathbf{u}|\mathbf{X}) = E \left[(\mathbf{u} - E[\mathbf{u}|\mathbf{X}]) (\mathbf{u} - E[\mathbf{u}|\mathbf{X}])^T \right] \\ &= \begin{pmatrix} \text{Var}(u_1|\mathbf{X}) & \text{Cov}(u_1, u_2|\mathbf{X}) & \cdots & \text{Cov}(u_1, u_n|\mathbf{X}) \\ \text{Cov}(u_2, u_1|\mathbf{X}) & \text{Var}(u_2|\mathbf{X}) & \cdots & \text{Cov}(u_2, u_n|\mathbf{X}) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(u_n, u_1|\mathbf{X}) & \text{Cov}(u_n, u_2|\mathbf{X}) & \cdots & \text{Var}(u_n|\mathbf{X}) \end{pmatrix}. \end{aligned} \quad (4.2)$$

Spezialfälle:

- Das einfache lineare Modell ist ein Spezialfall: $\mathbf{\Omega} = \sigma^2 \mathbf{I}$.
- Ist $\mathbf{\Omega}$ eine Diagonalmatrix mit $\omega_t^2 = \text{Var}(u_t|\mathbf{X}) \neq \omega_s^2$ für einige $s, t, s \neq t$,

$$\mathbf{\Omega} = \begin{pmatrix} \omega_1^2 & 0 & \cdots & 0 \\ 0 & \omega_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \omega_n^2 \end{pmatrix}, \quad (4.3)$$

sind die Fehler nicht korreliert, aber heteroskedastisch.

- **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:

$$\text{Var}(u_t|\Omega_t) = \omega_t^2 \neq \sigma^2, \quad (4.4a)$$

$$\text{Var}(y_t|\Omega_t) = E \left[(y_t - E[y_t|\Omega_t])^2 | \Omega_t \right] = E[u_t^2|\Omega_t] = \omega_t^2. \quad (4.4b)$$

– **Beispiel:**

- * 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:

$$\text{Var}(u_t|\Omega_t) = E[u_t^2|\Omega_t] = h(\delta + \mathbf{Z}_t\gamma), \quad \mathbf{Z}_t \in \Omega_t. \quad (4.5)$$

Drei Fälle sind zu unterscheiden:

- * Die Funktion $h(\cdot)$ ist inklusive aller Parameterwerte für δ, γ bekannt, dann Verwendung des GLS-Schätzers (4.7), siehe Abschnitt 4.1.
- * Die Funktion $h(\cdot)$ ist parametrisch, aber die Parameter δ, γ sind unbekannt, dann Verwendung des FGLS-Schätzers (4.17), siehe Abschnitt 4.2.1.

- * Die Funktion $h(\cdot)$ ist vollständig unbekannt, dann Verwendung von heteroskedastie-robusten Standardfehlern, siehe Abschnitt 4.3.

4.1. Verallgemeinerter Kleinst-Quadrate-Schätzer

- **Verallgemeinerter Linearer Kleinst-Quadrateschätzer (generalized least squares estimators (GLS estimator)):**

$$\hat{\beta}_{GLS} = (\mathbf{X}^T \Omega^{-1} \mathbf{X})^{-1} \mathbf{X}^T \Omega^{-1} \mathbf{y}.$$

- **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 (4.1) von links mit Ψ^T liefert

$$\underbrace{\Psi^T \mathbf{y}}_{\mathbf{y}^*} = \underbrace{\Psi^T \mathbf{X}}_{\mathbf{X}^*} \boldsymbol{\beta} + \underbrace{\Psi^T \mathbf{u}}_{\mathbf{u}^*}$$

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{u}^*, \quad (4.6)$$

wobei Ψ^T genau so gewählt wurde, dass $E \left[\mathbf{u}^* (\mathbf{u}^*)^T \mid \mathbf{X} \right] = \mathbf{I}$ (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:

$$\hat{\boldsymbol{\beta}}_{GLS} = \left((\mathbf{X}^*)^T \mathbf{X}^* \right)^{-1} (\mathbf{X}^*)^T \mathbf{y}^* \quad (4.7a)$$

$$= (\mathbf{X}^T \boldsymbol{\Psi} \boldsymbol{\Psi}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Psi} \boldsymbol{\Psi}^T \mathbf{y} \quad (4.7b)$$

$$= (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y}. \quad (4.7c)$$

- Der GLS-Schätzer lässt sich auch direkt aus den (theoretischen) Momentenbedingungen

$$\begin{aligned} \mathbf{X}^T \Psi (\Psi^T \mathbf{y} - \Psi^T \mathbf{X} \beta) &= \mathbf{0} \\ \mathbf{X}^T \Omega^{-1} (\mathbf{y} - \mathbf{X} \beta) &= \mathbf{0} \end{aligned} \quad (4.8)$$

ableiten, bzw. auch aus der Minimierung der SSR des Modells (4.6).

• Annahmen zur Bestimmung der Schätzeigenschaften

(vgl. Abschnitt 1.1 zu KQ-Annahmen)

- **(B1)** Das Modell ist korrekt, d.h. der DGP ist im Modell (4.1) 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**

$$E\left(\hat{\boldsymbol{\beta}}_{GLS}\right) = \boldsymbol{\beta}.$$

- Unter **(B1)**, **(B2')** und **(B3)** hat der GLS-Schätzer Kovarianzmatrix

$$\begin{aligned} \text{Var}\left(\hat{\boldsymbol{\beta}}_{GLS}|\mathbf{X}\right) &= \left(\left(\mathbf{X}^*\right)^T \mathbf{X}^*\right)^{-1} \\ &= \left(\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}\right)^{-1} \end{aligned} \quad (4.9)$$

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 $\mathbf{W}_t = \left(W_{t1} \ W_{t2} \ \cdots \ W_{tk}\right)$, $t = 1, \dots, n$, in der Matrix $\mathbf{W}^T = \left(\mathbf{W}_1^T \ \mathbf{W}_2^T \ \cdots \ \mathbf{W}_n^T\right)$ 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. (4.8)) 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 $\mathbf{A}\mathbf{X} = \mathbf{I}$ 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 Ω viel Speicherplatz ($n = 10000$ benötigt bspw. 1600 MB.) Deshalb besser: Vorheriges Anwenden von Ψ ohne Abspeichern von Ψ (sofern möglich).
- **Gewichteter Kleinst-Quadrate Schätzer (weighted least squares (WLS))**
 u_t heteroskedastisch und unkorreliert (d.h. Ω diagonal). Damit ist Ω diagonal (4.3) und der Ansatz (4.6) lautet

$$\frac{y_t}{\omega_t} = \frac{1}{\omega_t} \mathbf{X}_t \boldsymbol{\beta} + \frac{u_t}{\omega_t}$$

mit $Var(u_t/\omega_t | \mathbf{X}) = 1$. **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 (4.6) 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 $R^2 = \widehat{Corr}(\hat{y}_t, y_t)^2$.

• 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:

$$\sqrt{n} \left(\hat{\beta}_{GLS} - \beta_0 \right) \xrightarrow{d} N \left(\mathbf{0}, \mathbf{S}_{\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X}}^{-1} \right).$$

- Oft ist Ψ bzw. Ω 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.

4.2. Feasible GLS

- Ist die Fehlerkovarianzmatrix Ω unbekannt, muss diese modelliert werden.
- **Asymptotische Eigenschaften von FGLS**

– Kurz gesagt, der FGLS-Schätzer

$$\hat{\beta}_{FGLS} = \left(\mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^T \hat{\Omega}^{-1} \mathbf{y} \quad (4.10)$$

ist konsistent und asymptotisch normalverteilt, wenn die Fehlerkovarianzmatrix Ω 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

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \Omega^{-1} \mathbf{X}, \quad (4.11a)$$

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \hat{\Omega}^{-1} \mathbf{u} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \Omega^{-1} \mathbf{u}. \quad (4.11b)$$

- * I.A. erfordert dies, dass in der ersten Stufe der β -Vektor konsistent geschätzt wird, so dass aus den durch \hat{u} konsistent geschätzten Fehlern z.B. γ in (4.5) und damit Ω konsistent geschätzt werden kann.
- * **Achtung:** Ist Ω nicht diagonal, ist der KQ-Schätzer für die erste Stufe i.A. inkonsistent! In diesem Fall sind andere Schätzverfahren notwendig.

4.2.1. Modellierung heteroskedastischer Fehler

- Ein häufig geeignetes Modell zur Modellierung von $Var(u_t|\Omega_t) = \omega_t^2$ ist

$$E[u_t^2|\Omega_t] = e^{\delta + \mathbf{Z}_t\boldsymbol{\gamma}} = e^\delta e^{\mathbf{Z}_t\boldsymbol{\gamma}}. \quad (4.12)$$

Die Gleichung (4.12) spezifiziert die Funktion $h(\cdot)$ in (4.5) als $h(\cdot) = \exp(\cdot)$.

Wird für eine Zufallsvariable v_t festgelegt, dass

$$E[v_t|\Omega_t] = 0 \quad \text{und} \quad Var(v_t|\Omega_t) = 1,$$

lässt sich u_t^2 schreiben als

$$u_t^2 = e^{\delta + \mathbf{Z}_t\boldsymbol{\gamma}} v_t^2, \quad (4.13)$$

so dass (4.12) gilt. Um δ und $\boldsymbol{\gamma}$ mit einer linearen Regression schätzen zu können, wird (4.13) logarithmiert:

$$\ln u_t^2 = \delta + \mathbf{Z}_t\boldsymbol{\gamma} + \ln v_t^2. \quad (4.14)$$

Da $E[\ln v_t^2] \neq \ln E[v_t^2] = 0$ (Jensen-Ungleichung (A.15), siehe Abschnitt A.3), verwendet man folgenden Trick:

$$\ln u_t^2 = \underbrace{\delta + E[\ln v_t^2]}_{\delta'} + \mathbf{Z}_t \boldsymbol{\gamma} + \underbrace{(\ln v_t^2 - E[\ln v_t^2])}_{\eta_t \text{ wobei } E[\eta_t | \Omega_t] = 0}$$

$$\ln u_t^2 = \delta' + \mathbf{Z}_t \boldsymbol{\gamma} + \eta_t. \quad (4.15)$$

• **2-stufiger Schätzer:**

1. Schritt: Schätze das Modell (4.1) mit KQ und speichere die Residuen

$$\hat{\mathbf{u}} = \mathbf{M}_X \mathbf{y}.$$

Einsetzen der Residuen in (4.15) zur KQ-Schätzung von δ' und $\boldsymbol{\gamma}$

$$\ln \hat{u}_t^2 = \delta' + \mathbf{Z}_t \boldsymbol{\gamma} + \text{Fehler}. \quad (4.16)$$

2. Schritt: Schätze (4.6)-Ansatz mit $\hat{\omega}_t^2 = \exp(\mathbf{Z}_t \hat{\boldsymbol{\gamma}})$

$$\frac{y_t}{\hat{\omega}_t} = \frac{1}{\hat{\omega}_t} \mathbf{X}_t \boldsymbol{\beta} + \frac{u_t}{\hat{\omega}_t}. \quad (4.17)$$

Der Faktor $e^{\hat{\delta}'}$ kann weggelassen werden, da er für alle Beobachtungen konstant ist.

- **FGLS oder KQ mit heteroskedastierobusten Standardfehlern?** Die Frage ist wie gut Ω geschätzt werden kann. Je unpräziser, desto eher wird man den KQ-Schätzer mit heteroskedastie-robuster Varianz-Kovarianzmatrix, siehe Abschnitt 4.3 nehmen.
- Es ist möglich, den FGLS-Schätzer zu iterieren. Dies hat *keinen* Einfluss auf die asymptotischen Eigenschaften, jedoch auf die Schätzeigenschaften in endlichen Stichproben.

4.2.2. Modelle mit autokorrelierten Fehlern

Siehe Davidson und MacKinnon (2004, Sections 7.6-7.9).

4.3. Heteroskedastie-robuste Standardfehler bei KQ-Schätzung

- **Ableitung heteroskedastie-robuster Standardfehler**

Liegen heteroskedastische Fehler vor, dann ist die Varianz-Kovarianzmatrix des KQ-Schätzers gegeben durch (siehe Handout **Methoden der Ökonometrie**, Kapitel 9)

$$\text{Var}(\hat{\boldsymbol{\beta}}|\mathbf{X}) = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \text{Var}(\mathbf{u}|\mathbf{X}) \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \quad (4.18)$$

$$= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega} \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}. \quad (4.19)$$

Diese Varianz-Kovarianzmatrix wird oft als **sandwich covariance matrix** bezeichnet, wobei $(\mathbf{X}^T \mathbf{X})^{-1}$ die „Brot scheiben“ 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 (4.19) des KQ-Schätzers bei Heteroskedastie mittels

$$\widehat{Var}(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1} \left(\sum_{t=1}^n \hat{u}_t^2 \mathbf{X}_t^T \mathbf{X}_t \right) (\mathbf{X}'\mathbf{X})^{-1} \quad (4.20)$$

schätzen.

- Anmerkungen:

- Die Standardfehler, die man aus (4.20) 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 (4.20) verwendet.
- In Davidson und MacKinnon (2004, Section 5.5) wird erklärt, warum (4.20) ein konsistenter Schätzer von (4.19) 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 (4.19) und deren Bezeichnungen in Davidson und MacKinnon (2004, Section 5.5) und den R-Paketen car oder sandwich.
 - „HC0“: White-Standardfehler (4.20).
 - „HC1“: Multipliziert White-Standardfehler (4.20) mit $n/(n - k)$. (Default in EViews.)
 - „HC2“: Ersetzt \hat{u}_t^2 White-Standardfehler (4.20) 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 (4.20) 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 ??, zu korrigieren. Zur genaueren Begründung der jeweiligen Korrekturen siehe Davidson und 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.

4.4. Empirische Analyse von Handelsströmen: Teil 3

Fortsetzung der Analyse von Modell (??):

$$\begin{aligned} \ln(\text{Importe}_i) = & \beta_1 + \beta_2 \ln(\text{BIP}_i) + \beta_3 \ln(\text{Entfernung}_i) \\ & + \beta_4 \text{Offenheit}_i + \beta_5 \ln(\text{Fläche}) + u_i. \end{aligned} \quad (??)$$

- 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

```
mod_formula <- log(trade_0_d_o) ~ log(wdi_gdpusdcr_o) + log(cepii_dist)
  + ebrd_tfes_o + log(cepii_area_o)
```

```
# FGLS-Schätzung
```

```
# 1. Schritt
```

```
model_kq <- lm(mod_formula)
resids <- residuals(model_kq)
fits <- fitted(model_kq)
```

```
mod_formula_ln_u_squared <- log(resids^2) ~ log(wdi_gdpusdcr_o) + log(cepii_dist) + ebrd_tfes_o + log
```

```
# 2. Schritt
```

```
omega <- exp(fitted(lm(mod_formula_ln_u_squared)))
model_gls <- lm(mod_formula, weights=1/omega)
(summary(model_gls))
```

Man erhält:

Call:

```
lm(formula = mod_formula, weights = 1/omega)
```

Residuals:

Min	1Q	Median	3Q	Max
-4.8020	-1.1872	0.4694	1.2607	3.0480

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.02361	1.23595	1.637	0.108703	
log(wdi_gdpusdcr_o)	1.07937	0.05679	19.008	< 2e-16	***
log(cepii_dist)	-0.88760	0.10985	-8.080	3.1e-10	***
ebrd_tfes_o	0.26335	0.17878	1.473	0.147863	
log(cepii_area_o)	-0.20314	0.04777	-4.253	0.000108	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.884 on 44 degrees of freedom

Multiple R-squared: 0.9341, Adjusted R-squared: 0.9281

F-statistic: 155.8 on 4 and 44 DF, p-value: < 2.2e-16

- **Heteroskedastie-robuste KQ-Schätzer**

```
(coeftest(model_kq,vcov=hccm(model_kq,type="hc1")))
```

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: Outputtabelle für Modell (??) 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)
$\ln(\text{Fläche})$	-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

4.5. Paneldaten

- **Paneldaten** werden sowohl über die Zeit, als auch über einen Querschnitt erhoben und daher mit einem Doppelindex versehen:
 - **Querschnitt:** $i = 1, \dots, m$ (z.B. Individuen)
 - **Längsschnitt:** $t = 1, \dots, T$ (Zeit)

Die **Gesamtanzahl der Beobachtungen** ist $n = mT$.

- Das lineare Paneldatenmodell lautet:

$$y_{it} = \sum_{l=1}^k x_{it,l} \beta_l + u_{it}, \quad i = 1, \dots, m, \quad t = 1, \dots, T,$$

bzw.

$$y_{it} = \mathbf{X}_{it} \boldsymbol{\beta} + u_{it}, \quad i = 1, \dots, m, \quad t = 1, \dots, T \quad (4.21)$$

wobei \mathbf{X}_{it} ein $(1 \times k)$ -Vektor von Regressoren ist.

- Durch geeignete Anordnung erhält man wieder einen Ansatz $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, z.B. kann man die $n = mT$ Beobachtungen wie folgt stapeln:
 alle T Beobachtungen von Individuum 1,
 alle T Beobachtungen von Individuum 2,
 ...
 alle T Beobachtungen von Individuum m .

$$\underbrace{\begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ y_{31} \\ \vdots \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} x_{11,1} & x_{11,2} & \cdots & x_{11,k} \\ \vdots & \vdots & & \vdots \\ x_{1T,1} & x_{1T,2} & \cdots & x_{1T,k} \\ x_{21,1} & x_{21,2} & \cdots & x_{21,k} \\ \vdots & \vdots & & \vdots \\ x_{2T,1} & x_{2T,2} & \cdots & x_{2T,k} \\ x_{31,1} & x_{31,2} & \cdots & x_{31,k} \\ \vdots & \vdots & & \vdots \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}}_{\boldsymbol{\beta}} = \underbrace{\begin{pmatrix} u_{11} \\ \vdots \\ u_{1T} \\ u_{21} \\ \vdots \\ u_{2T} \\ u_{31} \\ \vdots \end{pmatrix}}_{\mathbf{u}} .$$

- Im Paneldatenkontext ist es unrealistisch, für den Ansatz

$$y_{it} = \mathbf{X}_{it}\boldsymbol{\beta} + u_{it}, \quad i = 1, \dots, m, \quad t = 1, \dots, T$$

von IID-Fehlern u_{it} auszugehen. Es ist daher wichtig, sich die Abhängigkeits- und Kovarianzstruktur des Fehlervektors \mathbf{u} inhaltlich (und exemplarisch) klarzumachen.

- **Mögliche Korrelationen in den Fehlern** u_{it} :

- a) Schocks beeinflussen **alle** Beobachtungen eines Individuums j , d.h. u_{j1}, \dots, u_{jT} sind korreliert.
- b) Schocks beeinflussen **alle** Beobachtungen zu einem Zeitpunkt t , d.h. u_{jt} und u_{it} sind korreliert.
- c) Es liegen sowohl Fall a) als auch Fall b) vor.
- d) Die Fälle a), b) bzw. c) gelten nur für einige Beobachtungen.

Ignoriert man die Kovarianzstruktur in den Residuen: ineffiziente Schätzer und inkonsistente Schätzung der Kovarianzmatrix des Schätzers.

Falls also nicht $E[\mathbf{u}\mathbf{u}^T | \mathbf{X}] = \sigma^2 \mathbf{I}$ gilt, ist **KQ ineffizient** (warum?)

- **Das Fehler-Komponenten-Modell (error-components model)**

Zerlegung der Fehler u_{it} in unabhängige Komponenten

$$u_{it} = e_t + v_i + \varepsilon_{it} \quad (4.22)$$

wobei der Fehler e_t alle abhängigen Variablen zum Zeitpunkt t beeinflusst, der Fehler v_i alle abhängigen Variablen des Individuums i beeinflusst und ε_{it} allein die abhängige Beobachtung des Individuums i zum Zeitpunkt t .

- Je nach Annahmen bezüglich der nicht-individuellen Fehler e_t und v_i ergeben sich unterschiedliche Paneldatenmodelle, die auch unterschiedliche Schätzverfahren erfordern.
- In diesem Abschnitt werden folgende **Vereinfachungen** angenommen:
 - * $e_t = 0, t = 1, \dots, T,$
 - * \mathbf{X}_{it} ist exogen. (Erlaubt man auch verzögerte endogene Variablen, wird alles komplizierter.)

- Unter Beachtung von (4.21) und (4.22) mit $e_t = 0$, $t = 1, \dots, T$, gehen wir also von folgendem Ansatz aus:

$$y_{it} = v_i + \mathbf{X}_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

- Beispiel: Lohngleichungen. Dabei könnte
- * y_{it} z.B. der log des Lohneinkommens des Haushaltsvorstandes von Haushalt i zum Zeitpunkt t sein,
 - * \mathbf{X}_{it} Variablen wie etwa Berufserfahrung, Lohnspreizung ($q_{.75}(y) - q_{.25}(y)$) und Arbeitslosenquote enthalten, und
 - * v_i der fixe Effekt sein, der alle haushaltsspezifischen, zeitinvarianten Effekte erfasst die nicht als Regressoren berücksichtigt werden, z.B. die nicht-beobachtbare Begabung des Haushaltsvorstandes von Haushalt i .

– Zur Modellierung der individuenspezifischen Fehler v_i existieren zwei weitverbreitete Modellansätze:

* **Dummy-Variablen-Modell (Fixed-Effects-Modell):**

Man interpretiert

$$v_i, \quad i = 1, \dots, m$$

als feste, jedoch unbekannte Parameter, die eine individuenspezifische Modellierung der Konstanten erlauben und mit Dummyvariablen geschätzt werden können.

* **Fehlerkomponentenmodell (Random-Effects-Modell):**

Die v_i werden als Zufallsvariablen betrachtet, so dass die

$$u_{it} = v_i + \varepsilon_{it}$$

für jedes Individuum über die Zeit hinweg korreliert sind. In der Schätzung werden sie im Rahmen von (F)GLS-Ansätzen berücksichtigt.

- **Das Dummy-Variablen-Modell (Fixed-Effects-Modell)**

Es enthalte \mathbf{X} keine Konstante und die Dummyvariablen d_{it} sind wie folgt definiert:

$$d_{it} = \begin{cases} 1 & \text{falls die } i\text{-te Beobachtung der Periode } t \text{ zur Gruppe } i \text{ gehört} \\ 0 & \text{sonst.} \end{cases}$$

Dann lässt sich das **Dummy-Variablen-Modell** schreiben als

$$\begin{aligned} y_{it} &= \mathbf{X}_{it}\boldsymbol{\beta} + d_{it}v_i + \varepsilon_{it} \\ \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \quad E[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T | \mathbf{X}, \boldsymbol{\eta}] = \sigma^2\mathbf{I}. \end{aligned} \quad (4.23)$$

- **Annahmen:**

Der Vektor $\boldsymbol{\eta}$ kann als Parametervektor aufgefasst werden. Wird der Vektor $\boldsymbol{\eta}$ jedoch als zufällig vorausgesetzt, dann muss folgende Annahme getroffen werden:

Der Vektor der fixed effects $\boldsymbol{\eta}$ ist unabhängig von ε_{it} . Der Vektor $\boldsymbol{\eta}$ kann jedoch mit den erklärenden Variablen \mathbf{X} korreliert sein.

- Die $(mT \times m)$ Matrix \mathbf{D} setzt sich aus den Dummyvariablen zusammen und lautet ausgeschrieben:

$$\mathbf{D} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}, \begin{matrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \\ 2 & 1 \\ \vdots & \vdots \\ 2 & T \\ 3 & 1 \\ \vdots & \vdots \\ m & T \end{matrix}, \text{ (} i \text{ läuft in der 1. Spalte und } t \text{ in der 2. Spalte)}$$

Der Vektor $\boldsymbol{\eta}$ fasst alle m fixed effects v_i zusammen.

– **Schätzung:**

Da $\varepsilon|\mathbf{X}, \boldsymbol{\eta} \sim \text{IID}$ ist, kann der KQ-Schätzer für $\boldsymbol{\beta}$ angewendet werden. Er ergibt sich bei gegebenen $\boldsymbol{\eta}$ aus den Momentenbedingungen

$$\begin{aligned} E [\mathbf{X}_{it}^T (y_{it} - \mathbf{X}_{it}\boldsymbol{\beta} - v_i)] &= \mathbf{0}, & i = 1, \dots, m, & , t = 1, \dots, T, \\ E [y_{it} - \mathbf{X}_{it}\boldsymbol{\beta} - v_i] &= \mathbf{0}, & i = 1, \dots, m, & , t = 1, \dots, T. \end{aligned}$$

Um das Dimensionalitätsproblem (i.e. jede Menge unbekannter Parameter) zu umgehen, wendet man das FWL-Theorem an und berechnet aus

$$\mathbf{M}_D \mathbf{y} = \mathbf{M}_D \mathbf{X} \boldsymbol{\beta} + \text{Residuen}$$

den KQ-Schätzer für $\boldsymbol{\beta}$ in (4.23)

$$\hat{\boldsymbol{\beta}}_{FE} = (\mathbf{X}^T \mathbf{M}_D \mathbf{X})^{-1} \mathbf{X}^T \mathbf{M}_D \mathbf{y}. \quad (4.24)$$

Dabei ist \mathbf{M}_D die idempotente Projektionsmatrix $\mathbf{M}_D = \mathbf{I}_n - \mathbf{D}(\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T$. Wegen $\mathbf{D}^T \mathbf{D} = T \mathbf{I}_m$ und $(\mathbf{D}^T \mathbf{D})^{-1} = T^{-1} \mathbf{I}_m$ gilt $\mathbf{M}_D = \mathbf{I}_n - T^{-1} \mathbf{D} \mathbf{D}^T$.

Der Fixed-Effects Schätzer für β hat eine einfache **Interpretation**, denn dieser Schätzer entspricht dem KQ-Schätzer des mittelwertbereinigten Modells

$$y_{it} - \bar{y}_i = \sum_{l=1}^k \beta_l (x_{it,l} - \bar{x}_{i,l}) + \varepsilon_{it} - \bar{\varepsilon}_i, \quad i = 1, \dots, m, \quad t = 1, \dots, T.$$

Dabei sei \mathbf{x}_l die l -te Spalte aus \mathbf{X} (mit den Elementen $x_{it,l}$) und $\bar{x}_{i,l}$ der Gruppendurchschnitt (group mean) der Gruppe i bei Variable l .

Man bezeichnet diesen Schätzer auch als **Within-Group-Schätzer**, da er alle Gruppendummyvariablen v_i eliminiert. Letztere können durch

$$\hat{v}_i = \bar{y}_i - \left(\bar{x}_{i,1} \quad \bar{x}_{i,2} \quad \dots \quad \bar{x}_{i,k} \right) \hat{\beta}_{FE}$$

geschätzt werden (verifizieren!).

– Algebraische Ableitung des Within-Group-Schätzers

$$(\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{x}_l = \frac{1}{T} \mathbf{D}^T \mathbf{x}_l = \frac{1}{T} \begin{pmatrix} \sum_{t=1}^T x_{1t,l} \\ \sum_{t=1}^T x_{2t,l} \\ \vdots \\ \sum_{t=1}^T x_{mt,l} \end{pmatrix} = \begin{pmatrix} \bar{x}_{1\cdot,l} \\ \bar{x}_{2\cdot,l} \\ \vdots \\ \bar{x}_{m\cdot,l} \end{pmatrix}$$

$$\mathbf{P}_D \mathbf{x}_l = \mathbf{D} (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{x}_l = \begin{pmatrix} \bar{x}_{1\cdot,l} \\ \bar{x}_{1\cdot,l} \\ \vdots \\ \bar{x}_{1\cdot,l} \\ \bar{x}_{2\cdot,l} \\ \vdots \\ \bar{x}_{2\cdot,l} \\ \bar{x}_{3\cdot,l} \\ \vdots \\ \bar{x}_{m\cdot,l} \end{pmatrix} \quad \text{und} \quad \mathbf{M}_D \mathbf{y} = \begin{pmatrix} y_{11} - \bar{y}_{1\cdot} \\ y_{12} - \bar{y}_{1\cdot} \\ \vdots \\ y_{1T} - \bar{y}_{1\cdot} \\ y_{21} - \bar{y}_{2\cdot} \\ \vdots \\ y_{2T} - \bar{y}_{2\cdot} \\ y_{31} - \bar{y}_{3\cdot} \\ \vdots \\ y_{mT} - \bar{y}_{m\cdot} \end{pmatrix}$$

D.h.

$$\mathbf{M}_D \mathbf{y} = (\mathbf{I}_n - \mathbf{P}_D) \mathbf{y}$$

bezeichnet die Abweichungen der abhängigen Variablen von den jeweiligen Gruppenmittelwerten. Entsprechend ergibt sich aus

$$\mathbf{M}_D \mathbf{y} = \mathbf{M}_D \mathbf{X} \boldsymbol{\beta} + \mathbf{M}_D \boldsymbol{\varepsilon}$$

das gruppenmittelwertbereinigte Modell

$$y_{it} - \bar{y}_i = \sum_{l=1}^k \beta_l (x_{it,l} - \bar{x}_{i,l}) + \varepsilon_{it} - \bar{\varepsilon}_{i..}$$

- Für den in (4.24) definierten Schätzer $\hat{\boldsymbol{\beta}}_{FE}$ (auch: **least squares dummy variables estimator (LSDV)**) gilt unter Beachtung von (4.23)

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{FE} &= (\mathbf{X}^T \mathbf{M}_D \mathbf{X})^{-1} \mathbf{X}^T \mathbf{M}_D \underbrace{(\mathbf{X} \boldsymbol{\beta}_0 + \mathbf{D} \boldsymbol{\eta}_0 + \boldsymbol{\varepsilon})}_{=\mathbf{y}} \\ &= \boldsymbol{\beta}_0 + (\mathbf{X}^T \mathbf{M}_D \mathbf{X})^{-1} \mathbf{X}^T \mathbf{M}_D \mathbf{D} \boldsymbol{\eta}_0 + (\mathbf{X}^T \mathbf{M}_D \mathbf{X})^{-1} \mathbf{X}^T \mathbf{M}_D \boldsymbol{\varepsilon}, \end{aligned}$$

d.h. er kann auch geschrieben werden als (verifizieren!)

$$\hat{\boldsymbol{\beta}}_{FE} = \boldsymbol{\beta}_0 + (\mathbf{X}^T \mathbf{M}_D \mathbf{X})^{-1} \mathbf{X}^T \mathbf{M}_D \boldsymbol{\varepsilon} \quad (4.25)$$

– Schätzeigenschaften in endlichen Stichproben

Für nicht stochastische η gilt:

* Erwartungstreue

Ist \mathbf{X} (streng) exogen ($E[\boldsymbol{\varepsilon}|\mathbf{X}] = \mathbf{0}$), dann ist

$$E \left[\hat{\boldsymbol{\beta}}_{FE} | \mathbf{X} \right] = \boldsymbol{\beta}_0 + (\mathbf{X}^T \mathbf{M}_D \mathbf{X})^{-1} \mathbf{X}^T \mathbf{M}_D E[\boldsymbol{\varepsilon} | \mathbf{X}] = \boldsymbol{\beta}_0.$$

* Varianz

$$Var \left(\hat{\boldsymbol{\beta}}_{FE} | \mathbf{X} \right) = \sigma_0^2 (\mathbf{X}^T \mathbf{M}_D \mathbf{X})^{-1}.$$

Schätzer:

$$\widehat{Var} \left(\hat{\boldsymbol{\beta}}_{FE} | \mathbf{X} \right) = \frac{\hat{\boldsymbol{\varepsilon}}^T \hat{\boldsymbol{\varepsilon}}}{m(T-1) - k} (\mathbf{X}^T \mathbf{M}_D \mathbf{X})^{-1}.$$

Beachte: im Nenner wird $m(T-1) - k$ statt $mT - k$ verwendet.

- * Ist \mathbf{X} exogen und η nicht stochastisch und gelten alle anderen üblichen Annahmen, dann ist der fixed-effects Schätzer **BLU**.

– Asymptotische Schätzeigenschaften

Was bedeutet $n \rightarrow \infty$, wenn $n = mT$? Hier sind die Schätzeigenschaften für β und η zu unterscheiden und die Bedingungen, welche(r) Index/Indizes gegen Unendlich streben.

- (i) m fest, $T \rightarrow \infty$ und $E[\varepsilon|\mathbf{X}] = \mathbf{0}$: Konsistenz und asymptotische Normalität des KQ-Schätzers gelten sowohl für β als auch für η . Man beachte, dass $E[\varepsilon|\mathbf{X}] = \mathbf{0}$ verletzt sein kann, wenn Gruppen weggelassen wurden, die jedoch mit den \mathbf{X}_{it} für die berücksichtigten Gruppen korreliert sind. Die Fehler ε_{it} können durch $\hat{\varepsilon}_{it}$ konsistent geschätzt werden und somit auch σ_ε^2 durch $s^2 = \hat{\varepsilon}^T \hat{\varepsilon} / (mT - m - k)$.
- (ii) $m \rightarrow \infty$, T fest und $E[\varepsilon|\mathbf{X}] = \mathbf{0}$: Nur β kann konsistent geschätzt werden, die fixed effects η jedoch nicht, da für Letztere, egal wie groß m ist, nur eine feste Anzahl an Beobachtungen zur Schätzung eines Gruppendummies vorliegen.
- (iii) $m \rightarrow \infty$, $T \rightarrow \infty$ und $E[\varepsilon|\mathbf{X}] = \mathbf{0}$: alles paletti.

– **Vor- und Nachteile von fixed effects Modellen:**

- + erfordert keine Berechnung von $(mT \times T)$ Matrizen.
- es können keine Regressoren berücksichtigt werden, die für eine Gruppe konstant sind (z.B. Ausbildungsdauer im Lohnbeispiel).
- großer Verbrauch an Freiheitsgraden, falls m groß.
- falls \mathbf{X} mit \mathbf{D} korreliert ist, dann weist $\hat{\beta}_{FE}$ hohe Schätzvarianz auf.

• **Random Effects Panelmodell**

- **Annahme:** Regressoren \mathbf{X} und **zufällige Effekte (random effects)** v_i sind stochastisch unabhängig von den unsystematischen Fehlern ε_{it} , für alle i, t , und $v_i | \mathbf{X} \sim IID(0, \sigma_v^2)$.

Und: Regressoren und random effects v_i sind stochastisch unabhängig:

$$E[u_{it} | \mathbf{X}] = E[v_i + \varepsilon_{it} | \mathbf{X}] = 0.$$

- * Diese Annahme ist nicht immer realistisch (z.B. im Lohnbeispiel ...).
- * Unter dieser Annahme ist der KQ-Schätzer unverzerrt, aber nicht effizient, da $u_{it} | \mathbf{X} \not\sim IID(0, \sigma^2)$.

– Kovarianzmatrix der u_{it}

$$\text{Var}(u_{it}|\mathbf{X}) = \sigma_\varepsilon^2 + \sigma_v^2$$

$$\text{Cov}(u_{it}, u_{is}|\mathbf{X}) = \sigma_v^2, \quad s, t = 1, \dots, T$$

$$\text{Cov}(u_{it}, u_{js}|\mathbf{X}) = 0, \quad \text{für alle } i \neq j, \quad i, j = 1, \dots, m \text{ und } s, t = 1, \dots, T.$$

Zusammenfassen dieser (Ko)Varianzen in einer Matrix für Gruppe i

$$\text{Var} \left(\begin{pmatrix} u_{i1} \\ \vdots \\ u_{iT} \end{pmatrix} \middle| \mathbf{X} \right) = \begin{pmatrix} \sigma_v^2 + \sigma_\varepsilon^2 & \sigma_v^2 & \cdots & \sigma_v^2 \\ \sigma_v^2 & \sigma_v^2 + \sigma_\varepsilon^2 & \cdots & \sigma_v^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_v^2 & \sigma_v^2 & \cdots & \sigma_v^2 + \sigma_\varepsilon^2 \end{pmatrix} = \Sigma.$$

bzw. für alle Gruppen $1 \leq i, j \leq m$

$$\Omega \equiv E(\mathbf{u}\mathbf{u}^T|\mathbf{X}) = \begin{pmatrix} \Sigma & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \Sigma & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \Sigma \end{pmatrix}.$$

– Es gilt

$$\Sigma = \sigma_v^2 \begin{pmatrix} 1 + \sigma_\varepsilon^2/\sigma_v^2 & 1 & \cdots & 1 \\ 1 & 1 + \sigma_\varepsilon^2/\sigma_v^2 & \cdots & 1 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & \cdots & 1 + \sigma_\varepsilon^2/\sigma_v^2 \end{pmatrix}$$

und ein konsistenter Schätzer für σ_ε^2 lässt sich immer finden (s.o.).

– Ist $Var(v_i|\mathbf{X}) = \sigma_v^2$ bekannt, lässt sich der GLS-Schätzer anwenden.

– Ist σ_v^2 unbekannt, muss es konsistent geschätzt werden und dann der FGLS-Schätzer angewendet werden. Man nennt den entsprechenden Schätzer auch **random effects Schätzer**. Um σ_v^2 konsistent zu schätzen, ist erforderlich, dass $m \rightarrow \infty$.

D.h. dass die **asymptotische Verteilung des Random Effects Schätzers bei kleiner Gruppenanzahl m schlechte Approximationseigenschaften aufweist!**

- **Darstellung des GLS-Schätzers:**

- Zur Wiederholung: Der **GLS-Schätzer** lautet bei bekannten σ_v^2 und σ_ε^2 :

$$\hat{\beta}_{RE} = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y}.$$

- Für große m und T ist diese Formel nicht günstig. Stattdessen sollte man $\boldsymbol{\Psi}$ aus $\boldsymbol{\Omega}^{-1} = \boldsymbol{\Psi} \boldsymbol{\Psi}^T$ bestimmen. Damit ergibt sich der GLS-Schätzer aus der Regression

$$(\mathbf{I}_n - \lambda \mathbf{P}_D) \mathbf{y} = (\mathbf{I}_n - \lambda \mathbf{P}_D) \mathbf{X} \boldsymbol{\beta} + \text{Residuen}.$$

- **Interpretation:**

- * Der GLS-Schätzer ist ein matrixgewichteter Durchschnitt des einfachen **KQ-Schätzers**

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

- und des **Between-Groups-Schätzers**

$$\hat{\beta}_{BG} = (\mathbf{X}^T \mathbf{P}_D \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_D \mathbf{y}.$$

- * Letzterer entspricht dem KQ-Schätzer für das Modell

$$\mathbf{P}_D \mathbf{y} = \mathbf{P}_D \mathbf{X} \boldsymbol{\beta} + \text{Residuen.}$$

Man beachte, dass dieses Modell genau m verschiedene Beobachtungen hat, da die Projektionsmatrix \mathbf{P}_D die Mittelwerte jeder Gruppe erzeugt (siehe oben).

- * Ist $\sigma_v^2 = 0$, dann gibt es keine Random Effects und der GLS-Schätzer entspricht dem KQ-Schätzer, da $\lambda = 0$ wegen

$$\lambda = 1 - \left(\frac{T\sigma_v^2}{\sigma_\varepsilon^2} + 1 \right)^{-1/2}. \quad (4.26)$$

- * Ist σ_v^2 im Vergleich zu σ_ε^2 sehr groß, ist im Extremfall also $\sigma_\varepsilon^2 = 0$, dann ist $\lambda = 1$ und der GLS-Schätzer des Random-Effects Modell entspricht dem Fixed-Effects Schätzer, da $\mathbf{M}_D = \mathbf{I} - \mathbf{P}_D$.
- * Für $T \rightarrow \infty$ erhält man ebenfalls $\lambda = 1$, d.h. für sehr große T sind der GLS-Schätzer und der Fixed-Effects-Schätzer gleich.

- Zur Wahl zwischen Fixed-Effects und Random-Effects-Modell:
 - Das Random-Effects-Modell wählt man, wenn die beobachteten Gruppen aus einer Grundgesamtheit zufällig gezogen wurden (man also beliebig viele Gruppen ziehen könnte) *und* die Random Effects von der Regressoren stochastisch unabhängig sind.
 - Das Fixed-Effects-Modell wählt man, wenn man die Schätzung von β *gegeben* die vorliegenden Gruppen anstrebt.

- **Hausman-Test**

Der Hausman-Test liefert basierend auf der Annahme der Unkorreliertheit zwischen den zufälligen Effekten und den exogenen Variablen einen Test, um zwischen fixen und zufälligen Effekten zu unterscheiden. Die Nullhypothese lautet: “Effekte” sind mit den übrigen Regressoren unkorreliert (Modell mit zufälligen Effekten).

Beachte: dieser Test liefert keine Aussage “fixed/random effects liegen vor”!

Zu Problemen des Hausman-Tests siehe aktuell [Guggenberger \(2008\)](#).

Zu lesen: [Davidson und MacKinnon \(2004\)](#), Kapitel 7.

4.6. Multivariate Regression

wird in Vorlesung behandelt und später hier ergänzt.

5. Instrumentvariablenschätzung

- Für die Konsistenz des KQ-Schätzers musste vorausgesetzt werden, dass für das dynamische Regressionsmodell (1.2)

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t$$

die Annahme **(C2a)** $E(u_t|\Omega_t) = 0$ gilt.

- Die Annahme **(C2a)** ist **verletzt**, wenn \mathbf{X}_t mindestens eine Variable enthält, die als mögliche Variable für die Informationsmenge Ω_t *nicht* in Frage kommt. Dann gilt $\{\mathbf{X}_t\} \not\subseteq \Omega_t$ und es ist mindestens eine Variable in \mathbf{X}_t nicht vorherbestimmt bezüglich des Fehlers u_t , siehe Abschnitt 1.1.

- Ist die Annahme **(C2a)** $E(u_t|\Omega_t) = 0$ verletzt,
 - dann gilt

$$E(u_t|\mathbf{X}_t) = h(\mathbf{X}_t) \neq 0 \quad \text{für mindestens einige } \mathbf{X}_t \in \mathbb{X}, \quad (5.1)$$

wobei \mathbb{X} den Träger der Dichte von \mathbf{X}_t bezeichnet.

- Dann ist möglich, dass

$$E(\mathbf{X}_t u_t) = E[\mathbf{X}_t E(u_t|\mathbf{X}_t)] = \mathbf{c}_t \neq \mathbf{0} \quad (5.2)$$

sodass dann der KQ-Schätzer inkonsistent ist, siehe XXX

- Zur Schätzung kann des linearen Regressionsmodells kann unter bestimmten Annahmen der im folgenden eingeführte IV-Schätzer verwendet werden.

5.1. Anwendungsbeispiele für den IV-Schätzer

- **Modellierungsfälle**, die im Allgemeinen den IV-Schätzer erfordern:
 - Vorliegen von Regressoren, die mit Fehlern gemessen werden.
 - Eine mit \mathbf{X}_t korrelierter Regressor steht nicht zur Verfügung.
 - Vorliegen eines simultanen Gleichungssystems.
 - Vorliegen von endogenen Regressoren.

5.1.1. Modelle mit Fehlern in den Variablen (error in variables)

- Fehler in den Variablen liegen vor, wenn sich die Messwerte von den (unbeobachteten) tatsächlichen Werten unterscheiden.
- Beispiel: die unbeobachteten Variablen y_t^\bullet und x_t^\bullet unterscheiden sich von den beobachteten Werten y_t und x_t durch die Messfehler v_{yt} und v_{xt}

$$y_t = y_t^\bullet + v_{yt}, \quad v_{yt} | x_t^\bullet, y_t^\bullet \sim IID(0, \omega_y^2), \quad (5.3a)$$

$$x_t = x_t^\bullet + v_{xt}, \quad v_{xt} | x_t^\bullet, y_t^\bullet \sim IID(0, \omega_x^2). \quad (5.3b)$$

Damit basiert das Regressionsmodell

$$y_t^\bullet = \beta_1 + \beta_2 x_t^\bullet + u_t^\bullet, \quad u_t^\bullet | x_t^\bullet, y_t^\bullet \sim IID(0, \sigma^2)$$

auf unbeobachtbaren Variablen. Wir müssen aber

$$y_t = \beta_1 + \beta_2 x_t + u_t,$$

schätzen, basierend auf

$$\begin{aligned} y_t &= \beta_1 + \beta_2(x_t - v_{xt}) + v_{yt} + u_t^\bullet \\ &= \beta_1 + \beta_2 x_t + \underbrace{u_t^\bullet + v_{yt} - \beta_2 v_{xt}}_{u_t}. \end{aligned}$$

- Auswirkungen der Messfehler gemäß (5.3a) bzw. (5.3b):
Sind die Fehler v_{xt} , v_{yt} , u_t^\bullet voneinander und von y_t^\bullet , x_t^\bullet stochastisch unabhängig, lautet die Varianz von u_t

$$\text{Var}(u_t | y_t^\bullet, x_t^\bullet) = \sigma^2 + \omega_y^2 + \beta_2^2 \omega_x^2.$$

– Zu (5.3a):

Der Messfehler in der abhängigen Variable y_t ist vernachlässigbar, falls ω_y^2 klein ist.

– Zu (5.3b):

Aber: v_{xt} verursacht eine Abhängigkeit zwischen u_t und x_t , da mit

$$u_t = u_t^\bullet + v_{yt} - \beta_2 v_{xt}$$

$$x_t = x_t^\bullet + v_{xt}$$

$$E(u_t x_t) = E[E(u_t x_t | x_t^\bullet)] = -\beta_2 E(v_{xt}^2) = -\beta_2 \omega_x^2 \neq 0$$

folgt ($\beta_2 \neq 0$ vorausgesetzt).

Daraus folgt auch, dass $E(u_t | x_t) \neq 0$ für einige x_t , also **(C2a)** verletzt ist. Denn würde $E(u_t | x_t) = 0$ für alle x_t gelten, müsste aus $E[E(u_t x_t | x_t)] = E[0] = 0$ folgen, was zu einem Widerspruch führt. Damit ist die Variable x_t bezüglich des Fehlers u_t **nicht vorherbestimmt!**

- Beachte, dass aus $E(u_t | x_t) \neq 0$ für einige x_t nicht folgt, dass $E(u_t) = 0$. Wieso?

5.1.2. Simultane Gleichungsmodelle

- **Klassisches Beispiel 1:**

Simultane Schätzung einer **Angebots- und Nachfragefunktion im Rahmen eines partiellen Gleichgewichtsmodells.**

$$q_t = \gamma_d p_t + \mathbf{X}_t^d \boldsymbol{\beta}_d + u_t^d \quad \text{Nachfragefunktion}$$

$$q_t = \gamma_s p_t + \mathbf{X}_t^s \boldsymbol{\beta}_s + u_t^s \quad \text{Angebotsfunktion.}$$

Da sowohl q_t als auch p_t endogen sind, ist es günstig, das Gleichungssystem umzuformen zu

$$\begin{pmatrix} 1 & -\gamma_d \\ 1 & -\gamma_s \end{pmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} = \begin{pmatrix} \mathbf{X}_t^d \boldsymbol{\beta}_d \\ \mathbf{X}_t^s \boldsymbol{\beta}_s \end{pmatrix} + \begin{pmatrix} u_t^d \\ u_t^s \end{pmatrix}$$

und, sofern $\gamma_d \neq \gamma_s$ ist, zu

$$\begin{pmatrix} q_t \\ p_t \end{pmatrix} = \begin{pmatrix} 1 & -\gamma_d \\ 1 & -\gamma_s \end{pmatrix}^{-1} \left[\begin{pmatrix} \mathbf{X}_t^d \boldsymbol{\beta}_d \\ \mathbf{X}_t^s \boldsymbol{\beta}_s \end{pmatrix} + \begin{pmatrix} u_t^d \\ u_t^s \end{pmatrix} \right].$$

Man sieht, dass die Menge q_t und der Preis p_t jeweils von beiden Fehlertermen u_t^d und u_t^s abhängen. Damit ist der Regressor in der Nachfragefunktion mit dem

Fehlerterm u_t^d und der Regressor in der Angebotsgleichung mit u_t^s korreliert. Damit ist p_t weder bezüglich des Fehlers u_t^d noch des Fehlers u_t^s vorherbestimmt!

- **Klassisches Beispiel 2:**

Schätzen einer **Konsumfunktion** in einem **einfachen Keynesianischen Modell** (Davidson, 2000, Section 4.2.3):

C_t : Konsum in Periode t ,

Y_t : Einkommen in Periode t ,

I_t : Investitionen in Periode t .

Modellannahmen:

– Es liegt eine lineare Konsumfunktion vor:

$$C_t = \beta_1 + \beta_2 Y_t + u_{1t} \quad (5.4a)$$

– Investitionen sind autonom (sehr vereinfachend...) und entsprechen einem AR(1)-Prozess

$$I_t = \delta_1 + \delta_2 I_{t-1} + u_{2t} \quad (5.4b)$$

- Der Markt ist im Gleichgewicht (ex-post eine Identität)

$$Y_t \equiv C_t + I_t \quad (5.4c)$$

- Gegeben die Informationsmenge $\Omega_t = \{Y_{t-1}, Y_{t-2}, \dots, C_{t-1}, C_{t-2}, \dots\}$ gilt

$$E(u_t | \Omega_t) = 0 \quad \Leftrightarrow \quad \begin{cases} E(u_{1t} | Y_{t-1}, C_{t-1}) = 0 \\ E(u_{2t} | Y_{t-1}, C_{t-1}) = 0 \end{cases} \quad (5.4d)$$

Die Fehler u_{it} und u_{js} sind gegeben Ω_t stochastisch unabhängig für alle s, t und $i \neq j$.

Berechnen von $E(Y_t u_{1t})$ (entspricht $E(\mathbf{X}_t u_t)$ in allgemeinem Modell, vgl. (5.6)):

- Berechnen einer Gleichung für Y_t : Einsetzen von $C_t = Y_t - I_t$ aus (5.4c) in (5.4a) und Umformen ergibt

$$\begin{aligned} Y_t &= \frac{\beta_1}{1 - \beta_2} + \frac{1}{1 - \beta_2} I_t + \frac{1}{1 - \beta_2} u_{1t} \\ &= \frac{\beta_1}{1 - \beta_2} + \frac{\delta_1}{1 - \beta_2} + \frac{\delta_2}{1 - \beta_2} I_{t-1} + \frac{1}{1 - \beta_2} u_{2t} + \frac{1}{1 - \beta_2} u_{1t} \end{aligned} \quad (5.5)$$

– Damit erhält man

$$\begin{aligned} E(Y_t u_{1t}) &= \text{const.} \underbrace{E(u_{1t})}_{=0} + \frac{\delta_2}{1 - \beta_2} \underbrace{E(I_{t-1} u_{1t})}_{=0} + \frac{1}{1 - \beta_2} \underbrace{E(u_{2t} u_{1t})}_{=0} + \frac{1}{1 - \beta_2} \text{Var}(u_{1t}) \\ &= \frac{1}{1 - \beta_2} \text{Var}(u_{1t}) \neq 0 \text{ für alle } t, \end{aligned}$$

so dass der KQ-Schätzer für β_1 und β_2 inkonsistent ist. Damit ist Y_t keine vorherbestimmte Variable bezüglich u_{1t} und kann somit auch nicht in der (potentiellen) Informationsmenge Ω_t sein.

– Fortsetzung siehe nächsten Abschnitt.

- Mehr zu simultanen Gleichungsmodellen finden Sie in den Folien der BA-Veranstaltung **Ökonometrie III, Abschnitt 5** und MA-Veranstaltung Quantitative Wirtschaftsforschung II.

5.2. Der einfache Instrumentvariablen-Schätzer

- Betrachtet wird im folgenden das lineare Modell (1.2)

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (1.2)$$

- Ist die Annahme **(C2a)** $E(u_t|\Omega_t) = 0$ verletzt, dann ist (5.2)

$$E(\mathbf{X}_t u_t) = \mathbf{c}_t \neq \mathbf{0}$$

möglich.

- Gilt $\mathbf{c}_t = \mathbf{c}_s = \mathbf{c}$ (z.B. bei Annahme einer Zufallsstichprobe) und ist die Varianz von $\mathbf{X}_t u_t$ für alle $t, s = 1, \dots, n$ beschränkt, gilt für $\frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T u_t$ ein Gesetz der Großen Zahlen, z. B. das Theorem von Tschebyscheff ([Davidson \(2000, Section 3.2.2\)](#), [Hayashi \(2000, Section 2.1\)](#))

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T u_t = \mathbf{c}.$$

- In diesem Fall ist der KQ- und GLS-Schätzer im Allgemeinen inkonsistent (und nicht erwartungstreu) ist, da auch bei korrekt spezifiziertem Modell (1.2) mit $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ gilt:

$$\begin{aligned} \mathbf{X}_t^T y_t &= \mathbf{X}_t^T \mathbf{X}_t \boldsymbol{\beta}_0 + \mathbf{X}_t^T u_t & (5.6) \\ \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T y_t &= \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t \boldsymbol{\beta}_0 + \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T u_t \\ \text{plim}_{n \rightarrow \infty} \left[\left(\frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t \right)^{-1} \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T y_t \right] &= \boldsymbol{\beta}_0 + \text{plim}_{n \rightarrow \infty} \left[\left(\frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T \mathbf{X}_t \right)^{-1} \frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T u_t \right], \\ \text{plim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}}_{KQ} &= \boldsymbol{\beta}_0 + \mathbf{S}_{\mathbf{X}^T \mathbf{X}}^{-1} \mathbf{c} \neq \boldsymbol{\beta}_0 \end{aligned}$$

• **Ausweg:**

Multipliziere (1.2) anstelle mit $(1 \times k)$ -Vektor \mathbf{X}_t wie in (5.6) mit einem $(1 \times k)$ -Vektor \mathbf{Z}_t für den $\{\mathbf{Z}_t\} \in \Omega_t$ für alle $t = 1, \dots, n$ gilt.

Dann gilt

$$E(u_t | \mathbf{Z}_t) = 0 \quad \Rightarrow \quad E(\mathbf{Z}_t^T u_t) = \mathbf{0}. \quad (5.7)$$

Gilt analog zu **(B3)**

$$\sum_{t=1}^n \mathbf{z}_t^T \mathbf{X}_t \quad \text{hat vollen Rang,} \quad (5.8)$$

folgt unter korrekter Spezifikation, $\boldsymbol{\beta} = \boldsymbol{\beta}_0$,

$$\mathbf{z}_t^T y_t = \mathbf{z}_t^T \mathbf{X}_t \boldsymbol{\beta}_0 + \mathbf{z}_t^T u_t \quad (5.9)$$

$$\frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^T y_t = \frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^T \mathbf{X}_t \boldsymbol{\beta}_0 + \frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^T u_t$$

$$\underbrace{\left(\frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^T \mathbf{X}_t \right)^{-1} \frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^T y_t}_{\hat{\boldsymbol{\beta}}_{IV}} = \boldsymbol{\beta}_0 + \left(\frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^T \mathbf{X}_t \right)^{-1} \frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^T u_t. \quad (5.10)$$

Gilt weiter für $\frac{1}{n} \sum_{t=1}^n \mathbf{X}_t^T u_t$ und $\frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^T \mathbf{X}_t$ analog zu analog zu **(A1)** und **(A2)** (siehe **Methoden der Ökonometrie**) jeweils ein Gesetz der Großen Zahlen

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^T u_t = \mathbf{0}, \quad (5.11a)$$

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^T \mathbf{X}_t = \mathbf{S}_{\mathbf{Z}^T \mathbf{X}}, \quad \text{rk}(\mathbf{S}_{\mathbf{Z}^T \mathbf{X}}) = k, \quad (5.11b)$$

erhält man

$$\text{plim}_{n \rightarrow \infty} \left[\left(\frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^T \mathbf{X}_t \right)^{-1} \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^T y_t \right] = \beta_0 + \text{plim}_{n \rightarrow \infty} \left[\left(\frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^T \mathbf{X}_t \right)^{-1} \frac{1}{n} \sum_{t=1}^n \mathbf{Z}_t^T u_t \right], \quad (5.12)$$

$$\text{plim}_{n \rightarrow \infty} \hat{\beta}_{IV} = \beta_0 + \mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^{-1} \mathbf{0}. \quad (5.13)$$

Der Schätzer $\hat{\beta}_{IV}$ ist also unter den getroffenen Voraussetzungen konsistent.

- **Fortsetzung: Beispiel eines einfachen Keynesianischen Modells (5.4):**
Multipliziere die Konsumgleichung mit I_{t-1} anstelle mit Y_t , da I_{t-1} in der zulässigen

Informationsmenge ist und $E(u_{1t}|I_{t-1}) = 0$ gilt.

- Mit diesen Überlegungen erhält man den sogenannten (einfachen) **Instrumentvariablenschätzer** (IV-Schätzer)

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{IV} &= \left[\left(\frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^T \mathbf{X}_t \right)^{-1} \frac{1}{n} \sum_{t=1}^n \mathbf{z}_t^T y_t \right] \\ &= (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y},\end{aligned}\tag{5.14}$$

wobei die Variablen \mathbf{Z}_t als **Instrumentvariablen** bezeichnet werden.

- **Notation:**

– Die Matrix

$$\mathbf{Z} = \begin{pmatrix} \mathbf{Z}_1 \\ \vdots \\ \mathbf{Z}_n \end{pmatrix}$$

wird als **Instrumentenmatrix** bezeichnet.

– **Beachte:** Davidson und MacKinnon (2004) verwenden zur Darstellung des einfachen IV-Schätzers anstelle der $(n \times k)$ -Instrumentenmatrix \mathbf{Z} das Symbol \mathbf{W} . Da im weiteren Verlauf \mathbf{W} auch $l \geq k$ Spalten haben kann, siehe Abschnitt 5.3, gilt folgende Notation:

- * \mathbf{Z} bezeichnet eine $(n \times k)$ -Instrumentenmatrix.
- * \mathbf{W} bezeichnet eine $(n \times l)$ -Instrumentenmatrix.

- Offene Fragen:
 - Sind alle Variablen in Ω_t brauchbare IV-Variablen?
 - Wie findet man brauchbare IV-Variable in der Praxis?
 - Wie sollen Instrumentvariablen aus Ω_t ausgewählt werden?
 - Auf welche Weise hängen die Schätzergebnisse des IV-Schätzers von der Auswahl der IV-Variablen ab?
 - Welche Eigenschaften (außer Konsistenz) hat der IV-Schätzer?

- **Alternative Ableitung des einfachen IV-Schätzers (5.14):**

Man erhält den **einfachen IV-Schätzer** als **Momentenschätzer** aus der Schätzung der Momentenbedingung (5.7) mit

$$\mathbf{Z}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = 0.$$

Umformen ergibt wiederum (5.14)

$$\hat{\boldsymbol{\beta}}_{IV} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y}.$$

- Unter der Annahme (5.7) erhält man für das lineare Modell (1.2)

$$E(y_t | \mathbf{Z}_t) = E(\mathbf{X}_t | \mathbf{Z}_t) \boldsymbol{\beta}$$

$$y_t - E(y_t | \mathbf{Z}_t) = [\mathbf{X}_t - E(\mathbf{X}_t | \mathbf{Z}_t)] \boldsymbol{\beta} + u_t - E(u_t | \mathbf{Z}_t)$$

$$u_t = y_t - E(y_t | \mathbf{Z}_t) - [\mathbf{X}_t - E(\mathbf{X}_t | \mathbf{Z}_t)] \boldsymbol{\beta},$$

so dass u_t gegeben \mathbf{Z}_t nicht vorhersagbar ist, bzgl. \mathbf{X}_t jedoch schon, falls $E(u_t | \mathbf{X}_t) \neq 0$.

- Gilt im linearen Modell (1.2) die Annahme **(C2a)** nicht, d. h. liegt $E(u_t | \mathbf{X}_t) \neq 0$ vor, dann bezeichnen einige Autoren (1.2) nicht als Regressionsgleichung, da

die Variablen auf der rechten Seite nicht vorherbestimmt sind (Davidson, 2000, Footnote 1, Chapter 8).

- **Identifizierbarkeit:**

- Der einfache IV-Schätzer $\hat{\beta}_{IV}$ ist in einer Stichprobe **identifiziert**, d. h. eindeutig, wenn $\mathbf{Z}^T \mathbf{X}$ invertierbar ist. Das setzt voraus, dass Z_{tk} und X_{tk} in der Stichprobe korreliert sein müssen. Für die Grundgesamtheit muss deshalb vorausgesetzt werden:

$$Cov(\mathbf{Z}_t, \mathbf{X}_t) \neq \mathbf{0} \quad \text{und} \quad Cov(\mathbf{Z}_t, \mathbf{X}_t) \text{ invertierbar.} \quad (5.15)$$

- Der einfache IV-Schätzer $\hat{\beta}_{IV}$ ist **asymptotisch identifiziert**, wenn $\mathbf{Z}^T \mathbf{X}$ für “unendlich große” Stichproben identifiziert ist, d. h. wenn (5.11b) erfüllt ist. Wieso?
- Für $\mathbf{Z} = \mathbf{X}$ erhält man den KQ-Schätzer!
- Alle Variablen, die Annahmen (5.7) und (5.15) bzw. (5.11b) erfüllen, sind (potentielle) **Instrumentvariablen, IV-Variablen** oder kurz **Instrumente**.
 - Instrumentvariablen können **vorherbestimmt** oder **exogen** sein. Soweit mög-

lich, sollte man alle vorherbestimmten und exogenen Variablen aus \mathbf{X} in \mathbf{Z} aufnehmen.

- Zur Auswahl **optimaler** Instrumente, siehe weiter unten. Werden optimale Instrumente verwendet, ist der IV-Schätzer asymptotisch effizient.

- **Schätzeigenschaften**

- **Konsistenz**: Ist das lineare Modell (1.2) korrekt spezifiziert, $\beta = \beta_0$, und gelten (5.8) und (5.11), ist der einfache IV-Schätzer (5.14) konsistent.

Beweisskizze: siehe (5.12). Dabei wird zur Berechnung des zweiten Terms auf der rechten Seite von (5.12) u.a. das Slutsky-Theorem (**Methoden der Ökonometrie**, Abschnitt 3.5 *Konvergenz in Verteilung*) verwendet.

– **Asymptotische Normalität:** Gilt (5.11) und zusätzlich

$$\frac{1}{\sqrt{n}} \mathbf{Z}^T \mathbf{u} \xrightarrow{d} \mathbf{z}_\infty \sim N(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{Z}^T \mathbf{Z}}), \quad (5.16)$$

ist der einfache IV-Schätzer (5.14) asymptotisch normalverteilt

$$\sqrt{n} (\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{Z}^T \mathbf{Z}} (\mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^T)^{-1}\right).$$

Beweisskizze: Berechnung der asymptotischen Varianz (Vgl. **Methoden der Ökonometrie**, Abschnitt 11.2 *Asymptotische Verteilung des KQ-Schätzers*)

$$\sqrt{n} (\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_0) = \left(\frac{\mathbf{Z}^T \mathbf{X}}{n} \right)^{-1} \frac{\mathbf{Z}^T \mathbf{u}}{\sqrt{n}} \xrightarrow{d} \mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^{-1} \mathbf{z}_\infty.$$

Somit erhält man die asymptotische Kovarianzmatrix

$$\begin{aligned} \text{Var}(\mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^{-1} \mathbf{w}_\infty) &= \mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^{-1} \text{Var}(\mathbf{z}_\infty) (\mathbf{S}_{\mathbf{Z}^T \mathbf{X}})^{-1} \\ &= \sigma_0^2 \mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{Z}^T \mathbf{Z}} (\mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^T)^{-1}. \end{aligned} \quad (5.17)$$

(Beachte: Die Gleichung (8.17) in Davidson und MacKinnon (2004) ist so nicht richtig, da $\text{plim}(\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_0)/\sqrt{n}$ nicht existiert, siehe Davidson (2000, Sections 3.3.1 and 3.5.4).)

– Darstellungen der **asymptotischen Kovarianzmatrix**

Für die asymptotische Kovarianzmatrix (5.17) gilt

$$\sigma_0^2 \mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{Z}^T \mathbf{Z}} (\mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^T)^{-1} = \sigma_0^2 \left\{ \mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^T \mathbf{S}_{\mathbf{Z}^T \mathbf{Z}}^{-1} \mathbf{S}_{\mathbf{Z}^T \mathbf{X}} \right\}^{-1}$$

und somit

$$\text{plim}_{n \rightarrow \infty} \left\{ \left(\frac{\mathbf{X}^T \mathbf{Z}}{n} \right) \left(\frac{\mathbf{Z}^T \mathbf{Z}}{n} \right)^{-1} \left(\frac{\mathbf{Z}^T \mathbf{X}}{n} \right) \right\}^{-1} = \left\{ \mathbf{S}_{\mathbf{Z}^T \mathbf{X}}^T \mathbf{S}_{\mathbf{Z}^T \mathbf{Z}}^{-1} \mathbf{S}_{\mathbf{Z}^T \mathbf{X}} \right\}^{-1}$$

bzw.

$$\text{plim}_{n \rightarrow \infty} \left\{ \frac{1}{n} (\mathbf{X}^T \mathbf{Z}) (\mathbf{Z}^T \mathbf{Z})^{-1} (\mathbf{Z}^T \mathbf{X}) \right\}^{-1} = \text{plim}_{n \rightarrow \infty} \left\{ \frac{1}{n} \mathbf{X}^T \mathbf{P}_Z \mathbf{X} \right\}^{-1}$$

$$\text{Var} (S_{\mathbf{Z}^T \mathbf{X}}^{-1} \mathbf{z}_\infty) = \sigma_0^2 \text{plim}_{n \rightarrow \infty} \left\{ \frac{1}{n} \mathbf{X}^T \mathbf{P}_Z \mathbf{X} \right\}^{-1} \quad (5.18)$$

Es bezeichne $\bar{\mathbf{X}}_t = E(\mathbf{X}_t | \Omega_t)$ die bedingten Erwartungswerte der Variablen \mathbf{X}_t gegeben die Informationsmenge Ω_t . Dann gilt

$$\mathbf{X} = \bar{\mathbf{X}} + \mathbf{V}$$

mit $E(\mathbf{V}_t|\Omega_t) = \mathbf{0}$. Es gilt nun für (5.11b)

$$\begin{aligned} S_{\mathbf{Z}^T \mathbf{X}} &= \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}^T \mathbf{X} \\ &= \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}^T \bar{\mathbf{X}} + \underbrace{\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}^T \mathbf{V}}_{=0 \text{ da } E(\mathbf{V}_t|\Omega_t) = 0} \\ &= \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}^T \bar{\mathbf{X}} = \mathbf{S}_{\mathbf{Z}^T \bar{\mathbf{X}}}. \end{aligned}$$

Nur der 'Teil' von \mathbf{X} , der mit den Instrumentvariablen \mathbf{Z} korreliert ist, geht in die asymptotische Momentenmatrix $\mathbf{S}_{\mathbf{Z}^T \bar{\mathbf{X}}}$ ein.

Einsetzen in die asymptotische Kovarianzmatrix liefert

$$\text{Var} \left(S_{\mathbf{Z}^T \mathbf{X}}^{-1} \mathbf{z}_\infty \right) = \sigma_0^2 \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \bar{\mathbf{X}}^T \mathbf{P}_Z \bar{\mathbf{X}} \right)^{-1}. \quad (5.19)$$

– **Asymptotische Effizienz und Wahl optimaler Instrumente:**

Der IV-Schätzer ist effizient, wenn die asymptotische Kovarianzmatrix (5.17) “so klein wie möglich” ist. Dies kann u.U. durch eine geschickte Auswahl der Instrumente erreicht werden.

Wählt man nun die **optimalen Instrumente** $\mathbf{Z} = \bar{\mathbf{X}} = E(\mathbf{X}_t | \Omega_t)$, dann vereinfacht sich die asymptotische Kovarianzmatrix zu

$$\sigma_0^2 \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \bar{\mathbf{X}}^T \bar{\mathbf{X}} \right)^{-1}. \quad (5.20)$$

Es lässt sich zeigen, dass (5.20) die “kleinstmögliche” Kovarianzmatrix ist, also $\mathbf{Z} = \bar{\mathbf{X}}$ die optimalen Instrumente sind.

Problem: $\bar{\mathbf{X}}$ ist nicht bekannt und muss konsistent geschätzt werden, **außer** für die Variablen in \mathbf{X} , die vorherbestimmt sind.

– **Schätzen von optimalen Instrumenten**

Die bedingten Erwartungswerte für die j -te Variable, $\bar{\mathbf{x}}_j = \begin{pmatrix} E(x_{1j}|\Omega_1) \\ \vdots \\ E(x_{nj}|\Omega_n) \end{pmatrix}$,

können auf Basis der Hilfsregression

$$\mathbf{x}_j = \mathbf{W} \cdot \text{Parametervektor} + \text{Fehler} \quad (5.21)$$

und dem geschätzten Parametervektor

$$\mathbf{J}_j = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{x}_j$$

mit $\hat{\bar{\mathbf{x}}}_j = \mathbf{W} \mathbf{J}_j = \mathbf{P}_W \mathbf{x}_j$ geschätzt werden. Dabei gilt für den $(l \times 1)$ -Regressorvektor $\mathbf{W}_t \in \Omega_t$, wobei $l \geq k$ gelten muss.

Alle k Regressionen lassen sich in Matrixschreibweise schreiben:

$$\hat{\bar{\mathbf{X}}} = \mathbf{P}_W \mathbf{X} = \mathbf{W} \mathbf{J} \quad \text{mit} \quad \mathbf{J} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X}. \quad (5.22)$$

Beachte: Die **Zahl** l der **Instrumentvariablen** in der Hilfsregression (5.21) darf größer k sein!

Eine **konsistente Schätzung**,

$$\text{plim}_{n \rightarrow \infty} \mathbf{W}_t \mathbf{J}_j = E(x_{tj} | \Omega_t), \quad \text{plim}_{n \rightarrow \infty} \mathbf{W}_t \mathbf{J} = E(\mathbf{X}_t | \Omega_t),$$

erfordert, dass die bedingte Hilfsregression (5.21) korrekt spezifiziert ist, d. h.

- * die bedingten Erwartungswerte durch ein lineares Modell beschrieben werden können und
- * der Vektor \mathbf{W}_t alle relevanten Variablen aus Ω_t enthält, wobei die Anzahl der Instrumente gleich oder größer k sein muss.

Werden die **geschätzten optimalen Instrumente**

$$\mathbf{Z} = \mathbf{P}_W \mathbf{X} \quad (5.23)$$

in den einfachen IV-Schätzer (5.14) eingesetzt, ergibt sich der **verallgemeinerte IV-Schätzer**

$$\hat{\boldsymbol{\beta}}_{IV} = (\mathbf{X}^T \mathbf{P}_W \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_W \mathbf{y} \quad (5.24)$$

Dabei wird der Variablenvektor \mathbf{W}_t wieder als **Instrumentenvektor** bezeichnet.

Ob die geschätzten optimalen Instrumente (5.23) tatsächlich optimal sind, hängt von der Wahl der Variablen in \mathbf{W}_t , also der Spezifikation der Hilfsregression (5.21) ab!

Also: Die Wahl des Instrumentenvektors $\mathbf{W}_t \in \Omega_t$ ist von entscheidender Bedeutung!

Nachtrag: Bedeutung des Index t der Beobachtungen

- **Zufallsstichprobe:** Index t hat keine Bedeutung, da Beobachtungen zufällig gezogen werden. Deshalb gilt für den *unbedingten* Erwartungswert $E(y_t) = \mu$ und $E(y_t) = \mu_t$ macht keinen Sinn. Beachte aber, dass $E(y_t|\mathbf{X}_t)$ nicht konstant ist. Entsprechend gilt bei Verletzung von Annahme **(C2a)** $E(u_t|\Omega_t) = 0$, dass $E(\mathbf{X}_t u_t) = \mathbf{c}$.
- **Zeitreihenbeobachtungen:** Index t hat Bedeutung, da Periode t fest und y_t eine Zufallsvariable für Periode t ist. Deshalb ist für den *unbedingten* Erwartungswert $E(y_t) = \mu_t$ sinnvoll. Entsprechend ist bei Verletzung von Annahme **(C2a)** möglich, dass $E(\mathbf{X}_t u_t) = \mathbf{c}_t$.

5.3. Der verallgemeinerte IV-Schätzer

- **Notation:** Wie im vorhergehenden Abschnitt wird der $(1 \times l)$ -Instrumentenvektor, bzw. die $(n \times l)$ -Instrumentenmatrix mit \mathbf{W}_t bzw. mit \mathbf{W} bezeichnet, wobei $l \geq k$ vorausgesetzt wird.
- **Zahl der Instrumente und Identifikation**
 - Zahl der Instrumente l und Zahl der Parameter k :
 - * $l > k$: IV-Model **überidentifiziert (overidentified)**
 - * $l = k$: IV-Model **exakt identifiziert (exactly identified)**
 - * $l < k$: IV-Model **unteridentifiziert (underidentified)**
 - Bei einem unteridentifizierten IV-Modell müssen weitere Instrumente gesucht werden oder $k - l$ Restriktionen auf die Parameter gelegt werden.
 - **Weitere Instrumente können erzeugt werden**, indem beliebige (stetige) deterministische Funktionen auf vorhandene Instrumente angewendet werden, z.B. Exponentialfunktion oder Quadratfunktion.

– Werden durch die ökonomische oder ökonometrische Modellierung mehr als k Instrumente vorgegeben (**überidentifiziertes IV-Modell**) müssen die Instrumente oder die darauf aufbauenden Funktionen geschickt gewählt werden.

* Eine optimale Wahl ist mit (5.22) bereits erfolgt — hier mit Instrumentenmatrix \mathbf{W} —

$$\hat{\mathbf{X}} = \mathbf{P}_{\mathbf{W}}\mathbf{X} = \mathbf{W}\mathbf{J} \quad \text{mit} \quad \mathbf{J} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{X}. \quad (5.22)$$

und lieferte den **verallgemeinerten IV-Schätzer** (5.24)

$$\hat{\beta}_{IV} = (\mathbf{X}^T\mathbf{P}_{\mathbf{W}}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_{\mathbf{W}}\mathbf{y}. \quad (5.24)$$

* Anstelle von \mathbf{J} in (5.22) kann jedoch **jede** $(l \times k)$ -Matrix \mathbf{J}^* verwendet werden, um die Zahl der Instrumente zu reduzieren und eine (asymptotisch) deterministische **Linearkombinationen** zu bilden:

$$\mathbf{W}\mathbf{J}^*.$$

Anforderungen an eine optimale Wahl \mathbf{J} :

- \mathbf{J} hat vollen Spaltenrang k ,
- \mathbf{J} ist asymptotisch deterministisch,
- \mathbf{J} minimiert die asymptotische Kovarianz (des IV-Schätzers).

Alle Anforderungen erfüllt $\mathbf{J} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X}$

- Der **verallgemeinerte IV-Schätzer (GIV estimator)** (5.24) lässt sich auch aus den **empirischen Momentenbedingungen**

$$\mathbf{X}^T \mathbf{P}_W (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0} \iff \mathbf{X}^T \mathbf{P}_W \mathbf{y} = \mathbf{X}^T \mathbf{P}_W \mathbf{X} \boldsymbol{\beta} \quad (5.25)$$

oder aus folgender **IV-Zielfunktion** ableiten

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{P}_W (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (5.26a)$$

$$(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \quad (5.26b)$$

• **Asymptotische Schätzeigenschaften**

– **Asymptotische Normalität:** Gilt (5.11) und zusätzlich

$$\frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u} \xrightarrow{d} \mathbf{w}_\infty \sim N(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{W}^T \mathbf{W}}), \quad \text{rk}(\mathbf{S}_{\mathbf{W}^T \mathbf{W}}) = l, \quad (5.27)$$

ist der verallgemeinerte IV-Schätzer (5.24) asymptotisch normalverteilt

$$\sqrt{n} (\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_0) \xrightarrow{d} N\left(\mathbf{0}, \sigma_0^2 \mathbf{S}_{\mathbf{W}^T \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{W}^T \mathbf{W}} (\mathbf{S}_{\mathbf{W}^T \mathbf{X}}^T)^{-1}\right).$$

Beweisskizze: Berechnung der asymptotischen Varianz

$$\begin{aligned} \sqrt{n} (\hat{\boldsymbol{\beta}}_{IV} - \boldsymbol{\beta}_0) &= \left(\frac{\mathbf{X}^T \mathbf{P}_\mathbf{W} \mathbf{X}}{n} \right)^{-1} \frac{\mathbf{X}^T \mathbf{P}_\mathbf{W} \mathbf{u}}{\sqrt{n}} \\ &\xrightarrow{d} (\mathbf{S}_{\mathbf{W}^T \mathbf{X}}^T \mathbf{S}_{\mathbf{W}^T \mathbf{W}}^{-1} \mathbf{S}_{\mathbf{W}^T \mathbf{X}})^{-1} \mathbf{S}_{\mathbf{W}^T \mathbf{X}}^T \mathbf{S}_{\mathbf{W}^T \mathbf{W}}^{-1} \mathbf{w}_\infty. \end{aligned}$$

Somit erhält man die asymptotische Kovarianzmatrix

$$\sigma_0^2 \mathbf{S}_{\mathbf{W}^T \mathbf{X}}^{-1} \mathbf{S}_{\mathbf{W}^T \mathbf{W}} (\mathbf{S}_{\mathbf{W}^T \mathbf{X}}^T)^{-1},$$

bzw. mit zum einfachen IV-Schätzer analoger Rechnung (siehe Ableitung (5.18))

$$\sigma_0^2 \text{plim}_{n \rightarrow \infty} \left\{ \frac{1}{n} \mathbf{X}^T \mathbf{P}_\mathbf{W} \mathbf{X} \right\}^{-1}. \quad (5.28)$$

- Der **verallgemeinerte IV-Schätzer** kann **als ein einfacher IV-Schätzer** mit $(n \times k)$ -Instrumentenmatrix $\mathbf{P}_W \mathbf{X}$ aufgefasst werden, da gilt

$$\begin{aligned}
 & \sigma_0^2 \operatorname{plim}_{n \rightarrow \infty} \left\{ \frac{1}{n} \mathbf{X}^T \mathbf{P}_{\mathbf{P}_W \mathbf{X}} \mathbf{X} \right\}^{-1} \\
 &= \sigma_0^2 \operatorname{plim}_{n \rightarrow \infty} \left\{ \frac{1}{n} \mathbf{X}^T \mathbf{P}_W \mathbf{X} (\mathbf{X}^T \mathbf{P}_W \mathbf{P}_W \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_W \mathbf{X} \right\}^{-1} \\
 &= \sigma_0^2 \operatorname{plim}_{n \rightarrow \infty} \left\{ \frac{1}{n} \mathbf{X}^T \mathbf{P}_W \mathbf{X} \right\}^{-1} \\
 &= \sigma_0^2 \operatorname{plim}_{n \rightarrow \infty} \left\{ \frac{1}{n} \widehat{\mathbf{X}}^T \widehat{\mathbf{X}} \right\}^{-1}.
 \end{aligned} \tag{5.29}$$

- **Kovarianzmatrix des verallgemeinerten IV-Schätzers**

- **Schätzung:** wie folgt möglich:

$$\widehat{Var} \left(\hat{\beta}_{IV} \right) = \hat{\sigma}^2 \left(\mathbf{X}^T \mathbf{P}_W \mathbf{X} \right)^{-1} \quad (5.30)$$

$$\hat{\sigma}^2 = \frac{1}{n} \left(\mathbf{y} - \mathbf{X} \hat{\beta}_{IV} \right)^T \left(\mathbf{y} - \mathbf{X} \hat{\beta}_{IV} \right) \quad (5.31)$$

- **Interpretation** der asymptotischen Kovarianzmatrix (5.28) für den Fall $k = 1$ anhand der Hilfsregression (5.21):

$$\begin{aligned} \widehat{\mathbf{X}}^T \widehat{\mathbf{X}} &= \sum_{t=1}^n \widehat{x}_t^2 = ESS = TSS - SSR \\ \text{plim}_{n \rightarrow \infty} \frac{1}{n} \widehat{\mathbf{X}}^T \widehat{\mathbf{X}} &= \text{plim}_{n \rightarrow \infty} \frac{TSS}{n} - \text{plim}_{n \rightarrow \infty} \frac{SSR}{n} \\ &= Var(x_t) - \begin{cases} (\sigma_{Fehler}^2 + \text{asympt. Bias}^2) & \text{falls (5.21) fehlsp.,} \\ \sigma_{Fehler}^2 & \text{sonst.} \end{cases} \end{aligned}$$

(Vgl. zu TSS , SSR **Methoden der Ökonometrie**, Abschnitt 7.1 *Die Geometrie des KQ-Schätzers*.) Argumentation lässt sich auf $k > 1$ ausweiten.

- * Enthält \mathbf{W} alle für eine korrekte Spezifikation von (5.21) notwendigen Regressoren, ist $\widehat{\mathbf{X}}^T \widehat{\mathbf{X}}$ größer als bei Fehlspezifikation. Damit ist dessen Inverse kleiner als bei Fehlspezifikation und somit auch die asymptotische Varianz. Eine Fehlspezifikation der Hilfsregression erhöht also die asymptotische Schätzvarianz.
- * Wird die Hilfsregression überspezifiziert, hat dies keine Auswirkungen auf die asymptotische Kovarianzmatrix. Warum?
- * Für eine **gegebene Stichprobengröße** führt die Hinzunahme eines zusätzlichen Regressors in der Hilfsregression jedoch zu einer Abnahme von SSR, so dass $\widehat{\mathbf{X}}^T \widehat{\mathbf{X}}$ steigt und die geschätzte Varianz (5.30) fällt (außer $\hat{\sigma}^2$ in (5.31) steigt). Wird also zu einer bereits korrekt spezifizierten Hilfsregression irrtümlich eine weitere Instrumentvariable hinzugenommen, wird die wahre Varianz unterschätzt.
- * Der KQ-Schätzer verwendet $\mathbf{W}_t = \mathbf{X}_t$ und hat wegen $SSR = 0$ die geringste geschätzte Varianz (ist aber bei $\mathbf{X}_t \notin \Omega_t$ inkonsistent).

Also: Verwende so wenige Instrumentvariablen wie möglich und dabei die richtigen!

- Der **verallgemeinerte IV-Schätzer** (5.24) wird auch als **2-stufiger KQ-Schätzer (2SLS)** bezeichnet, da er in 2 Schritten berechnet werden kann:

1. Stufe: Regressiere jedes \mathbf{x}_j aus \mathbf{X} auf \mathbf{W} , falls \mathbf{x}_j nicht exogen oder vorherbestimmt ist (also (5.21) mit \mathbf{W} anstelle von \mathbf{Z}). Damit ergeben sich die gefitteten Werte $\mathbf{P}_W \mathbf{x}_j$. Man beachte, dass für diejenigen \mathbf{x}_j , die gleichzeitig Instrumentvariablen sind, gilt: $\mathbf{P}_W \mathbf{x}_j = \mathbf{x}_j$, da ja \mathbf{x}_j bereits im Unterraum der Instrumente liegt. Zusammenfassen ergibt die $(n \times k)$ -Matrix $\mathbf{P}_W \mathbf{X}$ der Instrumente, die in der 2. Stufe anstelle von \mathbf{X} bei KQ verwendet werden.

2. Stufe: Schätzung von β mit KQ aus dem Ansatz

$$\mathbf{y} = \mathbf{P}_W \mathbf{X} \beta + \zeta$$

liefert den 2SLS-Schätzer

$$\hat{\beta}_{2SLS} = (\mathbf{X}^T \mathbf{P}_W \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_W \mathbf{y} = \hat{\beta}_{IV}.$$

- **Endliche Stichprobeneigenschaften**

Überraschung: Der Erwartungswert für $\hat{\beta}_{IV}$ im Fall

- exakt identifizierter Modelle **existiert nicht immer**.
- überidentifizierter Modelle **existiert**, weicht aber häufig substantiell von β_0 ab → **Gefahr von Fehlinterpretationen** (siehe Davidson und MacKinnon (2004, S. 324f.))

5.4. Asymptotische Tests

- ***t*-Tests**

$$t_{\beta_i} = \frac{\hat{\beta}_{iIV} - \beta_{i,H_0}}{\left(\widehat{Var}(\hat{\beta}_{iIV})\right)^{1/2}} \xrightarrow{d} N(0, 1).$$

- **Wald-Statistik**

$$\lambda_W = \left(\mathbf{R}\hat{\boldsymbol{\beta}}_{IV} - \mathbf{r}\right)^T \left(\widehat{Var}(\mathbf{R}\hat{\boldsymbol{\beta}}_{IV} - \mathbf{r})\right)^{-1} \left(\mathbf{R}\hat{\boldsymbol{\beta}}_{IV} - \mathbf{r}\right) \xrightarrow{d} \chi^2(q).$$

- Zu **Bootstrap-Tests** siehe [Davidson und MacKinnon \(2004, Section 8.8\)](#).

5.5. Testen von überidentifizierenden Restriktionen

- Sei $l - k > 0$: Der Unterraum der Instrumente $\delta(\mathbf{W}) = \delta(\mathbf{P}_W \mathbf{X}, \mathbf{W}^*)$ wird durch Linearkombinationen der (effektiven) Instrumente $\mathbf{P}_W \mathbf{X}$ und der (überidentifizierenden) Instrumente \mathbf{W}^* aufgespannt.
- Frage: Gilt die hinreichende Bedingung für die Konsistenz des IVE

$$E(u_t | W_t) = 0$$

auch für W_t^* ? Damit lautet das Hypothesenpaar

$$H_0 : E(u_t | W_t^*) = 0 \quad \text{versus} \quad H_1 : E(u_t | W_t^*) \neq 0.$$

Eine Ableitung der Teststatistik und ihrer Eigenschaften unter allgemeineren Bedingungen findet sich in Abschnitt 6.6.

Siehe auch Davidson und MacKinnon (2004, Section 8.6, p. 336f (**Sargan-Test**)) oder **Ökonometrie III**.

- **Durbin-Wu-Hausman-Tests** geben eine statistische Antwort auf die Frage, ob man einen IV-Schätzer verwenden muss. Ein Spezialfall ist der **Hausman-Test**

bei Panelmodellen (Davidson und MacKinnon (2004, Section 8.7, p. 338f)).

Zu lesen: Davidson und MacKinnon (2004), Kapitel 8.

6. Generalized Method of Moments (GMM)

6.1. Übersicht

Bisher betrachtete Modelle

1. Dynamisches Regressionsmodell mit homoskedastischen Fehlern (vgl. Abschnitt 1.1, S. 11)

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \quad E(y_t|\Omega_t) = \mathbf{X}_t\boldsymbol{\beta}, \quad \text{Var}(u_t|\Omega_t) = \sigma^2$$

$$\implies \text{OLS-Schätzer } \hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$$

2. Dynamisches Regressionsmodell mit heteroskedastischen Fehlern (vgl. Abschnitt 1.1, S. 11)

$$y_t = \mathbf{X}_t\boldsymbol{\beta} + u_t, \quad E(y_t|\Omega_t) = \mathbf{X}_t\boldsymbol{\beta}, \quad \text{Var}(u_t|\Omega_t) = \sigma_t^2, \quad \boldsymbol{\Omega} = \begin{pmatrix} \sigma_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

⇒ OLS-Schätzer mit heteroskedastierobusten Standardfehlern

$$\text{GLS-Schätzer } \hat{\beta}_{GLS} = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y}$$

$$\text{FGLS-Schätzer } \hat{\beta}_{FGLS} = (\mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}$$

3. Regressionsmodell mit *streng exogenen Regressoren* und heteroskedastischen und autokorrelierten Fehlern

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t, \quad E(y_t | \Omega_t) = \mathbf{X}_t \boldsymbol{\beta}, \quad \mathbf{u} | \mathbf{X} = (\mathbf{0}, \boldsymbol{\Omega})$$

$$\Rightarrow \text{GLS-Schätzer } \hat{\beta}_{GLS} = (\mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Omega}^{-1} \mathbf{y}$$

$$\text{FGLS-Schätzer } \hat{\beta}_{FGLS} = (\mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}$$

4. Nichtlineares Modell mit vorherbestimmten Variablen und homoskedastischen Fehlern (vgl. S. 16 und Kapitel 2)

$$y_t = x_t(\boldsymbol{\beta}) + u_t, \quad E(y_t | \Omega_t) = x_t(\boldsymbol{\beta}), \quad \text{Var}(u_t | \Omega_t) = \sigma^2$$

$$\Rightarrow \text{Momentenschätzer } \mathbf{Z}^T (\mathbf{y} - \mathbf{x}(\hat{\boldsymbol{\beta}})) = \mathbf{0}$$

$$\text{NLS-Schätzer } \mathbf{X}(\hat{\boldsymbol{\beta}})^T (\mathbf{y} - \mathbf{x}(\hat{\boldsymbol{\beta}})) = \mathbf{0}$$

5. Lineares Modell mit *endogenen Regressoren* und homoskedastischen Fehlern (vgl. S. 16 und Kapitel 5)

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t, \quad E(y_t | \Omega_t) = E(\mathbf{X}_t | \Omega_t) \boldsymbol{\beta} \neq \mathbf{X}_t \boldsymbol{\beta}, \quad \text{Var}(u_t | \Omega_t) = \sigma^2$$

$$\Rightarrow \text{IV-Schätzer } \hat{\beta}_{IV} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y}$$

$$\text{verallgemeinerter IV-Schätzer } \hat{\beta}_{IV} = (\mathbf{X}^T \mathbf{P}_W \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_W \mathbf{y}$$

mit $(n \times k)$ -Instrumentenmatrix \mathbf{Z} und $(n \times l)$ -Instrumentenmatrix \mathbf{W}

Was fehlt?

6. Lineares Modell mit *endogenen* Regressoren und heteroskedastischen und korrelierten Fehlern (vgl. S. 17)

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t, \quad E(y_t | \Omega_t) = E(\mathbf{X}_t | \Omega_t) \boldsymbol{\beta} \neq \mathbf{X}_t \boldsymbol{\beta}, \quad E(u_t u_s | \mathbf{W}_t, \mathbf{W}_s) = \omega_{ts}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \cdots & \omega_{nn} \end{pmatrix}$$

$$\begin{aligned} \implies \text{IV-Schätzer } \hat{\boldsymbol{\beta}}_{IV} &= (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y} \\ \text{verallgemeinerter IV-Schätzer } \hat{\boldsymbol{\beta}}_{IV} &= (\mathbf{X}^T \mathbf{P}_W \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_W \mathbf{y} \end{aligned}$$

mit $(n \times k)$ -Instrumentenmatrix \mathbf{Z} und $(n \times l)$ -Instrumentenmatrix \mathbf{W}

Statisches Regressionsmodell mit heteroskedastischen und korrelierten Fehlern

Allgemeines Schätzverfahren für die genannten Erweiterungen:

Verallgemeinerter Momentenschätzer

Generalized Method of Moments — GMM

Im Folgenden: GMM für lineare Modelle mit

- *endogenen* Regressoren und heteroskedastischen Fehlern,
- *endogenen* Regressoren und heteroskedastischen und autokorrelierten Fehlern,
- *statischen* Regressoren und heteroskedastischen und autokorrelierten Fehlern.

Entsprechende Erweiterungen auf Modelle, die nichtlinear in den Parametern sind, finden sich in [Davidson und MacKinnon \(2004, Sections 9.5, 9.6\)](#).

Notation wie Kapitel 5:

- \mathbf{Z} bezeichnet eine $(n \times k)$ -Instrumentenmatrix,
- \mathbf{W} bezeichnet eine $(n \times l)$ -Instrumentenmatrix, $l \geq k$.

- Für das lineare Modell (1.2)

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t,$$

mit möglicherweise endogenen Variablen wird angenommen, dass gilt:

$$E(y_t | \mathbf{W}_t) = E(\mathbf{X}_t | \mathbf{W}_t) \boldsymbol{\beta}, \quad \text{bzw.} \quad E(u_t | \mathbf{W}_t) = 0 \quad (6.1a)$$

$$E(u_t u_s | \mathbf{W}_t, \mathbf{W}_s) = \omega_{ts}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & \omega_{12} & \cdots & \omega_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ \omega_{n1} & \omega_{n2} & \cdots & \omega_{nn} \end{pmatrix}. \quad (6.1b)$$

Beachte: Ist das lineare Modell (6.1) ein Zeitreihenmodell, ist es **nicht dynamisch vollständig und korrekt spezifiziert**, wenn $\boldsymbol{\Omega}$ in (6.1b) keine Diagonalmatrix ist.

- **Wiederholung:** Ein lineares Modell ist **dynamisch vollständig und korrekt spezifiziert**, wenn (1.3) gilt.

- Angewendet auf das lineare Modell (6.1) impliziert die Bedingung für **dynamisch vollständige und korrekte** Spezifikation (1.3), dass

$$E(y_t|\Omega_t) = E(y_t|\mathbf{W}_t) = E(\mathbf{X}_t|\Omega_t) \boldsymbol{\beta} = E(\mathbf{X}_t|\mathbf{W}_t) \boldsymbol{\beta} \quad (6.2a)$$

und somit gilt:

$$E(u_t|\Omega_t) = 0 \quad \text{und deshalb} \quad E(u_t u_s|\Omega_t) = 0, \quad \text{falls } t > s, \quad \text{so dass} \quad (6.2b)$$

$$E(u_t^2|\Omega_t) = \omega_{tt}, \quad \boldsymbol{\Omega} = \begin{pmatrix} \omega_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \omega_{nn} \end{pmatrix}. \quad (6.2c)$$

Die **Fehler** in einem dynamisch vollständig und korrekt spezifizierten linearen Modell können **heteroskedastisch, jedoch nicht autokorreliert** sein!

- In beiden Fällen ist Endogenität der Regressoren möglich, also $E(\mathbf{X}_t|\Omega_t) \neq \mathbf{X}_t$.

Ein verallgemeinerter IV-Schätzer ist bei heteroskedastischen Fehlern (6.2c) und/oder (auto-)korrelierten Fehlern (6.1b) nicht effizient, wie im Folgenden gezeigt wird. Wie lauten geeignete Schätzer?

		Varianz-Kovarianzmatrix Ω						
		diagonal				nicht diagonal		
		Homoskedastie	Heteroskedastie			Heteroskedastie		
Regressoren	Instrumente	keine Korrelation	keine Korrelation			Korrelation		
		σ konstant	σ_t bekannt	σ_t -Funktion bekannt	σ_t völlig unbekannt	Ω bekannt	Ω -Funktion bekannt	Ω völlig unbekannt
$E(u_t \mathbf{X}) = 0$ streng exogen		OLS $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}\mathbf{y}$	GLS $(\mathbf{X}^T\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}\Omega^{-1}\mathbf{y}$	FGLS $(\mathbf{X}^T\hat{\Omega}^{-1}\mathbf{X})^{-1}\mathbf{X}\hat{\Omega}^{-1}\mathbf{y}$	OLS-HC	GLS	FGLS	OLS-HAC
$E(u_t \mathbf{X}_t) = 0$		OLS	GLS	FGLS	OLS-HC	GLS	-	OLS-HAC
$E(u_t \Omega_t) = 0$		OLS	GLS	FGLS	OLS-HC	nicht einschlägig		
$E(u_t \mathbf{X}_t) \neq 0$	$E(u_t \mathbf{Z}_t) = 0$	IV $(\mathbf{Z}^T\mathbf{X})^{-1}\mathbf{Z}\mathbf{y}$						
$E(u_t \mathbf{X}_t) \neq 0$	$E(u_t \mathbf{W}_t) = 0$	GIV $(\mathbf{X}^T\mathbf{P}_W\mathbf{X})^{-1}\mathbf{X}^T\mathbf{P}_W\mathbf{y}$						

Hinweise:

- OLS-HC: OLS-Schätzer mit heteroskedastierobusten Standardfehlern
- OLS-HAC: OLS-Schätzer mit heteroskedastie und autokorrelationsrobusten Standardfehlern

6.2. GMM-Schätzer für lineare Modelle

Ableitung — Teil 1: Bilden von empirischen Momentenbedingungen

- Da (5.7) — entspricht hier (6.1a) — gilt, erscheint ein Momentenschätzer möglich.
- Wegen (6.1a) gelten die $l \geq k$ **theoretischen Momentenbedingungen**

$$E(\mathbf{W}_t^T (y_t - \mathbf{X}_t \boldsymbol{\beta})) = \mathbf{0} \quad \Longleftrightarrow \quad E(W_{tj} (y_t - \mathbf{X}_t \boldsymbol{\beta})) = 0, \quad j = 1, \dots, l. \quad (6.3)$$

Die theoretischen Momente werden durch die l **empirischen Momente** geschätzt:

$$\bar{\mathbf{m}}(\boldsymbol{\beta}) = \begin{pmatrix} \bar{m}_1(\boldsymbol{\beta}) \\ \vdots \\ \bar{m}_l(\boldsymbol{\beta}) \end{pmatrix} \equiv \frac{1}{n} \mathbf{W}^T (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) = \begin{pmatrix} \frac{1}{n} \sum_{t=1}^n \mathbf{W}_{t1} (y_t - \mathbf{X}_t \boldsymbol{\beta}) \\ \vdots \\ \frac{1}{n} \sum_{t=1}^n \mathbf{W}_{tl} (y_t - \mathbf{X}_t \boldsymbol{\beta}) \end{pmatrix} \quad (6.4)$$

Die **empirischen Momentenbedingungen** lauten also

$$\mathbf{W}^T (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \mathbf{0} \quad \Longleftrightarrow \quad \mathbf{W}^T \mathbf{y} = \mathbf{W}^T \mathbf{X} \hat{\boldsymbol{\beta}} \quad (6.5)$$

- **Auflösen der empirischen Momentenbedingungen (6.5)**

$$\mathbf{W}^T \mathbf{y} = \mathbf{W}^T \mathbf{X} \hat{\boldsymbol{\beta}}$$

möglich?

- **Modell exakt identifiziert**, $l = k$: die empirischen Momentenbedingungen (6.5) entsprechen den empirischen Momentenbedingungen des einfachen IV-Schätzers, siehe S. 218. Hat die Matrix $\mathbf{W}^T \mathbf{X}$ vollen Rang $l = k$, dann erhält man als **Lösung** den **einfachen IV-Schätzer (5.14)**

$$\hat{\boldsymbol{\beta}}_{IV} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y}.$$

- **Modell überidentifiziert**, $l > k$: $\mathbf{W}^T \mathbf{X}$ in empirischer Momentenbedingung nicht invertierbar. Was tun? **Empirische Momentenbedingungen gewichten!**

Ableitung — Teil 2: Bilden einer Zielfunktion zum Gewichten der empirischen Momentenbedingungen

- Wähle eine symmetrische und positiv definite $(l \times l)$ -Gewichtungsmatrix Λ , bilde die **Zielfunktion (criterion function)**

$$Q_{\Lambda}(\beta | \mathbf{y}, \mathbf{X}, \mathbf{W}) = (\mathbf{y} - \mathbf{X}\beta)^T \mathbf{W} \Lambda \mathbf{W}^T (\mathbf{y} - \mathbf{X}\beta) = n^2 \bar{\mathbf{m}}^T(\beta) \Lambda \bar{\mathbf{m}}(\beta) \quad (6.6)$$

und minimiere diese bezüglich β .

Falls $\Lambda = (\mathbf{W}^T \mathbf{W})^{-1}$, entspricht diese Zielfunktion der Zielfunktion des verallgemeinerten IV-Schätzers (5.24).

- Die Zielfunktion (6.6) ist ein Beispiel für die allgemeine Form der Zielfunktionen von GMM-Schätzern

$$\bar{\mathbf{m}}^T(\boldsymbol{\theta}) \Lambda \bar{\mathbf{m}}(\boldsymbol{\theta}). \quad (6.7)$$

- Da die Zielfunktion (6.6) quadratisch ist, lässt sich das Minimum bezüglich β analytisch ableiten. Man erhält durch Ableiten

$$\frac{\partial Q_{\Lambda}(\beta | \mathbf{y}, \mathbf{X}, \mathbf{W})}{\partial \beta} = -2\mathbf{X}^T \mathbf{W} \Lambda \mathbf{W}^T \mathbf{y} + 2\mathbf{X}^T \mathbf{W} \Lambda \mathbf{W}^T \mathbf{X} \beta$$

und Nullsetzen

$$\mathbf{X}^T \mathbf{W} \mathbf{\Lambda} \mathbf{W}^T \mathbf{y} = \underbrace{\mathbf{X}^T \mathbf{W} \mathbf{\Lambda} \mathbf{W}^T \mathbf{X}}_{(k \times k)} \hat{\boldsymbol{\beta}}.$$

Hat $\mathbf{X}^T \mathbf{W} \mathbf{\Lambda} \mathbf{W}^T \mathbf{X}$ Rang k , erhält man

$$\hat{\boldsymbol{\beta}} = \left(\underbrace{\mathbf{X}^T \mathbf{W} \mathbf{\Lambda} \mathbf{W}^T \mathbf{X}}_{\equiv \mathbf{J}^T} \right)^{-1} \underbrace{\mathbf{X}^T \mathbf{W} \mathbf{\Lambda} \mathbf{W}^T \mathbf{y}}_{\equiv \mathbf{J}^T} = (\mathbf{J}^T \mathbf{W}^T \mathbf{X})^{-1} \mathbf{J}^T \mathbf{W}^T \mathbf{y} \quad (6.8)$$

Der Schätzer (6.8) wird als **Generalized-Method-of-Moments-Schätzer (GMM-Schätzer)** bezeichnet, da mehr Momentenbedingungen als Parameter, $l \geq k$, durch die Verwendung der recht frei wählbaren Gewichtungsmatrix $\mathbf{\Lambda}$ berücksichtigt werden können.

- **Beachte:** Ist $l = k$ erhält man aus den empirischen Momentenbedingungen $\mathbf{Z}^T (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \mathbf{0}$ die Formel des einfachen IV-Schätzers

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

und $\mathbf{\Lambda}$ spielt keine Rolle und muss nicht gewählt werden. Dieser Spezialfall ist mit $\mathbf{Z} = \mathbf{W} \mathbf{J}$ in (6.8) enthalten.

Schätzeigenschaften

- Für endliche Stichproben unter allgemeinen Bedingungen unbekannt.
- **Konsistenz:** Ist das Modell korrekt spezifiziert, d. h. gilt (6.1) mit $\beta = \beta_0$ und $\Omega = \Omega_0$ und existieren außerdem für alle zulässigen β folgende Wahrscheinlichkeitslimites (vgl. (5.11))

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{J}^T \mathbf{W}^T \mathbf{u} = \mathbf{0}, \quad \mathbf{J}^T = \mathbf{X}^T \mathbf{W} \Lambda \quad (6.9a)$$

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{J}^T \mathbf{W}^T \mathbf{X} \text{ existiert mit Rang } k \text{ (**Asymptotische Identifizierbarkeit**)} \quad (6.9b)$$

dann ist der Schätzer (6.8) konsistent.

(Beweisskizze wie im IV-Fall.)

- **Asymptotische Normalverteilung**

Es gilt wegen (6.1b)

$$\text{Var}(\mathbf{W}^T \mathbf{u}) = E(\mathbf{W}^T \mathbf{u} \mathbf{u}^T \mathbf{W}) = E\left(\sum_{t=1}^n \sum_{s=1}^n u_t u_s \mathbf{W}_t^T \mathbf{W}_s\right) \quad (6.10)$$

$$\begin{aligned} &= \sum_{t=1}^n \sum_{s=1}^n E(u_t u_s \mathbf{W}_t^T \mathbf{W}_s) \\ &= \sum_{t=1}^n \sum_{s=1}^n E(\mathbf{W}_t^T E(u_t u_s | \mathbf{W}_t, \mathbf{W}_s) \mathbf{W}_s) \\ &= \sum_{t=1}^n \sum_{s=1}^n E(\mathbf{W}_t^T \omega_{ts} \mathbf{W}_s) \\ &= E(\mathbf{W}^T \mathbf{\Omega} \mathbf{W}). \end{aligned} \quad (6.11)$$

Mit $\boldsymbol{\beta} = \boldsymbol{\beta}_0$ folgt aus (6.4), dass $\mathbf{W}^T \mathbf{u} = \mathbf{W}^T (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}_0) = n \bar{\mathbf{m}}(\boldsymbol{\beta}_0)$. Damit erhält man

$$\text{Var}(n \bar{\mathbf{m}}(\boldsymbol{\beta}_0)) = E(\mathbf{W}^T \mathbf{u} \mathbf{u}^T \mathbf{W}) = E(\mathbf{W}^T \mathbf{\Omega}_0 \mathbf{W}). \quad (6.12)$$

Asymptotische Normalverteilung: Gilt für $Var(\sqrt{n}\bar{\mathbf{m}}(\boldsymbol{\beta}_0))$ ein LLN

$$\begin{aligned}\boldsymbol{\Sigma}_{\bar{\mathbf{m}}} &= \lim_{n \rightarrow \infty} Var(\sqrt{n}\bar{\mathbf{m}}(\boldsymbol{\beta}_0)) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} E(\mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W}) = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W}\end{aligned}\quad (6.13a)$$

und zusätzlich ein CLT

$$\sqrt{n}\bar{\mathbf{m}}(\boldsymbol{\beta}_0) = \frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u} \xrightarrow{d} \mathbf{w}_\infty \sim N(\mathbf{0}, \boldsymbol{\Sigma}_{\bar{\mathbf{m}}}) \quad (6.13b)$$

(eine Verallgemeinerung von (5.27)), erhält man

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}_0) \quad (6.14)$$

$$\boldsymbol{\Sigma}_0 = \left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{J}^T \mathbf{W}^T \mathbf{X} \right)^{-1} \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{J}^T \mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W} \mathbf{J} \right) \left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \mathbf{W} \mathbf{J} \right)^{-1}. \quad (6.15)$$

Der mittlere plim entspricht $\text{plim}_{n \rightarrow \infty} (\mathbf{J}^T \boldsymbol{\Sigma}_{\bar{\mathbf{m}}} \mathbf{J})$.

Approximativ erhält man die Verteilung

$$\hat{\boldsymbol{\beta}} \approx N\left(\boldsymbol{\beta}_0, (\mathbf{J}^T \mathbf{W}^T \mathbf{X})^{-1} \mathbf{J}^T \mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W} \mathbf{J} (\mathbf{X}^T \mathbf{W} \mathbf{J})^{-1}\right). \quad (6.16)$$

6.3. Effizienter GMM-Schätzer für lineare Modelle

- Für $l > k$: Wahl des optimalen Λ und damit des optimalen \mathbf{J} :

- Verwende die Vorgehensweise des GLS-Schätzers, der berücksichtigt, dass die Fehler — hier die einzelnen Momentenbedingungen — unterschiedliche Varianzen haben und korreliert sein können, indem anstelle von $\mathbf{u}^T \mathbf{u}$ die quadratische Form $\mathbf{u}^T \text{Var}(\mathbf{u})^{-1} \mathbf{u}$ zur Minimierung verwendet wird.
- Entsprechend wäre es bei (6.6) wegen (6.12) optimal $(E(\mathbf{W}^T \Omega_0 \mathbf{W}))^{-1}$ für Λ zu wählen. Da der Erwartungswert unbekannt ist, schätzt man ihn durch $\mathbf{W}^T \Omega_0 \mathbf{W}$. Deshalb wählt man (vgl. auch Davidson und MacKinnon, 2004, Eq. (9.08))

$$\Lambda = (\mathbf{W}^T \Omega_0 \mathbf{W})^{-1} \quad (6.17a)$$

$$\mathbf{J} = \Lambda \mathbf{W}^T \mathbf{X} = (\mathbf{W}^T \Omega_0 \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X}. \quad (6.17b)$$

- Einsetzen von (6.17a) in die Zielfunktion (6.6) ergibt die sogenannte **GMM**

criterion function

$$Q(\boldsymbol{\beta}, \mathbf{y}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W})^{-1} \mathbf{W}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}). \quad (6.18)$$

- Mit (6.17) erhält man für gegebenes \mathbf{W} den **effizienten GMM-Schätzer**

$$\hat{\boldsymbol{\beta}}_{GMM} = \left(\mathbf{X}^T \mathbf{W} (\mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{W} (\mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W})^{-1} \mathbf{W}^T \mathbf{y} \quad (6.19)$$

mit asymptotischer Kovarianzmatrix (vgl. (6.15))

$$\boldsymbol{\Sigma}_{GMM} = \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^T \mathbf{W} (\mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X} \right)^{-1}. \quad (6.20)$$

- Der GMM-Schätzer (6.19) wird als **effizient** bezeichnet, da er unter allen zulässigen $\boldsymbol{\Lambda}$ die “kleinste” asymptotische Kovarianzmatrix aufweist, d. h. dass die Differenz der asymptotischen Kovarianzmatrizen der Schätzer mit beliebigem $\boldsymbol{\Lambda}$ und optimalen $\boldsymbol{\Lambda} = (\mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W})^{-1}$ positiv semidefinit ist.
- **Beachte:** Die Anwendung des effizienten GMM-Schätzers setzt die Kenntnis von $\mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W}$ voraus, nicht jedoch die Kenntnis von $\boldsymbol{\Omega}_0$. Erstere Matrix kann

mit den Methoden von Abschnitt 6.5 geschätzt werden. Ist es möglich, Ω konsistent zu schätzen und zusätzliche Annahmen zu treffen, dann lässt sich der effiziente GMM-Schätzer noch effizienter machen, siehe Abschnitt 6.4.

– Der **effiziente GMM-Schätzer** (6.19) ist sowohl eine **Verallgemeinerung des OLS-** als auch des **verallgemeinerten IV-Schätzers**:

* Sind **alle Regressoren \mathbf{X}_t vorherbestimmt**, d. h. $E(u_t|\mathbf{X}_t) = 0$, und gilt $E(u_t u_s | \mathbf{X}_t, \mathbf{X}_s) = \omega_{ts}$, dann wählt man $\mathbf{W} = \mathbf{X}$, so dass dann $\mathbf{J} = \mathbf{I}$. Dann erhält man den **OLS-Schätzer** mit **asymptotischer Kovarianzmatrix**

$$\left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \mathbf{X} \right)^{-1} \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \mathbf{X}^T \Omega_0 \mathbf{X} \right) \left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \mathbf{X} \right)^{-1}$$

* Für ein lineares Modell mit **homoskedastischen und unkorrelierten Fehlern** $\Omega_0 = \sigma_0^2 \mathbf{I}$ erhält man den **verallgemeinerten IV-Schätzer** (5.24).

• **Bei $l = k$: keine Wahl von Λ nötig:**

Wie bereits gezeigt, entspricht der effiziente GMM-Schätzer (6.19) dann mit $\mathbf{Z} = \mathbf{WJ}$ dem einfachen IV-Schätzer, allerdings mit einer Sandwich-Kovarianzmatrix,

so wie bei OLS mit allgemeiner Fehlerkovarianzmatrix:

$$\hat{\beta}_{GMM_Z} = (\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y} \quad (6.21)$$

$$\Sigma_{GMM_Z} = \left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \mathbf{Z} \right)^{-1} \left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}^T \Omega_0 \mathbf{Z} \right) \left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}^T \mathbf{X} \right)^{-1} \quad (6.22)$$

- Falls die **Fehler homoskedastisch** und **unkorreliert** sind, $\Omega_0 = \sigma_0^2 \mathbf{I}$, ist aus Abschnitt 5.2 bekannt, dass die **optimalen Instrumente** durch $\mathbf{Z}_t = \bar{\mathbf{X}}_t = E(\mathbf{X}_t | \Omega_t)$ gegeben sind und in diesem Fall die Kovarianzmatrix (5.20) des **einfachen IV-Schätzers mit optimalen Instrumenten**

$$\sigma_0^2 \left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \bar{\mathbf{X}}^T \bar{\mathbf{X}} \right)^{-1}$$

lautet, da wie ebendort gezeigt $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \bar{\mathbf{X}}^T \mathbf{X} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \bar{\mathbf{X}}^T \bar{\mathbf{X}}$.

- **Wie lassen sich für (6.21) im allgemeiner Fall mit heteroskedastisch und/oder korrelierten Fehlern die optimalen Instrumente bestimmen?** Hier ist es zunächst nötig, das Ausgangsmodell zu transformieren, so dass man ein Modell mit homoskedastischen und unkorrelierten Fehlern erhält. Dies geschieht in folgendem Abschnitt 6.4.

6.4. Vollständig effizienter GMM-Schätzer für lineare Modelle

- Zur Erinnerung: Sind alle Regressoren bezüglich u_t streng exogen und Ω_0 bekannt, gilt für den GLS-Schätzer das Gauss-Markov-Theorem.
- **Notation:** Davidson und MacKinnon (2004, P. 358) folgend, schreiben wir im Folgenden für Ω_0 auch Ω .
- Die Ableitung des vollständig effizienten GMM-Schätzers verläuft analog zur Ableitung des GLS-Schätzers, indem

- die $(n \times n)$ -Matrix Ψ , gewöhnlich als eine **obere Dreiecksmatrix**, durch

$$\Psi\Psi^T = \Omega^{-1} \quad (6.23)$$

definiert wird und

- das lineare Modell (6.1a) mit Ψ^T von links multipliziert wird

$$\Psi^T \mathbf{y} = \Psi^T \mathbf{X}\boldsymbol{\beta} + \Psi^T \mathbf{u} \quad (6.24a)$$

$$\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{u}^*. \quad (6.24b)$$

Für die Kovarianzmatrix Ω^* der transformierten Fehler \mathbf{u}^* gilt wegen (6.1b)

$$\begin{aligned}\Omega^* &= E(\mathbf{u}^*(\mathbf{u}^*)^T) = \Psi^T E(\mathbf{u}\mathbf{u}^T) \Psi \\ &= \Psi^T \Omega \Psi = \Psi^T (\Psi \Psi^T)^{-1} \Psi \\ &= \Psi^T (\Psi^T)^{-1} \Psi^{-1} \Psi = \mathbf{I}.\end{aligned}\tag{6.25}$$

Damit weist das transformierte Modell (6.24) homoskedastische und unkorrelierte Fehler auf.

– Gilt darüber hinaus

$$E(y_t^* | \Omega_t) = E(\mathbf{X}_t^* | \Omega_t) \boldsymbol{\beta} \quad \text{bzw.} \quad E(u_t^* | \Omega_t) = 0,\tag{6.26}$$

sind die **optimalen Instrumente** durch

$$E(\mathbf{X}_t^* | \Omega_t) = \bar{\mathbf{X}}_t^*\tag{6.27}$$

gegeben. Vgl. die Theorie für den einfachen IV-Schätzer aus Abschnitt 5.2.

Man erhält aus (6.26) und (6.27) $E((\bar{\mathbf{X}}_t^*)^T u_t^*) = \mathbf{0}$ und daraus den **vollständig effizienten GMM-Schätzer (fully efficient GMM)**

$$\hat{\boldsymbol{\beta}}_{EGMM} = \left((\bar{\mathbf{X}}^*)^T \mathbf{X}^* \right)^{-1} (\bar{\mathbf{X}}^*)^T \mathbf{y}^* \quad (6.28)$$

$$\boldsymbol{\Sigma}_{EGMM} = \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} (\bar{\mathbf{X}}^*)^T \bar{\mathbf{X}}^* \right)^{-1}. \quad (6.29)$$

– **Wie lautet der vollständig effiziente Schätzer in den ursprünglichen Variablen?**

Notation: $(\mathbf{AB})_t$ bezeichne die t -te Zeile der Matrix \mathbf{AB} .

Die optimalen Instrumentenbedingungen (6.27) lautet dann

$$E \left((\boldsymbol{\Psi}^T \mathbf{X})_t \mid \Omega_t \right) = \left(\boldsymbol{\Psi}^T \tilde{\mathbf{X}} \right)_t = \bar{\mathbf{X}}_t^*. \quad (6.30)$$

Beachte: Im Gegensatz zu Davidson und MacKinnon (2004, Eq. (9.24)) schreiben wir $\tilde{\mathbf{X}}$ anstelle von $\bar{\mathbf{X}}$, da erstere durch $E \left((\boldsymbol{\Psi}^T \mathbf{X})_t \mid \Omega_t \right)$ (6.27) und letztere durch $E(\mathbf{X}_t \mid \Omega_t)$ definiert sind.

Der **vollständig effiziente GMM-Schätzer** lautet dann

$$\begin{aligned}\hat{\beta}_{EGMM} &= ((\bar{\mathbf{X}}^*)^T \mathbf{X}^*)^{-1} (\bar{\mathbf{X}}^*)^T \mathbf{y}^* = \left(\tilde{\mathbf{X}}^T \Psi \Psi^T \mathbf{X} \right)^{-1} \tilde{\mathbf{X}}^T \Psi \Psi^T \mathbf{y} \\ &= \left(\tilde{\mathbf{X}}^T \Omega^{-1} \mathbf{X} \right)^{-1} \tilde{\mathbf{X}}^T \Omega^{-1} \mathbf{y}\end{aligned}\quad (6.31)$$

und hat **asymptotische Kovarianzmatrix**

$$\Sigma_{EGMM} = \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} (\bar{\mathbf{X}}^*)^T \bar{\mathbf{X}}^* \right)^{-1} = \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \tilde{\mathbf{X}}^T \Omega^{-1} \tilde{\mathbf{X}} \right)^{-1}, \quad (6.32)$$

da $\Psi^T \tilde{\mathbf{X}} \in \Omega_t$ und somit $\text{plim}_{n \rightarrow \infty} \frac{1}{n} \tilde{\mathbf{X}}^T \Psi \Psi^T \mathbf{X} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \tilde{\mathbf{X}}^T \Psi \Psi^T \tilde{\mathbf{X}}$, vgl. Abschnitt 5.2.

- Im Fall $\tilde{\mathbf{X}} = \mathbf{X}$ entspricht (6.32) gerade der asymptotischen Kovarianzmatrix des GLS-Schätzers.
- Die **empirischen Momentenbedingungen** lauten

$$\begin{aligned}(\bar{\mathbf{X}}^*)^T (\mathbf{y}^* - \mathbf{X}^* \hat{\beta}_{EGMM}) &= \mathbf{0}, \quad \text{bzw.} \\ \tilde{\mathbf{X}}^T \Psi \Psi^T (\mathbf{y} - \mathbf{X} \hat{\beta}_{EGMM}) &= \mathbf{0}, \\ \tilde{\mathbf{X}}^T \Omega^{-1} (\mathbf{y} - \mathbf{X} \hat{\beta}_{EGMM}) &= \mathbf{0}\end{aligned}\quad (6.33)$$

- **Beachte:** Ist das **lineare Modell dynamisch vollständig und korrekt spezifiziert**, gilt also (6.2), ist auch Ψ eine Diagonalmatrix und (6.30) vereinfacht sich zu

$$E\left((\Psi^T \mathbf{X})_t \mid \Omega_t\right) = \Psi_{tt} E(\mathbf{X}_t \mid \Omega_t) = \Psi_{tt} \bar{\mathbf{X}}_t. \quad (6.34)$$

- Gewöhnlich ist die Matrix der **optimalen Instrumente** $\bar{\mathbf{X}}^*$ **nicht bekannt** und muss geschätzt werden.
 - Eine Schätzung kann wie in (5.22) erfolgen, wenn eine $(n \times l)$ -Instrumentenmatrix \mathbf{W}^* vorliegt, für die

$$\mathbf{W}_t^* \in \Omega_t \quad \text{und} \quad E(u_t^* \mid \mathbf{W}_t^*) = 0 \quad (6.35)$$

gilt. Da Ψ^T invertiert werden kann, existiert eine Instrumentenmatrix \mathbf{W}

$$\mathbf{W} = (\Psi^T)^{-1} \mathbf{W}^*,$$

so dass alternativ $\Psi^T \mathbf{W}_t \in \Omega_t$ geschrieben werden kann.

Eine Schätzung erfolgt durch

$$\widehat{\bar{\mathbf{X}}}^* = \mathbf{P}_{\mathbf{W}^*} \mathbf{X}^*. \quad (6.36)$$

Ersetzen von $\bar{\mathbf{X}}^*$ in (6.33) durch (6.36) ergibt die empirischen Momentenbedingungen

$$(\mathbf{X}^*)^T \mathbf{P}_{\mathbf{W}^*} (\mathbf{y}^* - \mathbf{X}^* \boldsymbol{\beta}) = \mathbf{0}.$$

Der hierdurch definierte **vollständig effiziente GMM-Schätzer** entspricht für das **transformierte Modell** einem **verallgemeinerten IV-Schätzer** (vgl. (5.24))

$$\hat{\boldsymbol{\beta}} = ((\mathbf{X}^*)^T \mathbf{P}_{\mathbf{W}^*} \mathbf{X}^*)^{-1} (\mathbf{X}^*)^T \mathbf{P}_{\mathbf{W}^*} \mathbf{y}^* \quad (6.37)$$

und hat asymptotische Kovarianzmatrix

$$\begin{aligned} & \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} (\bar{\mathbf{X}}^*)^T \mathbf{P}_{\mathbf{W}^*} \bar{\mathbf{X}}^* \right)^{-1} \quad \text{bzw. in den ursprünglichen Variablen} \\ & \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} \tilde{\mathbf{X}}^T \boldsymbol{\Omega}^{-1} \mathbf{W} (\mathbf{W}^T \boldsymbol{\Omega}^{-1} \mathbf{W})^{-1} \mathbf{W}^T \boldsymbol{\Omega}^{-1} \tilde{\mathbf{X}} \right)^{-1}. \end{aligned} \quad (6.38)$$

Diese Kovarianzmatrix ist “größer oder gleich” der Kovarianzmatrix (6.32) (vgl. Davidson und MacKinnon, 2004, Exercise 9.9).

– Gilt

$$\bar{\mathbf{X}}^* \in \delta(\mathbf{W}^*) \quad \text{bzw.} \quad \tilde{\mathbf{X}} \in \delta(\mathbf{W}),$$

ist (6.36) ein **konsistenter** Schätzer und der GMM-Schätzer (6.37) **asymptotisch vollständig effizient**, da dessen asymptotische Kovarianzmatrix der des vollständig effizienten GMM-Schätzers (6.32) entspricht (Davidson und MacKinnon, 2004, Exercise 9.8).

- **Beispiele:**

1. **Einfaches Regressionsmodell mit vorherbestimmten Regressor und autokorrelierten Fehlern**

$$y_t = x_t\beta + u_t, \quad E(u_t|x_t) = 0, \quad (6.39a)$$

$$u_t = u_{t-1}\rho + v_t. \quad (6.39b)$$

Einsetzen von $u_t = y_t - x_t\beta$ in (6.39a) liefert

$$\underbrace{y_t - \rho y_{t-1}}_{=y_t^*} = \underbrace{x_t - \rho x_{t-1}}_{=x_t^*} + \underbrace{v_t}_{=u_t^*}.$$

x_t^* ist unter den bisherigen Annahmen kein Instrument, da

$$\begin{aligned} E(x_t^* u_t^*) &= E[(x_t - \rho x_{t-1})(u_t - \rho u_{t-1})] \\ &= E(x_t u_t) - \rho E(x_{t-1} u_t) - \rho E(u_{t-1} x_t) + \rho^2 E(x_{t-1} u_{t-1}) \\ &= -\rho [E(x_{t-1} u_t) + E(u_{t-1} x_t)] \end{aligned}$$

ungleich Null sein kann und dann die theoretische Momentenbedingung verletzt ist.

- Wird zusätzlich zu (6.39a)

$$E(u_t | x_t, x_{t-1}, \dots) = 0 \quad (6.40)$$

angenommen, dann ist $E(x_{t-1}u_t) = 0$, aber nicht notwendigerweise $E(u_{t-1}x_t)$.

- Eine mögliche Instrumentvariable ist x_{t-1}^* , denn dann erhält man mit (6.40)

$$\begin{aligned} E(x_{t-1}^* u_t^*) &= E((x_{t-1} - \rho x_{t-2})(u_t - \rho u_{t-1})) \\ &= E(x_{t-1}u_t) - \rho E(x_{t-2}u_t) - \rho E(u_{t-1}x_{t-1}) + \rho^2 E(x_{t-2}u_{t-1}) = 0. \end{aligned}$$

Ist die Instrumentvariable x_{t-1}^* optimal? Nur wenn $E(x_t^* | \Omega_t) = \delta_0 + \delta_1 x_{t-1}^*$.

- Ist x_t **streng exogen**

$$E(u_t | \dots, x_{t-1}, x_t, x_{t+1}, \dots) = 0,$$

dann ist x_t^* ein optimales Instrument und man erhält den **GLS-Schätzer**.

2. Einfaches lineares Regressionsmodell mit endogenem Regressor und autokorrelierten Fehlern

$$y_t = x_t\beta + u_t, \quad E(y_t|w_t) = E(x_t|w_t)\beta \quad (6.41)$$

$$u_t = u_{t-1}\rho + v_t, \quad (6.42)$$

Try it.

3. Einfaches lineares Regressionsmodell mit omitted variable bias und heteroskedastischen Fehlern

- Zur Erinnerung: Der vollständig effiziente GMM-Schätzer (6.31) und der GMM-Schätzer (6.37) mit geschätzten optimalen Instrumenten erfordern eine konsistente Schätzung von Ω !

Ist dies nicht möglich, z.B. weil die Form der Autokorrelation in den Fehlern unbekannt ist, dann bleibt die Alternative, den effizienten GMM-Schätzer (6.19) durch Schätzung von $\mathbf{W}^T\Omega\mathbf{W}$ anwendbar zu machen, siehe folgenden Abschnitt.

		Varianz-Kovarianzmatrix Ω						
		diagonal			nicht diagonal			
Regressoren	Instrumente	Homoskedastie	Heteroskedastie		Heteroskedastie			
		keine Korrelation	keine Korrelation		Korrelation			
		σ konstant	σ_t bekannt	σ_t -Funktion bekannt	σ_t -Fkt. unbek.	Ω bekannt	Ω -Funktion bekannt	Ω -Fkt. unb.
$E(u_t \mathbf{X}) = 0$ streng exogen		OLS $(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$						
		$s^2 (\mathbf{X}^T \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Omega \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Omega \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Omega \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Omega \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \Omega \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1}$	
		$\text{plim } \sigma_0^2 \left(\frac{1}{n} \mathbf{X}^T \mathbf{X}\right)^{-1}$	$\text{plim} \left(\frac{1}{n} \mathbf{X}^T \mathbf{X}\right)^{-1} \text{plim} \left(\frac{1}{n} \mathbf{X}^T \Omega_0 \mathbf{X}\right) \text{plim} \left(\frac{1}{n} \mathbf{X}^T \mathbf{X}\right)^{-1}$					
			GLS $\left((\mathbf{X}^*)^T \mathbf{X}^*\right)^{-1} (\mathbf{X}^*)^T \mathbf{y}^*$ $(\mathbf{X}^T \Omega^{-1} \mathbf{X})^{-1}$	FGLS $\left((\tilde{\mathbf{X}}^*)^T \tilde{\mathbf{X}}^*\right)^{-1} (\tilde{\mathbf{X}}^*)^T \tilde{\mathbf{y}}^*$ $(\tilde{\mathbf{X}}^T \hat{\Omega}^{-1} \tilde{\mathbf{X}})^{-1}$		GLS $\left((\mathbf{X}^*)^T \mathbf{X}^*\right)^{-1} (\mathbf{X}^*)^T \mathbf{y}^*$ $(\mathbf{X}^T \Omega^{-1} \mathbf{X})^{-1}$	FGLS $\left((\tilde{\mathbf{X}}^*)^T \tilde{\mathbf{X}}^*\right)^{-1} (\tilde{\mathbf{X}}^*)^T \tilde{\mathbf{y}}^*$ $(\tilde{\mathbf{X}}^T \hat{\Omega}^{-1} \tilde{\mathbf{X}})^{-1}$	
$E(u_t \mathbf{X}_t) = 0$		OLS	GLS	FGLS	OLS-HC	u.U. GLS	-	OLS-HAC
$E(u_t \Omega_t) = 0$		OLS	GLS	FGLS	OLS-HC	nicht einschlägig		
$E(u_t \mathbf{X}_t) \neq 0$	$E(u_t \mathbf{Z}_t) = 0$ $E(u_t \mathbf{Z}_t) = 0$	IV $(\mathbf{Z}^T \mathbf{X})^{-1} \mathbf{Z}^T \mathbf{y}$						
		$\hat{\sigma}^2 (\mathbf{X}^T \mathbf{P}_Z \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{Z})^{-1} \mathbf{Z}^T \Omega \mathbf{Z} (\mathbf{Z}^T \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{Z})^{-1} \mathbf{Z}^T \Omega \mathbf{Z} (\mathbf{Z}^T \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{Z})^{-1} \mathbf{Z}^T \Omega \mathbf{Z} (\mathbf{Z}^T \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{Z})^{-1} \mathbf{Z}^T \Omega \mathbf{Z} (\mathbf{Z}^T \mathbf{X})^{-1}$	$(\mathbf{X}^T \mathbf{Z})^{-1} \mathbf{Z}^T \Omega \mathbf{Z} (\mathbf{Z}^T \mathbf{X})^{-1}$	
		$\text{plim } \sigma_0^2 \left(\frac{1}{n} \mathbf{X}^T \mathbf{P}_Z \mathbf{X}\right)^{-1}$	$\left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{X}^T \mathbf{Z}\right)^{-1} \left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}^T \Omega_0 \mathbf{Z}\right) \left(\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{Z}^T \mathbf{X}\right)^{-1}$					
	$\tilde{\mathbf{X}}_t = E(\mathbf{X}_t \Omega_t)$ opt. Inst.	$(\tilde{\mathbf{X}}^T \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \mathbf{X}$ $\text{plim } \sigma_0^2 \left(\frac{1}{n} \tilde{\mathbf{X}}^T \tilde{\mathbf{X}}\right)^{-1}$						
	$\tilde{\mathbf{X}}_t^* = E(\mathbf{X}_t^* \Omega_t)$ opt. Instr. $(\Psi^T \tilde{\mathbf{X}})_t$ $= E\left(\left(\Psi^T \mathbf{X}\right)_t \Omega_t\right)$		vollst. eff. GMM $\left((\tilde{\mathbf{X}}^*)^T \mathbf{X}^*\right)^{-1} (\tilde{\mathbf{X}}^*)^T \mathbf{y}^*$ $(\tilde{\mathbf{X}}^T \Omega^{-1} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \Omega^{-1} \mathbf{y}$	feas. vollst. eff. GMM $(\tilde{\mathbf{X}}^T \hat{\Omega}^{-1} \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}^T \hat{\Omega}^{-1} \mathbf{y}$		vollst. eff. GMM falls IV-Bedingung erfüllbar	feas. vollst. eff. GMM	
			$\text{plim} \left(\frac{1}{n} (\tilde{\mathbf{X}})^T \Omega_0^{-1} \tilde{\mathbf{X}}\right)^{-1}$ $\text{plim} \left(\frac{1}{n} (\tilde{\mathbf{X}}^*)^T \mathbf{X}^*\right)^{-1}$			$\text{plim} \left(\frac{1}{n} (\tilde{\mathbf{X}})^T \Omega_0^{-1} \tilde{\mathbf{X}}\right)^{-1}$ $\text{plim} \left(\frac{1}{n} (\tilde{\mathbf{X}}^*)^T \mathbf{X}^*\right)^{-1}$		
	$E(u_t \mathbf{W}_t) = 0$ opt. Instr. geschätzt	verallg. IV $(\mathbf{X}^T \mathbf{P}_W \mathbf{X})^{-1} \mathbf{X}^T \mathbf{P}_W \mathbf{y}$	eff. GMM $(\mathbf{X}^T \mathbf{W} (\mathbf{W}^T \Omega \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X})^{-1}$ $\mathbf{X}^T \mathbf{W} (\mathbf{W}^T \Omega \mathbf{W})^{-1} \mathbf{W}^T \mathbf{y}$	feas. eff. GMM $(\mathbf{X}^T \mathbf{W} (\mathbf{W}^T \hat{\Omega} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X})^{-1}$ $\mathbf{X}^T \mathbf{W} (\mathbf{W}^T \hat{\Omega} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{y}$		eff. GMM $(\mathbf{X}^T \mathbf{W} (\mathbf{W}^T \Omega \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X})^{-1}$ $\mathbf{X}^T \mathbf{W} (\mathbf{W}^T \Omega \mathbf{W})^{-1} \mathbf{W}^T \mathbf{y}$	feas. eff. GMM $(\mathbf{X}^T \mathbf{W} (\mathbf{W}^T \hat{\Omega} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X})^{-1}$ $\mathbf{X}^T \mathbf{W} (\mathbf{W}^T \hat{\Omega} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{y}$	
		$\sigma_0^2 \text{plim} \left(\frac{1}{n} \mathbf{X}^T \mathbf{P}_W \mathbf{X}\right)^{-1}$	$\text{plim} \left(\frac{1}{n} \mathbf{X}^T \mathbf{W}\right)^{-1} \text{plim} \left(\frac{1}{n} \mathbf{W}^T \Omega_0 \mathbf{W}\right) \text{plim} \left(\frac{1}{n} \mathbf{W}^T \mathbf{X}\right)^{-1}$					
	$E(u_t^* \mathbf{W}_t^*) = 0$ $\mathbf{W} = (\Psi^T)^{-1} \mathbf{W}^*$ opt. Instr. geschätzt		vollst. eff. GMM $\left((\mathbf{X}^*)^T \mathbf{P}_{\mathbf{W}^*} \mathbf{X}^*\right)^{-1} (\mathbf{X}^*)^T \mathbf{P}_{\mathbf{W}^*} \mathbf{y}^*$ $\text{plim} \left(\frac{1}{n} (\mathbf{X}^*)^T \mathbf{P}_{\mathbf{W}^*} \mathbf{X}^*\right)^{-1}$	feas. vollst. eff. GMM		vollst. eff. GMM falls IV-Bedingung erfüllbar $\text{plim} \left(\frac{1}{n} (\mathbf{X}^*)^T \mathbf{P}_{\mathbf{W}^*} \mathbf{X}^*\right)^{-1}$	feas. vollst. eff. GMM	
			$\text{plim} \left(\frac{1}{n} \tilde{\mathbf{X}}^T \Omega_0^{-1} \mathbf{W} (\mathbf{W}^T \Omega_0^{-1} \mathbf{W})^{-1} \mathbf{W}^T \Omega_0^{-1} \tilde{\mathbf{X}}\right)^{-1}$			$\text{plim} \left(\frac{1}{n} \tilde{\mathbf{X}}^T \Omega_0^{-1} \mathbf{W} (\mathbf{W}^T \Omega_0^{-1} \mathbf{W})^{-1} \mathbf{W}^T \Omega_0^{-1} \tilde{\mathbf{X}}\right)^{-1}$		

6.5. HAC-Kovarianzschätzer und der anwendbare effiziente GMM-Schätzer

- Kann Ω nicht geschätzt und kann zumindest die Bedingung (6.36) erfüllt werden, kann keine schätzbare Variante des vollständig effizienten GMM-Schätzer in 6.4 verwendet werden.
- Lässt sich jedoch $\mathbf{W}^T \Omega \mathbf{W}$ konsistent schätzen, dann erhält man den **anwendbaren effizienten GMM-Schätzer (feasible efficient GMM)**. Zur Schätzung von $\mathbf{W}^T \Omega \mathbf{W}$ werden 2 Fälle unterschieden:
 1. Ω ist **diagonal**:

Das lineare Modell ist **dynamisch vollständig und korrekt spezifiziert**, vgl. (6.2), so dass **ausschließlich Heteroskedastie** vorliegen kann.
 2. Ω ist **nicht diagonal**:

Das lineare Modell ist möglicherweise **nicht dynamisch vollständig und korrekt spezifiziert**, vgl. (6.1), so dass Heteroskedastie und Autokorrelation vorliegen kann.

- Im Folgenden betrachten wir ein einzelnes Element der empirischen Momente (6.4) an der Stelle $\beta = \beta_0$

$$\frac{1}{n} \mathbf{W}^T (\mathbf{y} - \mathbf{X}\beta_0) (\mathbf{y} - \mathbf{X}\beta_0)^T \mathbf{W} = \frac{1}{n} \mathbf{W}^T \mathbf{u} \mathbf{u}^T \mathbf{W} = \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n u_t u_s \mathbf{W}_t^T \mathbf{W}_s$$

- Es bezeichne $\sigma_{\bar{m},ij}$ das ij -Element in $\Sigma_{\bar{m}}$ gegeben durch (6.13a).
Wenn Annahme (6.13b) gilt, gilt aufgrund des Continuous Mapping Theorems (**Methoden der Ökonometrie**, Abschnitt 3.5 *Konvergenz in Verteilung*)

$$\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\mathbf{W}_{tj} u_t}{\sqrt{\sigma_{\bar{m},jj}}} \xrightarrow{d} \frac{\mathbf{W}_{\infty,j}}{\sqrt{\sigma_{\bar{m},jj}}} \sim N(0, 1), \quad j = 1, \dots, l,$$

$$\frac{1}{n \sigma_{\bar{m},jj}} \sum_{t=1}^n \sum_{s=1}^n \mathbf{W}_{tj} \mathbf{W}_{sj} u_t u_s = \left(\frac{1}{\sqrt{n}} \sum_{t=1}^n \frac{\mathbf{W}_{tj} u_t}{\sqrt{\sigma_{\bar{m},jj}}} \right)^2 \xrightarrow{d} \left(\frac{\mathbf{W}_{\infty,jj}}{\sqrt{\sigma_{\bar{m},jj}}} \right)^2 \sim \chi^2(1),$$

d. h. die Diagonalelemente von (6.4) konvergieren für $n \rightarrow \infty$ gegen eine χ^2 -verteilte Zufallsvariable und nicht gegen eine Konstante! Dies lässt sich auch für alle Nicht-Diagonalelemente zeigen.

- Annahme (6.13b) setzt voraus, dass Annahme (6.13a)

$$\begin{aligned}
 \Sigma_{\bar{m}} &= \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{W}^T \Omega_0 \mathbf{W} \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} E(\mathbf{W}^T \Omega_0 \mathbf{W}) \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n E(u_t u_s \mathbf{W}_t^T \mathbf{W}_s) \\
 &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n E(\omega_{ts} \mathbf{W}_t^T \mathbf{W}_s) \tag{6.43}
 \end{aligned}$$

erfüllt ist. Welche zusätzlichen Annahmen sind hierfür erforderlich?

Aufgrund von Voraussetzung (6.1a) $E(u_t | \mathbf{W}_t) = 0$ gilt $E(u_t u_s \mathbf{W}_t^T \mathbf{W}_s) = \text{Cov}(u_t \mathbf{W}_t, u_s \mathbf{W}_s)$ und somit

$$\Sigma_{\bar{m}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \text{Cov}(u_t \mathbf{W}_t, u_s \mathbf{W}_s). \tag{6.44}$$

Es ist hilfreich, diese Doppelsumme mit alternativer Indexierung zu schreiben.

Es bezeichne $a_{ts} = Cov(u_t \mathbf{W}_t, u_s \mathbf{W}_s)$. Die Doppelsumme (6.44) entspricht der Aufsummierung aller Terme in der quadratischen Tabelle

Index t /Index s	1	2	...	t	...	n
1	a_{11}	a_{12}	...	a_{1t}	...	a_{1n}
2	a_{21}	a_{22}	...	a_{2t}	...	a_{2n}
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	...
t	a_{t1}	a_{t2}	...	a_{tt}	...	a_{tn}
\vdots	\vdots	\vdots	\ddots	\vdots	\ddots	...
n	a_{n1}	a_{n2}	...	a_{ns}	...	a_{nn}

$$\Sigma_{\bar{m}} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n Cov(u_t \mathbf{W}_t, u_s \mathbf{W}_s) \quad (6.45a)$$

$$= \lim_{n \rightarrow \infty} \left\{ \frac{1}{n} \sum_{t=1}^n Var(u_t \mathbf{W}_t) \right. \quad (6.45b)$$

$$+ \sum_{j=1}^{n-1} \frac{1}{n} \sum_{t=j+1}^n Cov(u_t \mathbf{W}_t, u_{t-j} \mathbf{W}_{t-j}) \quad (6.45c)$$

$$\left. + \sum_{j=-(n-1)}^{-1} \frac{1}{n} \sum_{t=-j+1}^n Cov(u_{t+j} \mathbf{W}_{t+j}, u_t \mathbf{W}_t) \right\}. \quad (6.45d)$$

Bemerkungen:

- Teile der schwarzen, blauen und roten Summe können als (Auto)kovarianzmatrizen interpretiert werden, da sie (beinahe) Durchschnitten von (Auto)kovarianzmatrizen für Lag j entsprechen ([Davidson und MacKinnon, 2004](#), Equation (9.34))

$$\mathbf{\Gamma}_n(j) \equiv \begin{cases} \frac{1}{n} \sum_{t=1}^n \text{Var}(u_t \mathbf{W}_t) & \text{falls } j = 0, \\ \frac{1}{n} \sum_{t=j+1}^n \text{Cov}(u_t \mathbf{W}_t, u_{t-j} \mathbf{W}_{t-j}) & \text{falls } j > 0, \\ \frac{1}{n} \sum_{t=-j+1}^n \text{Cov}(u_{t+j} \mathbf{W}_{t+j}, u_t \mathbf{W}_t) & \text{falls } j < 0. \end{cases} \quad (6.46)$$

Da außerdem $\mathbf{\Gamma}(j) = \mathbf{\Gamma}(-j)^T$ gilt, erhält man

$$\mathbf{\Sigma}_{\bar{m}} = \lim_{n \rightarrow \infty} \sum_{j=-(n-1)}^{n-1} \mathbf{\Gamma}_n(j) = \lim_{n \rightarrow \infty} \left(\mathbf{\Gamma}_n(0) + \sum_{j=1}^{n-1} (\mathbf{\Gamma}_n(j) + \mathbf{\Gamma}_n(j)^T) \right). \quad (6.47)$$

Offensichtlich kann $\mathbf{\Sigma}_{\bar{m}}$ nur existieren, wenn die Autokovarianzmatrizen $\mathbf{\Gamma}_n(j)$ für $j \rightarrow \infty$ gegen eine Nullmatrix gehen!

– **Beispiele:**

* Fehler $u_t \sim IID(0, \sigma^2)$ und Instrumente $\mathbf{W}_t \sim IID(\mu_{\mathbf{W}}, \Sigma_{\mathbf{W}})$: Dann gilt

$$E(\mathbf{W}_t^T \mathbf{W}_s) = \Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T \quad (6.48a)$$

$$\boldsymbol{\Gamma}_n(0) = \frac{1}{n} \sum_{t=1}^n \sigma^2 (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) = \sigma^2 (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \quad (6.48b)$$

$$\boldsymbol{\Gamma}_n(j) = \frac{1}{n} \sum_{t=j+1}^n 0 \cdot (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) = \mathbf{0}, \quad j > 0 \quad (6.48c)$$

$$\boldsymbol{\Gamma}_n(j) = \mathbf{0}, \quad j < 0 \quad (6.48d)$$

$$\boldsymbol{\Sigma}_{\bar{m}} = \lim_{n \rightarrow \infty} \boldsymbol{\Gamma}_n(0) = \sigma^2 (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \quad (6.48e)$$

* **Fehler sind unkorreliert und deren Varianzen hängen ausschließlich von der Zeit ab**, d. h. $Var(u_t) = \omega_{tt}(t)$ und Instrumente wie oben:
Dann gilt

$$\begin{aligned}\Gamma_n(0) &= \frac{1}{n} \sum_{t=1}^n \omega_{tt}(t) (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \\ &= (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \frac{1}{n} \sum_{t=1}^n \omega_{tt}(t)\end{aligned}\quad (6.49a)$$

$\Gamma_n(j)$ und $\Gamma_n(j)$ wie oben und somit

$$\Sigma_{\bar{m}} = \lim_{n \rightarrow \infty} \Gamma_n(0) = (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \omega_{tt}(t). \quad (6.49b)$$

Damit $\Gamma_n(j)$ existiert, dürfen also die Varianzen $\omega_{tt}(t)$ weder gegen Unendlich noch gegen Null gehen:

$$0 < \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{t=1}^n \omega_{tt}(t) < \infty. \quad (6.49c)$$

- * **Konstante Fehlerkovarianz** $\omega_{ts} = \omega$, Instrumente $\mathbf{W}_t \sim IID(\mu_{\mathbf{W}}, \Sigma_{\mathbf{W}})$ und unabhängig von allen u_t : Dann gilt

$$\Gamma_n(j) = \frac{1}{n} \sum_{t=1}^n \omega (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) = \omega (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T), \quad j \neq 0.$$

Dann gilt für die blaue und rote Summe in (6.45)

$$\sum_{j=1}^{n-1} \Gamma_n(j) = n \omega (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \xrightarrow{n \rightarrow \infty} \infty$$

$$\sum_{j=-n-1}^{-1} \Gamma_n(j) = n \omega (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \xrightarrow{n \rightarrow \infty} \infty,$$

so dass $\Sigma_{\bar{m}}$ nicht existieren kann.

\Rightarrow **Autokovarianzen müssen mit $j \rightarrow \infty$ gegen Null gehen, damit $\Sigma_{\bar{m}}$ existieren kann.**

- * **Fehler u_t folgen einem stationären AR(1)-Prozess** und Instrumente wie oben: $\omega_{ts} = \alpha^{|t-s|} \frac{\sigma^2}{1-\alpha^2}$:

$$\begin{aligned} \sum_{j=1}^{n-1} \Gamma_n(j) &= \sum_{j=1}^{n-1} \alpha^j \frac{\sigma^2}{1-\alpha^2} (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \\ &= \frac{\sigma^2}{1-\alpha^2} (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \underbrace{\sum_{j=1}^{n-1} \alpha^j}_{\frac{1-\alpha^n}{1-\alpha}-1} \\ &\xrightarrow{n \rightarrow \infty} \frac{\sigma^2}{1-\alpha^2} (\Sigma_{\mathbf{W}} + \boldsymbol{\mu}_{\mathbf{W}} \boldsymbol{\mu}_{\mathbf{W}}^T) \left(\frac{\alpha}{1-\alpha} \right) < \infty \end{aligned}$$

- * **Allgemeiner:** Die Summen existieren, wenn die Autokovarianzmatrizen $Cov(u_t \mathbf{W}_t, u_{t-j} \mathbf{W}_{t-j})$, $Cov(u_{t+j} \mathbf{W}_{t+1}, u_t \mathbf{W}_t)$ in (6.46) exponentiell schnell für $|j| \rightarrow \infty$ gegen Null gehen.

- **Schätzung:**

Ist eine Schätzung von (6.47)

$$\Sigma_{\bar{m}} = \text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{W}^T \Omega_0 \mathbf{W} \quad (6.50)$$

$$= \lim_{n \rightarrow \infty} \sum_{j=-(n-1)}^{n-1} \Gamma(j) = \lim_{n \rightarrow \infty} \left(\Gamma(0) + \sum_{j=1}^{n-1} (\Gamma(j) + \Gamma(j)^T) \right). \quad (6.51)$$

und damit auch von $\frac{1}{n} \mathbf{W}^T \Omega_0 \mathbf{W}$ durch

$$\hat{\Sigma} = \sum_{j=-(n-1)}^{n-1} \hat{\Gamma}_n(j) = \hat{\Gamma}(0) + \sum_{j=1}^{n-1} (\hat{\Gamma}(j) + \hat{\Gamma}(j)^T) \quad (6.52)$$

möglich, wobei die $\hat{\Gamma}(j)$ noch definiert werden müssen ?

- Sind die **Fehler unkorreliert und heteroskedastisch**, muss wegen (6.49b) ausschließlich $\Gamma_n(0)$ geschätzt werden.

Wie lässt sich $\Gamma_n(0)$ schätzen?

– Wegen (6.46)

$$\Gamma_n(0) = \frac{1}{n} \sum_{t=1}^n \text{Var}(u_t \mathbf{W}_t) = \frac{1}{n} \sum_{t=1}^n E(u_t^2 \mathbf{W}_t^T \mathbf{W}_t),$$

erhält man einen Schätzer für $\Gamma_n(0)$ durch Weglassen des Erwartungswertes und Ersetzen von u_t durch einen konsistenten Schätzer \hat{u}_t :

$$\begin{aligned} \hat{\Gamma}_n(0) &= \frac{1}{n} \sum_{t=1}^n \hat{u}_t^2 \mathbf{W}_t^T \mathbf{W}_t \\ &= \underbrace{\frac{1}{n} \sum_{t=1}^n \omega_{tt} \mathbf{W}_t^T \mathbf{W}_t}_{\xrightarrow{p} \Sigma_{\bar{m}}} + \underbrace{\frac{1}{n} \sum_{t=1}^n (\hat{u}_t^2 - \omega_{tt}) \mathbf{W}_t^T \mathbf{W}_t}_{\xrightarrow{p} 0}, \end{aligned}$$

wobei die plims unter geeigneten Annahmen folgen, u.a. z.B. (6.49c).

- Eine Schätzung von \hat{u}_t kann mit dem verallgemeinerten IV-Schätzer erfolgen, der wegen

$$\mathbf{\Lambda} = (\mathbf{W}^T \mathbf{W})^{-1} \quad \text{mit} \quad \mathbf{J} = (\mathbf{W}^T \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X}$$

ein Spezialfall von (6.8) ist und keine Kenntnis von $\mathbf{\Omega}$ erfordert.

- **Autokorrelierte und heteroskedastische Fehler:** Zusätzlich müssen $\mathbf{\Gamma}_n(j)$, $j \neq 0$ geschätzt werden.

- Als Schätzer wird

$$\hat{\mathbf{\Gamma}}_n(j) = \frac{1}{n} \sum_{t=j+1}^n \hat{u}_t \hat{u}_{t-j} \mathbf{W}_t \mathbf{W}_{t-j}^T$$

verwendet.

- Der Schätzer $\hat{\mathbf{\Gamma}}_n(j)$ ist nicht für alle j konsistent, da z.B. für $j = n - 2$ unabhängig von n immer genau 2 Beobachtungen vorliegen. Deshalb können in (6.52) im zweiten Term nicht alle j berücksichtigt werden. Stattdessen beschränkt man die Aufsummierung auf $p \ll n$ Terme.

- Alle konsistenten Schätzer von $\Sigma_{\bar{m}}$ werden als **HAC-Schätzer (heteroscedasticity and autocorrelation consistent)** bezeichnet.

HAC-Schätzer:

* Hansen-White-Schätzer

$$\hat{\Sigma} = \hat{\Gamma}(0) + \sum_{j=1}^p \left(\hat{\Gamma}(j) + \hat{\Gamma}(j)^T \right) \quad (6.53)$$

Der Lagparameter p sollte mit $n^{1/4}$ wachsen.

Nachteil: In endlichen Stichproben ist Schätzer nicht immer positiv semi-definit

* Newey-West-Schätzer

$$\hat{\Sigma} = \hat{\Gamma}(0) + \sum_{j=1}^p \left(1 - \frac{j}{p+1} \right) \left(\hat{\Gamma}(j) + \hat{\Gamma}(j)^T \right) \quad (6.54)$$

Der Lagparameter p sollte mit $n^{1/3}$ wachsen.

- * Weitere HAC-Schätzer finden sich in [Hayashi \(2000, Section 6.6\)](#) oder in dem Referenzartikel [Andrews \(1991\)](#).
- * Es gibt keine vollständig automatische Wahlprozedur für die Lagparameter p .

- Man erhält den **anwendbaren effizienten GMM-Schätzer (feasible efficient GMM estimator)**

$$\hat{\beta}_{FGMM} = \left(\mathbf{X}^T \mathbf{W} \hat{\Sigma}^{-1} \mathbf{W}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{W} \hat{\Sigma}^{-1} \mathbf{W}^T \mathbf{y} \quad (6.55)$$

$$= \left(\mathbf{X}^T \mathbf{W} \left(\mathbf{W}^T \hat{\Omega} \mathbf{W} \right)^{-1} \mathbf{W}^T \mathbf{X} \right)^{-1} \mathbf{X}^T \mathbf{W} \left(\mathbf{W}^T \hat{\Omega} \mathbf{W} \right)^{-1} \mathbf{W}^T \mathbf{y} \quad (6.56)$$

wobei in der zweiten Zeile die Notation von [Davidson und MacKinnon \(2004, S. 357\)](#) verwendet wurde $\hat{\Sigma} = \frac{1}{n} \mathbf{W}^T \hat{\Omega} \mathbf{W}$.

6.6. Tests auf Basis der GMM-Zielfunktion

In diesem Abschnitt:

1. Testen der überidentifizierenden Restriktionen
2. Testen linearer Restriktionen

im linearen Modell (1.2)

$$y_t = \mathbf{X}_t \boldsymbol{\beta} + u_t.$$

Die asymptotischen Ergebnisse dieses Abschnitts gelten unabhängig davon, ob $\mathbf{W}^T \boldsymbol{\Omega} \mathbf{W}$ bekannt ist oder mittels eines HAC-Schätzers geschätzt wurde.

(a) Testen der überidentifizierenden Restriktionen

- **Zur Erinnerung:** Der effiziente GMM-Schätzer (6.19) verwendet de facto k effektive Instrumente, die als Linearkombination $\mathbf{W}\mathbf{J} = \mathbf{W} (\mathbf{W}^T \boldsymbol{\Omega} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{X}$ aus den l Instrumenten \mathbf{W} bestimmt werden. Ist mindestens eines der $l > k$ Instrumente unzulässig, ist der effiziente GMM-Schätzer inkonsistent, da (6.1a)

verletzt wird.

- Umschreiben der GMM criterion function (6.18)

$$Q(\boldsymbol{\beta}, \mathbf{y}) = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \mathbf{W} (\mathbf{W}^T \boldsymbol{\Omega} \mathbf{W})^{-1} \mathbf{W}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

mit (6.23) und $\mathbf{A} \equiv \boldsymbol{\Psi}^{-1} \mathbf{W}$ liefert

$$\begin{aligned} Q(\boldsymbol{\beta}, \mathbf{y}) &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Psi} \boldsymbol{\Psi}^{-1} \mathbf{W} \left(\mathbf{W}^T (\boldsymbol{\Psi}^T)^{-1} \boldsymbol{\Psi}^{-1} \mathbf{W} \right)^{-1} \mathbf{W}^T (\boldsymbol{\Psi}^T)^{-1} \boldsymbol{\Psi}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Psi} \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \boldsymbol{\Psi}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Psi} \mathbf{P}_A \boldsymbol{\Psi}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\ &= [(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Psi} \mathbf{P}_A] [\mathbf{P}_A \boldsymbol{\Psi}^T (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})] \end{aligned} \tag{6.57}$$

- **Wäre der korrekte Parametervektor $\beta = \beta_0$ bekannt** (so dass β nicht geschätzt werden müsste), könnte man β_0 direkt in $Q(\beta, \mathbf{y})$ einsetzen und erhielte mit $\mathbf{u} = \mathbf{y} - \mathbf{X}\beta_0$

$$\begin{aligned}
 Q(\beta_0, \mathbf{y}) &= \mathbf{u}^T \mathbf{W} (\mathbf{W}^T \boldsymbol{\Omega} \mathbf{W})^{-1} \mathbf{W}^T \mathbf{u} \\
 &= \mathbf{u}^T \boldsymbol{\Psi} \mathbf{P}_A \boldsymbol{\Psi}^T \mathbf{u} \\
 &= (\mathbf{u}^*)^T \mathbf{P}_A \mathbf{u}^*
 \end{aligned} \tag{6.58}$$

– **Interpretation** von $Q(\beta_0, \mathbf{y})$:

$$\frac{1}{n} \mathbf{W}^T \mathbf{u} \tag{6.59}$$

ist ein Schätzer der theoretischen Momente $E(\mathbf{W}_t u_t)$ in (6.3). Ist Instrument j nicht zulässig und damit (6.1a) nicht erfüllt, gilt

$$\lim_{n \rightarrow \infty} E \left(\frac{1}{n} \sum_{t=1}^n \mathbf{W}_{tj} u_t \right) \equiv a_j \neq 0.$$

Wird Annahme (6.13b) entsprechend modifiziert, gelten

$$\frac{1}{n} \mathbf{W}^T \mathbf{u} - \mathbf{a} \xrightarrow{P} \mathbf{0}$$

$$\sqrt{n} \left(\frac{1}{n} \mathbf{W}^T \mathbf{u} - \mathbf{a} \right) = \frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u} - \sqrt{n} \mathbf{a} \xrightarrow{d} \mathbf{w}_\infty \sim N(\mathbf{0}, \Sigma_{\bar{m}}).$$

Damit lässt sich $\frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u}$ approximativ schreiben als

$$\frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u} \stackrel{a}{\sim} \mathbf{w}_\infty + \sqrt{n} \mathbf{a}.$$

Einsetzen in $Q(\beta_0, \mathbf{y})$ ergibt

$$\begin{aligned} Q(\beta_0, \mathbf{y}) &= \frac{1}{\sqrt{n}} \mathbf{u}^T \mathbf{W} \left(\frac{1}{n} \mathbf{W}^T \Omega \mathbf{W} \right)^{-1} \frac{1}{\sqrt{n}} \mathbf{W}^T \mathbf{u} \\ &\stackrel{a}{\sim} (\mathbf{w}_\infty + \sqrt{n} \mathbf{a})^T \Sigma_{\bar{m}}^{-1} (\mathbf{w}_\infty + \sqrt{n} \mathbf{a}) \\ &\stackrel{a}{\sim} \mathbf{w}_\infty^T \Sigma_{\bar{m}}^{-1} \mathbf{w}_\infty + 2\sqrt{n} \mathbf{a}^T \Sigma_{\bar{m}}^{-1} \mathbf{w}_\infty + n \mathbf{a}^T \Sigma_{\bar{m}}^{-1} \mathbf{a} \xrightarrow{n \rightarrow \infty} \infty. \end{aligned} \quad (6.60)$$

* **Ist also mindestens ein Instrument nicht zulässig**, geht der Term $n^2 \mathbf{a}^T \mathbf{a}$ und damit $Q(\beta_0, \mathbf{y})$ gegen unendlich

$$Q(\beta_0, \mathbf{y}) \xrightarrow{n \rightarrow \infty} \infty.$$

* **Sind alle l Instrumente zulässig**, ist $\mathbf{a} = \mathbf{0}$. Dann gilt (siehe (6.61))

$$Q(\boldsymbol{\beta}_0, \mathbf{y}) \xrightarrow{d} \chi^2(l).$$

– Damit würde sich $Q(\boldsymbol{\beta}_0, \mathbf{y})$ als Teststatistik eignen, wenn $\boldsymbol{\beta}$ bekannt wäre. Der Parametervektor $\boldsymbol{\beta}_0$ lässt sich jedoch mit $\hat{\boldsymbol{\beta}}_{GMM}$ schätzen. Allerdings ändern sich dadurch die Freiheitsgrade der asymptotischen χ^2 -Verteilung, siehe (6.67).

– **Asymptotische Verteilung von $Q(\boldsymbol{\beta}_0, \mathbf{y})$ unter der Nullhypothese “Alle l Instrumente sind zulässig”:**

Wenn außerdem Annahme (6.13b) gilt, folgt aufgrund des Continuous Mapping Theorems und den Eigenschaften der χ^2 -Quadratverteilung

$$\begin{aligned} Q(\boldsymbol{\beta}_0, \mathbf{y}) &= (\mathbf{u}^*)^T \mathbf{P}_A \mathbf{u}^* = (\mathbf{u}^*)^T \mathbf{A} (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{u}^* \\ &= \frac{1}{n} \mathbf{u}^T \mathbf{W} \left(\frac{1}{n} \mathbf{W}^T \boldsymbol{\Omega}_0 \mathbf{W} \right)^{-1} \mathbf{W}^T \mathbf{u} \\ &\xrightarrow{d} \mathbf{w}_\infty^T \boldsymbol{\Sigma}_{\bar{m}}^{-1} \mathbf{w}_\infty \sim \chi^2(l) \end{aligned} \tag{6.61}$$

- β ist unbekannt und wird mit $\hat{\beta}_{GMM}$ geschätzt:

- Entsprechend zu $Q(\beta, y)$ lässt sich der **effiziente GMM-Schätzer** (6.19) umschreiben zu

$$\begin{aligned}\hat{\beta}_{GMM} &= (\mathbf{X}^T \Psi \mathbf{P}_A \Psi^T \mathbf{X})^{-1} \mathbf{X}^T \Psi \mathbf{P}_A \Psi^T y \\ &= \left((\mathbf{X}^*)^T \mathbf{P}_A \mathbf{X}^* \right)^{-1} (\mathbf{X}^*)^T \mathbf{P}_A y^*.\end{aligned}$$

- Der effiziente GMM-Schätzer entspricht einem verallgemeinerten IV-Schätzer (5.24) für das transformierte Modell (6.24b)

$$y^* = \mathbf{X}^* \beta + \mathbf{u}^*$$

mit den l Instrumenten $\mathbf{A} = \Psi^{-1} \mathbf{W}$, die im allgemeinen jedoch nicht $\bar{\mathbf{X}}^*$ entsprechen und deshalb nicht optimal sind, vgl. (6.27).

- Damit entspricht der effiziente GMM-Schätzer der KQ-Schätzung von

$$y^* = \mathbf{P}_A \mathbf{X}^* \beta + \xi^* \quad (6.62)$$

\Leftrightarrow 2. Stufe des 2-stufigen KQ-Schätzers \Leftrightarrow verallgemeinerter IV-Schätzer

- Eine Voraussetzung für die Gültigkeit der Instrumente ist, dass die extra Regressoren im linearen Modell (6.62) nicht als Regressoren auftauchen. Genau dies kann mit $Q(\hat{\beta}_{GMM}, \mathbf{y})$ getestet werden, wie im folgenden gezeigt wird:
 - $\mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}$ und $\mathbf{M}_{\mathbf{P}_A \mathbf{X}^*}$ sind komplementäre Projektionen (**Methoden der Ökonometrie**, Abschnitt 7.1 *Die Geometrie des KQ-Schätzers*), da

$$\mathbf{I} = \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*} + \mathbf{M}_{\mathbf{P}_A \mathbf{X}^*}. \quad (6.63)$$

Der Unterraum $\delta(\mathbf{A})$ lässt sich deshalb in zwei orthogonale Unterräume aufspalten: den k -dimensionalen Unterraum der k effektiven Instrumente $\mathbf{P}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{A}$ und den $l - k$ -dimensionalen Unterraum der extra Instrumente $\mathbf{M}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{A}$:

$$\delta(\mathbf{A}) = \delta(\mathbf{P}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{A}, \mathbf{M}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{A}). \quad (6.64)$$

- Es bezeichne \mathbf{A}_e eine Matrix von $l - k$ extra Instrumenten, für die $\mathbf{A}_e \in \delta(\mathbf{M}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{A})$ gilt. Beeinflussen diese die zweite Stufe (6.62), gilt

$$\mathbf{y}^* = \mathbf{P}_A \mathbf{X}^* \boldsymbol{\beta} + \mathbf{A}_e \boldsymbol{\gamma} + \boldsymbol{\xi}^*, \quad (6.65)$$

wobei der Parametervektor $\boldsymbol{\gamma}$ nicht Null ist. Da für die erste Regressormatrix in (6.65) $\mathbf{P}_A \mathbf{X}^* \in \delta(\mathbf{P}_{\mathbf{P}_A \mathbf{X}^*})$ gilt, lässt sich (6.65) und damit das Modell

auch als

$$\mathbf{y}^* = \mathbf{A}\boldsymbol{\delta} + \boldsymbol{\xi}^*$$

schreiben.

– Hypothesen

H_0 : $l - k$ extra Instrumente sind zulässig $\Leftrightarrow \boldsymbol{\gamma} = \mathbf{0}$

H_1 : mindestens eins der $l - k$ extra Instrumente ist unzulässig $\Leftrightarrow \boldsymbol{\gamma} \neq \mathbf{0}$.

– Teststatistik:

Die Nullhypothese kann mit Hilfe einer F-Statistik für die Regressionen der 2. Stufe getestet werden:

$$\begin{aligned} (l - k)F &= \frac{SSR_0 - SSR_1}{1} = (\mathbf{y}^*)^T \mathbf{M}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{y}^* - (\mathbf{y}^*)^T \mathbf{M}_A \mathbf{y}^* \\ &= (\mathbf{y}^*)^T (\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{y}^* = Q(\hat{\boldsymbol{\beta}}_{GMM}, \mathbf{y}) \end{aligned}$$

wobei das letzte Gleichheitszeichen im folgenden gezeigt wird:

* Wird $\hat{\beta}_{GMM}$ in $Q(\beta, \mathbf{y})$ (6.57) eingesetzt, erhält man

$$\begin{aligned} \left[\mathbf{P}_A \Psi^T (\mathbf{y} - \mathbf{X} \hat{\beta}_{GMM}) \right] &= \left[\mathbf{P}_A (\mathbf{y}^* - \mathbf{X}^* \hat{\beta}_{GMM}) \right] \\ &= \mathbf{P}_A \mathbf{y}^* - \underbrace{\mathbf{P}_A \mathbf{X}^* \left((\mathbf{X}^*)^T \mathbf{P}_A \mathbf{X}^* \right)^{-1} (\mathbf{X}^*)^T \mathbf{P}_A \mathbf{y}^*}_{\equiv \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}} \\ &= (\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{y}^*. \end{aligned}$$

* Da $\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}$ idempotent und symmetrisch ist, erhält man

$$Q(\hat{\beta}_{GMM}, \mathbf{y}) = (\mathbf{y}^*)^T (\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{y}^* \quad (6.66)$$

was zu zeigen war.

– Asymptotische Verteilung

* Unter H_0 gilt $\mathbf{y}^* = \mathbf{X}^* \boldsymbol{\beta} + \mathbf{u}^*$, so dass

$$(\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{y}^* = \underbrace{(\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{X}^*}_{=0} \boldsymbol{\beta}_0 + (\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{u}^*$$

$$Q(\hat{\boldsymbol{\beta}}_{GMM}, \mathbf{y}) = (\mathbf{y}^*)^T (\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{y}^* = (\mathbf{u}^*)^T (\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{u}^*$$

* Multiplizieren von (6.63) mit \mathbf{P}_A ergibt

$$\mathbf{P}_A = \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*} + \mathbf{M}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{P}_A,$$

so dass die Projektionsmatrix

$$\mathbf{M}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{P}_A = \mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}$$

gerade in den Unterraum der extra $l - k$ Instrumente projiziert. In Analogie zu (6.61)

$$\begin{aligned} Q(\hat{\boldsymbol{\beta}}_{GMM}, \mathbf{y}) &= (\mathbf{u}^*)^T (\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{u}^* \\ &= (\mathbf{u}^*)^T \mathbf{M}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{P}_A \mathbf{u}^* \xrightarrow{d} \chi^2(l - k) \end{aligned} \quad (6.67)$$

– Die Teststatistik $Q(\hat{\boldsymbol{\beta}}_{GMM}, \mathbf{y})$ wird als **Hansen's overidentification statistic**, **Hansen's J statistic** oder als **Hansen-Sargan statistic** bezeichnet.

– **Interpretation:**

Vgl. (6.59). Muss β geschätzt werden, kann nur die Gültigkeit der extra $l - k$ Instrumente überprüft werden, da

$$\frac{1}{n} \mathbf{M}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{P}_A \mathbf{u}^*$$

in (6.67) gerade die $l - k$ theoretischen Momente $E \left(\left(\mathbf{M}_{\mathbf{P}_A \mathbf{X}^*} \mathbf{P}_A \right)_t u_t^* \right)$ überprüft.
Ursache: Durch Schätzung von β gehen k Freiheitsgrade verloren.

– **Beachte:**

Wird H_0 abgelehnt, kann das daran liegen, dass

- * die überidentifizierenden Instrumente nicht zulässig sind,
- * das lineare Modell (6.24b) nicht korrekt spezifiziert ist,
- * der HAC-Schätzer nicht gut gewählt ist,
- * die asymptotische Verteilung nicht gut approximiert.

(b) Testen linearer Restriktionen und von Nullrestriktionen

Gegeben sei wieder das lineare Modell (1.2)

$$\begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ &= \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u} \\ \mathbf{y}^* &= \mathbf{X}_1^*\boldsymbol{\beta}_1 + \mathbf{X}_2^*\boldsymbol{\beta}_2 + \mathbf{u}^* \end{aligned}$$

mit Instrumentenmatrix \mathbf{W} bzw. \mathbf{A} , die die Annahmen (6.1) erfüllt. Dabei weisen \mathbf{X}_1 und \mathbf{X}_2 jeweils k_1 und k_2 Spalten auf, wobei $k_1 + k_2 = k$.

- **Hypothesen**

$$H_0 : \boldsymbol{\beta}_2 = 0 \quad \text{versus} \quad H_1 : \boldsymbol{\beta}_2 \neq 0$$

- **Teststatistik:** Die Nullhypothese kann wieder mit einer F-Statistik getestet werden:

$$k_2 F = \frac{SSR_0 - SSR_1}{1}$$

Dabei ist SSR_1 wegen (6.66) durch

$$SSR_1 = Q(\hat{\boldsymbol{\beta}}_{GMM}, \mathbf{y}) = (\mathbf{y}^*)^T (\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}^*}) \mathbf{y}^*$$

gegeben.

Entsprechend lässt sich SSR_0 durch

$$SSR_0 = Q(\hat{\beta}_{1,GMM}, \mathbf{y}) = (\mathbf{y}^*)^T (\mathbf{P}_A - \mathbf{P}_{\mathbf{P}_A \mathbf{X}_1^*}) \mathbf{y}^*$$

Einsetzen in die obige Teststatistik ergibt

$$\begin{aligned} \frac{SSR_0 - SSR_1}{1} &= Q(\hat{\beta}_{1,GMM}, \mathbf{y}) - Q(\hat{\beta}_{GMM}, \mathbf{y}) \\ &= (\mathbf{y}^*)^T (\mathbf{P}_{\mathbf{P}_A \mathbf{X}^*} - \mathbf{P}_{\mathbf{P}_A \mathbf{X}_1^*}) \mathbf{y}^* \end{aligned}$$

- **Asymptotische Verteilung:**

Mit den oben dargestellten Überlegungen lässt sich zeigen, dass unter H_0

$$Q(\hat{\beta}_{1,GMM}, \mathbf{y}) - Q(\hat{\beta}_{GMM}, \mathbf{y}) \xrightarrow{d} \chi^2(k_2).$$

- **Beachte:**

- Die Teststatistik setzt voraus, dass die gleiche Instrumentenmatrix \mathbf{W} unter H_0 und H_1 verwendet wird.
- Es wird die gleiche HAC-Schätzung für $\mathbf{W}^T \boldsymbol{\Omega} \mathbf{W}$ verwendet.
- Es wird der effiziente GMM-Schätzer oder der anwendbare effiziente GMM-Schätzer verwendet.

A. Mehr zu Wahrscheinlichkeitstheorie

A.1. L^p -Raum und L^p -Norm

- Ein Wahrscheinlichkeitsraum (Ω, \mathcal{F}, P) , siehe [Eine kurze Einf. in die W'theorie](#) besteht aus einer Ergebnismenge Ω , einer σ -Algebra \mathcal{F} und einer Wahrscheinlichkeitsverteilung P .
- Um unterschiedliche Wahrscheinlichkeitsverteilungen zu klassifizieren, wird häufig untersucht, ob für die Zufallsvariable X das Moment p -ter Ordnung $E(|X|^p)$ existiert, d. h. ob gilt ([Davidson, 1994](#), Section 9.1, S. 132):

$$E(|X|^p) < \infty. \quad (\text{A.1})$$

Je größer [das maximale](#) p ist, [bei dem \(A.1\) noch erfüllt ist](#), desto weniger dick

sind die Flanken einer Verteilung. (Vgl. t -Verteilung mit m Freiheitsgraden und Normalverteilung.)

- Gilt (A.1), sagt man, dass X zu einem L^p -**Raum** (genauer $L^p(\Omega, \mathcal{F}, P)$) gehört, kurz

$$X \in L^p.$$

Eine Norm (vgl. **Methoden der Ökonometrie**, Abschnitt 1.3 *Euklidischer Raum*) für L^p -Räume ist durch die L^p -**Norm**

$$\|X\|_p \equiv (E(|X|^p))^{1/p} \quad (\text{A.2})$$

gegeben.

- L^2 -**Raum**:

$X \in L^2$ ist gleichbedeutend damit, dass $Var(X) < \infty$, also die Varianz von X existiert. Das innere Produkt eines L^2 -Raums ist durch

$$\langle X, Y \rangle = E(XY), \quad X, Y \in L^2,$$

gegeben ([Brockwell und Davis, 1991](#), Example 2.2.2). Zum inneren Produkt siehe **Methoden der Ökonometrie**, Abschnitt 1.5 *Weitere Operationen mit Matrizen*.

- Beispiel: Für eine normalverteilte Zufallsvariable $X \sim N(\mu, \sigma^2)$ gilt (Davidson, 1994, 9.7 Example, p. 131)

$$E [(X - \mu)^k] = \begin{cases} \frac{k! \sigma^k}{2^{k/2} ((k/2)!)} & k \text{ gerade,} \\ 0 & k \text{ ungerade.} \end{cases} \quad (\text{A.3})$$

- Mehr zu L^p -Räumen in Wikipedia unter Lp-Raum. Beachte, Davidson (1994, Section 9.1) schreibt L_p anstelle von L^p .

A.2. Konvergenz von Folgen von Zufallsvariablen

- Vgl. Methoden der Ökonometrie, Kapitel 3.
- **Konvergenz in der L^p -Norm**

Für eine Folge von Zufallsvariablen X_n , $n = 1, 2, \dots$, gelte für ein $p > 0$:

$$\|X_n\|_p < \infty.$$

Wenn für die (nicht-stochastische) Folge $\|X_n - X\|_p, n = 1, 2, \dots$, gilt, dass

$$\lim_{n \rightarrow \infty} \|X_n - X\|_p = 0, \quad (\text{A.4})$$

dann konvergiert $\{X_n\}$ in der L^p -Norm gegen die (Zufalls)variable X , kurz (L^p Konvergenz)

$$X_n \xrightarrow{L^p} X.$$

Vgl. Davidson (1994, Section 18.4).

- **Konvergenz im quadratischen Mittel**

Ist $p = 2$, spricht man auch von Konvergenz im quadratischen Mittel, da genau dann die Varianz von $X_n - X$ für $n \rightarrow \infty$ gegen 0 geht.

- **Eigenschaften**

- $0 < q < p$: L^p -Konvergenz $\Rightarrow L^q$ -Konvergenz
(Davidson, 1994, Section 18.4).

- $p > 0$: L^p -Konvergenz \Rightarrow Konvergenz in Wahrscheinlichkeit
folgt mit Hilfe von Markov-Ungleichung (Davidson, 1994, Section 18.4, Theorem 18.13).

– Für die Umkehrung siehe [Davidson \(1994, Section 18.4, Theorem 18.14\)](#).

- Ist der Grenzwert X unbekannt, lässt sich die Konvergenz in einer Norm $\|\cdot\|$ trotzdem überprüfen. Gilt für eine Folge $x_n, n = 1, 2, \dots$, dass

$$\|x_n - x_m\| \longrightarrow 0, \text{ für } m, n \rightarrow \infty, \quad (\text{A.5})$$

dann konvergiert die Folge $\{x_n\}$. Die Folge wird dann als **Cauchy-Folge** bezeichnet ([Brockwell und Davis, 1991, Definition 2.2.1](#)). Die Konvergenz erfordert, dass für jedes $\epsilon > 0$ eine positive ganze Zahl $N(\epsilon)$ existiert, so dass $\|x_n - x_m\| < \epsilon$ für alle $n, m > N(\epsilon)$.

- Beispiel: Es gelte $u_i \sim NID(0, 1)$ und

$$S_n = \sum_{i=1}^n a_i u_i.$$

Dann gilt auf Basis der L^2 -Norm

$$\|S_m - S_n\|^2 = \sum_{i=n+1}^m a_i^2, \quad m > n. \quad (\text{A.6})$$

Existiert für jedes $\epsilon > 0$ ein $N(\epsilon) > 0$, so dass $\sum_{i=n+1}^m a_i^2 < \epsilon$ für $m > n > N(\epsilon)$, dann konvergiert S_n gegen eine Zufallsvariable S . D. h. S_n konvergiert

im quadratischen Mittel, wenn $\sum_{i=1}^{\infty} a_i^2 < \infty$. Die Umkehrung gilt auch, siehe [Brockwell und Davis \(1991, Example 2.2.4\)](#).

A.3. Wahrscheinlichkeitsungleichungen

- **Markov-Ungleichung**

Für $\epsilon > 0$ und $p > 0$ und $\mu \equiv E(X)$

$$P(|X - \mu| \geq \epsilon) \leq \frac{E|X - \mu|^p}{\epsilon^p}. \quad (\text{A.7})$$

Für einen Beweis siehe [Davidson \(1994, Section 9.3, Theorem 9.9\)](#).

- **Tschebyscheff-Ungleichung**

Markov-Ungleichung für $p = 2$.

- **Hölder-Ungleichung** ([Davidson, 2000, Equation \(B.5.13\)](#)):

Für

$$\frac{1}{p} + \frac{1}{q} = 1 \quad (\text{A.8})$$

gilt

$$E(|XY|) \leq \|X\|_p \|Y\|_q = (E(|X|^p))^{1/p} (E(|Y|^q))^{1/q} \quad (\text{A.9})$$

Beweis: siehe [Casella und Berger \(2002, Section 4.7.1\)](#)

- Anwendung der Hölder-Ungleichung (Davidson, 2000, Equation (B.5.14)):
Für beliebige Paare reeller Zahlen (a_j, b_j) , $j = 1, \dots, n$ und $1/p + 1/q = 1$ gilt:

$$\left| \sum_{j=1}^n a_j b_j \right| \leq \left(\sum_{j=1}^n |a_j|^p \right)^{1/p} \left(\sum_{j=1}^n |b_j|^q \right)^{1/q} \quad (\text{A.10})$$

- **Cauchy-Schwarz-Ungleichung** (Davidson, 2000, Equation (B.5.10)):
Hölder-Ungleichung für $p = q = 2$:

$$E(XY) \leq \|X\|_2 \|Y\|_2 = (E(|X|^2))^{1/2} (E(|Y|^2))^{1/2} \quad (\text{A.11})$$

- **Ljapunow-Ungleichung (Ljapunow inequality)** (Davidson, 2000, Equation (B.4.16)):

$$\|X\|_p \leq \|X\|_q, \quad \text{für } 0 < p \leq q \quad (\text{A.12})$$

Beweis: mit Hölder-Ungleichung, siehe Casella und Berger (2002, Section 4.7.1).

- **Minkowski-Ungleichung** (Davidson, 2000, Equation (B.4.17)):

$$\|X + Y\|_p \leq \|X\|_p + \|Y\|_p. \quad (\text{A.13})$$

Beweis siehe [Casella und Berger \(2002, Section 4.7.1\)](#).

Endliches oder unendliches Iterieren ergibt ([Davidson, 2000, Equation \(B.4.19\)](#)):

$$\left\| \sum_{i=1} X_i \right\|_p \leq \sum_{i=1} \|X_i\|_p \quad (\text{A.14})$$

- **Dreiecksungleichung:**

Minkowski-Ungleichung für $p = 1$.

- **Jensen-Ungleichung**

Es sei $g(\cdot)$ eine konvexe Funktion und sowohl X als auch $g(X)$ integrierbar bezüglich der zugrundeliegenden Wahrscheinlichkeitsfunktion. Dann gilt

$$E[g(X)] \geq g(E(X)). \quad (\text{A.15})$$

Für eine konkave Funktion gilt die Umkehrung ([Davidson, 1994, Section 9.3, Theorem 9.12](#)). Beweis siehe [Casella und Berger \(2002, Section 4.7.2\)](#)

- **Modulus-Ungleichung** ([Davidson, 2000, Equation \(B.4.21\)](#)):

ergibt sich durch Jensen-Ungleichung, da $|\cdot|$ eine konvexe Funktion ist:

$$E(|X|) \geq |E(X)| \quad (\text{A.16})$$

Literaturverzeichnis

- Andrews, D. (1991), “Heteroscedasticity and autocorrelation consistent covariance matrix estimation,” *Econometrica*, 59, 913–931. 282
- Brockwell, P. J., und R. A. Davis (1991), *Time Series: Theory and Methods*, 2nd edn., Springer, New York, NY. 297, 300, 301
- Casella, G., und R. L. Berger (2002), *Statistical Inference*, 2nd edn., Duxbury - Thomson. 302, 303, 304
- Chakroun, M. (2009), “Health care expenditure and GDP: An international panel smooth transition approach,” *Munich Personal RePEc Archive*, 17559, 1–20, URL <http://mpa.ub.uni-muenchen.de/17559/>. 20

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press. 296, 298, 299, 300, 302, 304

Davidson, J. (2000), *Econometric Theory*, Blackwell Publishers. x, xi, 3, 5, 31, 55, 60, 105, 107, 111, 115, 118, 120, 124, 207, 210, 219, 221, 302, 303, 304

Davidson, R., und J. MacKinnon (1993), *Estimation and Inference in Econometrics.*, Oxford University Press, URL <http://www.oup.com/uk/catalogue/?ci=9780195060119>. 46, 83

Davidson, R., und J. G. MacKinnon (2004), *Econometric Theory and Methods*, Oxford University Press, Oxford. x, 3, 8, 10, 27, 28, 30, 46, 52, 58, 59, 81, 82, 83, 105, 111, 114, 118, 137, 138, 141, 151, 168, 171, 172, 198, 216, 221, 236, 237, 238, 239, 243, 254, 258, 260, 263, 264, 273, 282

Donauer, S., F. Heinen, und P. Sibbertsen (2010), “Identification problems in ESTAR models and a new model,” Tech. rep., Leibniz Universität Hannover Discussion Paper No. 444. 23

Gentle, J. E. (2007), *Matrix Algebra. Theory, Computations, and Applications in Statistics*, Springer Texts in Statistics, Springer, doi:10.1007/978-0-387-70873-7,

- URL <http://www.springerlink.com/content/x4rj03/>. 156
- Gourieroux, C., und A. Monfort (1995), *Statistics and Econometric Models*, vol. 2, Cambridge University Press. 129
- Granger, C. W., und T. Teräsvirta (1993), *Modelling Nonlinear Economic Relationships*, Advanced Texts in Econometrics, Oxford University Press, Oxford. 83
- Greene, W. (2018), *Econometric Analysis*, 8th edn., Pearson. x
- Guggenberger, P. (2008), “The Impact of a Hausman Pretest on the Size of a Hypothesis Test: the Panel Data Case,” , 1–24. 198
- Hassler, U. (2007), *Stochastische Integration und Zeitreihenmodellierung*, Springer, Berlin, Heidelberg. xi
- Haupt, H., J. Schnurbus, und R. Tschernig (2010), “On Nonparametric Estimation of a Hedonic Price Function,” *Journal of Applied Econometrics*, 25, 894–901, doi: 10.1002/jae.1186. 19
- Hayashi, F. (2000), *Econometrics*, Princeton University Press, Princeton, NJ [u.a.]. x, 210, 282

- Kirchgässner, G., und J. Wolters (2008), *Introduction To Modern Time Series Analysis*, Springer, Berlin, [u.a.]. xi
- Kirchgässner, G., J. Wolters, und U. Hassler (2013), *Introduction To Modern Time Series Analysis*, 2nd edn., Springer, Berlin, [u.a.]. xi
- Kleiber, C., und A. Zeileis (2008), *Applied Econometrics with R*, Springer, doi:10.1007/978-0-387-77318-6. x
- Lütkepohl, H., und M. Kräzig (eds.) (2004), *Applied Time Series Econometrics*, Cambridge University Press, Cambridge. xi
- Lütkepohl, H. und Kräzig, M. (ed.) (2008), *Applied Time Series Econometrics*, Cambridge University Press. xi
- Luukkonen, R., P. Saikkonen, und T. Teräsvirta (1988), “Testing linearity in univariate time series models,” *Scandinavian Journal of Statistics*, 15, 161–175. 92
- Lütkepohl, H. (1996), *Handbook of Matrices*, John Wiley & Sons, Chichester. 145, 156
- Lütkepohl, H. (2005), *New Introduction to Multiple Time Series Analysis*, Springer.

xi

- McMillan, D. G., und M. E. Wohar (2009), “Stock return predictability and dividend-price ratio: A nonlinear approach,” *International Journal of Finance and Economics*, 15, 351–365. 21
- Nelson, C., und A. F. Siegel (1987), “Parsimonious Modeling of Yield Curves,” *The Journal of Business*, 60, 473–489. 20
- Raaijmakers, J. (1987), “Statistical Analysis of the Michaelis-Menten Equation,” *Biometrics*, 43(4), 793–803. 24
- Sensier, M., D. R. Osborn, und N. Öcal (2002), “Asymmetric Interest Rate Effects for the UK Real Economy,” Tech. rep., CGBCR Discussion Paper Series, URL <http://www.ses.man.ac.uk/cgbcr/discussi.htm>. 21
- Teräsvirta, T. (1998), “Modelling Economic Relationships with Smooth Transition Regressions,” in: A. Ullah und D. Giles (eds.), *Handbook of Applied Economic Statistics*, Dekker, New York, pp. 507–552. 83
- Teräsvirta, T., und A.-C. Eliasson (2001), “Non-linear error correction and the UK

demand for broad money, 1878-1993,” *Journal of Applied Econometrics*, 16, 277–288. 22

Teräsvirta, T., D. Tjøstheim, und C. W. J. Granger (2010), *Modelling nonlinear economic time series*, Oxford University Press. 83

Tsionas, E., und D. K. Christopoulos (2003), “Maastricht convergence and real convergence: European evidence from threshold and smooth transition regression models,” *Journal of Policy Modeling*, 25, 43–52. 22

Wooldridge, J. M. (2009), *Introductory Econometrics. A Modern Approach*, 4th edn., Thomson South-Western, Mason. 8

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press. x