Folien für

Weiterführende Fragen der Ökonometrie (Bachelor)

Rolf Tschernig

Universität Regensburg

10. Oktober 2024 [1]

[1] Diese Version entspricht bis auf einige Korrekturen dem Foliensatz für Weiterführende Fragen der Ökonometrie (Bachelor) vom 12. Oktober 2021, welcher wiederum bis auf einige Korrekturen, dem Ersetzen von EViews-Output durch R-Output und dem Aktualisieren von Links den Folien für Ökonometrie III vom 16. Oktober 2013 entspricht. Der zuletzt genannte Foliensatz entspricht bis auf die Beseitigung einiger Tippfehler der Version vom 8. November 2012. Kathrin Kagerer, Joachim Schnurbus und Florian Brezina danke ich sehr herzlich für ihre Unterstützung und Korrekturen. Peter Zorn danke ich für die Aktualisierung der LaTeX-Pakete. Ich bitte etwaige Fehler an rolf.tschernig@wiwi.uni-regensburg.de zu schicken.

Bitte zitieren als: Rolf Tschernig, Folien für Weiterführende Fragen der Ökonometrie, Universität Regensburg, 10. Oktober 2024. Downloaded am [Tag Monat Jahr].

Inhaltsverzeichnis

Organisation

Kontakt

Prof. Dr. Rolf Tschernig
Gebäude RW(L), 5. Stock, Raum 514
Universitätsstr. 31, 93040 Regensburg
Tel. (+49) 941/943 2737, Fax (+49) 941/943 4917
Email: rolf.tschernig@wiwi.uni-regensburg.de
https://www.uni-regensburg.de/wirtschaftswissenschaften/vwl-tschernig/index.html

Zeiten, Räume und Kursleiter

siehe Kurshomepage

https://www.uni-regensburg.de/wirtschaftswissenschaften/vwl-tschernig/lehre/bachelor/weiterfuehrende-fragen-der-oekonometrie/index.html

Prüfungsmodalitäten und Notengebung

Der Kurs Weiterführende Fragen der Ökonometrie ist Bestandteil u. a. der Schwerpunktmodulgruppe **Empirische Wirtschaftsforschung**. Weitere Informationen enthält der Modulkatalog. Aktuelle Regelungen finden Sie auf der Kurshomepage

https://www.uni-regensburg.de/wirtschaftswissenschaften/vwl-tschernig/lehre/bachelor/weiterfuehrende-fragen-der-oekonometrie/index.html

Voraussetzung: Einführung in die Ökonometrie

Pflichtliteratur

Wooldridge, J.M. (2009). *Introductory Econometrics. A Modern Approach*, 4. Auflage, Thomson South-Western. Oder neuere Auflage (Chapter 13-17; Appendix B, C)

Angrist, J. & Pischke, J. (2009). *Mostly Harmless Econometrics. An Empiricist's Companion*, Princeton University Press.
(Kapitel 1,2,3 bis 3.2, 5.1 und 5.2).

Ergänzungsliteratur

Angrist, J. & Pischke, J. (2014). *Mastering 'Metrics. The Path from Cause to Effect*, Princeton University Press.

Weitere Literatur wird im Verlauf bekanntgegeben.

Software

Alle empirischen Beispiele sind in R (https://www.r-project.org) durchgeführt. Appendix 8.3 enthält das R-Programm, mit dem alle Outputs in diesen Folien erzeugt wurden. Diese R-Outputs sind teils editiert, um Platz zu sparen.

In der früheren Version der Folien von 2013 (Ökonometrie III) sind die empirischen Beispiele in EViews ausgeführt.

1 Einleitung

1.1 Wo stehen wir?

... wurde in **Zeitreihenökonometrie** behandelt:

- Schätz- und Testeigenschaften des OLS, GLS, FGLS-Schätzers, wenn die *Störterme nicht normalverteilt* sind
- Ökonometrie für Zeitreihendaten:
  - Modelle mit streng exogenen, möglicherweise verzögerten Regressoren, z.B.:

$$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t-1,1} + \beta_3 x_{t2} + u_t$$

  - *Autoregressive Modelle* der Ordnung p:

$$y_t = \nu + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + u_t$$

- Modellierung von Trends und Saisonkomponenten in Zeitreihendaten: deterministische und stochastische Trends
- Dynamische Regressionsmodelle: Modelle mit verzögerten endogenen Regressoren, z.B.: Phillipskurve
- Regressionsmodelle mit autokorrelierten und heteroskedastischen Fehlern
- Einheitswurzeltests
- Kointegration, Fehlerkorrekturmodelle, Kointegrationstests
- Prognosen

... und wird in **Quantitative Wirtschaftsforschung** behandelt.

- Vektorautoregressive Modelle und Vektorfehlerkorrekturmodelle:
  zur dynamischen Modellierung von Interaktionen zwischen ökonomischen Variablen.

... und Themen für **Weiterführende Fragen der Ökonometrie**?

1.2 Gliederung des Kurses

- Einleitung und Überblick
- Kausalitätsanalyse
- Gepoolte Querschnittsanalyse
- Paneldatenmethoden
- Instrumentvariablenmethoden und zweistufige LS-Schätzer
- Simultane Gleichungsmodelle
- Modelle für abhängige Variablen mit Beschränkung (Limited Dependent Variable Methods)
- Selektionsverzerrungen

1.3 Beispiele für Verwendung gepoolter Querschnittsdaten

Beispiel: Einflussgrößen der Geburtenrate

Frage: Hat das Ausbildungsniveau, gegeben ein Signifikanzniveau von 5%, einen Einfluss auf die Geburtenrate?

Daten: `fertil1` aus Wooldridge (2009, 2006)

Querschnittsregression auf Basis einer Erhebung von 1972

$$kids_i = \beta_0 + \beta_1 educ_i + \beta_2 age_i + \cdots + \beta_{10} town_i + \beta_{11} smcity_i + u_i,$$
$$i = 1, 2, \ldots, 156$$

```
Call:
lm(formula = kids ~ educ + age + I(age^2) + black + east + northcen +
    west + farm + othrural + town + smcity, data = fertil1, subset = year == 72)
Residuals:
    Min      1Q  Median      3Q     Max
-3.7067 -1.1763 -0.1576  1.0429  4.1648

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -7.342532  10.351665  -0.709   0.4793
educ        -0.070776   0.064525  -1.097   0.2745
age          0.530593   0.465099   1.141   0.2558
I(age^2)    -0.006549   0.005220  -1.255   0.2117
black        0.861039   0.573610   1.501   0.1355
east         0.677352   0.411884   1.645   0.1023
northcen     0.608414   0.417896   1.456   0.1476
west         0.570929   0.509231   1.121   0.2641
farm         0.264472   0.469443   0.563   0.5741
othrural     0.172848   0.572658   0.302   0.7632
town         0.304561   0.370206   0.823   0.4121
smcity       1.001307   0.496056   2.019   0.0454 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.798 on 144 degrees of freedom
Multiple R-squared:  0.1008,Adjusted R-squared:  0.03209
F-statistic: 1.467 on 11 and 144 DF,  p-value: 0.15
aic      hq       sc
4.085411 4.180697 4.320016
```

Antwort: Nein, da Ausbildung ($educ$) statistisch nicht signifikant ist.

Ergebnisse einer neuen Querschnittserhebung in 1974:

```
Call: lm(formula = kids ~ educ + age + I(age^2) + black + east + northcen +
    west + farm + othrural + town + smcity, data = fertil1, subset = year == 74)
Residuals:
    Min      1Q  Median      3Q     Max
-3.5247 -1.0472  0.0015  0.8070  3.6178
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.874651   7.633861   0.246   0.8063
educ        -0.083295   0.048606  -1.714   0.0885 .
age          0.138907   0.346506   0.401   0.6890
I(age^2)    -0.001908   0.003867  -0.494   0.6223
black        1.157618   0.504793   2.293   0.0231 *
east        -0.106182   0.343484  -0.309   0.7576
northcen     0.402746   0.298356   1.350   0.1789
west         0.383199   0.415032   0.923   0.3572
farm        -0.293758   0.361660  -0.812   0.4178
othrural    -0.635563   0.421008  -1.510   0.1331
town        -0.146612   0.304687  -0.481   0.6310
smcity      -0.424830   0.427708  -0.993   0.3221
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.491 on 161 degrees of freedom
Multiple R-squared:  0.0784,Adjusted R-squared:  0.01544
F-statistic: 1.245 on 11 and 161 DF,  p-value: 0.2615
aic      hq       sc
3.703984 3.792719 3.922709
```

Ergebnis bleibt unverändert. Liegt die Insignifikanz von β_1 lediglich an einer ineffizienten Ausnutzung der verfügbaren Daten?

Effizienzgewinn durch gemeinsame Verwendung der Daten von 1972 und 1974? $\longrightarrow$ gepoolte Querschnittsdaten, siehe Kapitel 3.

Beispiel: Auswirkung des Baus einer Müllverbrennungsanlage auf Immobilienpreise in der Umgebung (Example 13.3 in Wooldridge (2009)):

Frage: Hat der Bau einer Müllverbrennungsanlage einen Effekt auf die Preise von Wohngebäuden in der Umgebung der Anlage, gegeben ein Signifikanzniveau von 5%?

Daten: `kielmc` aus Wooldridge (2009, 2006)
`RPRICE` gibt die realen Preise an. Die Dummyvariable `NEARINC` nimmt den Wert 1 an, wenn die Immobilie in der Nähe der Müllverbrennungsanlage steht, ansonsten den Wert 0.

Regressionsergebnisse einer Querschnittserhebung in 1981 (nach Bau der Anlage) (Vgl. Gleichung (13.4) in Wooldridge (2009)):

```
Call:
lm(formula = rprice ~ nearinc, data = kielmc, subset = year == 1981)

Residuals:
   Min     1Q Median     3Q    Max
-60678 -19832  -2997  21139 136754

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   101308       3093  32.754  < 2e-16 ***
nearinc       -30688       5828  -5.266 5.14e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31240 on 140 degrees of freedom
Multiple R-squared:  0.1653,Adjusted R-squared:  0.1594
F-statistic: 27.73 on 1 and 140 DF,  p-value: 5.139e-07

aic      hq       sc
23.55065 23.56756 23.59228
```

Die Wirkung des Müllverbrennungsanlage ist demnach negativ und signifikant. Doch ist die Müllverbrennungsanlage *ursächlich* für den Unterschied in den Immobilienpreisen oder lässt sich dieser durch andere Faktoren erklären, d.h. ist unser Schätzer verzerrt?
$\longrightarrow$ Kausalität und Evaluationsstudien, siehe Kapitel 2 und folgende.

Für die Analyse von kausalen Zusammenhängen mit Hilfe von **natürlichen Experimenten** haben am 11. Oktober 2021 **David Card**, **Joshua Angrist** und **Guido Imbens** den **Nobelpreis für Wirtschaftswissenschaften** bekommen. Weitere Informationen zu den Preisträgern finden sich auf den Webseiten des Nobelpreiskomitees. Und die Grundlagen von natürlichen Experimenten werden in Kapitel 2 behandelt.

2 Grundlagen zu Kausalität und Evaluationsstudien

Hinweis: Dieser Abschnitt orientiert sich in weiten Teilen an Angrist und Pischke (2009) und unterscheidet daher in der Notation zwischen (groß geschriebenen) Zufallsvariablen, z. B. Y_i, und deren (klein geschriebenen) Ausprägungen, z.B. y_i.

Sehr häufig ist von Interesse, **ob und welchen Effekt** eine wirtschaftspolitische oder betriebliche oder sonstige **Maßnahme** (**treatment**) hat. Hat eine Maßnahme eine Wirkung auf bestimmte Größen, ist sie **kausal** für diese.

Beispiel (intensiv untersucht): Ob und welchen Einfluss die Klassengröße auf den Lernerfolg und später auf den beruflichen Erfolg hat.

Es geht also um die **Evaluation einer Maßnahme**.

Fortsetzung Beispiel Klassengröße: Wie wirkt die Klassengröße auf Individuum i? Ob die Maßnahme durchgeführt wird, gibt eine binäre Variable

C an:

$$C_i = \begin{cases} 1 & \text{Individuum } i \text{ nimmt an Maßnahme teil (in kleiner Klasse)}, \\ 0 & \text{Individuum } i \text{ nimmt an Maßnahme nicht teil (in großer Klasse)}, \end{cases}$$

Allgemeine Unterscheidung:

- **Maßnahmengruppe** oder **Teilnehmergruppe** (**treatment group**): enthält alle Individuen, die an der Maßnahme teilnehmen (ob freiwillig oder unfreiwillig ist egal), $C_i = 1$. In der Medizin wird von der Experimentalgruppe gesprochen.
- **Kontrollgruppe** (**control group**): enthält alle Individuen, die **nicht** an der Maßnahme teilnehmen, $C_i = 0$.

Potenzielles und beobachtbares Ergebnis

- Das **potentielle Ergebnis** (**potential outcome**) für Individuum i wird durch zwei Zufallsvariablen dargestellt:

  $$Y_{1i} \quad \text{Ergebnis, falls Ind. } i \text{ von Maßnahme betroffen, } C_i = 1, \tag{2.1a}$$

  $$Y_{0i} \quad \text{Ergebnis, falls Ind. } i \text{ nicht betroffen, } C_i = 0 \tag{2.1b}$$

  wobei **nur eines der beiden Ergebnisse beobachtbar** ist. Denn, wenn Individuum i an der Maßnahme teilnimmt ($C_i = 1$), ist Y_{0i} unbeobachtbar und, wenn Individuum i nicht teilnimmt ($C_i = 0$), ist Y_{1i} unbeobachtbar.

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

- Das **beobachtbare Ergebnis** für Individuum i ist

  $$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})C_i \tag{2.2}$$

Evaluationsgrößen

- **Individueller Erfolg der Maßnahme**:
  Der individuelle Erfolg der Maßnahme (individual treatment effect) für Individuum i ist

$$Y_{1i} - Y_{0i} \tag{2.3}$$

  und ist **immer unbeobachtbar**, da es immer eine kontrafaktische Größe enthält.

  Beispiel: Der Erfolg eines Schülers, der in der kleinen Klasse ist, $C_i = 1$, ist für den hypothetischen Fall, dass er in der großen Klasse wäre, nicht beobachtbar.

- **Durchschnittliche Erfolgsgrößen einer Maßnahme**:
  - **Durchschnittlicher Erfolg der Maßnahme** (**average treatment effect**, **ATE**) für eine Population
$$E[Y_{1i} - Y_{0i}] \tag{2.4}$$
  - **Durchschnittlicher Erfolg der Maßnahme für die Maßnahmeteilnehmer** (**average treatment on the treated**, **ATET**) für eine Population
$$E[Y_{1i} - Y_{0i} | C_i = 1] \tag{2.5}$$

Unter bestimmten Voraussetzungen können diese Größen, in der Literatur häufig als Parameter bezeichnet, aus Stichprobeninformation geschätzt werden.

Das Grundproblem: Selektionsverzerrungen

- Auf Basis einer Stichprobe, die sowohl Individuen enthält, die an der Maßnahme teilgenommen haben, $C_i = 1$, als auch Individuen, die nicht daran teilgenommen haben, können nur Erwartungswerte, die das beobachtbare Ergebnis enthalten, geschätzt werden:
  - **Durchschnittliches Ergebnis der Maßnahmengruppe**

$$E[Y_i|C_i = 1] = E[Y_{1i}|C_i = 1] \tag{2.6a}$$

  - **Durchschnittliches Ergebnis für die Kontrollgruppe**

$$E[Y_i|C_i = 0] = E[Y_{0i}|C_i = 0] \tag{2.6b}$$

- Damit lässt sich **für die Population** ein **direkter Vergleich der durchschnittlichen Ergebnisse** für die Maßnahmengruppe und die Kontrollgruppe durchführen (beachte, dass $E[Y_{1i}|C_i = 1] - E[Y_{0i}|C_i = 1] = E[Y_{1i} - Y_{0i}|C_i = 1]$):

$$\underbrace{E[Y_i|C_i = 1] - E[Y_i|C_i = 0]}_{\text{Beobachtbarer durchschnittlicher Unterschied}} = \underbrace{E[Y_{1i} - Y_{0i}|C_i = 1]}_{\text{Durch. Erfolg d. Maßnahme f. d. Maßnahmeteilnehmer (2.5)}} + \underbrace{E[Y_{0i}|C_i = 1] - E[Y_{0i}|C_i = 0]}_{\text{Selektionsverzerrung}} \tag{2.7}$$

- Der Term

$$E[Y_{0i}|C_i = 1] - E[Y_{0i}|C_i = 0] \tag{2.8}$$

verursacht eine **Selektionsverzerrung**. Sie misst die Differenz des durchschnittlichen potentiellen Ergebnisses bei Nichtteilnahme zwischen den Maßnahmeteilnehmern und der Kontrollgruppe.

- **Beispiel**: Vergleicht man die durchschnittliche Leistung aller Schüler in den kleinen Klassen und die durchschnittliche Leistung aller Schüler in den großen Klassen kommt es vermutlich zu einer Selektionsverzerrung, wenn beispielsweise Schüler in der kleinen Klasse deshalb in der kleinen Klasse sind, weil sie Lernstörungen haben. Dann ist vermutlich deren durchschnittliches Ergebnis für den Fall, dass sie in der großen Klasse gewesen wären, geringer als für Schüler, die in der großen Klasse sind, weil sie keine Lernstörungen haben. Es kommt zu einer negativen Selektionsverzerrung. Wird der beobachtbare durchschnittliche Erfolg $E[Y_i|C_i = 1] - E[Y_i|C_i = 0]$ positiv geschätzt, wird der durchschnittliche Erfolg für die Maßnahmeteilnehmer (2.5) unterschätzt.

2.1 Kontrollierte Zufallsexperimente

Wie lassen sich Selektionsverzerrungen vermeiden?

- **Kontrolliertes Zufallsexperiment** (**controlled randomized experiment**):
  Das zentrale Merkmal eines kontrollierten Zufallsexperiments ist, dass die Maßnahmen C_i den Individuen **zufällig** zugewiesen werden:
  - **(E1) Kontrolliertes Zufallsexperiment**

$$f(y_{0i}, y_{1i}, C_i) = f(y_{0i}, y_{1i})\, f(C_i), \tag{2.9a}$$

äquivalent dazu

$$f(y_{0i}, y_{1i}|C_i) = f(y_{0i}, y_{1i}). \tag{2.9b}$$

Die gemeinsame Dichte der potentiellen Ergebnisse und der Zuweisung der Maßnahme lässt sich aufgrund der stochastischen Unabhängigkeit als Produkt schreiben oder äquivalent dazu: die Zuweisung der Maßnahme C_i ist als Bedingung nicht informativ, so dass die bedingte Dichte der unbedingten Dichte entspricht.

Aus der Darstellung (2.9b) der Annahme **(E1)** “Kontrolliertes Zufallsexperiment” folgt direkt (vgl. Angrist und Pischke (2009, Sec. 2.2))

$$E[Y_{0i}|C_i = 0] = E[Y_{0i}|C_i = 1] = E[Y_{0i}], \tag{2.10a}$$

$$E[Y_{1i}|C_i = 0] = E[Y_{1i}|C_i = 1] = E[Y_{1i}]. \tag{2.10b}$$

Damit **verschwindet die Selektionsverzerrung** in (2.7), da

$$E[Y_{0i}|C_i = 1] - E[Y_{0i}|C_i = 0] = E[Y_{0i}|C_i = 0] - E[Y_{0i}|C_i = 0] = 0.$$

Bei einem **kontrollierten Zufallsexperiment** gilt deshalb

$$\underbrace{E[Y_i|C_i = 1] - E[Y_i|C_i = 0]}_{\text{Beobachtbarer durchschn. Unterschied}} = \underbrace{E[Y_{1i} - Y_{0i}|C_i = 1]}_{\text{Durch. Erf. d. Maßn. Maßn. teiln. (2.5)}} \tag{2.11a}$$

$$= \underbrace{E[Y_{1i} - Y_{0i}]}_{\text{Durch. Erfolg d. Maßnahme (2.4)}} \tag{2.11b}$$

Beispiel: Die Schüler (und Lehrer, etc.) einer Schule werden am Anfang des Schuljahres **zufällig** auf kleine und große Klassen aufgeteilt. Auf diese Weise wird vermieden, dass SchülerInnen mit bestimmten Eigenschaften vornehmlich in einer Klassengröße zu finden sind. Am Ende werden dann die durchschnittlichen Noten aus den kleinen Klassen mit denen der großen Klassen verglichen. Mit auf diese Weise erzeugten Stichproben können (2.4) oder (2.5) geschätzt werden.

- **Quasi-Experimente** oder **natürliche Experimente**:
  In vielen Fällen ist kein kontrolliertes Zufallsexperiment durchführbar. Man muss dann nach Quasi-Experimenten oder natürlichen Experimenten suchen, die Stichproben liefern, welche ein ideales kontrolliertes Zufallsexperiment möglichst gut approximieren. Dann sind allerdings zusätzliche Annahmen zu treffen, die üblicherweise **Identifikationsannahmen** genannt werden. Siehe hierzu Abschnitt 2.2.

Darstellung in Regressionsform

- Addieren und Subtrahieren von $E[Y_{0i}]$ zum beobachtbaren Ergebnis (2.2):

$$Y_i = \underbrace{E[Y_{0i}]}_{\alpha} + \underbrace{(Y_{1i} - Y_{0i})}_{\rho_i} C_i + \underbrace{Y_{0i} - E[Y_{0i}]}_{\eta_i} \tag{2.12}$$

- Unter der **Annahme** (zur Vereinfachung)
  - **(E2) Alle individuellen Erfolge der Maßnahme gleich**

$$\rho_i \equiv Y_{1i} - Y_{0i} = \rho \tag{2.13}$$

ergibt sich die einfache lineare Regressionsdarstellung

$$Y_i = \alpha + \rho C_i + \eta_i \tag{2.14a}$$

$$E[Y_i|C_i] = \alpha + \rho C_i + E[\eta_i|C_i] \tag{2.14b}$$

wobei $E[\eta_i|C_i]$ ungleich Null sein kann.

- **ATE** (2.4) und **ATET** (2.5) sind unter unter Annahme **(E2)** gleich.
- $E[\eta_i|C_i] \neq 0$ $\Leftrightarrow$ Annahme **MLR.4 verletzt**
  $\Leftrightarrow$ **Selektionsverzerrungen** treten auf.
  $E[\eta_i|C_i] = 0$ $\Leftarrow$ **MLR.4** erfüllt $\Leftarrow$ **(E1)** erfüllt.

  Also unter Annahmen **(E1)** und **(E2)** gilt:

$$E[Y_i|C_i] = \alpha + \rho C_i \tag{2.15}$$

- Um in einer KQ-Schätzung den **Standardfehler** für das geschätzte ρ zu **reduzieren**, ist es sinnvoll, in (2.14a) Regressoren $\mathbf{x}_i$ aufzunehmen, die einen Einfluss auf Y_i haben. Dann schätzt man

$$Y_i = \alpha + \rho C_i + \mathbf{x}_i'\gamma + \varepsilon_i \tag{2.16}$$

  und $Var(\varepsilon_i) \leq Var(\eta_i)$.

- **Selbst wenn man ein kontrolliertes Zufallsexperiment durchführen kann**, kann die **externe Validität** der Resultate verletzt sein, da eine Übertragung der Ergebnisse von der betrachteten Population auf andere Populationen nicht zulässig ist.

- **Beobachtbarer durchschnittlicher Unterschied** (2.7):
  - Annahme **(E2)** $\Rightarrow E\left[Y_{1i} - Y_{0i} | C_i = 1\right] = E\left[\rho | C_i = 1\right] = \rho$.
  - (2.7) wird zu

$$\underbrace{E[Y_i|C_i = 1] - E[Y_i|C_i = 0]}_{\text{Beobachtbarer durchschnittlicher Unterschied}} = \underbrace{\rho}_{\text{Durch. Erfolg d. Maßnahme (2.4)}} + \underbrace{E[\eta_i|C_i = 1] - E[\eta_i|C_i = 0]}_{\text{Selektionsverzerrung}} \qquad (2.17)$$

  - Annahme **(E1)** eliminiert die Selektionsverzerrung, siehe (2.11).
  - Annahme **(E1)** und Annahme **(E2)** implizieren

$$\underbrace{E[Y_i|C_i = 1] - E[Y_i|C_i = 0]}_{\text{Beobachtbarer durchschnittlicher Unterschied}} = \underbrace{\rho}_{\text{Durch. Erfolg d. Maßnahme (2.4)}} \qquad (2.18)$$

2.2 Evaluation ohne Zufallsexperimente

- Es wird weiterhin Annahme **(E2)** vorausgesetzt.
- Annahme **(E1)** ist verletzt, wenn die Daten nicht aus einem kontrollierten Zufallsexperiment stammen.

Die bedingte Unabhängigkeitsannahme

- Selektionsverzerrungen lassen sich für **Querschnittsstichproben**, die nicht aus kontrollierten Zufallsexperimenten stammen, vermeiden, wenn unterstellt werden kann, dass die Zuweisung der Individuen zur Maßnahme C_i zufällig erfolgt, sobald für jedes Individuum bestimmte beobachtbare Faktoren $\mathbf{x}$ mit berücksichtigt werden.

Formal lautet diese **Identifikationsannahme**:

– **(E3) Bedingte Unabhängigkeitsannahme (conditional independence assumption (CIA)**, **selection on observables)**

$$f(y_{1i}, y_{0i}, C_i|\mathbf{x}_i) = f(y_{1i}, y_{0i}|\mathbf{x}_i) \cdot f(C_i|\mathbf{x}_i) \quad (2.19a)$$

äquivalent dazu $$f(y_{1i}, y_{0i}|C_i, \mathbf{x}_i) = f(y_{1i}, y_{0i}|\mathbf{x}_i) \quad (2.19b)$$

Kurznotation: $$\{Y_{1i}, Y_{0i}\} \perp\!\!\!\perp C_i|\mathbf{x}_i,$$

wobei die relevanten beobachtbaren Faktoren im Vektor $\mathbf{x}_i$ zusammengefasst werden.

– Die Stichprobe ist dann unter Kenntnis der Bedingungen $\mathbf{x}_i$, $i = 1, \ldots, n$ "so gut" wie von einem kontrollierten Zufallsexperiment, denn aus der Darstellung (2.19b) der Annahme **(E3)** "Bedingte Unabhängigkeitsannahme" folgt in Analogie zu (2.10):

$$E[Y_{0i}|C_i = 1, \mathbf{x}_i] = E[Y_{0i}|C_i = 0, \mathbf{x}_i] = E[Y_{0i}|\mathbf{x}_i],$$
$$E[Y_{1i}|C_i = 1, \mathbf{x}_i] = E[Y_{1i}|C_i = 0, \mathbf{x}_i] = E[Y_{1i}|\mathbf{x}_i].$$

- Die Annahme lässt sich auch auf diskrete Variablen anwenden, die mehrere Werte annehmen können. Vgl. Angrist und Pischke (2009, Section 3.2.1).

- **Beoachtbarer durchschnittlicher Unterschied** (2.7):
  - Annahme **(E3)** und Annahme **(E2)** implizieren

$$\underbrace{E[Y_i|C_i = 1, \mathbf{x}_i] - E[Y_i|C_i = 0, \mathbf{x}_i]}_{\text{Beobachtbarer durchschnittlicher Unterschied}} = \underbrace{E[Y_{1i}|\mathbf{x}_i] - E[Y_{0i}|\mathbf{x}_i]}_{\text{Durch. Erfolg d. Maßnahme (2.4)}}$$
$$= E[Y_{1i} - Y_{0i}|\mathbf{x}_i] = \rho. \tag{2.20}$$

  - Im Gegensatz zu (2.17) unter Gültigkeit von Annahme **(E1)** muss der beobachtbare durchschnittliche Erfolg auf $\mathbf{x}_i$ **bedingt** sein, damit dieser ρ entspricht.

- Für die **Regressionsdarstellung** (2.14a) ergibt sich

$$Y_i = \alpha + \rho C_i + E[\eta_i|\mathbf{x}_i] + \underbrace{\eta_i - E[\eta_i|\mathbf{x}_i]}_{\nu_i} \tag{2.21a}$$

$$E[Y_i|C_i, \mathbf{x}_i] = \alpha + \rho C_i + E[\eta_i|\mathbf{x}_i], \tag{2.21b}$$

so dass nunmehr **MLR.4** erfüllt ist.

Allerdings muss noch $E[\eta_i|\mathbf{x}_i]$ spezifiziert werden!

Nimmt man an, dass $E[\eta_i|\mathbf{x}_i]$ linear in $\mathbf{x}_i$ ist, erhält man **Annahme**

– **(E4)**

$$E[\eta_i|\mathbf{x}_i] = \mathbf{x}_i'\boldsymbol{\gamma} \tag{2.22}$$

Damit ergibt sich das klassische **lineare multiple Regressionsmodell**

$$Y_i = \alpha + \rho C_i + \mathbf{x}_i'\boldsymbol{\gamma} + \nu_i \tag{2.23}$$

das unter den Annahmen **(E2)**, **(E3)** und **(E4)** eine **kausale Interpretation** für ρ hat!

- **Mögliche Probleme**:

1. In der Stichprobe liegen außer Y_i und C_i **keine** weiteren notwendigen **individuellen Informationen** vor. Mögliche Lösungen:

   – Einen möglichen Ausweg bietet der **Differenz-von-Differenzen-Ansatz**, wenn entsprechend gepoolte Querschnittsdaten vorliegen, siehe Abschnitt 3.3.

   – Verwendung des **Instrumentvariablen-Schätzers**, siehe Kap. 5.

2. Es **fehlen einzelne relevante individuelle Informationen**, z.B. die Fähigkeiten eines Individuums.

   Man sagt dann auch, dass Störfaktoren (**confounding factors**) vorliegen, die wiederum zur **Verletzung** von **MLR.4** führen, selbst wenn $\mathbf{x}_i$ berücksichtigt wird.

   Einen Ausweg bietet bei Vorliegen von Paneldaten unter bestimmten Voraussetzungen der **Fixed-Effekts-Schätzer**, siehe Kapitel 4.

3 Gepoolte Querschnittsdaten

3.1 Definitionen

- Siehe zu Datentypen Abschnitt 1.4 in Einführung in die Ökonometrie, August 2020.

- **Gepoolte Querschnittsdaten — Definition:**
  - Gepoolte Querschnittsdaten werden mehrmals in zeitlichen Abständen von einer Grundgesamtheit erhoben.
  - Dabei ist jede Erhebung eine Zufallsstichprobe, so dass sich die individuellen Einheiten im Allgemeinen von Periode zu Periode unterscheiden. Vgl. hierzu Paneldaten.
  - Die der Grundgesamtheit zugrundeliegende Verteilung kann sich über die Zeit verändern, z.B. Veränderung des Mittelwerts.

– Beispiel: 2 Perioden: 2005, 2006; 3 Beobachtungen

Jahr	Individuum	Konsum	Einkommen
2005	Frau Molier	30	50
2005	Herr Jung	20	60
2005	Herr Frank	50	60
2006	Herr Karl	20	70
2006	Frau Leng	40	50
2006	Herr Yang	50	50

– Allgemein: $y_{i_t,t}, \ldots, x_{i_t,t}, i_t = 1, \ldots, N_t, t = 1, 2, \ldots, T$

Jahr	Individuum	Konsum	Einkommen
$t = 2005$	$i_{2005} = 1$	$y_{i_{2005},2005} = y_{1,2005}$	$x_{1,2005}$
$t = 2005$	$i_{2005} = 2$	$y_{2,2005}$	$x_{2,2005}$
$t = 2005$	$i_{2005} = 3$	$y_{3,2005}$	$x_{3,2005}$
$t = 2006$	$i_{2006} = 1$	$y_{i_{2006},2006} = y_{1,2006}$	$x_{1,2006}$
$t = 2006$	$i_{2006} = 2$	$y_{2,2006}$	$x_{2,2006}$
$t = 2006$	$i_{2006} = 3$	$y_{3,2006}$	$x_{3,2006}$

- **Paneldaten — Definition**:
  - Einheiten (Individuen, Firmen, Länder, Kreise, etc.) bleiben über die Zeit gleich, d.h. $i_t = i$.
  - Die Beobachtungsperioden t werden auch als **Wellen** bezeichnet.
    - **'Balanced Panel'**: T ist für alle Einheiten $i = 1, \ldots, N$ gleich, d.h. es liegen für jede Periode für alle Einheiten Beobachtungen vor, bzw. können beobachtet werden.
    - **'Unbalanced Panel'**: T_i variiert mit i, d.h. es fehlen für einen oder mehrere Zeitpunkte Beobachtungen für eine oder mehrere Einheiten oder es gehen über die Zeit Einheiten verloren. Letzteres z.B. durch Tod oder Konkurs (attrition).

– Beispiel:

Jahr	Individuum	Konsum	Einkommen
2005	Frau Molier	30	50
2005	Herr Jung	20	60
2005	Herr Frank	50	60
2006	Frau Molier	20	70
2006	Herr Jung	40	50
2006	Herr Frank	50	50

– Allgemein: $y_{it}, x_{it}, i = 1, 2, \ldots, N, t = 1, 2, \ldots, T$

Jahr	Individuum	Konsum	Einkommen
$t = 2005$	$i = 1$	$y_{1,2005}$	$x_{1,2005}$
$t = 2005$	$i = 2$	$y_{2,2005}$	$x_{2,2005}$
$t = 2005$	$i = 3$	$y_{3,2005}$	$x_{3,2005}$
$t = 2006$	$i = 1$	$y_{1,2006}$	$x_{1,2006}$
$t = 2006$	$i = 2$	$y_{2,2006}$	$x_{2,2006}$
$t = 2006$	$i = 3$	$y_{3,2006}$	$x_{3,2006}$

3.2 Gepoolte Querschnittsregression

Unter welchen Voraussetzungen führt eine gepoolte Querschnittsanalyse zu sinnvollen Ergebnissen?

1. Die einzelnen Querschnittsregressionen weisen keine Verzerrung durch weggelassene Variablen auf (kein omitted variable bias).
2. Die einzelnen Querschnittsregressionen sind korrekt spezifiziert.
3. Die (unbekannten) wahren Parameter sind über die betrachteten Perioden konstant bzw. andernfalls flexibel modelliert.
4. Für jede Periode liegt eine Zufallsstichprobe vor.
5. Unterschiede in der Fehlerverteilung (z.B. Heteroskedastie), soweit nötig, modelliert.

- Fortsetzung Beispiel: Gepoolte Querschnittsanalyse für 1972 und 1974:

```
Call:
lm(formula = kids ~ educ + age + I(age^2) + black + east + northcen +
    west + farm + othrural + town + smcity + y74, data = fertil1,
    subset = year == 72 | year == 74)
Residuals:
    Min      1Q  Median      3Q     Max
-3.7754 -1.1562 -0.1145  1.0126  4.0715

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.478828   6.176011  -0.401   0.6884
educ        -0.068676   0.038723  -1.774   0.0771 .
age          0.311861   0.278294   1.121   0.2633
I(age^2)    -0.003951   0.003115  -1.268   0.2056
black        0.844100   0.372423   2.267   0.0241 *
east         0.312222   0.263421   1.185   0.2368
northcen     0.503366   0.247090   2.037   0.0425 *
west         0.485828   0.319397   1.521   0.1292
farm        -0.001483   0.287210  -0.005   0.9959
othrural    -0.214408   0.340825  -0.629   0.5297
town         0.105075   0.236747   0.444   0.6575
smcity       0.363103   0.322799   1.125   0.2615
y74          0.199983   0.184960   1.081   0.2804
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.639 on 316 degrees of freedom
Multiple R-squared:  0.06624,Adjusted R-squared:  0.03078
F-statistic: 1.868 on 12 and 316 DF,  p-value: 0.03757
aic     hq       sc
3.86481 3.924648 4.014806
```

Ergebnis: kaum eine Änderung

- Andere Verwendungsmöglichkeiten von gepoolten Daten?
  - **Testen auf Strukturveränderungen**, d.h. Veränderung eines, mehrerer, aller Parameter von erklärenden Variablen, siehe z.B. Example 13.2 in Wooldridge (2009, 2006) und **Chow-Test**.

    Beispiel für mögliche Veränderung des Einflusses von Ausbildung:

$$
\begin{aligned}
kids_{i_t,t} = \beta_0 + \delta_0 y74_t + \beta_1 educ_{i_t,t} + \delta_1 y74_t \cdot educ_{i_t,t} + \\
+ \beta_2 age_{i_t,t} + \cdots + \beta_{10} town_{i_t,t} + \beta_{11} smcity_{i_t,t} + u_{i_t,t}, \\
t = 1, 2 \quad i_1 = 1, 2, \ldots, 156 \quad i_2 = 1, 2, \ldots, 173
\end{aligned}
$$

$$
\begin{cases} H_0 : \delta_1 = 0 \\ H_1 : \delta_1 \neq 0 \end{cases} .
$$

  - **Testen auf Veränderung in Zeitdummies**, siehe z.B. Example 13.1 in Wooldridge (2009, 2006)

3.3 Evaluation mit Differenz-von-Differenzen-Ansatz (DD-Ansatz)

- **Voraussetzung** für den DD-Ansatz ist, dass für die Teilnehmergruppe und die Kontrollgruppe jeweils eine Querschnittsstichprobe **vor** und **nach** der Maßnahme vorliegt.

  Jede Beobachtung eines Individuums i ist entweder der Teilnehmergruppe $s = T$ oder der Kontrollgruppe zugeordnet $s = K$, sowie entweder Periode $t = V$ (vorher) oder $t = N$ (nachher). Deshalb wird jede Beobachtung zusätzlich mit Gruppen- und Zeitindex versehen: z.B. Y_{ist} oder Y_{0ist}.

- Aufgrund der Wahl der Gruppen und Zeitperioden gilt für die Teilnehmervariable $C_{ist} = C_{st}$:

$$C_{st} = \begin{cases} 1 & \text{falls } s = T, t = N, \\ 0 & \text{sonst.} \end{cases} \tag{3.1}$$

Damit ist der Wert von C_{ist} eindeutig durch s und t bestimmt, so dass

$$E[Y_{ist}|C_i, s, t] = E[Y_{ist}|s, t].$$

- Zusammen mit **(E2)** wird die Gleichung zum beobachtbaren Ergebnis (2.12) zu

$$Y_{ist} = \underbrace{E[Y_{0ist}]}_{\alpha_{st}} + \underbrace{(Y_{1ist} - Y_{0ist})}_{\rho} C_{st} + \underbrace{Y_{0ist} - E[Y_{0ist}]}_{\eta_{ist}}. \tag{3.2}$$

- Grundlage des DD-Ansatzes ist die **Identifikationsannahme für das potentielle Ergebnis der Nichtteilnehmer** Y_{0ist}:
  - **(E5) Linearität in Gruppen- und Zeitmittelwerten**

$$\alpha_{st} \equiv E[Y_{0ist}|s, t] = a_s + d_t, \quad s = T, K, \quad t = V, N. \tag{3.3}$$

Hieraus ergibt sich aus der Definition von $\eta_{ist} \equiv Y_{0ist} - E[Y_{0ist}]$, dass $E[\eta_{ist}|s, t] = 0$.

- Eine einfache Differenzenbildung für eine einzelne Periode reicht jetzt nicht mehr aus, denn

$$
\begin{align}
E[Y_{ist}|s = T, t = N] - E[Y_{ist}|s = K, t = N] &= a_T + \rho - a_K \tag{3.4}\\
E[Y_{ist}|s = T, t = V] - E[Y_{ist}|s = K, t = V] &= a_T - a_K. \tag{3.5}
\end{align}
$$

Die gruppenspezifischen Durchschnitte a_T und a_K können jedoch durch Differenzenbildung über die Zeit eliminiert werden:

$$
\begin{gathered}
E[Y_{ist}|s = T, t = N] - E[Y_{ist}|s = K, t = N] \\
-\left(E[Y_{ist}|s = T, t = V] - E[Y_{ist}|s = K, t = V]\right) = \rho,
\end{gathered} \tag{3.6}
$$

d. h. durch Differenzen von Differenzen. Deshalb heißt dieser Ansatz **Differenz-von-Differenzen-Ansatz** (**DD-Ansatz**).

- **Regressionsdarstellung des DD-Ansatzes**:
  - Definiere zwei Dummyvariablen D_s und D_t, die jeweils 1 sind, wenn Individuum i zu Gruppe $s = T$ gehört bzw. in der zweiten Periode $t = N$ beobachtet wird.

$$Y_{ist} = \alpha + \gamma D_s + \lambda D_t + \rho \underbrace{(D_s D_t)}_{=C_{st}} + \eta_{ist}. \qquad (3.7)$$

    Die Parameter α, γ, λ lassen sich aus verschiedenen Differenzenbildungen aus Annahme **(E5)** wie folgt erzeugen:
  - Die Mittelwerte zweier Gruppen lassen sich als Regression schreiben

$$t = V: \qquad Y_{isV} = (a_K + d_V)(1 - D_s) + (a_T + d_V) D_s + \eta_{isV}, \qquad (3.8)$$

$$t = N: \qquad Y_{isN} = (a_K + d_N)(1 - D_s) + (a_T + d_N + \rho) D_s + \eta_{isN}. \qquad (3.9)$$

    Zusammenfassen aller Terme mit Dummies

$$t = V: \qquad Y_{isV} = (a_K + d_V) + (a_T + d_V - (a_K + d_V)) D_s + \eta_{isV},$$

$$t = N: \qquad Y_{isN} = (a_K + d_N) + (a_T + d_N + \rho - (a_K + d_N)) D_s + \eta_{isN}.$$

und Kürzen von Zeitparametern

$$t = V: \qquad Y_{isV} = (a_K + d_V) + (a_T - a_K)\, D_s + \eta_{isV}, \tag{3.10}$$

$$t = N: \qquad Y_{isN} = (a_K + d_N) + (a_T + \rho - a_K) D_s + \eta_{isN}. \tag{3.11}$$

– Zusammenfassen beider Regressionen in einer Regression mit Dummyvariable D_t ergibt:

$$Y_{ist} = \underbrace{(a_K + d_V)}_{\alpha} + \underbrace{(a_T - a_K)}_{\gamma} D_s \tag{3.12}$$

$$+ \underbrace{\{(a_K + d_N) - (a_K + d_V)\}}_{d_N - d_V = \lambda} D_t \tag{3.13}$$

$$+ \underbrace{\{(a_T + d_N + \rho - (a_K + d_N)) - (a_T - a_K)\}}_{\rho} D_s D_t + \eta_{ist}. \tag{3.14}$$

- **Bemerkungen**:
  - **Vernachlässigung individueller erklärender Variablen**, die sich über die Zeit nicht ändern bzw. für alle gleich ändern (Alter),
    - erhöht unnötigerweise σ_η^2 und damit

$$Var(\hat{\rho}|\mathbf{X}) = \frac{\sigma_\eta^2}{SST_{D_s \cdot D_t}(1 - R^2_{D_s \cdot D_t})}$$

      (außer $R^2_{D_s \cdot D_t}$ würde sich hierdurch zu stark erhöhen),
    - reduziert unnötigerweise R^2.
  - Interpretation von ρ verändert sich bei **Logarithmierung** von Y_{ist}.

- In manchen Fällen ist Annahme **(E5)** nicht ausreichend, nämlich wenn weitere Einflussfaktoren $\mathbf{x}_{ist}$ eine Rolle spielen und $E[\mathbf{x}_{ist}|s,t]$ von s und t abhängt. Ist $\mathbf{x}_{ist}$ beobachtbar, und ist die **Annahme**
  - **(E5') Linearität in Gruppen- und Zeitmittelwerten und beobachtbaren (individuellen) Charakteristika**

$$E[Y_{0ist}|s,t,\mathbf{x}_{ist}] = a_s + d_t + \mathbf{x}_{ist}'\boldsymbol{\delta}, \quad s = T, K; t = V, N; \text{alle } i \tag{3.15}$$

erfüllt, erhält man statt (3.7) die Regression

$$Y_{ist} = \alpha + \gamma D_s + \lambda D_t + \rho \underbrace{(D_s D_t)}_{=C_{st}} + \mathbf{x}_{ist}\boldsymbol{\delta} + \eta_{ist} \tag{3.16}$$

- Bilden des auf s und t bedingten Erwartungswertes von (3.16) liefert unter Gültigkeit von **(E5')**

$$E[Y_{ist}|s,t] = \alpha + \gamma D_s + \lambda D_t + \rho \underbrace{(D_s D_t)}_{=C_{st}} + E[\mathbf{x}_{ist}|s,t]\,\boldsymbol{\delta}, \tag{3.17}$$

da $E\left[\eta_{ist}|s,t\right] = 0$. Solange $E\left[\mathbf{x}_{ist}|s,t\right] = konstant$, ändert sich bei Weglassen der $\mathbf{x}_{ist}$ lediglich die Konstante. In diesem Fall erhöht das Weglassen also nur die Varianz, siehe $Var(\hat{\rho}|\mathbf{X})$ oben.

- **Beispiel** : **Auswirkung des Baus einer Müllverbrennungsanlage auf die Hauspreise in der nahen und fernen Umgebung** (Fortsetzung von Abschnitt 1.3):

Abhängige Variable: $Realpreis$			
Gleichung	(3.7)	(3.16)	
Unabhängige Variable/Modell	(1)	(2)	(3)
Konstante	82,517.23	89,116.54	13,807.67
	(2,726.910)	(2,406.051)	(11,166.59)
$Y81$	18,790.29	21,321.04	13,928.48
	(4,050.065)	(3,443.631)	(2,798.747)
$NEARINC$	-18,824.37	9,397.936	3,780.337
	(4,875.322)	(4,812.222)	(4,453.415)
$Y81 * NEARINC$	-11,863.90	-21,920.27	-14,177.93
	(7,456.646)	(6,359.745)	(4,987.267)
andere Kontrollvar.	—	AGE, AGE^2	All
Stichprobengröße	321	321	321
R^2	0.173948	0.414448	0.660048
Standardfehler der Regression	30,242.90	25,543.29	19,619.02
Residuenquadratsumme	2.90 e+11	2.06 e+11	1.19 e+11
AIC	23.48429	23.15265	22.64005
HQ	23.50306	23.18080	22.69166
SC	23.53129	23.22315	22.76929

Anmerkungen: (3.7) schätzt Gleichung (13.7) und Spalte (1) in Table 13.2 in Wooldridge (2009). (3.16), d. h. Spalten (2) und (3) entsprechen Spalten (2) und (3) in Table 13.2 in Wooldridge (2009).

Durch Berücksichtigen der Kontrollvariablen wird der Standardfehler soweit reduziert, dass die Nullhypothese “Es liegt kein kausaler Effekt der Müllverbrennungsanlage vor”, $\rho = 0$, bei $\alpha = 0.05$ abgelehnt wird.

- **Mögliche Probleme**:
  - Entwickeln sich unbeobachtbare Einflüsse über die Zeit in Treatment- und Kontrollgruppe hinweg nicht gleich, ist Annahme **(E5)** verletzt.
  - Verursacht die Durchführung der Maßnahme eine Migration von Individuen von einer zur anderen Gruppe, kommt es zu Selektionsverzerrungen.
- Literatur: Angrist und Pischke (2009, Section 5.2).
- **Zur Erinnerung**: Die bisher diskutierten Annahmen sind Beispiele für Annahmen, damit ein Regressionsmodell eine **kausale** Interpretation hat.
- Der DD-Ansatz lässt sich natürlich auch verwenden, wenn “nur” auf ein Strukturbruch in den Parametern getestet werden soll — ohne

Durchführung einer Kausalitätsanalyse. Der DD-Ansatz ist ein Spezialfall einer gepoolten Querschnittsregression, wenn genau eine erklärende Dummyvariable vorliegt und diese eine Dummyvariable ist.

Beispiel: Auswirkung des Baus einer Müllverbrennungsanlage: Zwischen $t = V$ und $t = N$ wurde gar keine Müllverbrennungsanlage gebaut bzw. sind keine Veränderungen bekannt. Man möchte lediglich wissen, ob sich in den beiden Regionen die realen Hauspreise unterschiedlich entwickelt haben.

4 Paneldatenmethoden

- **Sprechweise**: Index i bezeichnet die betrachtete Einheit (panel unit) und kann Individuen, Firmen, Städte, Länder, Staaten, etc. bezeichnen. Im Folgenden wird stellvertretend nur von Individuen gesprochen.
- **Wiederholung**: Voraussetzung für die OLS-Schätzung gepoolter Querschnittsdaten ist: keine Verzerrung/Inkonsistenz durch vernachlässigte Variablen.
- Vernachlässigte Variable a_i bei Querschnittsanalyse:

$$y_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_j x_{ij} + \ldots + \beta_k x_{ik} + \underbrace{\delta a_i + u_i}_{\varepsilon_i}$$

  OLS verzerrt und inkonsistent, falls Fehlerterm ε_i mit x_{ij} korreliert, d.h. $Cov(a_i, x_{ij}) \neq 0$ für irgendein $j = 1, 2, \ldots, k$.
- Wird $Cov(a_i, x_{ij}) \neq 0$ vermutet und lässt sich die Variable a_i nicht

erheben und

– liegt nur **eine** Querschnittserhebung vor, dann bleibt nur die Instrumentvariablenschätzung, siehe Abschnitt 5. Im Zusammenhang mit Paneldaten wird die Verzerrung bei OLS-Schätzung auch als **'heterogeneity bias'** bezeichnet, siehe auch weiter unten.

– liegen mehrere Querschnittserhebungen für *dieselben* Individuen $i = 1, \ldots, N$ vor, dann ergeben sich neue Möglichkeiten. Unter der Annahme, dass $\beta_j, j = 0, \ldots, k, \delta$ über die Zeit konstant sind, ergibt sich (wir setzen im Folgenden auch o.E.d.A. $\delta = 1$)

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \underbrace{a_{it} + u_{it}}_{\varepsilon_{it}},$$

$$i = 1, 2, \ldots, N, \quad t = 1, \ldots, T$$

* **Fehlerterm** u_{it}:
  erfasst dann alle unsystematischen, individuellen, zeit-variablen Einflüsse **(idiosyncratic error)**
* **Fehlerterm** ε_{it}:
  wird als **zusammengesetzter Fehler (composite error)** bezeichnet.

$*$ **Variable** a_{it}:

$$a_{it} \begin{cases} = d_t + a_i & \text{Fall I} \\ \neq d_t + a_i & \text{Fall II} \end{cases}$$

Fall I:

Die Variable a_{it} lässt sich aufspalten in einen Faktor d_t, der über die Zeit variiert, aber über die Individuen hinweg konstant ist und einen Faktor a_i, der über die Zeit konstant ist, aber über die Individuen hinweg variiert.

Die Variable a_i erfasst alle unbeobachteten, zeit-konstanten individuellen (unobserved, time-constant) Faktoren. Da a_i über die Zeit hinweg konstant ist, wird die Variable auch als 'fixed effect' bezeichnet. In der Statistikliteratur wird a_i auch als 'incidental parameter' bezeichnet. Da sie über die Individuen hinweg variiert, also Heterogenität unter den Individuen erfasst, sagt man auch, dass unbeobachtete Heterogenität (unobserved heterogeneity) vorliegt.

Namen für a_i:
unbeobachteter (unobserved) Faktor, **unbeobachtete Heterogenität (unobserved heterogeneity)**, **incidental parameter**

Namen für Modelle mit a_i:
Modell mit unbeobachteten Effekten, **Modell mit fixen Effekten**, **Fixed-Effects-Modell**, **Fehler-Komponenten-Modell (error-components model)**.

Fall II: Der allgemeine Fall ist komplizierter und wird hier nicht behandelt.

– Die flexibelste Modellierung von d_t ist, für jede Periode (außer der ersten) eine Periodendummy zu definieren:

$$dtime_t = \begin{cases} 1 & \text{falls } t = time \\ 0 & \text{falls } t \neq time \end{cases}$$

z.B., falls $t = 2004, 2005, 2006$

$$d2006_t = \begin{cases} 1 & \text{falls } t = 2006 \\ 0 & \text{falls } t \neq 2006 \end{cases}$$

– Durch das Zulassen von fixen Effekten können u.U. individuelle beobachtbare, zeitkonstante Charakteristika (z.B. Geschlecht) nicht mehr berücksichtigt werden.

- Wird $Cov(a_i, x_{ij}) = 0$, $j = 1, \ldots, k$ vermutet, dann ist OLS weder verzerrt, noch inkonsistent. Aber wird nur ein Querschnitt verwendet, obwohl ein Panel vorliegt, wird Information verschenkt. Werden Paneldaten verwendet, muss man beachten, dass der zusammengesetzte Fehler ε_{it} Autokorrelation aufweist! Zu geeigneten Schätzverfahren kommen wir in Abschnitt 4.3.

4.1 Evaluation bei unbeobachtbarer Heterogenität mit Paneldaten

- In diesem Abschnitt wird die Notation von Kapitel 2 verwendet.
- Das **beobachtbare Ergebnis** für Individuum i in Periode t lautet, vgl. (2.2):

$$Y_{it} = Y_{0it} + (Y_{1it} - Y_{0it})C_{it}, \tag{4.1}$$

  wobei Y_{0it} das potentielle Ergebnis bei Nichtteilnahme bezeichnet.
- Die **Regressionsdarstellung** von (4.1) lautet, äquivalent zu (2.12),

$$Y_{it} = \underbrace{E[Y_{0it}]}_{d_t} + \underbrace{(Y_{1it} - Y_{0it})}_{\rho_{it}} C_i + \underbrace{Y_{0it} - E[Y_{0it}]}_{\eta_{it}} \tag{4.2}$$

- Erweitert man **(E2)** auf alle betrachteten Perioden, erhält man **Annahme**
  - **(E2') Alle individuellen Erfolge der Maßnahme gleich**

$$\rho_{it} \equiv Y_{1it} - Y_{0it} = \rho. \tag{4.3a}$$

Mit **(E2')** gilt für das **beobachtbare Ergebnis**

$$Y_{it} = Y_{0it} + \rho C_{it}.$$

Dessen **Regressionsdarstellung** ist:

$$Y_{it} = d_t + \rho C_{it} + \eta_{it}, \tag{4.3b}$$

$$E[Y_{it}|C_{it}] = d_t + \rho C_{it} + E[\eta_{it}|C_{it}]. \tag{4.3c}$$

Beachte: $E[\eta_{it}|C_{it}]$ **kann ungleich Null sein.** Wenn $E[\eta_{it}|C_{it}] \neq 0$, kommt es zu **Selektionsverzerrungen**, siehe Abschnitt 2.1.

- Falls $E[\eta_{it}|C_{it}] \neq 0$, kann man einen neuen Störterm ν_{it} definieren und erhält die **Regression**

$$Y_{it} = d_t + \rho C_{it} + E\left[\eta_{it}|C_{it}\right] + \underbrace{\eta_{it} - E\left[\eta_{it}|C_{it}\right]}_{\nu_{it}}, \tag{4.4}$$

dessen bedingter Erwartungswert bei Konditionierung auf C_{it} per Konstruktion Null ist.

- Dieser Schritt reicht jedoch nicht aus, da $E\left[\eta_{it}|C_{it}\right]$ in der Regression (4.4) prinzipiell unbeobachtbar ist, weil

$$E\left[\eta_{it}|C_{it}\right] = E\left[Y_{0it}|C_{it}\right] - E\left[Y_{0it}\right] \tag{4.5}$$

$E\left[Y_{0it}|C_{it} = 1\right]$ enthält, was prinzipiell unbeobachtbar ist.

- **Ausweg** (vgl. Abschnitt 2.2): Man muss Variablen finden, so dass die bedingte Unabhängigkeitsannahme erfüllt ist. Hierfür wird die bedingte Unabhängigkeitsannahme **(E3)** angepasst, indem individuelle Effekte a_i und Zeitdummies d_t berücksichtigt werden:
  - **(E3') Bedingte Unabhängigkeitsannahme (conditional independence assumption (CIA), selection on observables)**

$$\{Y_{1it}, Y_{0it}\} \perp\!\!\!\perp C_{it} | a_i, \mathbf{x}_{it}$$

woraus folgt:

$$E[Y_{0it}|a_i, \mathbf{x}_{it}, C_{it} = 1] = E[Y_{0it}|a_i, \mathbf{x}_{it}, C_{it} = 0] = E[Y_{0it}|a_i, \mathbf{x}_{it}],$$
$$E[Y_{1it}|a_i, \mathbf{x}_{it}, C_{it} = 1] = E[Y_{1it}|a_i, \mathbf{x}_{it}, C_{it} = 0] = E[Y_{1it}|a_i, \mathbf{x}_{it}].$$

Mit **(E3')** erhält man

$$\begin{aligned} E\left[\eta_{it}|a_i, \mathbf{x}_{it}, C_{it}\right] &= E\left[Y_{0it}|a_i, \mathbf{x}_{it}, C_{it}\right] - \overbrace{E\left[Y_{0it}\right]}^{d_t} \\ &= E\left[Y_{0it}|a_i, \mathbf{x}_{it}\right] - d_t = E\left[\eta_{it}|a_i, \mathbf{x}_{it}\right], \end{aligned} \tag{4.7}$$

so dass man η_{it} anstelle auf C_{it} auf $a_i, \mathbf{x}_{it}$ bedingen kann. Damit wird (4.4) zu

$$Y_{it} = d_t + \rho C_{it} + E\left[\eta_{it}|a_i, \mathbf{x}_{it}\right] + \underbrace{\eta_{it} - E\left[\eta_{it}|a_i, \mathbf{x}_{it}\right]}_{u_{it}}$$ & mit (4.7) zu

$$Y_{it} = \rho C_{it} + E\left[Y_{0it}|a_i, \mathbf{x}_{it}\right] + \underbrace{\eta_{it} - E\left[\eta_{it}|a_i, \mathbf{x}_{it}\right]}_{u_{it}}. \tag{4.8}$$

Jetzt benötigen wir noch die zu **(E4)** äquivalente **Annahme**

– **(E4')**

$$E[Y_{0it}|a_i, \mathbf{x}_{it}] = a_i + d_t + \mathbf{x}'_{it}\boldsymbol{\gamma}, \quad \text{bzw.} \tag{4.9a}$$

$$E[\eta_{it}|a_i, \mathbf{x}_{it}] = a_i + \mathbf{x}'_{it}\boldsymbol{\gamma} \tag{4.9b}$$

und erhalten folgende **Panelregression**

$$Y_{it} = d_t + a_i + \mathbf{x}'_{it}\boldsymbol{\gamma} + \rho C_{it} + u_{it}. \tag{4.10}$$

- **Ist MLR.4 erfüllt?**

$$E\left[Y_{it}|a_i, \mathbf{x}_{it}, C_{it}\right] = d_t + a_i + \mathbf{x}'_{it}\boldsymbol{\gamma} + \rho C_{it} + E\left[u_{it}|a_i, \mathbf{x}_{it}, C_{it}\right].$$

Unter den getroffenen Annahmen ist **MLR.4 ist erfüllt**, da

$$E\left[u_{it}|a_i, \mathbf{x}_{it}, C_{it}\right] = E\left[\eta_{it}|a_i, \mathbf{x}_{it}, C_{it}\right] - E\left[\eta_{it}|a_i, \mathbf{x}_{it}\right] = 0,$$

wobei das zweite Gleichheitszeichen wegen (4.7) gilt.

- Fazit: Unter den Annahmen **(E2')**, **(E3')** und **(E4')** treten keine Selektionsverzerrungen auf und ρ misst den **durchschnittlichen Erfolg einer Maßnahme (average treatment effect, ATE)** (2.4).
- Obwohl wir auf die **individuellen Effekte** a_i bedingt haben, **müssen** diese **nicht beobachtbar sein, wenn Paneldaten vorliegen**, weil es dann möglich ist, a_i aus der Schätzgleichung zu eliminieren, siehe folgende Abschnitte. Über die Zeit konstante individuelle Effekte sind dann letztlich keine Störfaktoren (**confounding factors**) mehr.
- Literatur: Angrist und Pischke (2009, Section 5.3).

4.2 Differenzen-Schätzer und Fixed-Effects-Schätzer: Paneldatenschätzer für den Fall unbeobachteter Faktoren, die mit Regressoren korrelieren

4.2.1 Differenzen-Schätzer

- Im Folgenden wird Fall I $a_{it} = d_t + a_i$ angenommen.
- Bei Vorliegen von **zwei Wellen**, lässt sich der fixe Effekt in

$$\begin{aligned} y_{it} = \beta_0 + \delta_0 d2_t + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it}, \\ i = 1, \ldots, N, \quad t = t_1, t_2, \end{aligned} \tag{4.11}$$

durch das **Bilden erster Differenzen (first-differenced equation)** beseitigen:

$$\begin{aligned} y_{it_2} - y_{it_1} &= \delta_0 + \beta_1 \left(x_{it_21} - x_{it_11}\right) + \cdots + \beta_k \left(x_{it_2k} - x_{it_1k}\right) \\ &\quad + \left(u_{it_2} - u_{it_1}\right), \quad i = 1, \ldots, N \\ \Delta y_{it_2} &= \delta_0 + \beta_1 \Delta x_{it_21} + \cdots + \beta_k \Delta x_{it_2k} + \Delta u_{it_2} \end{aligned} \tag{4.12}$$

- Der OLS-Schätzer für (4.12) wird häufig als **Differenzen-Schätzer (first-difference estimator)** bezeichnet.

- Das Bilden erster Differenzen ist auch im Fall von **mehr als zwei Wellen** möglich (zur Vereinfachung der Notation setzen wir $t = 1, 2, \ldots, T$)

$$\begin{aligned} y_{it} = \beta_0 + \delta_2 d2_t + \delta_3 d3_t + \cdots + \delta_T dT_t + \\ + \beta_1 x_{it1} + \cdots + \beta_k x_{itk} + a_i + u_{it}, \\ t = 1, 2, 3, \ldots, T. \end{aligned} \tag{4.13}$$

  Das Bilden erster Differenzen ergibt

$$\begin{aligned} \Delta y_{it} = \delta_2 \Delta d2_t + \delta_3 \Delta d3_t + \cdots + \delta_T \Delta dT_t + \\ + \beta_1 \Delta x_{it1} + \cdots + \beta_k \Delta x_{itk} + \Delta u_{it}, \\ t = 2, 3, \ldots, T. \end{aligned}$$

  Beachte: Konstante ist verloren gegangen.

$\sharp$ **Hinweis**: Falls Konstante gewünscht: Man bildet Linearkombinationen aus Zeitdummies, die Konstante ergeben. Beispiel für $T = 3$: Die Parameter δ_2 und δ_3 geben den Einfluss von

$$\begin{pmatrix} \Delta d2_2 & \Delta d3_2 \\ \Delta d2_3 & \Delta d3_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}$$

an. Gewünscht ist jedoch der Einfluss von

$$\begin{pmatrix} const & \Delta d3_2 \\ const & \Delta d3_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Man muss demnach, gegeben δ_2, δ_3,

$$\begin{pmatrix} const & \Delta d3_2 \\ const & \Delta d3_3 \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} \Delta d2_2 & \Delta d3_2 \\ \Delta d2_3 & \Delta d3_3 \end{pmatrix} \begin{pmatrix} \delta_2 \\ \delta_3 \end{pmatrix}$$

nach α_0, α_3 auflösen. Dies ergibt

$$\begin{pmatrix} \alpha_0 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} const & \Delta d3_2 \\ const & \Delta d3_3 \end{pmatrix}^{-1} \begin{pmatrix} \Delta d2_2 & \Delta d3_2 \\ \Delta d2_3 & \Delta d3_3 \end{pmatrix} \begin{pmatrix} \delta_2 \\ \delta_3 \end{pmatrix}$$

bzw.

$$\begin{pmatrix} \alpha_0 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} \delta_2 \\ \delta_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix} \begin{pmatrix} \delta_2 \\ \delta_3 \end{pmatrix}.$$

Entsprechend ist das Vorgehen für $T > 3$, so dass man immer

$$\begin{aligned}\Delta y_{it} = \alpha_0 + \alpha_3 \Delta d3_t + \cdots + \alpha_T \Delta dT_t \\ + \beta_1 \Delta x_{it1} + \cdots + \beta_k \Delta x_{itk} + \Delta u_{it}, \\ t = 2, 3, \ldots, T\end{aligned} \tag{4.13'}$$

betrachten kann. Man beachte, dass die α_j's nicht direkt zu interpretieren sind.

- **Voraussetzungen für OLS-Differenzen-Schätzer**:
  Im Prinzip müssen MLR.1 - MLR.5 für neue Querschnitts-regression (4.12) bzw. (4.13') gelten. Für den Fall von zwei Wellen erhält man, $t = t_1, t_2$:
  - MLR.1: Gleichung (4.12) korrekt spezifiziert.
  - MLR.2: $\left\{(\Delta x_{it_2 1}, \ldots, \Delta x_{it_2 k}, \Delta y_{it_2}\right\}$ ist Zufallsstichprobe.
  - MLR.3: Keine perfekte Multikollinearität, d.h. $\Delta x_{it_2 1}, \ldots, \Delta x_{it_2 k}$ müssen (genügend) variieren, dürfen also insbesondere nicht konstant sein!
  - MLR.4: Bedingter Erwartungswert Null
    $$E\left[u_{it_j} | x_{it_1 1}, \ldots, x_{it_1 k}, x_{it_2 1}, \ldots, x_{it_2 k}, a_i\right] = 0, \quad j = 1, 2,$$
    $$\Longrightarrow \quad E\left[\Delta u_{it_2} | \Delta x_{it_2 1}, \ldots, \Delta x_{it_2 k}\right] = 0, \quad \text{d. h.}$$
    die Regressoren müssen also **gegeben den unbeobachtbaren Effekt streng exogen** sein.

– MLR.5: Bedingte Homoskedastie.

Übersetzt in die ursprünglichen Gleichungen erhält man (für beliebige T):

Annahmen für Konsistenz

– **FD.1**: Gleichung (4.13) bzw. (4.13') korrekt spezifiziert für alle $i = 1, \ldots, N$ und $t = 1, \ldots, T$ ($\Leftarrow$ MLR.1).

– **FD.2**: Die Querschnittsstichprobe (hinsichtlich Auswahl der Individuen) ist eine Zufallsstichprobe ($\Leftarrow$ MLR.2).

– **FD.3**: Die Regressoren $x_{it1}, \ldots, x_{itk}$ sind nicht konstant, auch nach Differenzenbildung nicht und sind nicht perfekt kollinear ($\Leftarrow$ MLR.3).

– **FD.4**: Bedingter Erwartungswert Null

$$\begin{aligned} & E\left[u_{it}|x_{i11}, \ldots, x_{iTk}, a_i\right] = 0, \quad t = 1, \ldots, T, \\ \Longrightarrow \quad & E\left[\Delta u_{it}|\Delta x_{is1}, \ldots, \Delta x_{isk}, s = 2, \ldots, T\right] = 0, \quad t = 2, \ldots, T, \end{aligned}$$

d. h. die Regressoren müssen also **gegeben den unbeobachtbaren Effekt streng exogen** sein ($\Leftarrow$ MLR.4).

Fasst man für Individuum i die Regressoren für alle Zeitpunkte t in einer Matrix $\mathbf{X}_i$ zusammen, lässt sich die Bedingung **FD.4** auch schreiben als

$$E\left[u_{it}|\mathbf{X}_i, a_i\right] = 0, \quad t = 1, \ldots, T.$$

Annahmen für (asymptotisch) gültige Standardfehler und Teststatistiken:

– **FD.5**: Bedingte Homoskedastie ⇒ unbedingte Homoskedastie

$$Var\left(\Delta u_{it}|\mathbf{X}_i\right) = \sigma^2, \quad t = 2, \ldots, T \quad (\Leftarrow \text{MLR.5}),$$

– **FD.6**: Keine Autokorrelation in den Fehlern

$$Cov\left(\Delta u_{it}, \Delta u_{is}|\mathbf{X}_i\right) = 0, \quad s, t = 2, \ldots, T, \quad t \neq s \quad (\Leftarrow \text{TS.5}).$$

– **FD.7** Normalverteilte Fehler

$$\Delta u_{it}|\mathbf{X}_i \sim N(0, \sigma^2)$$

Kommentare:

- Unter **FD.1** bis **FD.4** ist der Differenzen-Schätzer **unverzerrt** und für festes T und $N \to \infty$ **konsistent**.
- Unter **FD.1** bis **FD.6** ist der Differenzen-Schätzer **BLUE** und für festes T und $N \to \infty$ **asymptotisch normalverteilt**.
- Unter **FD.1** bis **FD.7** ist der Differenzen-Schätzer **normalverteilt**.
- Beachte die Implikationen von **FD.5** und **FD.6**: Aus ihnen folgt:

$$v_{it} \equiv \Delta u_{it} \sim WN(0, \sigma^2).$$

Und somit ergibt sich mit $u_{i0} = 0$, vgl. Zeitreihenökonometrie, Abs. 4.2

$$\Delta u_{it}1 = v_{it}$$

$$u_{it} = u_{i,t-1} + v_{it}.$$

Die N stochastischen Prozesse der **Fehler** $\{u_{it}, t = 1, 2, \ldots, T\}$, $i = 1, \ldots, N$, sind also alle **Random Walks** und genau dann ist der Differenzen-Schätzer BLUE!

– Sind die Fehler selbst White Noise, dann ist **FD.6** verletzt.

∗ Man kann dann den GLS-Schätzer anwenden, denn die $Cov(\Delta u_{it}, \Delta u_{is})$ sind dann alle bekannt.

∗ Im allgemeinen wird jedoch der Fixed-Effects-Schätzer verwendet, siehe folgenden Abschnitt 4.2.2.

- Schätzung des **Average Treatment Effect**:
  - Für die Regresssion (hier ist im Gegensatz zu Abschnitt 4.1 das beobachtbare Ergebnis klein geschrieben)

$$y_{it} = d_t + a_i + \mathbf{x}'_{it}\boldsymbol{\gamma} + \rho C_{it} + u_{it}, \tag{4.10}$$

    gelten die Annahmen **(E2')**, **(E3')** und **(E4')**, siehe Abschnitt 4.1, so dass **FD.1** und **FD.4** erfüllt sind und ρ als durchschnittlicher Erfolg einer Maßnahme (2.4) interpretiert werden kann.
  - Durch Differenzenbildung ergibt sich

$$\Delta y_{it} = \Delta d_t + \rho \Delta C_{it} + \Delta \mathbf{x}'_{it}\boldsymbol{\gamma} + \Delta u_{it}. \tag{4.14}$$

    Der KQ-Schätzer alias FD-Schätzer $\hat{\rho}$ schätzt den **durchschnittlichen Erfolg einer Maßnahme** (2.4). Vgl. hierzu den DD-Ansatz (3.6). Der FD-Schätzer $\hat{\rho}$ hat die beschriebenen Eigenschaften, wenn für die Stichprobe **FD.2** und **FD.3** gelten.
  - Gilt im Fall von zwei Wellen, dass $C_{i1} = 0, C_{i2} \in \{0, 1\}$ sowie $\boldsymbol{\gamma} = \mathbf{0}$,

so lässt sich der Differenzen-Schätzer für ρ alternativ berechnen durch

$$\hat{\rho} = \overline{\Delta y}_{treatment} - \overline{\Delta y}_{control}. \tag{4.15}$$

Ableitung: Bilden erster Differenzen liefert

$$\Delta y_{i2} = \Delta d_2 + \rho C_{i2} + \Delta \mathbf{x}'_{i2}\boldsymbol{\gamma} + \Delta u_{i2},$$
$$\Delta y_{i2} = \Delta d_2(1 - C_{i2}) + (\Delta d_2 + \rho) C_{i2} + \Delta \mathbf{x}'_{i2}\boldsymbol{\gamma} + \Delta u_{i2}.$$

Ist $\boldsymbol{\gamma} = \mathbf{0}$, dann ergeben sich folgende Schätzer

$$\widehat{\Delta d_2} = \frac{1}{\sharp \in \text{Kontroll-Gr.}} \sum_{i \in \text{Kontroll-Gr.}} \Delta y_{i2} \equiv \overline{\Delta y}_{control}$$
$$\widehat{\Delta d_2} + \hat{\rho} = \frac{1}{\sharp \in \text{Treatment-Gr.}} \sum_{i \in \text{Treat.-Gr.}} \Delta y_{i2} \equiv \overline{\Delta y}_{treatment}$$

– Ist $\boldsymbol{\gamma} \neq \mathbf{0}$, dann geht die einfache Berechnungsmöglichkeit (4.15) verloren.

4.2.2 Fixed-Effects-Schätzer

- Im Folgenden wird im Gegensatz zur allgemeinen Darstellung auf Folie 55 im Modell keine Konstante β_0 verwendet. Sie würde beim beim Fixed-Effects-Schätzer eliminiert werden. Siehe auch 88.

- Darstellung des Modells mit unbeobachtbaren Effekten mit Vektoren

$$\mathbf{x}_{it} = \begin{pmatrix} x_{it1} & \cdots & x_{itk} \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdots \\ \beta_k \end{pmatrix}$$

  ergibt

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + a_i + u_{it}, \tag{4.16}$$

  wobei zur Vereinfachung die Zeitdummies weggelassen werden.

- **Idee** des Fixed-Effects-Schätzer: Anstatt die Beobachtung der Vorperiode von der gegenwärtigen Beobachtung abzuziehen, wird der jeweilige Mittelwert aller Perioden abgezogen; in beiden Fällen ist der fixe Effekt konstant und kann durch Subtraktion eliminiert werden. Bildet man den Mittelwert von (4.16) über die Zeit, erhält man

$$\bar{y}_i = \bar{x}_i \boldsymbol{\beta} + a_i + \bar{u}_i,$$

wobei $\bar{\mathbf{x}}_i = \begin{pmatrix} \bar{x}_{i1} & \cdots & \bar{x}_{ik} \end{pmatrix}$ und

$$\bar{y}_i \equiv \bar{y}_{i\cdot} \equiv \frac{1}{T} \sum_{t=1}^{T} y_{it}, \quad \bar{x}_{ij} \equiv \bar{x}_{i \cdot j} \equiv \frac{1}{T} \sum_{t=1}^{T} x_{itj}$$

$$\bar{u}_i \equiv \bar{u}_{i\cdot} \equiv \frac{1}{T} \sum_{t=1}^{T} u_{it}$$

Durch Subtraktion der Zeitmittelwerte (anstelle der Beobachtungen aus der Vorperiode) eliminiert man den unbeobachtbaren Effekt

$$\begin{aligned} y_{it} - \bar{y}_i &= (\mathbf{x}_{it} - \bar{\mathbf{x}}_i)\,\boldsymbol{\beta} + (u_{it} - \bar{u}_i) \\ \ddot{y}_{it} &= \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{u}_{it}, \\ & \qquad i = 1, \ldots, N, \quad t = 1, \ldots, T, \end{aligned} \tag{4.17}$$

wobei z.B. $\ddot{y}_{it} \equiv y_{it} - \bar{y}_i$. Schätzt man $\boldsymbol{\beta}$ auf Basis von (4.17), erhält man den **Fixed-Effects-Schätzer**.

- Die Variablen $\ddot{y}_{it}$, $\ddot{\mathbf{x}}_{it}$ variieren *innerhalb* der Beobachtungen für Individuum i. Diese Variation liegt dem OLS-Schätzer zugrunde. Der **Fixed-Effects-Schätzer** wird deshalb als **Within Estimator** oder **Within-Group-Schätzer** bezeichnet.

- Man beachte, dass aufgrund der Definition der $\ddot{u}_{it}$ gilt

$$\frac{1}{T}\sum_{t=1}^{T}\ddot{u}_{it} = 0, \quad i = 1, \dots, N.$$

Durch diese N **zusätzlichen Restriktionen verringert sich die Zahl der Freiheitsgrade** von $df = NT - k$ auf

$$df = NT - k - N = N(T-1) - k.$$

Die zugrundeliegende Intuition ist, dass beim Within-Schätzer implizit N Mittelwerte geschätzt werden.

Man beachte die korrekte Berechnung der Freiheitsgrade bei der Berechnung von Standardfehlern und Teststatistiken.

- **Voraussetzungen für den Fixed-Effects-Schätzer** für $\boldsymbol{\beta}$:

  Annahmen für Konsistenz

  – T fest, $N \to \infty$

  – **FE.1** ($=$ **FD.1**): Gleichung (4.17) korrekt spezifiziert für alle $i = 1, \ldots, N$ und $t = 1, \ldots, T$ ($\Leftarrow$ MLR.1).

  – **FE.2** ($=$ **FD.2**): Die Querschnittsstichprobe (hinsichtlich Auswahl der Individuen) ist eine Zufallsstichprobe ($\Leftarrow$ MLR.2).

  – **FE.3** ($=$ **FD.3**): Die Regressoren $\mathbf{x}_{it}$ sind nicht konstant und für kein Individuum konstant über die Zeit und sie sind nicht perfekt kollinear ($\Leftarrow$ MLR.3).

  – **FE.4** ($=$ **FD.4**): Bedingter Erwartungswert Null

$$E\left[u_{it}|\mathbf{X}_i, a_i\right] = 0, \quad t = 1, \ldots, T.$$

Zusätzliche **Annahmen für (asymptotisch) gültige Standardfehler und Teststatistiken für $\boldsymbol{\beta}$** unter der Bedingung $N \to \infty$ und T fest:

– **FE.5**: Homoskedastie

$$Var\left(u_{it}|\mathbf{X}_i, a_i\right) = \sigma^2, \quad t = 1, \ldots, T, \quad (\Leftarrow \text{MLR.5}).$$

– **FE.6**: Keine Autokorrelation in den Fehlern

$$Cov\left(u_{it}, u_{is}|\mathbf{X}_i, a_i\right) = 0, \quad s, t = 1, \ldots, T, \quad t \neq s \quad (\Leftarrow \text{TS.5}).$$

Zusätzliche **Annahme für exakte Normal- und F-Verteilung** (für T und N beliebig):

– **FE.7** Normalverteilte Fehler

$$u_{it}|\mathbf{X}_i, a_i \sim N(0, \sigma^2).$$

- **Vergleich der Annahmen für Differenzen- und Fixed-Effects-Schätzer**:
  - Annahmen für Konsistenz gleich.
  - Annahmen für Standardfehler beziehen sich bei

$$\begin{cases} \text{Differenzen-Schätzer auf} & \Delta u_{it}, \\ \text{Fixed-Effects-Schätzer auf} & u_{it}, \end{cases}$$

    d.h. Fixed-Effects-Schätzer ist BLUE unter **FE.1** bis **FE.6**, aber nicht falls u_{it} ein Random Walk ist! D.h. Differenzen- und Fixed-Effects-Schätzer sind für unterschiedliche DGPs BLUE.
  - Konsistenz erfordert in beiden Fällen bei festem T, dass $N \to \infty$.

• Dummy-Variablen-Regression

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

$$\mathbf{y} = \begin{pmatrix} y_{11} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ y_{31} \\ \vdots \\ y_{NT} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} x_{111} & x_{112} & \cdots & x_{11k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1T1} & x_{1T2} & \cdots & x_{1Tk} \\ x_{211} & x_{212} & \cdots & x_{21k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{2T1} & x_{2T2} & \cdots & x_{2Tk} \\ x_{311} & x_{312} & \cdots & x_{31k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{NT1} & x_{NT2} & \cdots & x_{NTk} \end{pmatrix} = \begin{pmatrix} \mathbf{x}_{11} \\ \vdots \\ \mathbf{x}_{1T} \\ \mathbf{x}_{21} \\ \vdots \\ \mathbf{x}_{2T} \\ \mathbf{x}_{31} \\ \vdots \\ \mathbf{x}_{NT} \end{pmatrix}, \quad \mathbf{u} = \begin{pmatrix} u_{11} \\ \vdots \\ u_{1T} \\ u_{21} \\ \vdots \\ u_{2T} \\ u_{31} \\ \vdots \\ u_{NT} \end{pmatrix},$$

$$\mathbf{D}_a = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \qquad \underbrace{\begin{matrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & T \\ 2 & 1 \\ \vdots & \vdots \\ 2 & T \\ 3 & 1 \\ \vdots & \vdots \\ N & T \end{matrix}}_{\text{wobei } i \text{ in der ersten Spalte und } t \text{ in der zweiten Spalte läuft}} \qquad , \qquad \mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}.$$

Fasst man beide Regressormatrizen in $\mathbf{W} = \begin{pmatrix} \mathbf{X} & \mathbf{D}_a \end{pmatrix}$ zusammen, erhält man

$$\mathbf{y} = \mathbf{W} \begin{pmatrix} \boldsymbol{\beta} \\ \mathbf{a} \end{pmatrix} + \mathbf{u}$$

und den OLS-Schätzer für die Fixed-Effekts $\mathbf{a}$ plus die $\boldsymbol{\beta}$

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{a}} \end{pmatrix} = \left(\mathbf{W}'\mathbf{W}\right)^{-1} \mathbf{W}'\mathbf{y}.$$

Beachte (Wooldridge, 2015, Section14.1-1a):

– Der Dummy-Variablen-Schätzer $\hat{\boldsymbol{\beta}}$ ist mit dem Fixed-Effects-Schätzer/Within-Group-Schätzer (4.17) numerisch identisch.

– Im Dummy-Variablen-Schätzer ist automatisch die Zahl der Freiheitsgrade korrekt!

– Gelten **FE.1** bis **FE.7**, dann lässt sich z.B. mit einem F-Test die gemeinsame Signifikanz aller Fixed-Effects testen. Wie?

– Ist **FE.7** nicht erfüllt, existiert unter den gegebenen Annahmen nicht die übliche asymptotische Verteilung für $\hat{\mathbf{a}}$. Warum?

– Ist N groß, kann die Invertierung von $\left(\mathbf{W}'\mathbf{W}\right)$ numerisch instabil sein und es ist besser den Within-Schätzer zu verwenden.

– Beachte: die Schätzung einer Konstanten β_0 gemeinsam mit allen Fixed-Effects a_i, $i = 1, \ldots, N$ ist nicht möglich, da dann die Spalten in $\mathbf{W}$ linear abhängig werden würden und MLR.3 für den KQ-Schätzer der Dummy-Variablen Regression verletzt sein würde.

– **Alternative Berechnung für $\hat{\mathbf{a}}$:**

Wurde $\boldsymbol{\beta}$ mit dem Fixed-Effects-Schätzer $\hat{\boldsymbol{\beta}}_{FE}$ via (4.17) geschätzt, lässt sich $\hat{\mathbf{a}}$ direkt mit

$$\hat{a}_i = \bar{y}_i - \bar{\mathbf{x}}_i \hat{\boldsymbol{\beta}}_{FE}, \quad i = 1, \ldots, N,$$

berechnen, wie bspw. im `R`-Paket `plm` mittels `fixef()`.

- **Alternative Modellierung der Fixed-Effects**:
  (z.B. in **EViews**): Definiere

$$\beta_0 \equiv \frac{1}{N}\sum_{i=1}^{N} a_i, \quad \ddot{a}_i \equiv a_i - \beta_0.$$

Dann lässt sich Modell (4.16) schreiben als

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + \underbrace{a_i - \beta_0}_{\ddot{a}_i} + u_{it} \qquad (4.16')$$

und es gilt

$$\sum_{i=1}^{N} \ddot{a}_i = \sum_{i=1}^{N} a_i - N\beta_0 = N\beta_0 - N\beta_0 = 0.$$

EViews gibt anstelle der **Fixed-Effects** a_i die **Cross-Section-Effects** $\ddot{a}_i$ an. Man beachte, dass sich durch diese Lineartransformation die Zahl der Freiheitsgrade nicht ändert!

- **Schätzeigenschaften für Fixed-Effects**:
  - Gelten **FE.1** bis **FE.4**, ist $\hat{a}_i$ unverzerrt, aber nicht konsistent für festes T und $N \to \infty$. Warum?
  - Gelten **FE.1** bis **FE.4**, ist $\hat{a}_i$ (ebenso wie $\hat{\boldsymbol{\beta}}$) unverzerrt und konsistent für N fest und $T \to \infty$. Warum?
  - Gelten **FE.1** bis **FE.4**, ist $\hat{a}_i$ (ebenso wie $\hat{\boldsymbol{\beta}}$) unverzerrt und konsistent für $N \to \infty$ und $T \to \infty$.
  - Gelten **FE.1** bis **FE.7**, ist $\hat{a}_i$ exakt normalverteilt für T fest und N fest).

Für Paneldatensätze, in denen T und N groß ist, kann eine andere Asymptotik zugrunde gelegt werden: $T \to \infty, N \to \infty$. Dieser Fall ist für Fortgeschrittene.

- **Zeitdummies**:

$$\mathbf{d}_t = \begin{pmatrix} d1_t & d2_t & \cdots & dT_t \end{pmatrix}, \quad \boldsymbol{\delta} = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_T \end{pmatrix}$$

können nicht direkt als $\mathbf{d}_t\boldsymbol{\delta}$ in (4.16) bzw. (4.16') integriert werden, da es sonst zu perfekter Korrelation bei den Regressoren kommt, weil durch die Fixed-Effects bereits implizit oder explizit eine Konstante im Modell enthalten ist. Es muss also 1 Restriktion auf $\boldsymbol{\delta}$ auferlegt werden.

Zwei **Möglichkeiten** sind:

– **EViews**: Man betrachtet in Analogie zur Vorgehensweise bei Fixed-Effects die Abweichungen $\ddot{\delta}_j$, $j = 1, \ldots, T$ von einem (bei Vorhandensein von Fixed-Effects nicht bestimmbaren) Mittelwert, so dass für die $\ddot{\delta}_t$ die Restriktion $\frac{1}{T}\sum_{j=1}^{T} \ddot{\delta}_j = 0$ gilt:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + a_i + \mathbf{d}_t\ddot{\boldsymbol{\delta}} + u_{it}, \quad \sum_{j=1}^{T} \ddot{\delta}_j = 0 \tag{4.18}$$

oder

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + \ddot{a}_i + \mathbf{d}_t\ddot{\boldsymbol{\delta}} + u_{it}, \quad \sum_{j=1}^{T} \ddot{\delta}_j = 0. \tag{4.19}$$

In EViews werden die $\ddot{\delta}_j$ als **Period Effects** bezeichnet. Man kann die Restriktion auch leicht direkt in die Schätzgleichung einsetzen durch

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + \ddot{a}_i + \mathbf{d}_t \begin{pmatrix} \ddot{\delta}_1 \\ \ddot{\delta}_2 \\ \vdots \\ \ddot{\delta}_{T-1} \\ -\sum_{j=1}^{T-1} \ddot{\delta}_j \end{pmatrix} + u_{it}. \qquad (4.19')$$

– Man setzt eine Nullrestriktion, z.B: $\delta_1 = 0$ und schätzt die verbleibenden Parameter α_j, $j = 2, \ldots, T$,

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + \ddot{a}_i + d2_t\alpha_2 + \cdots + dT_t\alpha_T + u_{it}, \qquad (4.20)$$

wobei

$$\alpha_j = \delta_j - \beta_0 = \ddot{\delta}_j - \ddot{\delta}_1, \quad j = 2, \ldots, T.$$

Vgl. z.B. Wooldridge (2009, 2006, Example 14.1).

Beide Gleichungen (4.19') und (4.20) können mit OLS entweder als Dummy-Variablen-Regression oder mit dem Within-Group-Schätzer geschätzt werden. Für (4.20) lautet der Within-Group-Schätzer

$$\ddot{y}_{it} = \ddot{\mathbf{x}}_{it}\boldsymbol{\beta} + \ddot{d2}_t\alpha_2 + \cdots + \ddot{dT}_t\alpha_T + \ddot{u}_{it}. \tag{4.21}$$

Beachte:

- Bei Berücksichtigung von Zeitdummies reduziert sich Zahl der Freiheitsgrade um $T-1$. Warum?
- Bei Verwendung von Zeitdummies können gleichzeitig keine Variablen wie Alter oder Erfahrung verwendet werden. Wieso?

- **Interpretation der Parameter**:
  - Geschätzte Parameter $\boldsymbol{\beta}$ häufig leichter in Ausgangsgleichungen (4.16), (4.19), bzw. (4.20) statt in zeitmittelwertbereinigter Gleichung (4.17) bzw. (4.21) zu interpretieren.
  - Verwendung von Logs wie üblich.
  - Verwendung von verzögerten streng exogenen Variablen wie üblich.

- **Berechnung des Bestimmtheitsmaßes** R^2:

  Welche Gleichung soll zugrundegelegt werden?

  – die ursprüngliche Regression (4.16) alias Dummy-Variablen-Regression. Dies macht z.B. EViews.

    Problem: Ein großer Anteil des R^2 wird im Allgemeinen durch die Fixed-Effects bzw. Zeitdummies verursacht. Um den Einfluss der variablen Regressoren zu bewerten ist es deshalb besser,

  – die Regressionsgleichung (4.17), (4.21) mit den um die Zeitmittelwerte bereinigten Variablen zugrunde zu legen.

- **Beispiel — Datensatz `jtrain`: Effekt von Jobtraining auf die Ausschussrate**

  Vgl. Wooldridge (2009, 2006, Example 14.1).

  1. Erstellen eines Panel-Dataframes aus einem normalen Dataframe. Wähle als `cross-section identifier fcode` und als `date identifier year`.
  2. Berechnen des Within-Group-Schätzers ohne plm-Paket

```
    # Umwandlung der Variablen fcode von "factor" zu "character".
    # Letzteres ist notwendig, um Mittelwerte über die Zeit mit
    # tapply() zu berechnen
fcode <- as.character(jtrain$fcode)
    # tapply: wendet Funktion "mean" (drittes Argument) auf
    # Variable (erstes Argument) an, wobei die Einheiten über den gleichlangen
    # Charakter-Vektor (zweites Argument) identifiziert werden.
    # Das Ergebnis wird dann angewendet auf den Charakter-Vektor in [].
lscrap_2p  <- jtrain$lscrap  - tapply(jtrain$lscrap,  fcode, mean)[fcode]
grant_2p   <- jtrain$grant   - tapply(jtrain$grant,   fcode, mean)[fcode]
grant_1_2p <- jtrain$grant_1 - tapply(jtrain$grant_1, fcode, mean)[fcode]
d88_2p     <- jtrain$d88     - tapply(jtrain$d88,     fcode, mean)[fcode]
d89_2p     <- jtrain$d89     - tapply(jtrain$d89,     fcode, mean)[fcode]
    # selbsterstellter Within-group-Schätzer
eq_table14_1 <- lm(lscrap_2p ~ 0 + grant_2p + grant_1_2p + d88_2p + d89_2p)
```

```
summary(eq_table14_1)
# BEACHTE: Std. Error, t value und Pr() sind im Vergleich zum plm()-Befehl
# falsch, da die Freiheitsgrade nicht richtig berechnet werden. Es werden
# zu wenig verwendet.
```

3. Direkte Berechnung des Fixed-Effect-Schätzers auf Basis von (4.21):

```
eq_table14_1_fullR <- plm(lscrap ~ grant + grant_1 + d88 + d89,
                          data = jtrain_pan, model = "within", effect = "individual")
summary(eq_table14_1_fullR)
# Ausgabe der geschätzten fixen Effekte
fixef(eq_table14_1_fullR, effect="individual")
# Umrechnung in Cross-section Effects a la EViews
# Berechnung Konstante als Mittelwert der fixen Effekte
hat_beta_0 <- mean(fixef(eq_table14_1_fullR))
# Cross-section Fixed Effects a la EViews
fixef(eq_table14_1_fullR) - hat_beta_0
```

4. Direkte Berechnung des Fixed-Effect-Schätzers inklusive Period Effects (Basis ist (4.19)):

```
# Fixe Effekte und Period Effects mit plm()
eq_table14_1_cs_per <- plm(lscrap ~ grant + grant_1,
                           data = jtrain_pan, model = "within", effect = "twoways")
summary(eq_table14_1_cs_per)

# Durchscnitt der fixen Effekte
```

```
hat_beta_0 <- mean(fixef(eq_table14_1_cs_per, effect="twoways"))

# Cross-section Effects a la EViews
fixef(eq_table14_1_cs_per, effect="individual") - mean(fixef(eq_table14_1_cs_per,
                      effect="individual"))
# Period Effects a la EViews
fixef(eq_table14_1_cs_per, effect="time") - mean(fixef(eq_table14_1_cs_per, effect="time"))
```

5. Berechnung des Fixed-Effect-Schätzers mit Restriktion auf Zeitdummyparameter (Basis ist (4.19')):

```
d87 <- as.numeric(jtrain_pan$year == 1987)
eq_fixed_time_effects <- lm(lscrap_2p ~ grant_2p + grant_1_2p + I(d87-d89) +
                                I(d88-d89), data = jtrain_pan)
summary(eq_fixed_time_effects)
```

- **Residuenanalyse**
  - Differenzen-Schätzer: die Δu_{it} können geschätzt werden und entsprechende Autokorrelationstests durchgeführt werden.
  - Fixed-Effects-Schätzer: es können nur die $\ddot{u}_{it}$, nicht die u_{it} geschätzt werden, so dass eine Überprüfung der Autokorrelation etwas aufwändiger ist.
  - Eine gute Idee: Verwendung beider Schätzer und Überprüfen der Abhängigkeit der Ergebnisse von der Schätzmethode.

- **Beschränkungen** des Differenzen- und Fixed-Effects-Schätzers:
  - großer Verlust an Freiheitsgraden bei großem N.
  - N klein relativ zu T:
    - Differenzen-Schätzer: funktioniert besser als
    - Fixed-Effects-Schätzer: eine Verletzung von **FE.1** bis **FE.7** kann zu stark fehlerhafter Inferenz führen (z.B. spurious regression, falls u_{it} Random Walks); anfälliger bei Heteroskedastie, Autokorrelation in den Fehlern u_{it}.
  - Verzögerte endogene Variablen: werden hier nicht behandelt, siehe hierzu z.B. Wooldridge (2002, 2010).
  - Regressorvariablen dürfen über die Zeit für ein Individuum nicht konstant sein, z.B. Geschlecht, geographische Lage. Extremfall: Regressorvariablen unterscheiden sich nur hinsichtlich der Individuen.

Aber: Zeitkonstante Variablen können als Interaktionsvariablen mit zeit-variierenden Variablen genutzt werden, z.B. Geschlecht mit Zeitdummies.

– Wie bereits erwähnt, können bei Verwendung aller möglichen Zeitdummies keine Variablen wie Alter verwendet werden.

- **'Unbalanced' Panel**
  - **Fehlende Beobachtungen (missing values)**: für ein oder mehrere Individuen liegen für eine Welle keine Beobachtungen vor, für alle anderen Individuen schon.
  - Ein Paneldatensatz mit fehlenden Beobachtungen heißt 'unbalanced' Panel.
  - Fixed-Effects-Schätzer: Es fehlen z.B. die Beobachtungen für das dritte Individuum in der vierten Welle, $i = 3, t = 4$.

Dann werden die jeweiligen Zeilen aus $\mathbf{y}$, $\mathbf{X}$, $\mathbf{u}$, bzw. in der Dummy-Variablen-Regression $\mathbf{D}_a$ gestrichen. Die Zahl der Beobachtungen reduziert sich von TN auf $T(N-1)+(T-1)$, bzw. allgemein auf $(T_1+\ldots+T_N) \leq TN$, wobei T_i die Anzahl aller Beobachtungen für Individuum i ist.

Bei der Berechnung der zeitlichen Mittel $\bar{y}_i$, $\bar{\mathbf{x}}_i$ werden nur die für Individuum i verfügbaren Beobachtungen verwendet.

Wichtig: Falls das Auftreten von fehlenden Werten nicht mit den Fehlern u_{it} korreliert ist, ändern sich die Schätzeigenschaften nicht.

– Häufiger Fall: Ab einem bestimmten Zeitpunkt t liegen für ein Individuum i keine Daten mehr vor, z.B. wegen Konkurs (Firmen), Umzug, Tod (Individuen), etc. Dieses Phänomen wird als **'attrition'** (wörtliche Übersetzung: Verschleiß, Abrieb) bezeichnet.

– Ist die 'attrition' korreliert

  * mit den Fehlern u_{it}, kann es zu verzerrten Schätzern kommen.

  * mit den unbeobachtbaren Effekten a_i, ist alles ok.

4.3 Random-Effects-Schätzer: Paneldatenschätzer für den Fall unbeobachteter Faktoren, die mit Regressoren nicht korrelieren

- Differenzen-Schätzer und Fixed-Effects-Schätzer wurden für den Fall abgeleitet, dass die unbeobachtbaren Effekte mit einem oder mehreren Regressoren korreliert sind, d.h. $Cov(a_i, x_{itj}) \neq 0$ für mindestens ein $j = 1, \ldots, k$.
- Gilt $Cov(a_i, x_{itj}) = 0$ für alle $j = 1, \ldots, k$, $t = 1, \ldots, T$, $i = 1, \ldots, N$, ist
  - OLS basierend auf einer einzelnen Welle t konsistent, aber nicht effizient,
  - OLS für gepoolte Querschnittsdaten konsistent, aber produziert falsche Standardfehler.

Ausweg: Random-Effects-Schätzer.

- **Random-Effects-Schätzer**:
  - In der Panelregression

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + \underbrace{a_i + u_{it}}_{v_{it}},$$

wobei o.E.d.A. angenommen wurde, dass $E[a_i] = 0$ und deshalb eine Konstante aufgenommen wurde, müssen die unbeobachtbaren Effekte a_i nicht mehr explizit eliminiert werden, so dass man die Regression

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + v_{it}$$

mit dem zusammengesetzten Fehler $v_{it} = a_i + u_{it}$ betrachten kann.

Es gilt:

$$Corr(v_{it}, v_{is}) = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_u^2}, \quad t \neq s,$$

wobei $Var(a_i) = \sigma_a^2 \Rightarrow$ Standardfehler des gepoolten OLS-Schätzers inkorrekt, da dieser unter der Annahme unkorrelierter und homoskedastischer Fehler abgeleitet wurde.

– Lösung: Verwende GLS:

GLS-Schätzer für Random-Effects-Modell (Ableitung erfordert aufwändige Matrixalgebra)

$$\lambda = 1 - \left(\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2}\right)^{1/2} = 1 - \left(\frac{1}{1 + T\frac{\sigma_a^2}{\sigma_u^2}}\right)^{1/2}, \qquad (4.22)$$

$$y_{it} - \lambda\bar{y}_i = \beta_0(1-\lambda) + (\mathbf{x}_{it} - \lambda\bar{\mathbf{x}}_i)\boldsymbol{\beta} + (v_{it} - \lambda\bar{v}_i). \qquad (4.23)$$

– Variablen, die über die Zeit hinweg konstant sind, können verwendet werden. Warum?

– Zeitdummies können wie im Fixed-Effects-Schätzer berücksichtigt werden.

– Man kann $\boldsymbol{\beta}$ auch auf Basis der N Gruppendurchschnitte

$$\bar{y}_i = \bar{\mathbf{x}}_i \boldsymbol{\beta} + \bar{v}_i$$

mittels OLS schätzen. Dies liefert den **Between-Group-Schätzer**, der jedoch ineffizient ist.

– Random-Effects-Schätzer (4.23) ist gewichtetes Mittel von Within-Group-Schätzer (4.17) und gepooltem OLS-Schätzer, da

$$v_{it} - \lambda \bar{v}_i = \lambda (v_{it} - \bar{v}_i) + (1 - \lambda) v_{it}.$$

– Quasi-Zeitmittelwertbereinigung für $0 < \lambda < 1$, da

$$v_{it} - \lambda \bar{v}_i = a_i (1 - \lambda) + u_{it} - \lambda \bar{u}_i.$$

– Anteil der Zeitmittelwertbereinigung abhängig von T, σ_u^2 und σ_a^2, da

$$1 - \lambda = \left(\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_a^2} \right)^{1/2}.$$

Spezialfälle:

* Keine Random-Effects, d.h. $\sigma_a^2 = \lambda = 0$:
  GLS = **Gepoolte OLS**.
* Varianz der unbeobachtbaren Effekte σ_a^2 groß relativ zur Varianz der idiosynkratischen Fehler σ_u^2: GLS entspricht beinahe **Fixed-Effects-Schätzer**
* Für $T \to \infty$ erhält man $\lambda = 1$ und somit als Grenzfall den **Fixed-Effects-Schätzer**.

– Ist a_i mit mindestens einem x_{itj} korreliert, ist die asymptotische Verzerrung des RE-Schätzers umso kleiner, je näher λ an Eins ist.

- **Random-Effects-Schätzer** ist FGLS-Schätzer mit konsistentem Schätzer $\hat{\lambda}$. Ersetze unbekannte Größen σ_a^2 und σ_u^2 in λ durch konsistente Schätzer $\hat{\sigma}_a^2$ und $\hat{\sigma}_u^2$

$$\hat{\lambda} = 1 - \left(\frac{1}{1 + T\frac{\hat{\sigma}_a^2}{\hat{\sigma}_u^2}} \right)^{1/2}. \tag{4.24}$$

  – **Schätzung**:

    * Gepoolte OLS-Schätzung von

$$y_{it} = \beta_0 + \mathbf{x}_{it}\boldsymbol{\beta} + v_{it} \tag{4.25}$$

      liefert $\hat{v}_{it}$'s und $\hat{\sigma}_v^2$,

    * $\hat{\sigma}_a^2 = \frac{1}{NT(T-1)/2-(k+1)} \sum_{i=1}^{N} \sum_{t=1}^{T-1} \sum_{s=t+1}^{T} \hat{v}_{it}\hat{v}_{is}$,

    * $\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_a^2$.

Es gibt alternative Möglichkeiten v_{it} und u_{it} zu schätzen: den Swamy-Arora Schätzer (R und EViews), Wansbeek and Kapteyn Schätzer (Eviews), Wallace-Hussain Schätzer (R und EViews), etc.

– **Annahmen für Konsistenz** und **asymptotische Normalität** unter der Bedingung $N \to \infty$ und T fest (wie in Abschnitt 4.2.2 über Fixed-Effects-Schätzer), d. h. es wird im Gegensatz zu Folie 106 die Darstellung $y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + a_i + u_{it}$ ohne Konstante β_0 betrachtet:

  * **FE.1**: Korrekte Spezifikation.

  * **FE.2**: Zufallsstichprobe.

  + **RE.3**: Die Regressoren sind nicht perfekt kollinear. Beachte: RE.3 ist schwächer als FE.3.

  - **RE.4**: Es gilt **FE.4** und

$$E[a_i|\mathbf{X}_i] = \beta_0, \quad i = 1, \ldots, N,$$

impliziert $Cov(a_i, \mathbf{x}_{it}) = 0$ und eine Interpretation für β_0. (Würde eine Konstante im Modell enthalten sein, müsste $E[a_i|\mathbf{X}_i] = 0$ vorausgesetzt werden.)

– Zusätzliche **Annahmen für asymptotisch korrekte Standardfehler** und **effiziente Schätzer**:

* **RE.5**: Es gilt **FE.5** und

$$Var(a_i|\mathbf{X}_i) = \sigma_a^2.$$

* **FE.6**: Keine Autokorrelation in den idiosynkratischen Fehlern u_{it}.

– **Diskussion**:

* Ist **RE.4** erfüllt, ist Random-Effects-Schätzer wesentlich effizienter als Fixed-Effects-Schätzer. **Tradeoff zwischen Robustheit und Effizienz**.

* Für konsistente Schätzung von σ_a^2 ist $N \to \infty$ essentiell. Deshalb funktioniert Random-Effekts-Schätzer für kleine N im Allgemeinen nicht gut.

* Für N fest und $T \to \infty$ ($\Leftrightarrow$ in der Praxis: kleines N, großes T) können zwar die unbeobachtbaren Effekte a_i mit Fixed-Effects-Schätzer konsistent geschätzt werden, nicht jedoch σ_a^2, da nur endlich viele, nämlich N Beobachtungen $\hat{a}_i$ hierfür vorhanden sind. Genaue Eigenschaften unbekannt.

– Output (für RE-Spezifikation von Wooldridge (2009, Table 14.2), Datensatz: `wagepan`):

∗
```
Effects:
                  var std.dev share
idiosyncratic 0.1232  0.3510 0.539
individual    0.1054  0.3246 0.461
theta: 0.6429
```

bezeichnet:

$$\text{Erste Zeile: } \hat{\sigma}_u^2, \hat{\sigma}_u \text{ und } \hat{\sigma}_u^2 / \left(\hat{\sigma}_a^2 + \hat{\sigma}_u^2 \right).$$

$$\text{Zweite Zeile: } \hat{\sigma}_a^2, \hat{\sigma}_a \text{ und } \widehat{Corr}(v_{it}, v_{is}) = \hat{\sigma}_a^2 / \left(\hat{\sigma}_a^2 + \hat{\sigma}_u^2 \right).$$

∗ Ergebnisse beziehen sich auf das FGLS Modell (4.23).

∗ $\lambda = 0.642911$ (im R-Output als `theta` bezeichnet.)

- **Wahl zwischen Fixed-Effects und Random-Effects-Schätzer**
  - Berücksichtigung von Variablen, die für Individuen konstant sind, nur in Gepoolte OLS oder Random-Effects-Schätzer möglich. Voraussetzung: **RE.4**, $E[a_i|\mathbf{X}_i] = \beta_0$, gilt!
    Beispiel: Ausbildung.
  - Test auf Gültigkeit von **RE.4**:
    Hausman Test überprüft, ob statistisch signifikante Unterschiede zwischen Parameterschätzungen bezüglich der zeitvariierenden Variablen aus dem RE- und FE-Schätzer vorliegen:
    * Falls ja, Annahmen des Random-Effect Schätzers verletzt und Fixed-Effects-Schätzer wird verwendet.
    * Falls nein, entweder Power zu schwach oder Annahmen des Random-Effect-Schätzers erfüllt.
- **Pflichtlektüre**: Wooldridge (2009, 2006, Sect. 14.1, 14.2+App. 14A)

- **Cluster-Stichproben** und **Matched Pairs**:
  - Mechanik des Fixed-Effect-Schätzers oder Differenzen-Schätzers auch anwendbar, wenn Individuen innerhalb eines Clusters, z.B. Schule, Betrieb, Verein unbeobachtbare Gemeinsamkeiten aufweisen und deshalb die unbeobachtbaren Effekte mit den beobachtbaren Variablen korreliert sein können.
  - Werden zwei Individuen mit sehr ähnlichen unbeobachtbaren Effekten, z.B. eineiige Zwillinge (Beispiel von **matched pairs**) verglichen, Anwendung des Differenzen-Schätzers möglich.
  - Mehr Details finden sich in Wooldridge (2009, Section 14.3) bzw. etwas weniger ausführlich in Wooldridge (2006, Section 14.3).

5 Instrumentvariablenmethoden und zweistufige LS-Schätzer (two-stage least squares (2SLS))

Motivation für Instrumentvariablenmethoden:

- Anwendung des Fixed-Effect-Schätzers erlaubt, Inkonsistenz durch zeitkonstante unbeobachtbare Effekte zu vermeiden.
- Was tun, wenn
  - keine Paneldaten verfügbar?
  - eine zeitvariierende Variable unbeobachtbar ist *und* durch keine ähnliche, aber beobachtbare Variable ersetzbar ist?
  - eine Variable nur mit Beobachtungsfehler verwendet werden kann (**error-in-variables Problematik**)?

- **Beispiel**:

$$log(wage) = \beta_0 + \beta_1 \cdot educ + \beta_2 \cdot abil + e.$$

  - **Proxy** für $abil$: z.B.: IQ.
    Beachte: $Corr(IQ, abil)$ sollte nahe 1 sein!
  - Die Variable $abil$ wird weggelassen

$$log(wage) = \beta_0 + \beta_1 \cdot educ + u.$$

OLS-Schätzer verzerrt und inkonsistent, wenn

$$Cov(educ, abil) \neq 0.$$

5.1 Instrumentvariablenschätzung im einfachen Regressionsmodell

Einfaches Regressionsmodell

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

- **Zur Erinnerung**:

$$Cov(y, x) = \beta_1 Var(x) + Cov(u, x).$$

  Gilt MLR.4, $E[u|x] = 0$, dann ist

$$\beta_1 = \frac{Cov(y, x)}{Cov(x, x)}.$$

- Analog gilt für eine beliebige Variable z

$$Cov(y, z) = \beta_1 Cov(x, z) + Cov(u, z). \tag{5.1}$$

Kann man annehmen, dass $Cov(u, z) = 0$ und $Cov(x, z) \neq 0$, erhält man analog

$$\beta_1 = \frac{Cov(y, z)}{Cov(x, z)}.$$

Eine Variable z, für die

$$Cov(u, z) = 0 \quad \text{und} \quad Cov(x, z) \neq 0 \tag{5.2}$$

gilt, heißt **Instrumentvariable (instrumental variable)** für x.

- Eine Instrumentvariable erlaubt die **Identifikation** von β_1, d.h. im gegebenen Fall, dass der Parameter β_1 durch Größen der Grundgesamtheit (z.B. Momente) eindeutig bestimmt ist. Dies ist eine Voraussetzung für die Konsistenz eines Schätzers. Warum?

- Der Parameter β_1 kann mit Hilfe des Momentenschätzers, vgl. Abschnitt 2.3 in Einführung in die Ökonometrie, August 2020, geschätzt werden, indem die Momente in (5.1) durch Stichprobenmittel geschätzt werden

$$\hat{\beta}_{1,IV} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}, \tag{5.3}$$

$$\hat{\beta}_{0,IV} = \bar{y} - \bar{x}\hat{\beta}_{1,IV}. \tag{5.4}$$

  Die Schätzer $\hat{\beta}_{0,IV}$ und $\hat{\beta}_{1,IV}$ heißen **Instrumentvariablenschätzer (IV-Schätzer) (instrumental variables (IV) estimator)**.

- Sind die Annahmen des einfachen Regressionsmodells erfüllt, wählt man x als Instrumentvariable und erhält den OLS-Schätzer.

- **Schätzeigenschaften**:
  - **Konsistenz**:
    - Sind die Voraussetzungen (5.2) und entsprechende Regularitätsbedingungen erfüllt, ist der IV-Schätzer **konsistent**. Beweisskizze wie in Zeitreihenökonometrie, Abschnitt 5.3.5.
    - Gilt $Cov(z, u) \neq 0$, ist der IV-Schätzer **verzerrt**.

– Asymptotische Normalverteilung

* Unter der Annahme bedingt homoskedastischer Fehler

$$Var(u|z) = \sigma^2 \tag{5.5}$$

ist der Instrumentvariablenschätzer (5.3) **asymptotisch normalverteilt**

$$\sqrt{n}(\hat{\beta}_{1,IV} - \beta_1) \xrightarrow{d} N\left(0, \frac{\sigma^2}{\sigma_x^2 \rho_{x,z}^2}\right)$$

mit asymptotischer Varianz

$$\frac{\sigma^2}{\sigma_x^2 \rho_{x,z}^2},$$

wobei $\rho_{x,z}$ die Korrelation zwischen dem Regressor x und der Instrumentvariablen z in der Grundgesamtheit ist. Beweisskizze wie in Zeitreihenökonometrie, Abschnitt 5.3.5.

Beachte: Je größer die Korrelation $\rho_{x,z}$ ist, desto geringer ist die Schätzvarianz, d.h. der OLS-Schätzer hat die geringste Schätzvarianz, *wenn* er angewendet werden kann.

* Der Begriff **asymptotische Varianz** wird in Wooldridge (2009, 2006), Gleichung (15.12) abweichend verwendet und bezeichnet die Approximation der $Var(\hat{\beta})$, die sich durch die asymptotische Verteilung ergibt, anstatt wie allgemein üblich $\lim_{n\to\infty} Var\left(\sqrt{n}(\hat{\beta}-\beta)\right)$
* Die exakte Verteilung des IV-Schätzers für Stichprobengröße n ist i.a. unbekannt.

– **Schätzung der asymptotischen Varianz**

- Varianz des Regressors in der Grundgesamtheit σ_x^2: kann durch die Stichprobenvariation SST_x/n geschätzt werden.
- Varianz des Fehlers σ^2: wie üblich — durch
$$\hat{\sigma}^2 = \frac{1}{n-2}\sum_{i=1}^{n}(y_i - \hat{\beta}_{0,IV} - \hat{\beta}_{1,IV}x_i)^2.$$
- Korrelation $\rho_{x,z} = Corr(x,z)$: zu schätzen durch $\sqrt{R^2_{x,z}}$ für Hilfsregression
$$x_i = b_0 + b_1 z_i + v_i, \quad i = 1, \ldots, n.$$
Da $R^2_{x,z} = \widehat{Corr}^2(x,\hat{x})$ (siehe Kursmaterial zu Einführung in die Ökonometrie, August 2020, Abschnitte 3.2 und 3.3), ergibt sich $\widehat{Corr}^2(x,z) = R^2_{x,z}$, da (geschätzte) Korrelationen unabhängig von Skalierungen sind.

* Insgesamt:

$$
\begin{aligned}
Var(\hat{\beta}_{1,IV}) &\approx \frac{\sigma^2}{n\sigma_x^2\rho_{x,z}^2}, \\
\widehat{Var}(\hat{\beta}_{1,IV}) &\approx \frac{\hat{\sigma}^2}{SST_x R_{x,z}^2} \\
&= \frac{\hat{\sigma}^2}{SSE_x}. \qquad (5.6)
\end{aligned}
$$

(5.6) ergibt sich aus $R_{x,z}^2 = 1 - \frac{SSR_x}{SST_x}$ und $SST_x = SSE_x + SSR_x$ und entsprechenden Umformungen.

– **Asymptotische Verzerrung** des IV-Schätzers bei Verletzung der Voraussetzungen (5.2) des IV-Schätzers :

$$\text{plim } \hat{\beta}_{1,IV} = \text{plim } \frac{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})}$$

— mit Regeln von plims (vgl. Zeitreihenökonometrie, Abschnitt 5.1)

—

$$= \frac{\text{plim } \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})}{Cov(z,x)}(\beta_0 - \beta_0) + \beta_1 \frac{Cov(z,x)}{Cov(z,x)} + \frac{\text{plim } \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(u_i - \bar{u})}{\text{plim } \frac{1}{n}\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})},$$

$$\text{plim } \hat{\beta}_{1,IV} = \beta_1 + \frac{Cov(z,u)}{Cov(z,x)} = \beta_1 + \frac{Corr(z,u)\sigma_z\sigma_u}{Corr(z,x)\sigma_z\sigma_x},$$

$$\text{plim } \hat{\beta}_{1,IV} = \beta_1 + \frac{Corr(z,u)}{Corr(z,x)}\frac{\sigma_u}{\sigma_x}.$$

Zum Vergleich: Asymptotische Verzerrung des OLS-Schätzers

$$\text{plim } \hat{\beta}_1 = \beta_1 + Corr(x,u)\frac{\sigma_u}{\sigma_x},$$

da $Corr(z,x) = Corr(z,z) = 1$.

Ergo: **IV nur besser, wenn**

$$\left|\frac{Corr(z,u)}{Corr(z,x)}\right| < |Corr(x,u)|.$$

Ist die IV-Variable nur (sehr) schwach mit Regressor korreliert, liefert der IV-Schätzer möglicherweise keine Verbesserung!

$\Longrightarrow$ Schätze immer $Corr(x,z)$.

- **Bestimmtheitsmaß**: R^2 bei IV-Schätzung:
  - kann negativ sein. Warum?
  - R^2 hat keine Interpretation, da Zerlegung $Var(y) = \beta_1^2 Var(x) + Var(u)$ nicht gilt. Was fehlt?
  - kann nicht zur Berechnung von F-Tests verwendet werden.
  - IV-Schätzung maximiert nicht R^2. Welcher Schätzer tut das?

5.2 Wann ist der OLS-Schätzer inkonsistent?

- **Inkonsistenz eines Schätzers durch fehlende Variablen**: (Wiederholung)
  Gegeben sei der wahre Zusammenhang

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon, \tag{5.7}$$

$$E[\varepsilon | x_1, x_2, x_3] = 0, \quad Cov(x_1, x_3) \neq 0, \tag{5.8}$$

  der mit dem Modell

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \tag{5.9}$$

  geschätzt werden soll.

– Der Fehler des Modells ist: $u = \beta_3 x_3 + \varepsilon$.

– MLR.4 verletzt, da

$$E[u|x_1, x_2] = \beta_3 E[x_3|x_1, x_2] \neq 0,$$

da wegen $Cov(x_1, x_3) \neq 0$ folgen muss, dass $E[x_3|x_1] \neq 0$. Denn würde $E[x_3|x_1] = 0$ folgen, wäre im Umkehrschluss $Cov(x_1, x_3) = 0$, was zu einem Widerspruch führt. Weiter folgt aus $E[x_3|x_1] \neq 0$, dass $E[x_3|x_1, x_2] \neq 0$, denn $E[\underbrace{E[x_3|x_1, x_2]}_{\neq 0}|x_1] = E[x_3|x_1] \neq 0$.

– Deshalb sind die OLS-Schätzer für die Parameter $\beta_0, \beta_1, \beta_2$ im Modell (5.9) inkonsistent.

– Lässt sich $E[x_3|x_1,x_2]$ durch ein lineares Modell darstellen

$$x_3 = \pi_0 + \pi_1 x_1 + \pi_2 x_2 + v,$$

lassen sich durch Berechnung $E[y|x_1,x_2]$ modifizierte Parameterwerte berechnen, für die der OLS-Schätzer konsistent ist. Diese werden häufig als **pseudo wahre Werte (pseudo true values)** bezeichnet. Hier ergibt sich

$$\begin{aligned} E[y|x_1,x_2] &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 E[x_3|x_1,x_2] \\ &= \underbrace{\beta_0 + \beta_3\pi_0}_{\beta_0^*} + (\underbrace{\beta_1 + \beta_3\pi_1}_{\beta_1^*})x_1 + (\underbrace{\beta_2 + \beta_3\pi_2}_{}\, y = \beta_0^* + \beta_1^* x_1 + \beta \end{aligned}$$

bzw. allgemein

$$y = E[y|x_1,x_2] + (\underbrace{y - E[y|x_1,x_2]}_{w}).$$

Beachte, dass

* per Konstruktion $E[w|x_1, x_2] = 0$;
* die Parameter β_1^* und β_2^* jeweils die **partiellen Korrelationen** zwischen y und x_1 bzw. x_2 angeben.

– Trotz $\beta_3 \neq 0$ kann in einer gegebenen Stichprobe der OLS-Schätzer für (5.9) unverzerrt sein, **wenn** die jeweiligen Korrelationen in der Stichprobe zwischen x_{i3} und x_{i1} bzw. x_{i2} Null sind, siehe Kursmaterial zu Einführung in die Ökonometrie, August 2020, Abschnitt 3.4.1.

- **Mehrgleichungssystem**: (neu)

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1, \tag{5.10}$$

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2, \tag{5.11}$$

$$E[u_1|z_1, z_2] = 0, \quad E[v_2|z_1, z_2] = 0. \tag{5.12}$$

Zeige, dass im Allgemeinen gilt:
$E[u_1|z_1, y_2] \neq 0$, da $Cov(u_1, v_2) = 0$ *nicht* vorausgesetzt wurde.

  – In diesem Gleichungssystem werden

    * z_1 und z_2 als **exogene** Variablen bezeichnet, da für sie MLR.4 gilt,
    * y_1 und y_2 als **endogene** Variablen bezeichnet, da für sie MLR.4 *nicht* gilt.

– Einteilung von Regressionsgleichungen:

  * Die Gleichung (5.10) wird als **strukturelle Gleichung (structural equation)** bezeichnet, da sie den Zusammenhang angibt, an dem man sachlich interessiert ist. Aber sie kann endogene Regressoren enthalten.

  * Gleichungen mit ausschließlich exogenen Variablen als Regressoren werden als **Gleichungen in reduzierter Form** bezeichnet, z.B. (5.11).

  * Natürlich kann auch eine Regressionsgleichung in reduzierter Form sachlich interessant sein, z.B. alle korrekt spezifizierten Regressionsmodelle, die MLR.1 bis MLR.4 erfüllen.

5.3 Einfacher IV-Schätzer

Ableitung des einfachen IV-Schätzers für einfache strukturelle Gleichungen

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + u_1. \tag{5.13}$$

- **Voraussetzungen**

  1. Voraussetzung: Vorliegen einer IV-Variablen z_k, die mit z_1,. . .,z_{k-1} nicht perfekt kollinear ist und zusammen für alle z_j, $j = 1, \ldots, k$, gilt:

$$E[u_1|z_1, \ldots, z_k] = 0. \tag{5.14}$$

  2. Voraussetzung: In der reduzierten Form

$$y_2 = \pi_0 + \pi_1 z_1 + \ldots + \pi_k z_k + v_2 \tag{5.15}$$

  muss $\pi_k \neq 0$ gelten. $Cov(y_2, z_k) \neq 0$ reicht alleine nicht aus. Wieso?

- Zeige, dass aus (5.14) folgt:

$$E[u_1] = 0, \quad Cov(u_1, z_j) = 0, \quad j = 1, \ldots, k. \tag{5.16}$$

- Anwenden des Momentenschätzers auf die Momente (5.16) liefert $k+1$ Gleichungen

$$\sum_{i=1}^{n} \left(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} - \ldots - \hat{\beta}_k z_{ik-1} \right) = 0,$$

$$\sum_{i=1}^{n} z_{ij} \left(y_{i1} - \hat{\beta}_0 - \hat{\beta}_1 y_{i2} - \hat{\beta}_2 z_{i1} - \ldots - \hat{\beta}_k z_{ik-1} \right) = 0,$$

$$j = 1, \ldots, k,$$

deren Lösung $\hat{\beta}$ den **einfachen IV-Schätzer** liefern. Zur Vereinfachung wurde $_{IV}$ weggelassen.

- **Überprüfung der Voraussetzungen**
  - Warum kann (5.14) bzw. (5.16) nicht überprüft werden?
  - Zum Überprüfen von $\pi_k \neq 0$ wird ein t-Test durchgeführt (Signifikanzniveau $\leq 5\%$).

    Weshalb sind hier vernachlässigte Variablen keine Gefahr?
- Die Variablen $z_1, \ldots, z_{k-1}$ sind ihre eigenen Instrumente (wie bei OLS).
- Ist z_k optimales Instrument? Im Allgemeinen nein, denn das **optimale Instrument** für y_2 ist $E[y_2|z_1, \ldots, z_k, z_{k+1}, \ldots]$, wobei z_{k+j}, $j = 1, 2, \ldots$ weitere denkbare exogene Variablen bezeichnen

$$y_2 = \underbrace{E[y_2|z_1, \ldots, z_k, \ldots]}_{y_2^*(z_1,\ldots,z_k,\ldots)} + \underbrace{(y_2 - E[y_2|z_1, \ldots, z_k, \ldots])}_{\text{Fehler}}.$$

Wie erhält man das optimale Instrument? Man kann es lediglich auf der Basis endlich vieler vorhandener exogener Variablen bzw. Instrumentvariablen schätzen. Ist die Zahl der vorliegenden Instrumente gerade k, dann schätzt der einfache IV-Schätzer automatisch das optimale Instrument, falls $E[y_2|z_1, \ldots, z_k, \ldots] = \pi_0 + \pi_1 z_1 + \cdots + \pi_k z_k$. Beweis mit Matrixalgebra nicht so schwer. Ist die Zahl der Instrumente größer, dann wird der allgemeine IV-Schätzer benötigt, siehe Abschnitt 5.4.

- **Matrixschreibweise**
  - $\mathbf{X}$: $(n \times (k+1))$ Matrix mit allen Beobachtungen für alle erklärenden Variablen der Schätzgleichung, also sowohl exogene als auch ggf. endogene Variable (hier: z_{i1}, ..., z_{ik-1}, y_{i2}).
  - $\mathbf{y}$: $(n \times 1)$ Vektor mit der endogenen Variablen der betrachteten Gleichung (hier: y_{i1}).
  - $\mathbf{Z}$: $(n \times (k+1)$ Matrix mit allen IV-Variablen, inkl. Konstante und exogene Variablen (hier: z_{i1}, ..., z_{ik}).
  - Beachte, dass sowohl $\mathbf{X}$ als auch $\mathbf{Z}$ genau $k+1$ Spalten aufweisen müssen und dass die Instrumentvariable zur endogenen Variable in der gleichen Spalte steht!

– Momentengleichungen in Matrixschreibweise
(analog zu Einführung in die Ökonometrie, August 2020, Abschnitt 3.3)

$$\mathbf{Z}'\mathbf{X}\hat{\boldsymbol{\beta}}_{IV} = \mathbf{Z}'\mathbf{y}.$$

– **Einfacher IV-Schätzer in Matrixschreibweise**

$$\hat{\boldsymbol{\beta}}_{IV} = \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\mathbf{y}, \tag{5.17}$$

wenn folgende Voraussetzung gegeben ist:

3. Voraussetzung: $\mathbf{Z}'\mathbf{X}$ invertierbar.

- **Probleme des einfachen IV-Schätzers**:
  - Optimales Instrument kann möglicherweise nicht geschätzt werden.
  - Möglicherweise liegen mehr Kandidaten als IV-Variable vor. Wie erfolgt eine optimale Auswahl?
  - Es liegen mehrere endogene Variablen als Regressoren vor.

Ausweg: allgemeiner IV-Schätzer.

5.4 Allgemeiner IV-Schätzer bzw. zweistufiger LS-Schätzer

5.4.1 Eine endogene erklärende Variable

- Die Einschränkung, dass für eine endogene erklärende Variable in

$$y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + u_1 \tag{5.13}$$

  genau eine Instrumentvariable ausgewählt werden muss, kann aufgegeben werden, so dass für y_2 mehrere Instrumentvariablen $z_k, z_{k+1}, \ldots, z_m$ verwendet werden können.

- Wird z.B. nur eine der Instrumentvariablen $z_k, z_{k+1}, \ldots, z_m$ im einfachen IV-Schätzer verwendet, ist dieser im Allgemeinen ineffizient, da in keinem Fall das optimale Instrument geschätzt wird.

- Die IV-Variablen, die *nicht* in der Strukturgleichung vorkommen ('excluded from the structural equation'), also $z_k, z_{k+1}, \ldots, z_m$, erfüllen die **Ausschlussrestriktionen (exclusion restrictions)**.
- Um die Zahl aller $(m - k + 1)$ Instrumentvariablen auf eine IV-Variable zu reduzieren, kann jede Linearkombination verwendet werden. Welche ist auszuwählen?

- Beste Linearkombination: Größtmögliche Korrelation zwischen Linearkombination aller IV-Variablen und y_2. Definiere

$$\mathbf{z} = \begin{pmatrix} 1 & z_1 & \cdots & z_m \end{pmatrix}, \quad \boldsymbol{\pi} = \begin{pmatrix} \pi_0 & \pi_1 & \cdots & \pi_m \end{pmatrix}'$$

und schätze mit OLS.

$$\begin{aligned} y_2 &= \pi_0 + \pi_1 z_1 + \ldots + \pi_k z_k + \pi_{k+1} z_{k+1} + \ldots + \pi_m z_m + v_2 \\ &= \mathbf{z}\boldsymbol{\pi} + v_2. \end{aligned}$$

Verwende

$$\hat{y}_2 = \mathbf{z}\hat{\boldsymbol{\pi}}$$

als IV-Variable im einfachen IV-Schätzer.

Ergebnis: 2-stufiger LS-Schätzer.

- **2-stufiger LS-Schätzer (two-stage least squares (2SLS) estimator)**:

  1. Stufe: Berechne $\hat{\boldsymbol{\pi}} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{y}_2$.

  2. Stufe: Ersetze in (5.13) y_2 durch $\hat{y}_2 = \mathbf{z}\hat{\boldsymbol{\pi}}$ aus der 1. Stufe und schätze

  $$y_1 = \beta_0 + \beta_1 \hat{y}_2 + \beta_2 z_1 + \cdots + \beta_k z_{k-1} + Fehler \tag{5.18}$$

  mit OLS.

  Der 2-stufige LS-Schätzer ist gleichzeitig der **allgemeine IV-Schätzer**.

- **2-stufiger LS-Schätzer in Matrixschreibweise**
  - Darstellung der Strukturgleichung (5.13) in Matrixform. Da später u.a. die Matrixdarstellung der 2. Stufe erleichtert wird, wenn die *endogene erklärende Variable in* (5.13) *als letzte erklärende Variable* aufgeführt wird, schreiben wir

$$y_1 = \beta_0 + \beta_1 z_1 + \cdots + \beta_{k-1} z_{k-1} + \beta_k y_2 + u_1 \qquad (5.13\text{'})$$

und definieren

$$\mathbf{x} = \begin{pmatrix} 1 & z_1 & \cdots & z_{k-1} & y_2 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 & \beta_1 & \cdots & \beta_k \end{pmatrix}',$$

$$y_1 = \mathbf{x}\boldsymbol{\beta} + u_1.$$

– Für die Matrixdarstellung des Regressionsmodell der Stichprobe definiere

$$\mathbf{y}_1 = \begin{pmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{n1} \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & z_{11} & \cdots & z_{1k-1} & y_{12} \\ 1 & z_{21} & \cdots & z_{2k-1} & y_{22} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & z_{n1} & \cdots & z_{nk-1} & y_{n2} \end{pmatrix}, \mathbf{u}_1 = \begin{pmatrix} u_{11} \\ u_{21} \\ \vdots \\ u_{n1} \end{pmatrix},$$

$$\mathbf{y}_1 = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_1.$$

– Zur Darstellung der 2. Stufe des 2-stufigen LS-Schätzers ersetze in $\mathbf{x}$ die letzte Spalte mit y_2 durch $\hat{y}_2$ und entsprechend in $\mathbf{X}$ die Spalte mit $\mathbf{y}_2$ durch $\hat{\mathbf{y}}_2 = \mathbf{Z}\hat{\boldsymbol{\pi}}$

$$\hat{\mathbf{X}} = \begin{pmatrix} 1 & z_{11} & \cdots & z_{1k-1} & \hat{y}_{12} \\ 1 & z_{21} & \cdots & z_{2k-1} & \hat{y}_{22} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & z_{n1} & \cdots & z_{nk-1} & \hat{y}_{n2} \end{pmatrix}.$$

Somit ergibt sich für (5.18)

$$\mathbf{y}_1 = \hat{\mathbf{X}}\boldsymbol{\beta} + Fehler$$

und der LS-Schätzer für die 2. Stufe lautet

$$\hat{\boldsymbol{\beta}}_{2SLS} = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}_1. \tag{5.19}$$

– Eine alternative Schreibweise ist (Beweis in Abschnitt 5.4.3)

$$\hat{\boldsymbol{\beta}}_{2SLS} = \left(\hat{\mathbf{X}}'\mathbf{X}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}_1.$$

Deshalb ist der 2-stufige LS-Schätzer ein einfacher IV-Schätzer mit $\hat{\mathbf{X}}$ als IV-Variablen und wird als **allgemeiner IV-Schätzer** bezeichnet.

– Im Folgenden wird der 2-stufige LS-Schätzer mit dem Index '2SLS' (2-stage least squares) versehen, um eine mit Wooldridge (2009, 2006) einheitliche Notation zu gewährleisten.

5.4.2 Mehrere endogene erklärende Variablen

- In Matrixschreibweise lässt sich leicht der Fall mit mehreren endogenen erklärenden Variablen darstellen.
  - **Beispiel**: 2 endogene erklärende Variablen y_2 und y_3, 3 Ausschlussrestriktionen (exclusion restrictions) z_3, z_4, z_5. Die Strukturgleichung lautet

$$y_1 = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 y_2 + \beta_4 y_3 + u_1. \qquad (5.20)$$

    Analog zu oben, schätzt man in der 2. Stufe mit OLS die Regression

$$y_1 = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 \hat{y}_2 + \beta_4 \hat{y}_3 + u_1,$$

    wobei für jede endogene erklärende Variable eine Regressionsgleichung in reduzierter Form vorliegt, die mit OLS geschätzt wird:

$$y_2 = \mathbf{z}\boldsymbol{\pi}_2 + v_2, \quad y_3 = \mathbf{z}\boldsymbol{\pi}_3 + v_3,$$
$$\boldsymbol{\pi}_j = \begin{pmatrix} \pi_{0j} & \pi_{1j} & \pi_{2j} & \cdots & \pi_{5j} \end{pmatrix}', \quad j = 2, 3.$$

– **Allgemein**:

1. Stufe: Jede endogene erklärende Variable y_j wird mit OLS geschätzt:

$$\hat{\mathbf{y}}_j = \mathbf{Z} \underbrace{\left(\mathbf{Z}'\mathbf{Z}\right)^{-1} \mathbf{Z}'\mathbf{y}_j}_{\hat{\boldsymbol{\pi}}_j} = \mathbf{Z}\hat{\boldsymbol{\pi}}_j. \tag{5.21}$$

2. Stufe: Jede endogene erklärende Variable y_j wird in $\mathbf{X}$ durch $\hat{y}_j$ ersetzt und die resultierende Matrix mit $\hat{\mathbf{X}}$ bezeichnet. Der 2-stufige LS-Schätzer entspricht (5.19)

$$\hat{\boldsymbol{\beta}}_{2SLS} = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1} \hat{\mathbf{X}}'\mathbf{y}_1.$$

- **Identifikation der Parameter in den Gleichungen**
  - Wie im Fall einer OLS-Regression auf $\mathbf{X}$ dürfen auch bei einer OLS-Regression auf $\hat{\mathbf{X}}$ die **Spalten in $\hat{\mathbf{X}}$ nicht linear abhängig** sein, denn **sonst ist $\hat{\mathbf{X}}'\hat{\mathbf{X}}$ nicht invertierbar** (vgl. MLR.3) und damit $\boldsymbol{\beta}$ nicht schätzbar.

    Das **Abzählkriterium** ist hierfür eine notwendige Bedingung: Es muss für jede endogene erklärende Variable mindestens eine weitere IV-Variable geben, die nicht in der Strukturgleichung vorkommt, d.h. die Zahl der exclusions restrictions ist größer gleich der Zahl der endogen erklärenden Variablen.

– Das Abzählkriterium ist nicht hinreichend, damit $\boldsymbol{\beta}$ in jedem Fall mit 2SLS geschätzt werden kann. Man betrachte in obigem Beispiel folgenden Fall mit $m = 5$:

$$y_2 = \mathbf{z} \begin{pmatrix} \pi_{02} \\ \pi_{12} \\ \pi_{22} \\ \pi_{32} \\ \pi_{42} \\ \pi_{52} \end{pmatrix} + v_2, \quad y_3 = \mathbf{z} \begin{pmatrix} \pi_{03} \\ \pi_{13} \\ \pi_{23} \\ 2\pi_{32} \\ 2\pi_{42} \\ 2\pi_{52} \end{pmatrix} + v_3.$$

Dann liegen für y_2 und y_3 de facto nur 3 IV-Variablen vor, da die Parameter für die exclusion restrictions (in Rot) linear abhängig sind. Die Folge ist, dass $\mathbf{z}\boldsymbol{\pi}_2$ und $\mathbf{z}\boldsymbol{\pi}_3$ keine zwei zusätzlichen IV-Variablen darstellen, da sie lediglich von z_1, z_2 und $z_3\pi_{32} + z_4\pi_{42} + z_5\pi_{52}$

abhängen. Um diesen Fall auszuschließen, muss im vorliegenden Fall gelten, dass für eine beliebige Konstante c und für alle (z_3, z_4, z_5) gilt:

$$z_3\pi_{32} + z_4\pi_{42} + z_5\pi_{52} \neq c(z_3\pi_{33} + z_4\pi_{43} + z_5\pi_{53})$$

Mit anderen Worten, es darf *nicht* gelten, dass

$$\frac{\pi_{32}}{\pi_{33}} = \frac{\pi_{42}}{\pi_{43}} = \frac{\pi_{52}}{\pi_{53}}.$$

Genau dann liegen 2 zusätzliche IV-Variablen für die beiden endogen erklärenden Variablen vor. Diese Bedingung wird als **Rangkriterium** bezeichnet. Diese Bedingung ist hinreichend und garantiert, dass die Parameter der Strukturgleichung und der Gleichungen in reduzierter Form **identifiziert** sind, d.h. eindeutig bestimmt sind.

– Um solche Fälle allgemein auszuschließen, fasst man alle Regressionsgleichungen für die endogenen erklärenden Variablen y_j in Matrixschreibweise zusammen (hier wieder für das Beispiel mit y_2 und y_3):

$$\begin{pmatrix} y_2 & y_3 \end{pmatrix} = \begin{pmatrix} 1 & z_1 & z_2 & z_3 & z_4 & z_5 \end{pmatrix} \begin{pmatrix} \pi_{02} & \pi_{03} \\ \pi_{12} & \pi_{13} \\ \pi_{22} & \pi_{23} \\ \pi_{32} & \pi_{33} \\ \pi_{42} & \pi_{43} \\ \pi_{52} & \pi_{53} \end{pmatrix} + \begin{pmatrix} v_2 & v_3 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & z_1 & z_2 & z_3 & z_4 & z_5 \end{pmatrix} \begin{pmatrix} \mathbf{\Pi}_s \\ \mathbf{\Pi}_e \end{pmatrix} + \begin{pmatrix} v_2 & v_3 \end{pmatrix}.$$

Die Parametermatrix $\mathbf{\Pi}_s$ bezieht sich auf die Instrumentvariablen, die in der Strukturgleichung vorkommen, deshalb der Index s, die Parametermatrix $\mathbf{\Pi}_e$ bezieht sich auf alle aus der Strukturgleichung ausgeschlossenen Variablen, also die exclusion restrictions, deshalb der Index e. Das **Rangkriterium** fordert nun, dass die zwei Spalten von $\mathbf{\Pi}_e$ nicht linear abhängig sind. Damit werden Fälle wie im obigen Beispiel vermieden.

– **Allgemein** lautet das **Rangkriterium**:
Im Fall k_y endogener erklärender Variablen und k_z exogener erklärender Variablen, $k = k_z + k_y$, muss gelten: Durch die zusätzlichen IV-Variablen (exclusion restrictions) $z_{k_z+1}, \ldots, z_m$ sind alle Parameter identifiziert, wenn die Parametermatrix $\mathbf{\Pi}_e$ mindestens Rang k_y hat, d.h. die Zahl der linear unabhängigen Spalten der Anzahl k_y endogener erklärender Variablen entspricht.

- Der allgemeine IV-Schätzer, alias 2-SLS Schätzer ist **optimal**, wenn für alle endogenen erklärenden Variablen gilt

$$
\begin{aligned}
E[y_j|z_1,\ldots,z_k,z_{k+1},\ldots,z_m,\ldots] = \\
E[y_j|z_1,\ldots,z_k,z_{k+1},\ldots,z_m] = \mathbf{z}\boldsymbol{\pi}_j,
\end{aligned} \tag{5.22}
$$

so dass bei Kenntnis von $\boldsymbol{\pi}_j$ das optimale Instrument für $\mathbf{y}_j$ vorliegen würde. Bei Unkenntnis von $\boldsymbol{\pi}_j$ wird das mittels OLS geschätzte optimale Instrument $\hat{y}_j$ verwendet (sofern das Rangkriterium erfüllt ist).

- **Voraussetzungen** für die Gültigkeit der IV-Variablen
  - **1. Voraussetzung**:

$$E[u_1|z_1,\ldots,z_k,z_{k+1},\ldots,z_m]=0. \tag{5.23}$$

Alle IV-Variablen sind exogen.

2. Voraussetzung: **Identifikationsvoraussetzung**:
Es liegen mindestens k_y Ausschlussrestriktionen vor. Das **Rangkriterium** ist erfüllt:

* Eine endogene erklärende Variable: Gegeben dass die z_j, $j=k,\ldots,m$, die Ausschlussrestriktionen erfüllen, gilt für mindestens ein
  $j=k,k+1,\ldots,m$, dass $\pi_j\neq 0$. (Die Nullhypothese $\pi_j=0$ für alle $j=k,\ldots,m$ kann mit einem F-Test getestet werden.)
* Mehrere endogene erklärende Variablen: $\mathbf{\Pi}_e$ hat Rang k_y.

- **Darstellung des 2-SLS-Schätzers als einstufiger Schätzer** in Matrixschreibweise möglich :

  Es lässt sich zeigen (siehe Abschnitt 5.4.3),

$$\hat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}. \tag{5.24}$$

  Definiere (die Projektionsmatrix)

$$\mathbf{P}_Z = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'.$$

  Dann lässt sich leicht zeigen, dass

$$\mathbf{P}'_Z = \mathbf{P}_Z = \mathbf{P}'_Z\mathbf{P}_Z.$$

  Somit gelten folgende äquivalente Darstellungen des 2-stufigen LS-

Schätzers

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{2SLS} &= \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}_1 \\
&= \left(\mathbf{X}'\mathbf{P}_Z'\mathbf{P}_Z\mathbf{X}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}_1 \\
&= \left(\mathbf{X}'\mathbf{P}_Z'\mathbf{X}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}_1 \\
&= \left(\hat{\mathbf{X}}'\mathbf{X}\right)^{-1}\hat{\mathbf{X}}'\mathbf{y}_1 \quad \text{(allg. IV-Schätzer)} \\
&= \left(\mathbf{X}'\mathbf{P}_Z\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y}_1 \quad \text{(einstufiger Schätzer).}
\end{aligned}
$$

- Die **Kovarianzmatrix** lässt sich schätzen durch

$$\widehat{Var}(\hat{\boldsymbol{\beta}}_{2SLS}) = \hat{\sigma}^2_{u_1} \left(\mathbf{X}'\mathbf{P_Z}\mathbf{X}\right)^{-1}$$
$$\hat{\sigma}^2_{u_1} = \hat{\mathbf{u}}'_{1,2SLS}\hat{\mathbf{u}}_{1,2SLS}/(n-k-1),$$

wobei

$$\hat{\mathbf{u}}_{1,2SLS} = \mathbf{y}_1 - \mathbf{X}\hat{\boldsymbol{\beta}}_{2SLS}$$

die 2SLS-Residuen bezeichnet. Man beachte, dass hier $\mathbf{X}$ und nicht $\hat{\mathbf{X}}$ verwendet wird!

Beweis: siehe Advanced Econometrics.

Beachte: Damit sind die Standardfehler des OLS-Schätzers in der 2. Stufe (5.19) falsch, denn sie basieren auf $\hat{\mathbf{X}}$!

Falls $m = k$, also genau eine IV-Variable für jede endogene erklärende Variable y_j vorliegt, sind der **einfache IV-Schätzer und der 2-stufige LS-Schätzer identisch**. Beweis siehe Abschnitt 5.4.3.

- Das j-te Diagonalelement von $\widehat{Var}(\hat{\boldsymbol{\beta}}_{2SLS})$ kann ähnlich wie bei OLS folgendermaßen berechnet werden:

$$\widehat{Var}(\hat{\boldsymbol{\beta}}_{j,2SLS}) = \frac{\hat{\sigma}^2_{u_1}}{SSE_j(1 - R^2_j)}, \tag{5.25}$$

wobei im Gegensatz zur OLS-Schätzung hier R^2_j das Bestimmtheitsmaß von $\hat{x}_j$ auf exogene Variable in der Strukturgleichung (z_l's, $\hat{y}_j$'s), nicht jedoch von x_j selbst bezeichnet. Hieraus folgt i.a. ein wesentlich höheres Bestimmtheitsmaß, so dass die Schätzvarianz wesentlich vergrößert wird. Das Problem der Multikollinearität verschärft sich beim IV-Schätzer.

- **Anzahl an IV-Variablen**:

	Model	Schätzer	Testen auf überidentifizierende Restriktionen
$m > k$	überidentifiziert	2SLS	möglich
$m = k$	exakt identifiziert	einf. IV-Schätzer	nicht möglich
$m < k$	unteridentifiziert	-	-

- **Annahmen für Konsistenz** und **asymptotische Normalität**

  2SLS.1 ($\Leftrightarrow$ **MLR.1**): Das Modell ist linear in den Parametern und korrekt spezifiziert

$$y_1 = \mathbf{x}\boldsymbol{\beta} + u_1.$$

  2SLS.2 ($\Rightarrow$ **MLR.2**): Zufallsstichprobe für y_1, y_j, $j = 2, \ldots, k_y$, $\mathbf{x}$,$\mathbf{z}$.

  2SLS.3

  * Die IV-Variablen sind nicht linear abhängig.
  * Das Rangkriterium ist erfüllt.

  2SLS.4: (5.23) gilt: $E[u_1|z_1, \ldots, z_k, z_{k+1}, \ldots, z_m] = 0$.

- Zusätzliche Annahme für asymptotisch korrekte Standardfehler und effizienten Schätzer

  2SLS.5: $E[u_1^2|\mathbf{z}] = \sigma^2_{u_1}$.

- Schwächere Annahmen für **Zeitreihendaten**: $(y_{t1}, \mathbf{x}_t, \mathbf{z}_t)$ sind schwach abhängig (weakly dependent) und die Fehler der Strukturgleichung sind nicht autokorreliert

$$E[u_{s1}u_{t1}|z_{s1},\ldots,z_{sk},z_{s,k+1},\ldots,z_{sm},z_{t1},\ldots,z_{tk},z_{t,k+1},\ldots,z_{tm}] = 0 \quad \text{für } s \neq t.$$

Testen mehrerer Restriktionen bezüglich β

F-Teststatistik kann nicht wie bei OLS berechnet werden, wird jedoch in Standardökonometriepaketen korrekt berechnet, siehe Wooldridge (2009, 2006) für Literaturhinweise.

Messfehler (Error-in-variables)

Statt der erklärenden Variable x_j^* kann nur x_j

$$x_j = x_j^* + e_j$$

beobachtet werden, wobei e_j als **Messfehler** bezeichnet wird. Wird x_j in einer Regression verwendet, ist x_j endogen, da mit e_j im Fehlerterm korreliert.

IV-Schätzer bzw. 2SLS anwendbar.

Mehr dazu in Wooldridge (2009, Sections 9.4 und 15.4: 2SLS-Schätzer) bzw. Wooldridge (2006, Sections 9.3 und 15.4: 2SLS-Schätzer).

5.4.3 Appendix: Beweise (optional)

Beweis von (5.24) $\hat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}$:

- Da für die Inverse gilt

$$\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{Z} = \mathbf{I}_{m+1},$$

folgt

$$\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\begin{pmatrix} 1 & z_{11} & \cdots & z_{1j} & \cdots & z_{1m} \\ 1 & z_{21} & \cdots & z_{2j} & \cdots & z_{2m} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 1 & z_{n1} & \cdots & z_{nj} & \cdots & z_{nm} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}.$$

Für die j-te IV-Variable gilt also

$$\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\begin{pmatrix} z_{1j} \\ z_{2j} \\ \vdots \\ z_{nj} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}.$$

Multipliziert man diese Gleichung von links mit $\mathbf{Z}$ erhält man

$$\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{z}_j = \mathbf{z}_j.$$

- Damit erhält man

$$\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\begin{pmatrix} z_{1j} \\ z_{2j} \\ \vdots \\ z_{nj} \end{pmatrix} = \begin{pmatrix} z_{1j} \\ z_{2j} \\ \vdots \\ z_{nj} \end{pmatrix}$$

und zusammen mit (5.21) ergibt sich

$$\begin{pmatrix} 1 & z_{11} & \cdots & z_{1k-1} & \hat{y}_{12} \\ 1 & z_{21} & \cdots & z_{2k-1} & \hat{y}_{22} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & z_{n1} & \cdots & z_{nk-1} & \hat{y}_{n2} \end{pmatrix} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\begin{pmatrix} 1 & z_{11} & \cdots & z_{1k-1} & y_{12} \\ 1 & z_{21} & \cdots & z_{2k-1} & y_{22} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & z_{n1} & \cdots & z_{nk-1} & y_{n2} \end{pmatrix}$$

$$\hat{\mathbf{X}} = \mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X},$$

wobei dies auch für mehrere y_j gilt. q.e.d.

Beweis: Einfacher IV-Schätzer gleich 2SLS-Schätzer bei $m = k$:

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{2SLS} &= \left(\mathbf{X}'\mathbf{P}_Z\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y}_1 \quad \text{(2SLS-Schätzer)} \\
&= \left(\underbrace{\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}}_{\mathbf{W}}\mathbf{Z}'\mathbf{X}\right)^{-1} \underbrace{\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}}_{\mathbf{W}}\mathbf{Z}'\mathbf{y}_1 \\
&= \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\underbrace{\mathbf{W}^{-1}\mathbf{W}}_{=\mathbf{I}_{k+1}}\mathbf{Z}'\mathbf{y}_1 \\
&= \left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'\mathbf{y}_1 \quad \text{(einfacher IV-Schätzer).}
\end{aligned}
$$

Beachte, dass $\mathbf{W}$ eine quadratische Matrix mit $k = m$ Spalten und Zeilen ist. Gilt das Rangkriterium, ist damit $\mathbf{W}$ invertierbar und fällt aus der Gleichung heraus. Ist $m > k$, ist $\mathbf{W}$ nicht quadratisch und nicht invertierbar und die beiden Schätzer unterscheiden sich.

5.5 Endogenität und überidentifizierende Restriktionen testen

- Sind alle erklärenden Variablen exogen, ist OLS effizient und 2SLS ineffizient. Verwende dann OLS!

- **Testen auf Endogenität**

  – Zur Erinnerung: Strukturgleichung mit einer endogenen erklärenden Variable

$$y_1 = \beta_0 + \beta_1 z_1 + \ldots + \beta_{k-1} z_{k-1} + \beta_k y_2 + u_1, \tag{5.26}$$

$$y_2 = \mathbf{z}\boldsymbol{\pi} + v_2. \tag{5.27}$$

  Das Endogenitätsproblem tritt nur auf, wenn u_1 und v_2 korreliert sind, also $Cov(u_1, v_2) \neq 0$, da $\mathbf{z}\boldsymbol{\pi}$ exogen ist.

Wenn $Cov(u_1, v_2) \neq 0$, muss in der Regressionsgleichung

$$u_1 = \delta_1 v_2 + e_1, \quad E[e_1|v_2] = 0$$

gelten, dass $\delta_1 \neq 0$. Sonst wäre die Kovarianz ja Null.

– Einsetzen von u_1 in Regressionsgleichung ergibt

$$y_1 = \beta_0 + \beta_1 z_1 + \beta_2 z_2 + \ldots + \beta_k z_{k-1} + \beta_k y_2 + \delta_1 v_2 + e_1.$$

Da v_2 unbeobachtbar ist, wird v_2 durch $\hat{v}_2 = y_2 - \mathbf{z}\hat{\boldsymbol{\pi}}$ aus 1. Stufe ersetzt und ein (asymptotischer) t-Test mit $H_0 : \delta_1 = 0$ vs. $H_1 : \delta_1 \neq 0$ für die Regression

$$y_1 = \beta_0 + \beta_1 z_1 + \ldots + \beta_{k-1} z_{k-1} + \beta_k y_2 + \delta_1 \hat{v}_2 + \tilde{e}_1 \qquad (5.28)$$

durchgeführt. Da $\hat{v}_2$ eingesetzt wurde, liegt nicht mehr e_1 vor, sondern ein modifizierter Fehler $\tilde{e}_1$.

– Beachte: Der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ in (5.28) entspricht gerade $\hat{\boldsymbol{\beta}}_{2SLS}$. (Ohne Beweis.)

– Vergleich der OLS-Schätzungen für β_k auf Basis von (5.26) und (5.28) zeigt, ob Berücksichtigung von potentieller Endogenität praktisch relevant ist.

Beachte: Unter $H_0 : \delta_1 = 0$ sind sowohl OLS als auch 2SLS-Schätzer für (5.26) konsistent, jedoch nur OLS effizient.

Der **Hausman-Test** (Hausman, 1978) nutzt diese Eigenschaft und testet allgemein, ob die Differenz zwischen $\hat{\boldsymbol{\beta}}$ und $\hat{\boldsymbol{\beta}}_{2SLS}$ 'genügend' groß ist, auch im Fall mehrerer endogener erklärender Variablen.

– Liegen mehrere endogene erklärende Variablen vor, dann wird die gemeinsame $H_0 : \delta_1 = \ldots = \delta_{m-k+1} = 0$ in

$$y_1 = \mathbf{x}\boldsymbol{\beta} + \delta_1 \hat{v}_2 + \cdots + \delta_{m-k+1}\hat{v}_{m-k+2} + e$$

getestet.

– Der Hausman-Test ist ein so genannter Pretest, da von dessen Ausgang die Wahl des Schätzverfahrens abhängt. Guggenberger (2009) analysiert, wann dieses Verfahren auch asymptotisch unzuverlässig ist.

- **Testen der 'overidentifying restrictions'**
  - Da $\hat{\boldsymbol{\beta}}_{2SLS} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}_1$, entspricht $\hat{\mathbf{b}}$ von

    $$\hat{\mathbf{u}}_1 = \hat{\mathbf{X}}\mathbf{b} + Fehler$$

    gerade Null. Ist $m = k$, gilt dies auch für

    $$\hat{\mathbf{u}}_1 = \mathbf{Z}\mathbf{d} + Fehler, \tag{5.29}$$

    da $\hat{\mathbf{X}}$ und $\mathbf{Z}$ linear abhängig sind. (Siehe Beweise in Abschnitt 8.4.)
  - Ist jedoch $m > k$, dann sind $\hat{\mathbf{X}}$ und $\mathbf{Z}$ nicht linear abhängig und man kann die Regression (5.29) durchführen und testen, ob $H_0 : \mathbf{d} = 0$ gilt. Ist dies nicht der Fall, sind einige der überidentifizierenden IV-Variablen mit u_1 korreliert und man sollte die IV-Variablen eliminieren, für die $\hat{d}_j \neq 0$.

6 Simultane Gleichungsmodelle

- Ursachen für **Endogenität** einer erklärenden Variable:
  - fehlende/vernachlässigte erklärende Variablen (Abschnitt 5.2)
  - Messfehler in den Variablen (Abschnitt 5.4)
  - **Simultanität (simultaneity)** (dieses Kapitel): mindestens zwei endogene Variablen sind jeweils in einer Gleichung abhängige Variable und in einer anderen Gleichung erklärende Variable.

- **Beispiel**: Modellierung des Milchkonsums:

$$\begin{aligned} \text{Angebotsgleichung:} \quad q_A &= \alpha_1 p + \beta_1 z_1 + \tilde{u}_1. \\ \text{Nachfragegleichung:} \quad q_N &= \alpha_2 p + \beta_2 z_2 + \tilde{u}_2. \\ \text{Gleichgewichtsbedingung:} \quad q_{iA} &= q_{iN} = q_i \end{aligned}$$

mit q_A, q_{iA} : Angebot, bzw. Angebot in Landkreis i

q_N, q_{iN} : Nachfrage, bzw. Nachfrage in Landkreis i

p_i : Preis für Milch

z_1, z_{i1} : Preis für Tierfutter

z_2, z_{i2} : Haushaltseinkommen

ergibt nach Einsetzen der Gleichgewichtsbedingung ein **simultanes Gleichungssystem (simultaneous equation model) (SEM)**:

$$\text{Angebotsgleichung: } q_i = \alpha_1 p_i + \beta_1 z_{i1} + \tilde{u}_{i1}.$$

$$\text{Nachfragegleichung: } q_i = \alpha_2 p_i + \beta_2 z_{i2} + \tilde{u}_{i2}.$$

- **Beobachtbare** q_i und p_i ergeben sich hier durch Interaktion von Angebot und Nachfrage.
- **Strukturelle Gleichungen**
  - sind aus (ökonomischer) Theorie abgeleitet (z.B. Gewinn- oder Nutzenmaximierung oder Produktionsfunktion) $\Rightarrow$ **Verhaltensgleichung** oder **technischer Zusammenhang**.
  - beschreiben häufig **nichtbeobachtbare Zustände (counterfactual situations)** und geben damit Antworten auf 'counterfactual questions': 'Was wäre wenn' ohne dass jeweiliger Zusammenhang empirisch beobachtbar ist. (Im Beispiel: z.B.: Höhe des Milchangebots bei verdoppeltem Preis.)
  - erlauben **kausale Interpretation**, d.h. ceteris-paribus Interpretation der Parameter möglich (z.B. α_1 Angebotselastizität, falls p und q_A in logs).

6.1 Alternative Darstellungen von simultanen Gleichungssystemen

- **Alternative Darstellungen** von

$$q_i = \alpha_1 p_i + \beta_1 z_{i1} + \tilde{u}_{i1}, \tag{6.1a}$$
$$q_i = \alpha_2 p_i + \beta_2 z_{i2} + \tilde{u}_{i2}. \tag{6.1b}$$

Alle endogenen Variablen auf der linken Seite

$$\begin{aligned} q_i - \alpha_1 p_i &= \beta_1 z_{i1} + \qquad\qquad \tilde{u}_{i1}, \\ q_i - \alpha_2 p_i &= \qquad\qquad \beta_2 z_{i2} + \tilde{u}_{i2}. \end{aligned}$$

In beiden Gleichungen ist der Parameter vor q_i Eins. Man kann die beiden Gleichungen anders **normieren**, so dass in jeder Gleichung eine andere endogene Variable eine Eins als 'Parameter' stehen hat. Dies ist vorteilhaft, wenn die Gleichungen in Matrixform geschrieben werden. Hier gibt es zwei Möglichkeiten:

I. Angebotsgleichung bleibt unverändert :

$$
\begin{array}{rrcll}
q_i & -\alpha_1 p_i & = & \beta_1 z_{i1} & +\tilde{u}_{i1}, \\
-\frac{1}{\alpha_2} q_i & +p_i & = & -\frac{\beta_2}{\alpha_2} z_{i2} & -\frac{1}{\alpha_2}\tilde{u}_{i2},
\end{array}
$$

bzw. mit jeweils einer endogenen Variablen auf der linken Seite

$$
q_i = \alpha_1 p_i + \beta_1 z_{i1} + \tilde{u}_{i1}, \tag{6.2a}
$$

$$
p_i = \frac{1}{\alpha_2} q_i - \frac{\beta_2}{\alpha_2} z_{i2} - \frac{1}{\alpha_2}\tilde{u}_{i2}. \tag{6.2b}
$$

II. Nachfragegleichung bleibt unverändert:

$$\begin{aligned} -\frac{1}{\alpha_1}q_i \quad + p_i &= -\frac{\beta_1}{\alpha_1}z_{i1} \qquad -\frac{1}{\alpha_1}\tilde{u}_{i1}, \\ q_i - \alpha_2 p_i &= \qquad \beta_2 z_{i2} \qquad + \tilde{u}_{i2}, \end{aligned}$$

bzw. mit jeweils einer endogenen Variablen auf der linken Seite

$$p_i = \frac{1}{\alpha_1}q_i \qquad -\frac{\beta_1}{\alpha_1}z_{i1} \qquad -\frac{1}{\alpha_1}\tilde{u}_{i1}, \tag{6.3a}$$

$$q_i = \qquad \alpha_2 p_i \qquad + \beta_2 z_{i2} \qquad + \tilde{u}_{i2}. \tag{6.3b}$$

Beachte:

– Welche Variante man wählt, hängt davon ab, welche Parameter geschätzt werden sollen. Es macht z.B. keinen Sinn Gleichung (6.3a) zu schätzen, wenn man an α_1 interessiert ist, da α_1 hier im Nenner steht.

– Die Matrixform ist wichtig, falls man alle Gleichungen gleichzeitig schätzen will. Das ist mit dem 3SLS-Schätzer möglich, der hier nicht besprochen wird.

- **Darstellung von I. in Matrixform** mit $u_{i1} = \tilde{u}_{i1}, u_{i2} = -\tilde{u}_{i2}/\alpha_2$:

$$\begin{array}{ccccccc} q_i & -\alpha_1 & p_i & = & \beta_1 \; z_{i1} & + & u_{i1} \\ \frac{-1}{\alpha_2} \; q_i & + & p_i & = & \frac{-\beta_2}{\alpha_2} \; z_{i2} & + & u_{i2} \end{array}$$

$$\underbrace{\begin{pmatrix} q_i & p_i \end{pmatrix}}_{\mathbf{y}_i} \underbrace{\begin{pmatrix} 1 & \frac{-1}{\alpha_2} \\ -\alpha_1 & 1 \end{pmatrix}}_{\mathbf{A}} = \underbrace{\begin{pmatrix} z_{i1} & z_{i2} \end{pmatrix}}_{\mathbf{z}_i} \underbrace{\begin{pmatrix} \beta_1 & 0 \\ 0 & \frac{-\beta_2}{\alpha_2} \end{pmatrix}}_{\mathbf{B}} + \underbrace{\begin{pmatrix} u_{i1} & u_{i2} \end{pmatrix}}_{\mathbf{u}_i}$$

$$\mathbf{y}_i \mathbf{A} = \mathbf{z}_i \mathbf{B} + \mathbf{u}_i \quad \text{(strukturelle Gleichungen)}$$

$$\mathbf{y}_i = \mathbf{z}_i \underbrace{\mathbf{B}\mathbf{A}^{-1}}_{\mathbf{\Pi}} + \underbrace{\mathbf{u}_i \mathbf{A}^{-1}}_{\mathbf{v}_i}$$

$$\mathbf{y}_i = \mathbf{z}_i \mathbf{\Pi} + \mathbf{v}_i \quad \text{(reduzierte Form).}$$

Anmerkungen:

– **Jede Gleichung** erhält in der Matrixdarstellung des Systems eine **eigene Spalte**.

– Wie in Abschnitt 5.4 gibt es **für jede endogene Variable** eine Gleichung. Im Gegensatz zu Abschnitt 5.4 ist im Allgemeinen jede Gleichung **eine strukturelle Gleichung**.

– **Doppelfunktion endogener Variablen**: sie können im Gleichungssystem sowohl die Funktion einer abhängigen *als auch* erklärenden Variablen haben.

– Das Gleichungssystem in **reduzierter Form**

* ist nicht mehr simultan (bis auf die mögliche gleichzeitige Korrelation zwischen den Fehlern),
* enthält keine endogenen Variablen als erklärenden Variablen und
* hat Parameter, die nichtlineare Funktionen der Parameter der strukturellen Form und normalerweise (ökonomisch) nicht interpretierbar sind;
* existiert nur, wenn $\mathbf{A}$ invertierbar ist! Im Beispiel heißt das, dass $det(\mathbf{A}) = 1 - \frac{\alpha_1}{\alpha_2} \neq 0$, also $\alpha_1 \neq \alpha_2$,
* ist unabhängig von der Normierung.

– **Endogene** Variablen $\mathbf{y}$ lassen sich in Abhängigkeit der **exogenen** Variablen $\mathbf{z}$ und der Fehler $\mathbf{v}$ eindeutig bestimmen, wenn reduzierte Form existiert.

– Für jede strukturelle Gleichung lassen sich das Abzählkriterium und das Rangkriterium anhand von $\mathbf{A}$ und $\mathbf{B}$ überprüfen und somit bestimmen, ob die jeweilige Gleichung identifiziert ist, siehe folgenden Abschnitt.

– Ist $A_{kj} \neq 0$ und $Cov(v_k, u_j) \neq 0$, ist Variable y_k endogen für die j-te Strukturgleichung $\Rightarrow$ OLS-Schätzer verzerrt durch **'simultaneity bias'**, siehe folgendes Beispiel.

– **Indexierung der Parametermatrizen $\mathbf{A}$ und $\mathbf{B}$:**

 * Erster Index: Nummerierung der Parameter bezüglich der endogenen bzw. exogenen Variablen.

 * Zweiter Index: Nummer der strukturellen Gleichung.

 Beispiel:

 1. Gleichung: $A_{11} = 1, \quad A_{21} = -\alpha_1, \quad B_{11} = \beta_1, \quad B_{21} = 0.$
 2. Gleichung: $A_{12} = -\dfrac{1}{\alpha_2}, \quad A_{22} = 1, \quad B_{12} = 0, \quad B_{22} = -\dfrac{\beta_2}{\alpha_2}.$

– Die Matrizen $\mathbf{A}$ und $\mathbf{B}$ sind nicht eindeutig bestimmt, denn man hätte auch die Reihenfolge der Gleichungen vertauschen können und dann durch $-\alpha_1$ dividieren können.

- **Beispiel für 'simultaneity bias'**: Milchkonsum
  - Reduzierte Form für Preis (ableiten entweder durch Invertieren von $\mathbf{A}$ oder durch Einsetzen der ersten in die zweite Gleichung):

$$p = \overbrace{\frac{\beta_1}{\alpha_2 - \alpha_1}}^{\pi_{12}} z_1 - \overbrace{\frac{\beta_2}{\alpha_2 - \alpha_1}}^{\pi_{22}} z_2 + \underbrace{\frac{1}{\alpha_2 - \alpha_1} u_1 + \frac{\alpha_2}{\alpha_2 - \alpha_1} u_2}_{v_2},$$

$$v_2 = \frac{1}{\alpha_2 - \alpha_1} u_1 + \frac{\alpha_2}{\alpha_2 - \alpha_1} u_2.$$

Daraus ergibt sich

$$Cov(v_2, u_1) = \frac{1}{\alpha_2 - \alpha_1} \sigma_{u_1}^2 + \frac{\alpha_2}{\alpha_2 - \alpha_1} Cov(u_1, u_2),$$

wobei der zweite Term Null ist, falls u_1 und u_2 unkorreliert sind. Damit ist i.a. $Cov(v_2, u_1) \neq 0$ und die OLS-Schätzer der Parameter der Angebotsgleichung verzerrt.

– Möchte man die Angebotsgleichung schätzen (Fall I.) und nimmt an, dass $\beta_1 = 0$, erhält man die Gleichungen

$$
\begin{aligned}
q &= \alpha_1 p + u_1, \\
p &= -\frac{\beta_2}{\alpha_2 - \alpha_1} z_2 + \underbrace{\frac{1}{\alpha_2 - \alpha_1} u_1 + \frac{\alpha_2}{\alpha_2 - \alpha_1} u_2}_{v_2}.
\end{aligned}
$$

Nimmt man weiter an, dass $Cov(u_1, u_2) = 0$, lässt sich das Ausmaß und die Richtung der asymptotischen Verzerrung des OLS-Schätzers von α_1 (vgl. Abschnitt 5.1) angeben durch

$$
\begin{aligned}
\text{plim } \hat{\alpha}_1 &= \frac{Cov(q, p)}{Cov(p, p)} \left(= \frac{Cov(y_1, y_2)}{Cov(y_2, y_2)} \right) \\
&= \alpha_1 + \frac{Cov(u_1, p)}{Cov(p, p)}
\end{aligned}
$$

$$= \alpha_1 + \frac{Cov(u_1, p)}{Var(p)}$$
$$= \alpha_1 + \frac{\sigma_{u_1}^2}{(\alpha_2 - \alpha_1)Var(p)}.$$

Interpretieren Sie dieses Ergebnis. Allgemeinere Fälle erfordern Matrixalgebra.

– Hinweis: In Section 16.2 in Wooldridge (2009, 2006) werden die Bezeichnungen α_1 und α_2 abweichend vom Beispiel des Milchkonsums verwendet. Um Verwirrung mit der bisher verwendeten Definition zu vermeiden, werden die Parameter entsprechend Wooldridge als a_1 und a_2 bezeichnet. Sie sind definiert als $a_1 = -A_{21} = \alpha_1$ und $a_2 = -A_{12} = 1/\alpha_2$:

$$\text{plim } \hat{a}_1 = a_1 + \frac{a_2 \sigma_{u_1}^2}{(1 - a_1 a_2)Var(y_2)}.$$

6.2 Identifikation von simultanen Gleichungsmodellen

Zweigleichungssystem

- Illustration anhand eines einfachen Falls:
  Beispiel Milchkonsum (6.1a), (6.1b) mit $\beta_2 = 0$:

$$q = \alpha_1 p + \beta_1 z_1 + \tilde{u}_1, \tag{6.1a}$$

$$q = \alpha_2 p + \tilde{u}_2. \tag{6.1b'}$$

  Die Preisvariable p ist in beiden Gleichungen endogen. Welche IV-Variable kommt jeweils in Frage? (Vgl. Kapitel 5.)

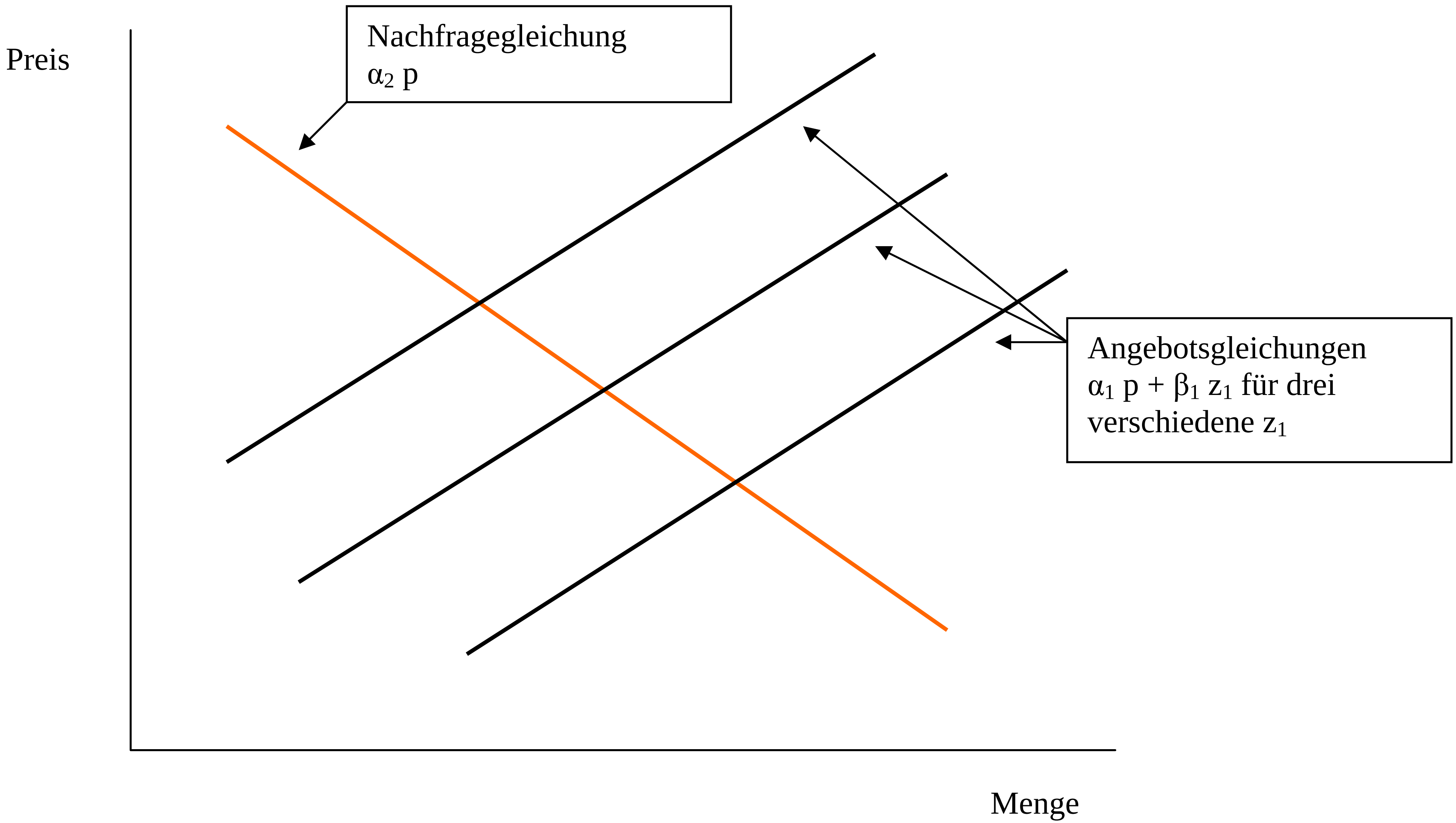
  – Angebotsgleichung (6.1a): die exogene Variable des Systems, z_1, ist bereits in Angebotsgleichung enthalten; für Anwendbarkeit des IV-Schätzers müsste *zusätzliche* exogene Variable z_2 vorhanden sein, die Voraussetzungen als IV-Variable für p erfüllt.

– Nachfragegleichung (6.1b'): z_1 ist potentielle IV-Variable für p, da diese wegen **Ausschlussrestriktion** $B_{21} = 0$ nicht in Nachfragegleichung vorkommt (Abzählkriterium). Damit z_1 verwendbare IV-Variable ist, ist notwendig, dass diese in andere Gleichung eingeht, d.h. $\beta_1 \neq 0$ (Rangkriterium).

– Im Gegensatz zur Schätzung von Einzelgleichungen muss man bei simultanen Gleichungssystemen nicht unbedingt nach potentiellen IV-Variablen *suchen*: Alle exogenen Variablen, die in einer Gleichung nicht vorkommen, sind potentielle IV-Variablen für etwaige endogene Variablen.

– weitere Voraussetzung: $\alpha_1 \neq \alpha_2$, d.h. Angebots- und Nachfragegleichung reagieren auf Preisänderung ceteris paribus unterschiedlich!

– ökonomische Interpretation



* z_1 **verschiebt Angebotskurve in beobachtbarer Weise**. Dies erlaubt Identifikation der Nachfragekurve (in perfekter Weise, wenn Fehler Null sind).

* Für Nachfragekurve keine beobachtbare Verschiebung vorhanden. Damit Angebotskurve nicht identifizierbar.

* Unbeobachtbare Verschiebungen durch $\tilde{u}_1$ und $\tilde{u}_2$ helfen für die Identifikation nichts, führen aber zu Schätzfehlern in der Bestimmung der strukturellen Parameter (hier α_1).

– Alternative Darstellung von (6.1a), (6.1b') (mit $\beta_2 = 0$)

$$q = \alpha_1 p + \beta_1 z_1 + u_1 \tag{6.2a}$$

$$p = \frac{1}{\alpha_2} q \qquad\qquad + u_2 \tag{6.2b'}$$

macht deutlich, dass im **simultanen Gleichungssystem für jede endogene Variable und damit auch für jede endogen erklärende Variable eine strukturelle Gleichung vorliegt.** Für 2SLS-Schätzung einer Gleichung ist jedoch nur erforderlich, dass die jeweils in einer Gleichung vorkommenden endogen erklärende Variablen durch eine reduzierte Form schätzbar sind. Vgl. hierzu Abschnitt 5.4.

– 2SLS-Schätzung von α_2 in der Nachfragefunktion (6.1b') lässt sich nicht gut mit (6.2b') durchführen - warum? - sondern direkt auf Basis der strukturellen Gleichung (6.1b') und einer Gleichung für p in reduzierter Form

$$p = \mathbf{z}\pi_p + v_p. \tag{6.5a}$$

Im vorliegenden Fall einer IV-Variable entspricht dies dem einfachen IV-Schätzer mit z_1 als IV-Variable.

– Wäre die Angebotsgleichung identifiziert, würde man analog vorgehen und

$$q = \mathbf{z}\pi_q + v_q \tag{6.5b}$$

mit OLS schätzen.

- Enthalten die strukturellen Gleichungen **mehr exogene Variablen**, dann bleibt das obige Vorgehen gleich. Die Identifikation für jede strukturelle Gleichung wird mit dem **Rangkriterium** bzw. zumindest mit dem **Abzählkriterium** überprüft. Vgl. Abschnitt 5.4.

Mehrgleichungssystem

- Im Fall von mehr als zwei Gleichungen können mehr als eine endogen erklärende Variable in einer spezifischen strukturellen Gleichung vorkommen. Die Identifikation jeder strukturellen Gleichung wird wieder mit dem **Rangkriterium** bzw. zumindest mit dem **Abzählkriterium** überprüft. Vgl. Abschnitt 5.4.

- Wie im Zweigleichungsfall kann es sein, dass nicht die Parameter aller strukturellen Gleichungen identifiziert sind.

- Zur Schätzung jeder strukturellen Gleichung wird der 2SLS-Schätzer verwendet.

- Es ist möglich, alle Gleichungen gleichzeitig mit Hilfe des **3SLS-Schätzers** zu schätzen, (siehe hierzu **Methoden der Ökonometrie** im Master-Studium). Dies kann sinnvoll sein, wenn die Fehler der jeweiligen Gleichungen korreliert sind.

- **Grundproblematik simultaner Gleichungssysteme in Matrixschreibweise**:
  - die Parameter der reduzierten Form $\mathbf{\Pi}$ lassen sich, gegeben deren korrekte Spezifikation, per OLS konsistent schätzen. Aus $\hat{\mathbf{\Pi}}$ lassen sich jedoch nicht notwendigerweise $\hat{\mathbf{A}}$ und $\hat{\mathbf{B}}$ gewinnen. Hierfür ist Voraussetzung, dass
  $$\mathbf{\Pi A} = \mathbf{B} \tag{6.6}$$
  eindeutig lösbar ist!
  - Es kann vorkommen, dass (6.6) nicht allgemein eindeutig lösbar ist, jedoch für einzelne Spalten, d.h. einzelne Strukturgleichungen.
  - Beachte, dass nur die Matrix $\mathbf{A}$ quadratisch ist und damit invertierbar.

- **Beispiel**: Gleichungen (16.27)-(16.29) in Wooldridge (2009, 2006).

  Beachte: Indexierung der Variablen und Parameter in Gleichungssystemen in diesen Folien unterscheidet sich von der in Wooldridge (2009, 2006):

$$y_1 = -y_2 A_{21} - y_3 A_{31} + z_1 B_{11} + u_1, \tag{6.7a}$$
$$y_2 = -y_1 A_{12} + z_1 B_{12} + z_2 B_{22} + z_3 B_{32} + u_2, \tag{6.7b}$$
$$y_3 = -y_2 A_{23} + z_1 B_{13} + z_2 B_{23} + z_3 B_{33} + z_4 B_{43} + u_3. \tag{6.7c}$$

  Grund: Verwendung von Matrixschreibweise.

  – Zur **Schätzung der strukturellen Gleichung für** y_1 muss die reduzierte Form für y_2 *und* y_3 existieren.

Zur Überprüfung des Rangkriteriums muss geprüft werden, ob $\mathbf{\Pi}_e$ Rang **zwei** hat, da zwei endogen erklärende Variablen vorhanden sind:

$$
\begin{aligned}
\begin{pmatrix} y_2 & y_3 \end{pmatrix} &= \begin{pmatrix} z_1 & z_2 & z_3 & z_4 \end{pmatrix} \begin{pmatrix} \pi_{12} & \pi_{13} \\ \pi_{22} & \pi_{23} \\ \pi_{32} & \pi_{33} \\ \pi_{42} & \pi_{43} \end{pmatrix} + \begin{pmatrix} v_2 & v_3 \end{pmatrix} \\
&= \begin{pmatrix} z_1 & z_2 & z_3 & z_4 \end{pmatrix} \begin{pmatrix} \mathbf{\Pi}_s \\ \mathbf{\Pi}_e \end{pmatrix} + \begin{pmatrix} v_2 & v_3 \end{pmatrix}.
\end{aligned}
$$

– Zur **Schätzung der strukturellen Gleichung für** y_2 muss die reduzierte Form für y_1 existieren. Zur Überprüfung des Rangkriteriums muss geprüft werden, ob $\mathbf{\Pi}_e$ Rang **eins** hat, da *eine* endogen erklärende Variable vorliegt. Im vorliegenden Fall muss π_{41} ungleich Null sein:

$$\begin{aligned}\begin{pmatrix} y_1 & y_3 \end{pmatrix} &= \begin{pmatrix} z_1 & z_2 & z_3 & z_4 \end{pmatrix} \begin{pmatrix} \pi_{11} & \pi_{13} \\ \pi_{21} & \pi_{23} \\ \pi_{31} & \pi_{33} \\ \pi_{41} & \pi_{43} \end{pmatrix} + \begin{pmatrix} v_1 & v_3 \end{pmatrix} \\ &= \begin{pmatrix} z_1 & z_2 & z_3 & z_4 \end{pmatrix} \begin{pmatrix} \mathbf{\Pi}_s \\ \mathbf{\Pi}_e \end{pmatrix} + \begin{pmatrix} v_1 & v_3 \end{pmatrix}.\end{aligned}$$

- Einteilung von strukturellen Gleichungen in überidentifiziert, exakt identifiziert und unteridentifiziert wie beim allgemeinen IV-Schätzer (Abschnitt 5.4).

6.3 Spezifikationstests

- **Endogenitätstests**:
  können wie in Abschnitt 5.5 durchgeführt werden.
  Zu **Beispiel** (6.7a)-(6.7c): Überprüfung der Endogenität von y_2 und y_3 in erster struktureller Gleichung (6.7a). Schätze reduzierte Form für y_2 und y_3 mit OLS und ergänze (6.7a) um $\hat{v}_2$ und $\hat{v}_3$

$$y_1 = y_2 A_{21} + y_3 A_{31} + z_1 B_{11} + \delta_2 \hat{v}_2 + \delta_3 \hat{v}_3 + Fehler.$$

  Schätze mit OLS und teste mit asymptotischem F-Test, ob $\delta_2 = \delta_3 = 0$.

- **Tests auf überidentifizierende Restriktionen**
  können wie in Abschnitt 5.5 durchgeführt werden.
  Zu **Beispiel** (6.7a)-(6.7c): Erste Gleichung (6.7a) überidentifiziert, da drei IV-Variable für 2 endogene Variablen vorliegen. Teste in

$$\hat{u}_1 = \mathbf{z}\mathbf{d} + Fehler,$$

  ob $\mathbf{d} = 0$. Alle signifikanten IV-Variablen sollten eliminiert werden. Dann wird allerdings die Gleichung möglicherweise unteridentifiziert.

6.4 SEMs für Zeitreihendaten

- **SEM**s mit **verzögert endogenen** Variablen: simultane Gleichungssysteme in struktureller Form mit endogenen und exogenen Variablen als erklärenden Variablen, wobei beide auch in verzögerter Form vorkommen können.

- **Vektorautoregressive (VAR) Modelle**: Gleichungssystem in reduzierter Form mit ausschließlich verzögert endogenen Variablen als erklärenden Variablen

- **strukturelle vektorautoregressive (SVAR) Modelle**: simultane Gleichungssysteme in struktureller Form mit ausschließlich endogenen Variablen als erklärenden Variablen. SVAR-Modelle **erfordern Prüfung, ob strukturelle Parameter identifiziert** sind.

- SEMs mit verzögert endogenen Variablen, VAR- und SVAR-Modelle:
  - **Erfordern** zusätzlich zu statischen SEMs die **Überprüfung der Stabilitätseigenschaften** der Dynamik des Systems, siehe Veranstaltung **Quantitative Wirtschaftsforschung I** und **Quantitative Wirtschaftsforschung II** im Master-Studium.
  - Können **Kointegrationsbeziehungen** aufweisen, wenn die Variablen Random Walks aufweisen, vgl. Zeitreihenökonometrie, Abschnitt 10.
  - Verzögert endogene Variablen werden als **vorherbestimmte (predetermined)** bezeichnet, da sie im Allgemeinen nicht exogen sind, siehe Zeitreihenökonometrie, Abschnitt 5.3.
  - Annahme der Exogenität von Variablen häufig problematisch, deshalb werden heutzutage VAR- und SVAR-Modelle vorgezogen.
  - Solche Modelle können komfortabel mit der freien Software JMulTi geschätzt werden, die menügeführt ist.

7 Modelle für abhängige Variablen mit Beschränkung (Limited Dependent Variable Methods)

Abhängige Variablen mit Beschränkung (Limited Dependent Variables):

- **Definition**:
  Eine abhängige Variable wird als **abhängige Variable mit Beschränkung (Limited Dependent Variable)** bezeichnet, wenn ihr Definitionsbereich substantiell eingeschränkt ist.

- **Arten**:
  - **Diskrete Variablen**:
    * binäre Variablen, also Dummyvariablen (Kaufen – Nichtkaufen, Arbeit anbieten – nicht anbieten, etc.),

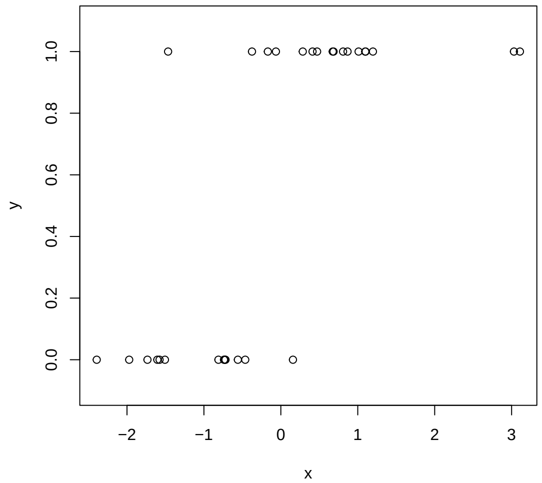
  * ordinale Variablen, z.B. Noten, Bewertungen,
  * kategoriale Variablen, z.B. Codierungen von Obstsorten,
  * Zähldaten, z.B. Zahl der Kinder.
- **Variablen mit Ecklösungen (corner-solution-response-Variablen)**: Der untere oder obere Wert eines Intervalls wird mehrfach beobachtet, typischerweise 0. Beispiel: Zahl der Arbeitsstunden 0 oder positiv.
- **Gestutzte oder zensierte Variablen**.
  * Gestutzte Variablen: Stichprobe nicht zufällig, da Beobachtungen mit Werten über/unter eines Schwellenwerts nicht in der Stichprobe vorliegen, z.B. keine Frauen mit Null-Arbeitsangebot.
  * Zensierte Variablen: Eigentlich interessierende Werte können nicht immer beobachtet werden, z.B. Einkommen über einem bestimmten Schwellenwert.

7.1 Modelle für binäre Daten (Binary Response Models)

- Ohne Einschränkung der Allgemeinheit wird angenommen, dass die abhängige Variable y die Werte 0 oder 1 annimmt:

$$y \in \{0, 1\}$$

- **Beispiel** für bivariate Daten: x: stetig, y: binär



Lässt sich hier überhaupt eine lineare Regression

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

zur Modellierung anwenden, da $\beta_0 + \beta_1 x$ stetig ist, y jedoch binär, und so ε praktisch ungleich Null sein **muss**?

Antwort: Nicht direkt, aber indirekt ... .

- zur **Modellierung der bedingten Wahrscheinlichkeit** $P(y = 1|x)$, die im Gegensatz zu y eine stetige Variable sein kann (außer z.B. x ist eine diskrete Variable)
  $\Longrightarrow$ **Lineares Wahrscheinlichkeitsmodell**.

- Man beachte, dass für beliebige $\mathbf{x}$ aufgrund

$$y = \begin{cases} 0 & \text{mit } P(y=0|\mathbf{x}) \\ 1 & \text{mit } P(y=1|\mathbf{x}) \end{cases},$$

allgemein gilt:

$$E[y|\mathbf{x}] = 0 \cdot P(y=0|\mathbf{x}) + 1 \cdot P(y=1|\mathbf{x}) = P(y=1|\mathbf{x}).$$

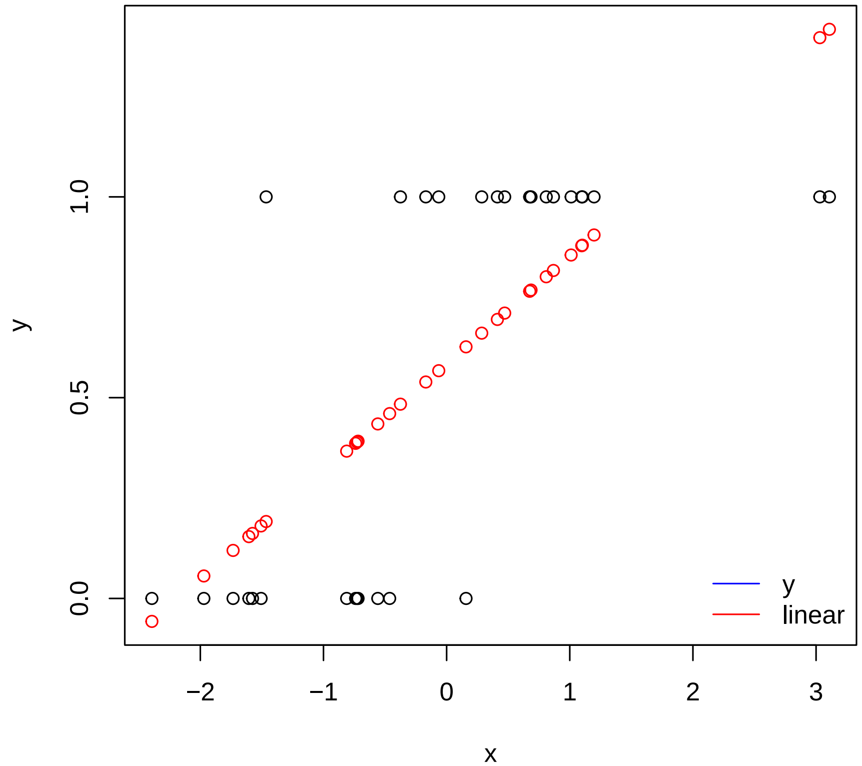
Der bedingte Erwartungswert entspricht der bedingten Wahrscheinlichkeit für $y = 1$ gegeben die Regressorvariablen.

- Beachte außerdem: y ist **bedingt heteroskedastisch**

$$\begin{aligned} Var(y|\mathbf{x}) &= E\left[(y - E[y|\mathbf{x}])^2|\mathbf{x}\right] \\ &= (1 - P(y=1|\mathbf{x}))^2 P(y=1|\mathbf{x}) + (0 - P(y=1|\mathbf{x}))^2 P(y=0|\mathbf{x}) \\ &= (1 - P(y=1|\mathbf{x}))P(y=1|\mathbf{x}). \end{aligned}$$

Das lineare Wahrscheinlichkeitsmodell (Linear Probability Model (LPM))

- $P(y = 1|\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$.
- Eine OLS-Schätzung für das Beispiel ergibt ($\texttt{pf} = \hat{\beta}_0 + \hat{\beta}_1 x$):



- Offensichtliches **Problem**: Die prognostizierten Wahrscheinlichkeiten $pf_t = P(y_t = 1|x_t)$ können Werte kleiner Null oder größer Eins annehmen!

  Abhilfen:

  – Stückweise lineares Modell: beschneide $\hat{\beta}_0 + \hat{\beta}_1 x$ auf $[0, 1]$.

  – Wähle **nichtlineares Modell**.

- Der **Standard für nichtlineare Modelle** sind Modelle folgender Art

$$P(y = 1|\mathbf{x}) = G(\mathbf{x}\boldsymbol{\beta}), \tag{7.1}$$

  wobei für alle reellen z gefordert wird, dass $G(\cdot)$ stetig ist, $0 < G(z) < 1$,

  – $\mathbf{x}\boldsymbol{\beta}$ als **lineare Indexfunktion** und

  – $G(\cdot)$ als **Linkfunktion (link function)** bezeichnet wird.

- Als **Linkfunktion** kann z.B. jede **stetige kumulative Wahrscheinlichkeitsverteilung** gewählt werden! Dann gilt: $G(\cdot)$ ist streng monoton steigend $\Longrightarrow$ Logit- und Probitmodelle.

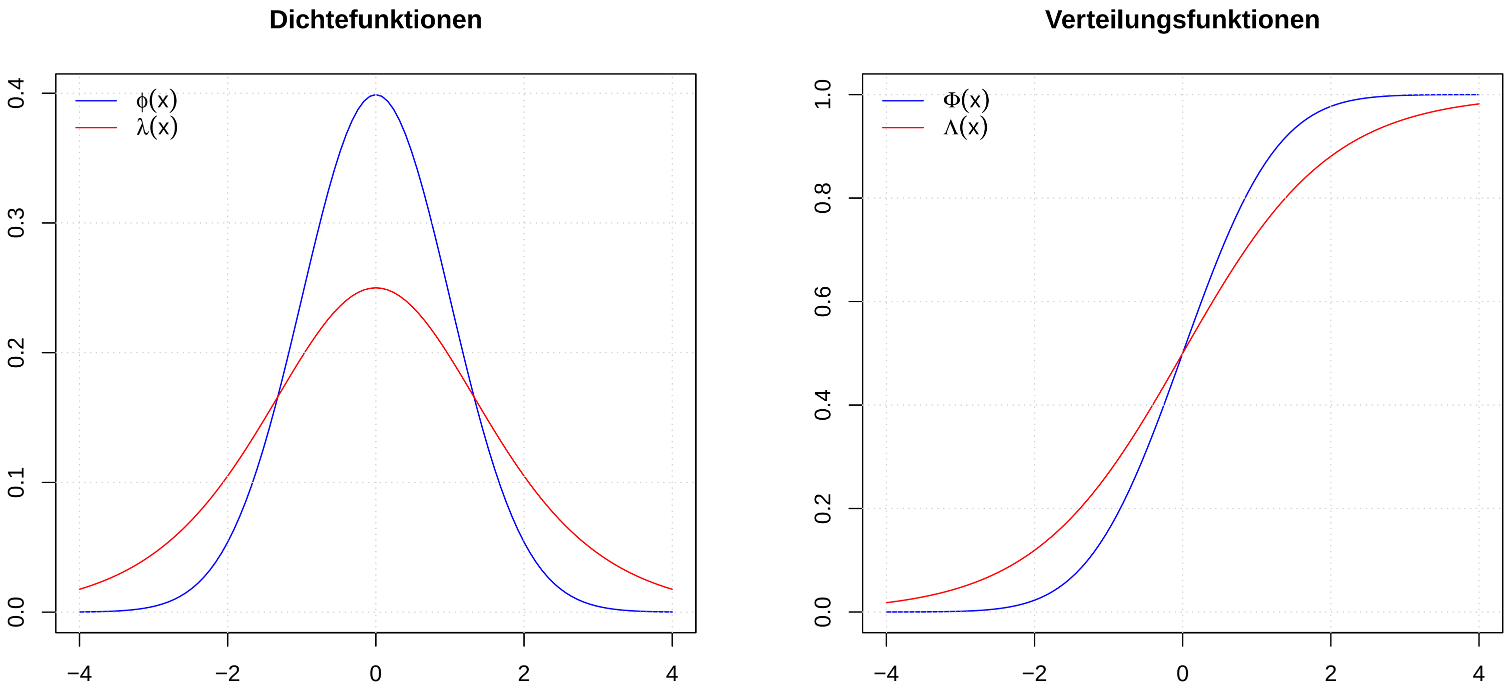
Probit- und Logit-Modelle

- Besonders geeignet: die Wahrscheinlichkeitsverteilung der
  - Standardnormalverteilung

$$\Phi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}\, dx, \tag{7.2}$$

  - logistischen Verteilung

$$\Lambda(z) = \frac{e^z}{1 + e^z}. \tag{7.3}$$



- Gemeinsame **Eigenschaften**:
  - streng monoton steigend,
  - Steigung maximal bei $z = 0$,
  - $G(z) \to 0$ für $z \to -\infty$, $G(z) \to 1$ für $z \to \infty$.

- Ableitung aus zugrundeliegendem **Latenten Variablenmodell (Latent Variable Model)**
  - **Unbeobachtbare (latente) Variable** y^*,
    z.B. Nutzen, wird durch lineares Regressionsmodell bestimmt

    $$y^* = \mathbf{x}\boldsymbol{\beta} + \varepsilon,$$

    und beeinflusst z.B. die Kaufentscheidung. Die beobachtbare Variable y ist Eins, falls $y^* > 0$, und sonst Null

    $$y = \mathbf{1}(y^* > 0).$$

    Dabei gibt $\mathbf{1}(\cdot)$ die Indikatorfunktion an und der Schwellenwert wurde auf 0 gesetzt, damit β_0 identifizierbar ist.

– Wird für **die Verteilung der Fehler ε die stetige Wahrscheinlichkeitsverteilung**

$$P(\varepsilon < z|\mathbf{x}) \overset{\mathbf{x}\text{ unabhängig v. }\varepsilon}{=} P(\varepsilon < z) = G(z)$$

angenommen, erhält man

$$\begin{aligned} P(y=1|\mathbf{x}) &= P(y^* > 0|\mathbf{x}) \\ &= P(\mathbf{x}\boldsymbol{\beta} + \varepsilon > 0|\mathbf{x}) = P(\varepsilon > -\mathbf{x}\boldsymbol{\beta}|\mathbf{x}) \\ &= 1 - P(\varepsilon < -\mathbf{x}\boldsymbol{\beta}|\mathbf{x}) \\ &= 1 - G(-\mathbf{x}\boldsymbol{\beta}) \qquad (7.4) \\ &= G(\mathbf{x}\boldsymbol{\beta}), \quad \text{falls } G(\cdot) \text{ symmetrisch um 0} \qquad (7.5) \end{aligned}$$

und somit (7.1).

– Sind die Fehler ε standardnormalverteilt, dann erhält man das **Probit-Modell**, sind sie logistisch verteilt, das **Logit-Modell**.

- **Interpretation** von $\boldsymbol{\beta}$:
  - **Marginaler Effekt** im latenten Regressionsmodell: wie üblich. Nicht besonders hilfreich, da z.B. marginaler Effekt auf Nutzen.
  - **Marginaler Effekt** auf die Veränderung der bedingten Wahrscheinlichkeiten

$$
\begin{aligned}
\frac{\partial P(y=1|\mathbf{x})}{\partial x_j} &= \frac{dG(\mathbf{x}\boldsymbol{\beta})}{d\mathbf{x}\boldsymbol{\beta}}\frac{\partial \mathbf{x}\boldsymbol{\beta}}{\partial x_j} \\
&= g(\mathbf{x}\boldsymbol{\beta})\beta_j, \quad \text{wobei } g(z) = \frac{dG(z)}{dz}.
\end{aligned}
$$

    Der marginale Effekt

    * hängt nichtlinear vom Parameter β_j und der Indexfunktion, also $\mathbf{x}\boldsymbol{\beta}$, ab,
    * weist die **gleiche Richtung/Vorzeichen wie** β_j auf,
    * ist für $\mathbf{x}\boldsymbol{\beta} = 0$ am größten,

$*$ wird für binäre Regressoren besser durch — hier für $x_1 = \{0, 1\}$

$$G(\beta_0 + \beta_1 + x_2\beta_2 + \ldots + x_k\beta_k) - G(\beta_0 + x_2\beta_2 + \ldots + x_k\beta_k) \quad (7.6)$$

berechnet. Entsprechend für diskrete Daten.

– **Relativer Effekt**

$$\frac{\frac{\partial P(y=1|\mathbf{x})}{\partial x_j}}{\frac{\partial P(y=1|\mathbf{x})}{\partial x_l}} = \frac{\beta_j}{\beta_l}.$$

– Wie im linearen Regressionsmodell werden die marginalen Effekte komplizierter, wenn die Indexfunktion linear in den Parametern, aber nichtlinear in den Variablen ist (quadrierte oder logarithmierte Variablen).

– **Welche marginalen/partiellen Effekte sollen angegeben werden?**

* bei stetigem x_j: Wahl des Skalierungsfaktors $g(\mathbf{x}\boldsymbol{\beta})$.
  Möglichkeiten:

  · Berechnung an interessanten Werten von $\mathbf{x}$.

  · Berechnung z.B. an Mittelwerten, Quantilen, Minima, Maxima.

  · Berechnung am Mittelwert $\bar{\mathbf{x}}$ ⇔ entspricht dies einer 'Durchschnittsperson'? Problematisch z.B. bei Dummyvariablen, nichtlinearen Regressoren.

  · Berechnung von **durchschnittlichen marginalen bzw. partiellen Effekten**

$$\frac{1}{n}\sum_{i=1}^{n} g(\mathbf{x}_i\boldsymbol{\beta})\beta_j$$

* bei diskretem x_j auf Basis von (7.6).

– Vergleich partieller Effekte zwischen LPM, Probit- und Logit-Modell:

$$g(0) \begin{cases} \approx 0.4 & \text{Probit} \\ = 0.25 & \text{Logit} \\ = 1 & \text{LPM} \end{cases}$$

erlaubt groben Vergleich geschätzter Koeffizienten $\hat{\boldsymbol{\beta}}$, wenn man davon ausgeht, dass die partiellen Effekte an $\mathbf{x}\boldsymbol{\beta} = 0$ für alle Modelle nahezu gleich sind. Dann ist z.B.

$$g_{Logit}(0)\,\hat{\boldsymbol{\beta}}_{Logit} \approx g_{Probit}(0)\,\hat{\boldsymbol{\beta}}_{Probit},$$

$$\hat{\boldsymbol{\beta}}_{Logit} \approx \frac{g_{Probit}(0)}{g_{Logit}(0)}\,\hat{\boldsymbol{\beta}}_{Probit} = 1.6\,\hat{\boldsymbol{\beta}}_{Probit}.$$

7.2 Maximum-Likelihood-Schätzung

Einführendes Beispiel: diskrete Zufallsvariablen

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 Würfen k mal 'Kopf' zu erhalten, durch die Binomialverteilung

$$P('k \text{ mal 'Kopf' bei } n \text{ Würfen'}|p) = \frac{n!}{(n-k)!k!} p^k (1-p)^{(n-k)} \quad (7.7)$$

  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:

$$p = 1/2 \quad \rightarrow P(\text{'9 mal 'Kopf' bei } 10 \text{ Würfen'}|p = 1/2) \approx 0.01$$
$$p = 3/4 \quad \rightarrow P(\text{'9 mal 'Kopf' bei } 10 \text{ Würfen'}|p = 3/4) \approx 0.19$$
$$p = 9/10 \quad \rightarrow P(\text{'9 mal 'Kopf' bei } 10 \text{ Würfen'}|p = 9/10) \approx 0.39$$

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

- Man kann nun die Verwendung der Wahrscheinlichkeitsfunktion (7.7) ändern und sie benutzen, um einem **gegebenen** Ereignis eine Wahrscheinlichkeit auf Basis eines gewählten Wertes für p zuzuweisen. Mit dieser Interpretation nennt man (7.7) eine **Likelihood-Funktion**, um so die Verwendung von der als Wahrscheinlichkeitsfunktion (gegebenes p) unterscheiden zu können. Für den gegebenen Fall erhält man

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

- Da man die Likelihood $L(p|\text{'}k \text{ mal 'Kopf' bei } n \text{ Würfen'})$ für ein gegebenes Ereignis, z.B. '9 mal 'Kopf' bei 10 Würfen', für jedes beliebige p berechnen kann, kann man die Likelihood $L(p|\text{'}k$ mal 'Kopf' bei n Würfen'$)$ 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 erste partielle Ableitung von (7.7) 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

$$\mathcal{L}(p|'k \text{ mal 'Kopf' bei } n \text{ Würfen'}) = \ln L(p|'k \text{ mal 'Kopf' bei } n \text{ Würfen'}) \tag{7.8}$$

$$= \ln\left(\frac{n!}{(n-k)!k!}\right) + k\ln p + (n-k)\ln(1-p). \tag{7.9}$$

Die erste Ableitung ist

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

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 zweite partielle Ableitung in der Umgebung von $\hat{p}$ negativ sein.)

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. $P(Y = y) = 0$. Denn es gibt eine unendliche Anzahl an möglichen Werten an reellen Zahlen innerhalb eines Intervalls und die Summe von unendlich vielen positiven Werten ist unendlich und nicht 1 — wie erforderlich.

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

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

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

Dies ergibt die absolute Veränderung in der Wahrscheinlichkeit

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

und die relative Veränderung

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

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

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

die an einigen y positiv sein muss, denn ansonsten würde sich bei einer Änderung der Intervalllänge keine Veränderung der Wahrscheinlichkeit ergeben.

Da

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

erhält man, salopp gesprochen,

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

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

- Entsprechend hat man, falls die Wahrscheinlichkeit und die Dichte vom Parametervektor $\boldsymbol{\theta}$ abhängen,

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

Die Maximierung der Likelihood, Y in einem winzigen Intervall um y herum zu beobachten, kann deshalb erfolgen, indem man die Dichte bezüglich $\boldsymbol{\theta}$ 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 $\mathbf{y} = \begin{pmatrix} y_1 & y_2 & \dots & y_n \end{pmatrix}'$ 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 unabhängig und identisch verteilte (IID) Beobachtungen ist das Produkt der n marginalen Dichten. 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**

$$\mathcal{L}(\boldsymbol{\theta}|\mathbf{y}) = \ln f(\mathbf{y}|\boldsymbol{\theta}) = \sum_{i=1}^{n} \ln f(y_i|\boldsymbol{\theta}).$$

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

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

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

  Nach Logarithmieren erhält man die Summe

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

- Wenn der Term $\ln f(y_1|\boldsymbol{\theta})$ vernachlässigt wird, erhält man die **bedingte Likelihood-Funktion** bedingt auf y_1. Ihre Maximierung ergibt einen **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. Ist auch $\boldsymbol{\theta}$ bekannt, hat man eine vollständige Beschreibung eines datengenerierenden Prozesses (DGP) und man kann $\mathbf{y}$ und gegebenenfalls $\mathbf{X}$ simulieren.

- **Schätzeigenschaften**: Unter sehr allgemeinen Bedingungen (z.B. korrekte Modellspezifikation, identifizierte Parameter, Stationarität, keine linearen Abhängigkeiten, etc.) ist der ML-Schätzer **konsistent**, **asymptotisch normalverteilt** und **asymptotisch effizient**. Zur Ableitung siehe z.B. Davidson und MacKinnon (2004, Chapter 10), Wooldridge (2002, Chapter 13) oder die Masterveranstaltung Advanced Econometrics.

Ergänzung: Multiples Regressionsmodell

Sind die Fehler im **multiplen Regressionsmodell** unabhängig normalverteilt mit Mittelwert 0 und Varianz σ^2, lautet die multivariate Dichte

$$f(u_i|\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2}\frac{u_i^2}{\sigma^2}\right), \quad i = 1, 2, \ldots, n, \text{ bzw.}$$

$$f(\mathbf{u}|\sigma^2) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left(-\frac{1}{2\sigma^2}\mathbf{u}'\mathbf{u}\right).$$

Da $\mathbf{u} = \mathbf{y} - \mathbf{X}\boldsymbol{\beta}$, $Var(\mathbf{y}|\mathbf{X}) = \sigma^2\mathbf{I}$ und damit

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

lautet die Likelihood-Funktion demnach

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

Logarithmieren ergibt die Log-Likelihood-Funktion

$$
\begin{aligned}
\mathcal{L}(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, \mathbf{X}) &= \sum_{i=1}^{n} \left\{ -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{(y_i - \mathbf{x}_i \boldsymbol{\beta})^2}{2\sigma^2} \right\} \\
&= const - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mathbf{x}_i \boldsymbol{\beta})^2 \\
&= const - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \\
&= const - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} SSR(\boldsymbol{\beta}).
\end{aligned}
$$

Die Maximierung der Log-Likelihood bezüglich $\boldsymbol{\beta}$ im Fall des normalen linearen Regressionsmodells ist identisch mit dem OLS-Schätzer. Bei streng exogenen Regressoren sind deshalb auch die Schätzeigenschaften identisch. Gilt die Normalverteilungsannahme nicht, ist der ML-Schätzer unter bestimmten Bedingungen asymptotisch normalverteilt (wie der OLS-Schätzer).

Numerische Optimierung

- Im Allgemeinen ist die (partielle) erste Ableitung der Log-Likelihood-Funktion nicht analytisch bezüglich der Parameter lösbar. Dann
  - werden die Nullstellen der partiellen ersten Ableitung numerisch bestimmt, oder
  - die Log-Likelihood-Funktion wird direkt numerisch optimiert.
- Numerische Verfahren funktionieren iterativ, d.h. gegeben einen Wert für den geschätzten Parametervektor $\hat{\boldsymbol{\theta}}_{(j)}$ aus der j-ten Iteration wird in der $j+1$-ten Iteration ein neuer Schätzer $\hat{\boldsymbol{\theta}}_{(j+1)}$ des Parametervektors bestimmt, dessen Funktionswert näher am Optimum liegt.
- Für die numerische Optimierung von Funktionen gibt es eine Reihe von **Algorithmen**. In EViews stehen zur Verfügung:
  - Quadratic-Hill climbing,

– Newton-Raphson,

– Berndt-Hall-Hall-Hausman.

Alle numerischen Verfahren benötigen:

– **Startwerte** $\hat{\boldsymbol{\theta}}_{(0)}$:

* Ist die Zielfunktion **streng konvex** in den Parametern und existiert ein Minimum, ist dieses eindeutig. Ist die Zielfunktion **streng konkav** und existiert ein Maximum ist dieses **eindeutig**.

* In allen anderen Fällen können mehrere bzw. sogar beliebig viele Optima existieren! Zum Auffinden des globalen Optimums ist die Wahl der Startwerte im Allgemeinen bedeutsam, bzw. sollten verschiedene Startwerte ausprobiert werden.

– ein Konvergenzkriterium, z.B.

$$\frac{\sqrt{(\hat{\boldsymbol{\theta}}_{(j+1)} - \hat{\boldsymbol{\theta}}_{(j)})'(\hat{\boldsymbol{\theta}}_{(j+1)} - \hat{\boldsymbol{\theta}}_{(j)})}}{\sqrt{\hat{\boldsymbol{\theta}}'_{(j)}\hat{\boldsymbol{\theta}}_{(j)}}} < tol,$$

wobei tol vom Anwender vorgegeben wird.

– eine Vorgabe für die maximale Anzahl $jmax$ an Iterationen, für den Fall, dass kein eindeutiges Optimum vorliegt. Damit gilt $j \leq jmax$.

Details finden sich z.B. in der Masterveranstaltung Advanced Econometrics.

Allgemeine Testprinzipien

Es gebe q die Anzahl der Restriktionen an. Diese können nichtlinear sein. Auf Basis des ML-Schätzers gibt es drei Testprinzipien:

- **Likelihood-Ratio-Test**

$$LR = 2\left(\mathcal{L}_{ur} - \mathcal{L}_r\right) \xrightarrow{d} \chi^2(q).$$

- **Wald-Test**: ähnlich wie LR-Test, jedoch muss $\mathcal{L}_r$ nicht berechnet werden, sondern wird auf Basis von $\mathcal{L}_{ur}$ approximiert.

- **Lagrange-Multiplikator-Test**: ähnlich wie LR-Test, jedoch muss $\mathcal{L}_{ur}$ nicht berechnet werden, sondern wird auf Basis von $\mathcal{L}_r$ approximiert.

Alle drei Tests sind unter den üblichen Bedingungen asymptotisch $\chi^2(q)$-verteilt. (Ableitung der Teststatistiken und asymptotischen Verteilungen in Advanced Econometrics..)

7.3 Schätzung von Probit- und Logit-Modellen

ML-Schätzung

- Anstelle von 'Kopf' und 'Zahl', wie in (7.8), ist y entweder 0 oder 1. Und anstelle der (unbedingten, konstanten) Wahrscheinlichkeit p in (7.8) hat man bei Modellen mit binären Variablen eine bedingte Wahrscheinlichkeit

$$p(\mathbf{x}_i) = P(y_i = 1|\mathbf{x}_i; \boldsymbol{\beta}) = G(\mathbf{x}_i\boldsymbol{\beta}).$$

- Da $p(\mathbf{x}_i)$ im Allgemeinen variiert, muss (7.8) für jede Beobachtung i bestimmt werden. Damit ist $n = 1$ und man erhält

$$\begin{aligned} P(y_i|\mathbf{x}_i; \boldsymbol{\beta}) &= P(y_i = 1|\mathbf{x}_i; \boldsymbol{\beta})^{y_i} P(y_i = 0|\mathbf{x}_i; \boldsymbol{\beta})^{1-y_i} \\ &= G(\mathbf{x}_i\boldsymbol{\beta})^{y_i}(1 - G(\mathbf{x}_i\boldsymbol{\beta}))^{1-y_i}. \end{aligned}$$

- Bei unabhängigen Beobachtungen lautet die Likelihood-Funktion

$$L(\boldsymbol{\beta}|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^{n} G(\mathbf{x}_i\boldsymbol{\beta})^{y_i}(1 - G(\mathbf{x}_i\boldsymbol{\beta}))^{1-y_i}$$

und die **Log-Likelihood-Funktion**

$$\mathcal{L}(\boldsymbol{\beta}|\mathbf{y},\mathbf{X}) = \sum_{i=1}^{n} \left\{ y_i \ln G(\mathbf{x}_i\boldsymbol{\beta}) + (1-y_i)\ln(1-G(\mathbf{x}_i\boldsymbol{\beta})) \right\}. \quad (7.10)$$

- Der ML-Schätzer $\hat{\boldsymbol{\beta}}$ wird schließlich bestimmt, indem

$$\max_{\boldsymbol{\beta}} \mathcal{L}(\boldsymbol{\beta}|\mathbf{y},\mathbf{X})$$

bestimmt wird:

  – Probit-Schätzer $\hat{\boldsymbol{\beta}}$: Ersetze $G(\cdot)$ in (7.10) durch kumulative Normalverteilung (7.2).

  – Logit-Schätzer $\hat{\boldsymbol{\beta}}$: Ersetze $G(\cdot)$ in (7.10) durch logistische Verteilung (7.3).

- Da für Probit- und Logit-Modelle $0 < G(\cdot) < 1$, ist $\boldsymbol{\beta}$ identifizierbar (solange $\mathbf{X}$ nicht linear abhängig ist). Es kann gezeigt werden, dass $\mathcal{L}(\boldsymbol{\beta}|\mathbf{y},\mathbf{X})$ für diese Modelle streng konkav ist.

- **Schätzeigenschaften**:
  (Wooldridge, 2010, Equations (13.28), (13.29), (15.19), (15.20))

$$\sqrt{n}\left(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) \xrightarrow{d} N(0, \mathbf{Q}^{-1})$$

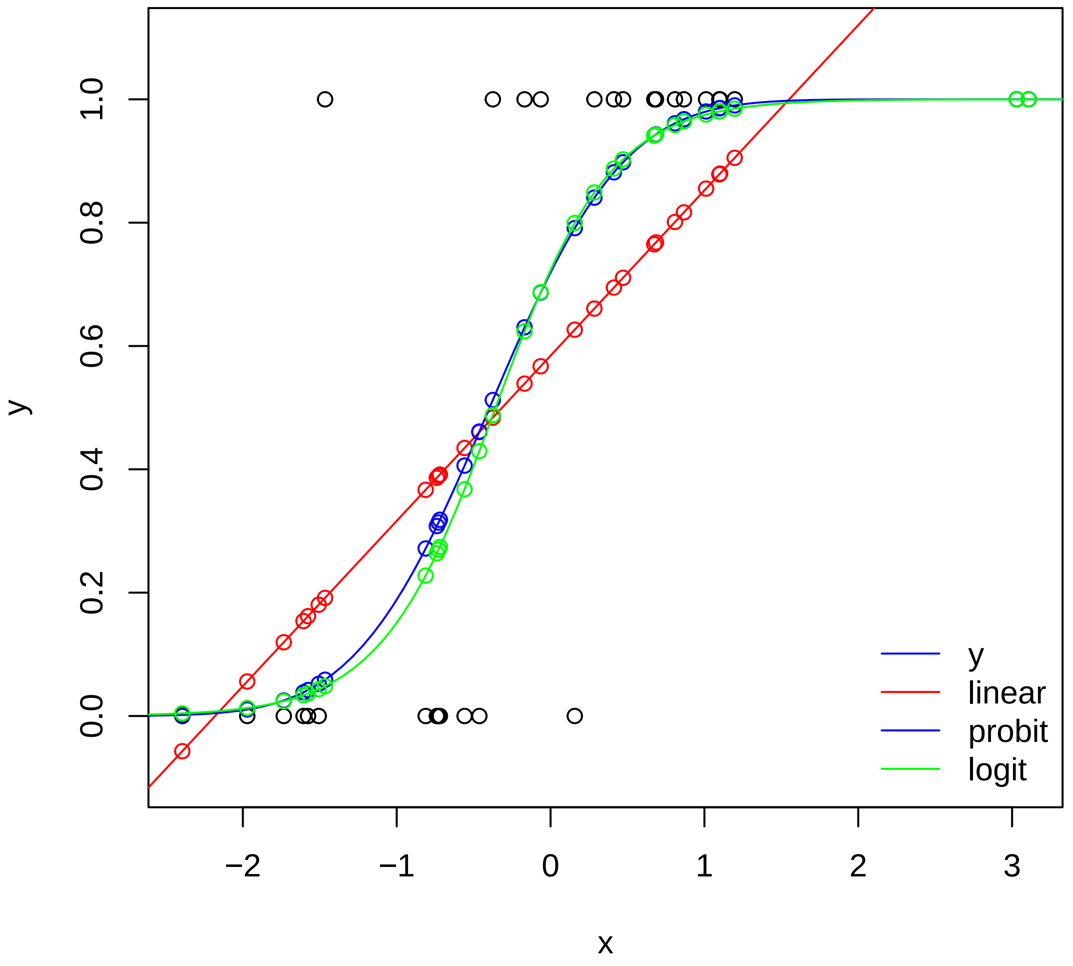
$$\mathbf{Q} = \text{plim}\frac{1}{n}\sum_{i=1}^{n}\frac{(g(\mathbf{x}_i\boldsymbol{\beta}))^2\mathbf{x}_i'\mathbf{x}_i}{G(\mathbf{x}_i\boldsymbol{\beta})(1 - G(\mathbf{x}_i\boldsymbol{\beta}_i))}$$

bzw. approximativ

$$\hat{\boldsymbol{\beta}} \approx N\left(\boldsymbol{\beta}, \left(\sum_{i=1}^{n}\frac{(g(\mathbf{x}_i\hat{\boldsymbol{\beta}}))^2\mathbf{x}_i'\mathbf{x}_i}{G(\mathbf{x}_i\hat{\boldsymbol{\beta}})(1 - G(\mathbf{x}_i\hat{\boldsymbol{\beta}}_i))}\right)^{-1}\right) \tag{7.11}$$

— wird von Softwarepaketen ausgegeben. (Ableitung mit Hilfe der Techniken aus Methoden der Ökonometrie möglich.)

• Beispiel:



Tests

- **t-Tests**: wie bisher, jedoch unter Verwendung der Diagonalelemente von (7.11) anstelle von $\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}$.
- Anstelle von **F-Tests** wird der **Wald**-Test oder der **Likelihood-Ratio**-Test verwendet. Ersterer kann standardmäßig in Softwarepaketen berechnet werden. Letzterer ist leicht selbst zu berechnen.

Goodness-of-fit Maße

- **Anteil korrekter Prognosen**

  Die Prognose von y_i für jede Beobachtung $\mathbf{x}_i$ wird durch

$$\tilde{y}_i = \mathbf{1}(G(\mathbf{x}_i\hat{\boldsymbol{\beta}}) > \tau)$$

  bestimmt. Dann gibt

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\tilde{y}_i = y_i) \tag{7.12}$$

  den Anteil der korrekten Prognosen an.

  Wahl des Schwellenwerts τ:

  – $\tau = 0.5$.

  – Anteil der '$y_i = 1$' Beobachtungen in der Stichprobe.

  – Wähle τ so, dass $\sum_{i=1}^{n} \tilde{y}_i = \sum_{i=1}^{n} y_i$.

  Der Quotient (7.12) sollte auch separat jeweils für alle Beobachtungen mit $y_i = 1$ und $y_i = 0$ berechnet werden. Warum?

- **Pseudo R^2**
  - **McFadden (1974)**:

$$0 < 1 - \frac{\mathcal{L}_{ur}}{\mathcal{L}_o} < 1,$$

    wobei $\mathcal{L}_{ur}$ die Log-Likelihood des betrachteten Modells und $\mathcal{L}_o$ eines Modells ausschließlich mit Konstante angibt.

  - Es gibt eine Reihe Alternativen, die hier nicht besprochen werden.

Beispiel: Arbeitsmarktteilnahme von verheirateten Frauen

(Example 17.1 in Wooldridge (2009, 2006).)

- Definition der Variablen: siehe Wooldridge (2009, 2006, Section 7.5).

- ML-Schätzung eines Probit-Modells:

```
Call: glm(formula = inlf ~ nwifeinc + educ + exper + expersq + age +
    kidslt6 + kidsge6, family = binomial(link = "probit"), data = mroz)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.2156  -0.9151   0.4315   0.8653   2.4553

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.2700736  0.5080782   0.532  0.59503
nwifeinc    -0.0120236  0.0049392  -2.434  0.01492 *
educ         0.1309040  0.0253987   5.154 2.55e-07 ***
exper        0.1233472  0.0187587   6.575 4.85e-11 ***
expersq     -0.0018871  0.0005999  -3.145  0.00166 **
age         -0.0528524  0.0084624  -6.246 4.22e-10 ***
kidslt6     -0.8683247  0.1183773  -7.335 2.21e-13 ***
kidsge6      0.0360056  0.0440303   0.818  0.41350
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1029.7  on 752  degrees of freedom
Residual deviance:  802.6  on 745  degrees of freedom
AIC: 818.6
Number of Fisher Scoring iterations: 4
log Lik.’ -401.3022 (df=8)
AIC a la EViews  1.087124
BIC a la EViews  1.136251
BIC a la EViews 1.136251
HQ a la EViews 1.10605
McFadden Pseude-R^2 ’log Lik.’ 0.2205805 (df=8)
LR-Statistik ’log Lik.’ 227.142 (df=8)             p-Wert ’log Lik.’ 0 (df=8)
```

- ML-Schätzung eines Logit-Modells:

```
Call: glm(formula = inlf ~ nwifeinc + educ + exper + expersq + age +
    kidslt6 + kidsge6, family = binomial(link = "logit"), data = mroz)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.1770  -0.9063   0.4473   0.8561   2.4032

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.425452   0.860365   0.495  0.62095
nwifeinc    -0.021345   0.008421  -2.535  0.01126 *
educ         0.221170   0.043439   5.091 3.55e-07 ***
exper        0.205870   0.032057   6.422 1.34e-10 ***
expersq     -0.003154   0.001016  -3.104  0.00191 **
age         -0.088024   0.014573  -6.040 1.54e-09 ***
kidslt6     -1.443354   0.203583  -7.090 1.34e-12 ***
kidsge6      0.060112   0.074789   0.804  0.42154
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1029.75  on 752  degrees of freedom
Residual deviance:  803.53  on 745  degrees of freedom
AIC: 819.53
Number of Fisher Scoring iterations: 4
log likelihood for estimated model 'log Lik.' -401.7652 (df=8)
AIC a la EViews 1.088354
BIC a la EViews 1.137481
HQ a la EViews 1.10728
McFadden Pseude-R^2 'log Lik.' 0.2196814 (df=8)
LR-Statistik 'log Lik.' 226.2161 (df=8)      p-Wert 'log Lik.' 0 (df=8)
```

- Goodness-of-fit für Probit-Modell:
  - Prognose von $P(y_i = 1|\mathbf{x}_i) = \Phi(\mathbf{x}_i\boldsymbol{\beta})$ mit Hilfe von $\widehat{P}(y_i = 1|\mathbf{x}_i) = \Phi(\mathbf{x}_i\hat{\boldsymbol{\beta}})$:

    R: `fitted(probit_model)`, wobei `probit_model` das Objekt des geschätzten Probit-Modells ist
  - Prognose von y_i mit Hilfe von $\tilde{y}_i = \mathbf{1}\left(\widehat{P}(y_i = 1|\mathbf{x}_i) > \tau\right)$, wobei im Folgenden $\tau = 0.5$ gewählt wird:

    R: `y_tilde <- (fitted(probit_model) > 0.5)`
  - Bewerten des Anteils korrekter Prognosen: Berechnen von $\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}(\tilde{y}_i = y_i)$:

    R: `mean(y_tilde == mroz$inlf)` liefert 0.7343958.
  - Durch Einschränken des Samples auf alle Beobachtungen mit $y_i = 1$ bzw. $y_i = 0$ können die Anteile korrekter Prognosen für die jeweiligen

Ausprägungen bestimmt werden:

$$\frac{1}{n_0}\sum_{i=1,y_i=0}^{n} \mathbf{1}(\tilde{y}_i = y_i), \quad \frac{1}{n_1}\sum_{i=1,y_i=1}^{n} \mathbf{1}(\tilde{y}_i = y_i), \quad n_0 + n_1 = n.$$

R: Tabelle mit korrekten und falschen Prognosen: `table(true = mroz$inlf, pred = round(fitted(probit_model)))`. Dividieren durch Anzahl der jeweiligen Beobachtungen ergibt jeweils den Anteil der korrekten Prognosen für alle arbeitenden, bzw. nicht arbeitenden Frauen: 0.6307 bzw. 0.8131.

- **Durchschnittlicher partieller Effekt**:
  bezüglich educ ist $\frac{1}{n}\sum_{i=1}^{n} \phi(\mathbf{x}_i\hat{\boldsymbol{\beta}})\hat{\beta}_{educ} = 0.039.$

  R: `mean(dnorm(predict(probit_model, type = "link"))) * coef(probit_model)`

  Zum Vergleich: $\hat{\beta}_{educ}$ im linearen Wahrscheinlichkeitsmodell, das direkt den partiellen Effekt angibt, ist 0.038.

Mögliche Probleme und Erweiterungen

- Fehlende Variablen.
- Endogene erklärende Variablen: es existieren Methoden in Anlehnung an 2SLS.
- Fehler ε sind weder normal- noch logistisch verteilt.
- Heteroskedastie.
- Gepoolte Querschnittsdaten.
- Zeitreihendaten.

Siehe Wooldridge (2002, 2010); Cameron und Trivedi (2005, Chapter 15).

Multinomiale Modelle

- Diskrete abhängige Variable hat mehr als 2 Ausprägungen:
  - geordnete Daten, z.B. Noten, Bewertungen,
  - ungeordnete Daten, z.B. Obstsorten.
- Modellansätze: multinomiale Logit- und Probitmodelle, Discrete Choice-Modelle.

Siehe Wooldridge (2002, 2010); Cameron und Trivedi (2005).

7.4 Daten mit Ecklösungen (Corner Solution Responses)

Nochmal ein Überblick

Stichprobenwerte $(y_i, \mathbf{x}_i)$ sind in dem für die Fragestellung relevanten Bereich

- vollständig beobachtbar oder
- nicht vollständig beobachtbar.

Vollständig beobachtbar bedeutet:

- y ist stetig mit unbegrenztem Definitionsbereich oder
- diskret oder
- aber nur auf Teilintervall stetig, z.B. für $y > 0$ und sonst Null. Es liegen dann **Daten mit Ecklösungen (Corner Solution Responses)** vor. Beispiel: Ausgaben für Bücher.

Nicht vollständig beobachtbar bedeutet:
Es liegt das Problem **fehlender Beobachtungen (missing values)** vor.

- **Zensierte Daten**: y ist für einen Teil der Stichprobenwerte nur mit ungenauen Werten verfügbar, z.B. ist für $y_i > c$ nicht das interessierende y_i sondern lediglich $y_i = c$ bekannt. Es liegt jedoch eine **Zufallsstichprobe** vor.
- **Gestutzte Daten**: Beispiel: Es können nur $\mathbf{x}_i$ für $y_i < c$ beobachtet werden, aber keine $\mathbf{x}$ mit $y = c$ oder $y > c$. Damit liegt **keine Zufallsstichprobe** vor.

Tobit-Modell

- **Modell der Grundgesamtheit** lautet für latente Variable y^*

$$y^* = \mathbf{x}\boldsymbol{\beta} + u, \quad u|\mathbf{x} \sim N(0, \sigma^2). \tag{7.13}$$

- **Stichprobenmodell** für Corner-Response-Stichprobenbeobachtungen $(y_i, \mathbf{x}_i)$ mit Eckwert 0

$$y_i = \begin{cases} 0 & \text{falls } y_i^* \leq 0 \\ y_i^* & \text{sonst} \end{cases},$$
$$y_i^* = \mathbf{x}_i\boldsymbol{\beta} + u_i, \quad u_i|\mathbf{x}_i \sim N(0, \sigma^2).$$

- OLS-Schätzung des bedingten Erwartungswertes $E[y_i|\mathbf{x}_i]$ kann negative Werte liefern. Ausweg: Tobit-Modell.

• Dichte

– $y = 0$: Wenn $y = 0$ beobachtet wird, ist $y^* \leq 0$. Aufgrund der **Annahme normalverteilter Fehler** mit Varianz σ^2 ist die Wahrscheinlichkeit hierfür

$$\begin{aligned} P(y = 0|\mathbf{x}) &= P(y^* \leq 0|\mathbf{x}) \\ &= P(\mathbf{x}\boldsymbol{\beta} + u \leq 0|\mathbf{x}) = P(u \leq -\mathbf{x}\boldsymbol{\beta}|\mathbf{x}) \\ &= P\left(\frac{u}{\sigma} \leq \frac{-\mathbf{x}\boldsymbol{\beta}}{\sigma}\middle| \mathbf{x}\right) \\ &= \Phi\left(\frac{-\mathbf{x}\boldsymbol{\beta}}{\sigma}\right). \end{aligned}$$

Alternativ lässt sich dies auch ausrechnen, indem man die Wahrscheinlichkeit hierfür über die Integration der Dichte berechnet:

$$P(y = 0|\mathbf{x}) = \int_{-\infty}^{0} f(y^*|\mathbf{x})dy^*.$$

Da aufgrund der Normalverteilungsannahme

$$f(y^*|\mathbf{x}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(y^*-\mathbf{x}\boldsymbol{\beta})^2}{\sigma^2}} = \frac{1}{\sigma}\phi\left(\frac{y^* - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right)$$

gilt, ergibt sich mit $z = (y^* - \mathbf{x}\boldsymbol{\beta})/\sigma$

$$P(y = 0|\mathbf{x}) = \frac{1}{\sigma}\int_{-\infty}^{b_{y^*}=0} \phi\left(\frac{y^* - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right) dy^*$$

$$= \int_{-\infty}^{b_z} \phi(z) dz = \Phi(b_z) = \Phi\left(\frac{0 - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right),$$

wobei die letzte Zeile aus $dz = \frac{1}{\sigma}dy^*$ und damit $dy^* = \sigma dz$ sowie $b_z = (b_{y^*} - \mathbf{x}\boldsymbol{\beta})/\sigma$ folgt.

– $y > 0$: somit gilt: $y^* = y$ bzw. $f(y|\mathbf{x}) = f(y^*|\mathbf{x})$ und die Dichte lautet

$$f(y|\mathbf{x}) = \frac{1}{\sigma}\phi\left(\frac{y - \mathbf{x}\boldsymbol{\beta}}{\sigma}\right).$$

- **Log-Likelihood**

$$\mathcal{L} = \sum_{i=1}^{n} \mathbf{1}(y_i = 0) \ln \left[1 - \Phi\left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) \right] + \mathbf{1}(y_i > 0) \ln \left[\frac{1}{\sigma} \phi \left(\frac{y_i - \mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) \right]$$

  – Numerische Optimierung wie bei Logit- und Probit-Modellen.

  – Bei korrekter Spezifikation, etc. ist ML-Schätzer **konsistent** und **asymptotisch normalverteilt**.

- **Bedingte Erwartungswerte**:
  - Es sei $z \sim N(0,1)$. Dann gilt (siehe z.B. Abschnitt 8.1 im Anhang):
  $$E[z|z > c] = \frac{\phi(c)}{1 - \Phi(c)} = \frac{\phi(-c)}{\Phi(-c)} \quad \text{für beliebiges } c,$$
  wobei das letzte '$=$' aufgrund der Symmetrie der Normalverteilung gilt. Man unterscheidet $E[y|\mathbf{x}, y > 0]$ und $E[y|\mathbf{x}]$. Beides sind bedingte Erwartungswerte, da auf $\mathbf{x}$ bedingt. Jedoch wird ersterer häufig als 'bedingter' und letzterer als 'unbedingter' Erwartungswert (eben hinsichtlich $y > 0$) bezeichnet.

– **'Bedingter' Erwartungswert**:

$$\begin{aligned}
E[y|\mathbf{x}, y > 0] &= E[\mathbf{x}\boldsymbol{\beta} + u|\mathbf{x}, \mathbf{x}\boldsymbol{\beta} + u > 0] \\
&= \mathbf{x}\boldsymbol{\beta} + E[u|\mathbf{x}, u > -\mathbf{x}\boldsymbol{\beta}] \\
&= \mathbf{x}\boldsymbol{\beta} + E[\sigma z|\mathbf{x}, \sigma z > -\mathbf{x}\boldsymbol{\beta}] \\
&= \mathbf{x}\boldsymbol{\beta} + \sigma E[z|\mathbf{x}, z > \underbrace{-\mathbf{x}\boldsymbol{\beta}/\sigma}_{-c}] \\
&= \mathbf{x}\boldsymbol{\beta} + \sigma \frac{\phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)} \\
&= \mathbf{x}\boldsymbol{\beta} + \sigma\lambda\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)
\end{aligned}$$

mit

$$\lambda(c) = \phi(c)/\Phi(c). \tag{7.14}$$

Die Funktion $\lambda(c)$ wird als **inverse Mills-Ratio** bezeichnet.

– **'Unbedingter' Erwartungswert**:

$$\begin{aligned} E[y|\mathbf{x}] &= P(y>0|\mathbf{x})E[y|\mathbf{x}, y>0] \\ &= \Phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\left[\mathbf{x}\boldsymbol{\beta} + \sigma\lambda\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\right] \\ &= \Phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\mathbf{x}\boldsymbol{\beta} + \sigma\phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right) > 0, \end{aligned}$$

wobei das Ungleichheitszeichen hier nicht bewiesen wird.

- **Interpretation** der Parameter:

  Betrachtung **partieller Effekte**: Hier unterscheidet man bei stetigen Regressoren die partiellen Effekte $\frac{\partial E[y|\mathbf{x},y>0]}{\partial x_j}$ und $\frac{\partial E[y|\mathbf{x}]}{\partial x_j}$. Bei **diskreten Regressoren** ist der partielle Effekt nicht “marginal” und man berechnet besser die Differenzen, z.B. bei einer Dummyvariablen x_j

$$E[y|x_1,\ldots,x_j=1,\ldots,x_k] - E[y|x_1,\ldots,x_j=0,\ldots,x_k].$$

– **'Bedingte' partielle Effekte — auf $y > 0$ bedingt**:

$$\frac{\partial E[y|\mathbf{x}, y > 0]}{\partial x_j} = \beta_j + \beta_j \frac{d\lambda}{dc}\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)$$
$$= \beta_j \left\{1 - \lambda\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\left[\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma} + \lambda\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right)\right]\right\}.$$

Das zweite Gleichheitszeichen folgt aus der Quotientenregel und $\frac{d}{dc}\phi(c) = -c\phi(c)$:

$$\frac{d}{dc}\lambda(c) = \frac{d}{dc}\frac{\phi(c)}{\Phi(c)}$$
$$= \frac{d\phi(c)/dc\ \Phi(c) - d\Phi(c)/dc\ \phi(c)}{\Phi(c)^2}$$
$$= \frac{-c\phi(c)\Phi(c) - \phi(c)^2}{\Phi(c)^2} = \frac{\phi(c)}{\Phi(c)}\left(-c - \frac{\phi(c)}{\Phi(c)}\right)$$
$$= -\lambda(c)(c + \lambda(c)).$$

– **'Unbedingte' partielle Effekte**:

$$\frac{\partial E[y|\mathbf{x}]}{\partial x_j} = \beta_j \Phi\left(\frac{\mathbf{x}\boldsymbol{\beta}}{\sigma}\right).$$

(Beweis der Leserin, dem Leser überlassen.)

Der **durchschnittliche 'unbedingte' partielle Effekt**

$$\beta_j \frac{1}{n}\sum_{i=1}^{n} \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)$$

kann gut mit dem partiellen Effekt $\hat{\beta}_{j,OLS}$ aus der OLS-Schätzung von $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ verglichen werden.

Man kann den partiellen Effekt für einen speziellen Vektor $\mathbf{x}$ berechnen, z.B. $\bar{\mathbf{x}}$. Man beachte dann jedoch, dass $\beta_j \bar{x}_j + \beta_{j+1}\overline{x_j}^2 \neq \overline{\beta_j x_j + \beta_{j+1} x_j^2}$ ist.

- R^2: wird am besten durch $Corr(y_i, \hat{y}_i)^2$ berechnet.

- **Probleme**: Das Tobit-Modell ist fehlspezifiziert, wenn
  - der Term $\mathbf{x}\boldsymbol{\beta}$ fehlerhaft ist,
  - die Fehler u heteroskedastisch sind,
  - die Fehler u nicht normalverteilt sind,
  - die Variablen, die y^* und die inverse Mills-Ratio bestimmen, unterschiedlich sind. Dies lässt sich überprüfen, indem man für $(y_i > 0)$ und $(y_i = 0)$ ein Probit-Modell schätzt und die geschätzten Probit-Parameter mit den geschätzten Tobit-Parametern $\hat{\beta}_j/\hat{\sigma}$ vergleicht (insbesondere hinsichtlich möglicher Vorzeichenunterschiede).

7.5 Zensierte Daten

- **Zensierte Daten** liegen vor, wenn
  - die abhängige Variable y_i nicht über den gesamten interessierenden Datenbereich vollständig beobachtbar ist, z.B. durch eine obere Schranke bei Einkommenserhebungen und damit die Stichprobe (in diesem Bereich) von **zensierten Werten (censored values)** charakterisiert ist.
  - (trotzdem) eine Zufallsstichprobe vorliegt.

- **Modell der Grundgesamtheit**:

$$y = \mathbf{x}\boldsymbol{\beta} + u, \quad u|\mathbf{x} \sim N(0, \sigma^2), \tag{7.15}$$

  y ist (ökonomisch) interessierende Variable.

- **Stichprobenmodell** für rechtszensierte zufällige Stichprobenbeobachtungen $(w_i, \mathbf{x}_i)$ mit Zensierungsvariable c_i

$$w_i = \begin{cases} c_i, & \text{falls } y_i \geq c_i \\ y_i & \text{sonst,} \end{cases}$$

$$y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i, \quad u_i|\mathbf{x}_i, c_i \sim N(0, \sigma^2).$$

  – Zensierung darf mit $\mathbf{x}_i$ variieren, z.B. Einkommensschranken mit Haushaltsgröße.

  – Linkszensiertes Stichprobenmodell analog.

  – Auch doppelseitige Zensierung möglich.

- OLS inkonsistent, wenn
  - nur $(y_i, \mathbf{x}_i)$ für $y_i < c_i$ verwendet werden $\Rightarrow$ Modelle für gestutzte Daten,
  - $(w_i, \mathbf{x}_i)$ verwendet wird, da dann formal zu Tobit-Modell für Daten mit Ecklösungen (Corner-Solution-Responses) gleich.
- in Literatur wird im Allgemeinen das Modell für zensierte Daten als Tobit-Modell bezeichnet, so z.B. auch in dem Lehrbuch für den Master-Kurs „Methoden in Ökonometrie“: Davidson und MacKinnon (2004).

- **Dichte und Log-Likelihood**:
  - bei linksseitiger Zensierung mit $c = 0$ identisch zur Dichte des Tobit Modells für Ecklösungen in Abschnitt 7.4.
  - bei rechtsseitiger Zensierung:

$$
\begin{aligned}
P(w_i = c_i|\mathbf{x}_i) &= P(y_i \geq c_i|\mathbf{x}_i) \\
&= P(\mathbf{x}_i\boldsymbol{\beta} + u_i \geq c_i|\mathbf{x}_i) = P(u_i \geq c_i - \mathbf{x}_i\boldsymbol{\beta}|\mathbf{x}_i) \\
&= P\left(\frac{u_i}{\sigma} \geq \frac{c_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\middle| \mathbf{x}_i\right) \\
&= 1 - P\left(\frac{u_i}{\sigma} \leq \frac{c_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\middle| \mathbf{x}_i\right) \\
&= 1 - \Phi\left(\frac{c_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right), \text{für } w_i = c_i, \\
f(y_i|\mathbf{x}_i) &= \frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right), \quad \text{für } w_i < c_i.
\end{aligned}
$$

Log-Likelihood

$$\mathcal{L} = \sum_{i=1}^{n} \mathbf{1}(w_i < c_i) \ln \left[\frac{1}{\sigma} \phi \left(\frac{w_i - \mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) \right] + \mathbf{1}(w_i = c_i) \ln \left[1 - \Phi \left(\frac{c_i - \mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) \right].$$

- **Interpretation**
  - Da man am Zusammenhang von y und $\mathbf{x}$ ($=$ Modell der Grundgesamtheit) (7.15) und nicht am Zusammenhang von w und $\mathbf{x}$ interessiert ist, kann β **wie im OLS-Fall** interpretiert werden!
  - Unterschied zu Tobit-Modell für Ecklösungen, wo man nicht am Zusammenhang von y^* und $\mathbf{x}$, sondern von y und $\mathbf{x}$ interessiert ist!

- **Probleme**:
  - wie im Tobit-Modell.
  - Zensierung ist 'teuer', da Robustheit der Konsistenz des OLS-Schätzers bei Heteroskedastie oder nichtnormalverteilten Fehlern verlorengeht!

7.6 Gestutzte Daten

- Klassischer Fall: Zusammenhang in Grundgesamtheit erfüllt normales Regressionsmodell

$$y = \mathbf{x}\boldsymbol{\beta} + u, \quad u|\mathbf{x} \sim N(0, \sigma^2),$$

  aber in der Stichprobe gilt $y_i \leq c_i$ und somit ist $\mathbf{x}_i$ bzw. $(y_i, \mathbf{x}_i)$ nur beobachtbar, wenn $y_i \leq c_i$.

- Kann also die abhängige Variable in einem bestimmten Datenbereich nicht beobachtet werden, obwohl diese Variablenwerte in der Grundgesamtheit vorkommen, ist eine erhobene Stichprobe $(y_i, \mathbf{x}_i)$ **keine Zufallsstichprobe**!

- **Dichte**

  – Für eine gestutzte normalverteilte Zufallsvariable $Z \sim N(0,1)$. (Hier ist Z Zufallsvariable und z eine Realisation davon.) Dann gilt

$$f(z|Z < a) = \frac{\phi(z)}{\Phi(a)} \quad \text{für beliebiges } a$$

  (für den Fall $Z > a$ siehe z.B. Abschnitt 8.1 im Anhang).

  – Mit (wie bereits geübt)

$$\begin{aligned} \mathbf{x}\boldsymbol{\beta} + u &< c \\ u &< c - \mathbf{x}\boldsymbol{\beta} \\ \frac{u}{\sigma} &< \frac{c - \mathbf{x}\boldsymbol{\beta}}{\sigma} \end{aligned}$$

lautet die Dichte des normalverteilten gestutzten Regressionsmodells

$$f(y_i|\mathbf{x}_i, c_i) = \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{c_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}.$$

- **Maximum-Likelihood-Schätzer**
  - sind konsistent und asymptotisch normalverteilt, wenn Annahmen des Modells erfüllt (insbesondere: korrekte Spezifikation, keine Heteroskedastie, normalverteilte Fehler) sind.

- **Beispiel**: Zur Schätzung einer Lohngleichung liegen nur Beobachtungen für Individuen mit Stundenlöhnen zwischen 0 und 8 Euro vor. Datensatz: `wagepan`

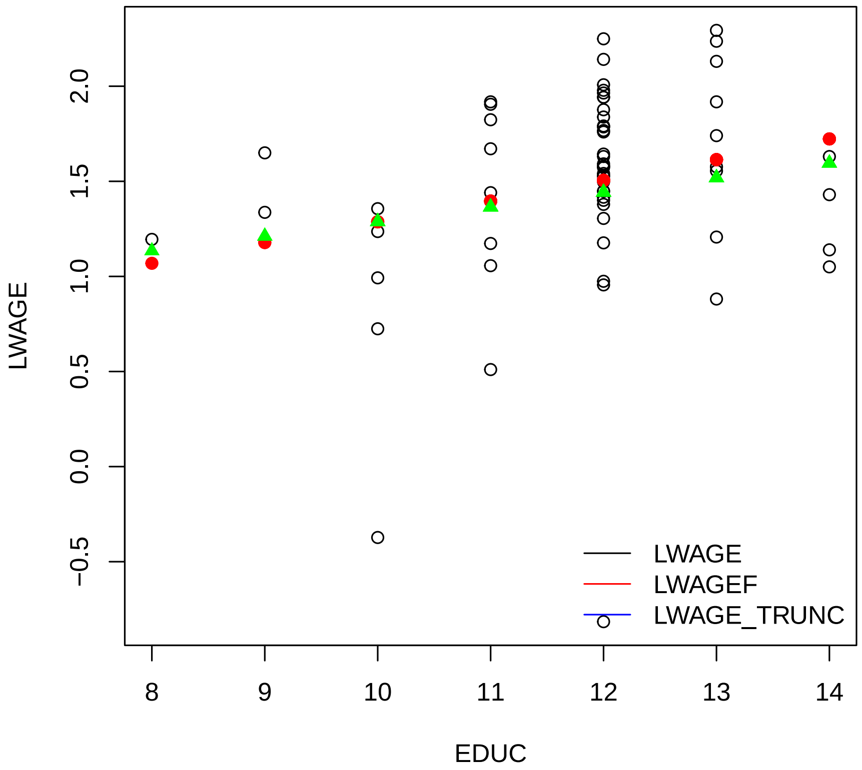
  OLS-Schätzung mit gestutzten Daten:

```
Call:
lm(formula = lwage ~ educ, data = wagepan, subset = (year ==
    1980 & black == 0 & hisp == 0 & married == 0 & exper == 3 &
    lwage <= log(8)))

Residuals:
     Min       1Q   Median       3Q      Max
-2.25740 -0.15877  0.07908  0.33927  0.56558

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.51977    0.67028   0.775    0.441
educ         0.07686    0.05672   1.355    0.181

Residual standard error: 0.5093 on 56 degrees of freedom
Multiple R-squared:  0.03175,Adjusted R-squared:  0.01446
F-statistic: 1.836 on 1 and 56 DF,  p-value: 0.1809

 aic      hq       sc
1.522225 1.549901 1.593275
```

OLS-Schätzung mit allen Daten:

```
Call:
lm(formula = lwage ~ educ, data = wagepan, subset = (year ==
    1980 & black == 0 & hisp == 0 & married == 0 & exper == 3))

Residuals:
     Min       1Q   Median       3Q      Max
-2.32032 -0.16314  0.06922  0.31864  0.74440

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.1970     0.6773   0.291   0.7721
educ          0.1090     0.0570   1.912   0.0605 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5257 on 61 degrees of freedom
Multiple R-squared:  0.05656,Adjusted R-squared:  0.04109
F-statistic: 3.657 on 1 and 61 DF,  p-value: 0.06053

 aic      hq       sc
1.583146 1.609904 1.651182
```

$\Rightarrow$ OLS-Schätzer bei gestutzten Daten gegen Null verzerrt.



Im Folgenden sind die Schätzergebnisse für ein gestutztes normales Regressionsmodell und ein gestutztes logistisches Regressionsmodell angegeben.

ML-Schätzung eines gestutzten normalen Regressionsmodells:

```
Call:
trch(formula = lwage ~ educ, data = wagepan, subset = (year == 1980 & black == 0 & hisp == 0 & married == 0 & exper == 3 & lwage <=
    log(8)), link.scale = "identity", dist = "gaussian", right = log(8), type = "ml")

Standardized residuals:
    Min      1Q  Median      3Q     Max
-3.5664 -1.0540 -0.7699 -0.4944  0.0717

Coefficients (location model):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.2734     1.7750  -0.154    0.878
educ          0.2069     0.1669   1.240    0.215

Coefficients (scale model with identity link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   0.8482     0.2295   3.695  0.00022 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Distribution: gaussian
Log-likelihood: -30.36 on 3 Df
Number of iterations in BFGS optimization: 19
>     # Berechne Informationskriterien mit ganz oben definierter Funktion
> IC.loglik(tobit_gestutzt_norm$loglik, length(tobit_gestutzt_norm$coefficients),
+         tobit_gestutzt_norm$n)

AIC 1.150269
BIC 1.256843
HQ  1.191781
```

ML-Schätzung eines gestutzten logistischen Regressionsmodells:

```
Call:
trch(formula = lwage ~ educ, data = wagepan, subset = (year == 1980 & black == 0 & hisp == 0 & married == 0 & exper == 3 & lwage <=
    log(8)), link.scale = "identity", dist = "logistic", right = log(8), type = "ml", x = TRUE, y = TRUE)

Standardized residuals:
    Min      1Q  Median      3Q     Max
-7.7011 -1.2503 -0.4521  0.2902  1.0173

Coefficients (location model):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.47667    0.93786   0.508    0.611
educ         0.10097    0.08273   1.220    0.222

Coefficients (scale model with identity link):
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.32510    0.05863   5.545 2.94e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Distribution: logistic
Log-likelihood: -27.63 on 3 Df
Number of iterations in BFGS optimization: 18
> IC.loglik(tobit_gestutzt_logit$loglik, length(tobit_gestutzt_logit$coefficients),
+           tobit_gestutzt_logit$n)

AIC 1.056256
BIC 1.16283
HQ 1.097769
```

Man sieht beim Vergleich der Parameterschätzungen für das gestutzte normale und das gestutzte logistische Regressionsmodell deutlich die Bedeutung der Verteilungsannahme für die Fehler! Das Schätzergebnis für β_1 des gestutzten normalen Regressionsmodells ist weiter weg vom unzensierten OLS-Schätzer als der OLS-Schätzer unter Ignorieren der Stutzung.

Die Betrachtung des Scatterplots macht deutlich, dass die Normalverteilungsannahme wohl nicht gültig ist, da für educ-Werte 10 und 12 jeweils ein sehr niedriger Lohn beobachtet wird. Damit ist die Fehlerverteilung sicher nicht homoskedastisch und wohl auch nicht normalverteilt. Die Schätzung für β_1 ist beim logistischen Modell näher an der unzensierten OLS-Schätzung. Insgesamt gilt, dass bei Fehlspezifikation der Fehlerverteilung im Vergleich zum inkonsistenten OLS-Schätzer nichts gewonnen wird.

Man sieht außerdem, dass durch die Stutzung die Signifikanz von educ verlorengeht. Die p-Werte steigen von 0.06 auf zwischen 0.18 und 0.22. Durch die Selektion von Stichprobenbeobachtungen kann also wichtige Stichprobeninformation verlorengehen!

Eine Residuenanalyse ist bei zensierten und gestutzten Regressionen schwieriger. Siehe Wooldridge (2002, 2010, Chapter 16 und 17).

7.7 Stichprobenauswahlverzerrungen (Sample Selection Bias)

Außer bei Stichproben von Zeitreihendaten und von gestutzten Daten wurde bisher angenommen, dass eine Zufallsstichprobe vorliegt. In der Praxis ist dies häufig nicht der Fall. Wann ist dies ein Problem und welche Lösungen gibt es ggf.?

Kategorisierung von Stichprobenauswahlproblemen

- In der Grundgesamtheit gilt

$$y = \mathbf{x}\boldsymbol{\beta} + u, \quad E[u|\mathbf{x}] = 0. \tag{7.16}$$

- Im Folgenden sei s_i eine binäre Variable

$$s_i = \begin{cases} 1 & \text{wenn } (y_i, \mathbf{x}_i) \text{ beobachtbar ist,} \\ 0 & \text{wenn } y_i \text{ oder/und } \mathbf{x}_i \text{ nicht beobachtbar ist.} \end{cases}$$

- Bei n Ziehungen liegt als Stichprobe

$$s_i y_i = s_i \mathbf{x}_i \boldsymbol{\beta} + s_i u_i, \quad i = 1, \ldots, n, \tag{7.17}$$

vor, wobei man für $s_i = 1$ für alle $i = 1, \ldots, n$ gerade das Standardmodell erhält.

- **Konsistenz des OLS-Schätzers erfordert** für (7.17) ((TS.3') in Zeitreihenökonometrie, Abschnitt 5.3)

$$E[su|s\mathbf{x}] = 0. \tag{7.18}$$

Daraus folgt

$$E[(s\mathbf{x})(su)] = 0. \tag{7.19}$$

Wann ist der OLS-Schätzer konsistent? D.h. wann gilt (7.18)?

– **Zufallsstichprobe**:

$$s_i y_i = s_i \mathbf{x}_i \boldsymbol{\beta} + s_i u_i, \quad i = 1, \ldots, n.$$

Ist s_i stochastisch unabhängig von $(y_i, \mathbf{x}_i)$, dann gilt (7.18)

$$E[(s\mathbf{x})(su)] = E[s\mathbf{x}u] \overset{\text{unabh.}}{=} E[s]E[u\mathbf{x}] = 0.$$

De facto hat man in diesem Fall eine Zufallsstichprobe mit möglicherweise weniger als n Beobachtungen.

Im vorliegenden Fall darf s durchaus von Zufallsvariablen abhängen, allerdings müssen diese von u und $\mathbf{x}$ stochastisch unabhängig sein!

– **Exogene Stichprobenauswahl** $s = s(\mathbf{x})$: dann gilt

$$s(\mathbf{x})E[u|s(\mathbf{x})\mathbf{x}] = sE[u|s\mathbf{x}] = sE[u|\mathbf{x}] = 0.$$

– Für IV-Schätzer wird (7.18) zu $E[su|s\mathbf{z}] = 0$ umgewandelt und überprüft.

- **Typische Fälle von Inkonsistenz**

– **gestutzte Daten**: $s = s(y, c)$, z.B. von oben gestutzte Daten

$$\begin{aligned} s_i = s(y_i, c_i) &= \mathbf{1}(y_i < c_i) \\ = s(u_i, \mathbf{x}_i, c_i) &= \mathbf{1}(u_i < c_i - \mathbf{x}_i\boldsymbol{\beta}). \end{aligned}$$

– **Incidental Truncation**: $s = s(y, d^*)$: die Beobachtbarkeit s hängt von einem Schwellenwert d ab, der nicht direkt beobachtbar ist

$$s_i = \mathbf{1}(y_i \geq d_i^*). \tag{7.20}$$

Bekanntes **Beispiel**: Die Schätzung einer **Lohnangebotsfunktion**: Unter der Annahme, dass der tatsächlich empfangene individuelle Lohn dem individuellen Lohnangebot w_i^A entspricht, für das jemand bereit ist zu arbeiten, ist w^A nur beobachtbar, wenn w^A größer oder gleich dem individuellen Reservationslohn w^R ist, also $s_i = \mathbf{1}(w_i^A \geq w_i^R)$ und damit (7.20) gilt, da man $d_i^* = w_i^R$ nicht direkt beobachten kann.

Lösung des Problems im nächsten Teil.

Incidental Truncation

- Gemäß (7.20) gilt

$$s_i = s(y_i, d_i^*) = \mathbf{1}(y_i \geq d_i^*),$$

wobei d_i^* unbeobachtbar ist.

- Man kann jedoch **unterstellen, dass die beobachtbare binäre Variable s_i einem Probit-Modell** entspricht und man erhält

$$s_i = \mathbf{1}(\mathbf{z}\boldsymbol{\gamma} + v > 0), \quad v|\mathbf{z} \sim N(0,1), \tag{7.21}$$

wobei $\mathbf{z}$ alle Variablen von $\mathbf{x}$ enthält plus mindestens eine zusätzliche exogene Variable, sowie

$$E[u|\mathbf{z}] = 0 \tag{7.22}$$

gilt.

- Wann ist nun (7.18), d.h. $E[su|s\mathbf{z}] = 0$ verletzt, wobei $\mathbf{x}$ durch $\mathbf{z}$ ersetzt wird?

– Beachte, dass die Beobachtung von s und $\mathbf{z}$ keinen Rückschluss auf v zulässt. Würden wir annehmen, dass wir auch v beobachten könnten, u jedoch nicht, gilt wenn u und v gemeinsam unabhängig von $\mathbf{z}$ und gemeinsam normalverteilt sind, dass

$$E[u|\mathbf{z}, v] \overset{u,\, v \text{ gemeinsam unabh. v. } \mathbf{z}}{=} E[u|v] \overset{u, v \text{ gemeins. normalvert.}}{=} \rho v.$$

– Ist $\rho = 0$, **sind die Störterme u und v unkorreliert** (und wegen der Normalverteilungsannahme unabhängig), gilt

$$E[u|\mathbf{z}, v] = E[u|\mathbf{z}] = 0$$

und somit auch

$$E[u|s] = 0,$$

da ja s bekannt ist, wenn $\mathbf{z}$ und v bekannt sind.

Für $\rho = 0$ ist OLS also konsistent und es kommt zu keiner Verzerrung durch eine selektive Stichprobenerhebung.

- **Korrekte Prognose von y bei selektiver Stichprobe im Fall $\rho \neq 0$**
  - Wegen $E[u|v] = \rho v$ ist

$$\begin{aligned} E[y|\mathbf{z}, v] &= \mathbf{x}\boldsymbol{\beta} + E[u|v] \\ &= \mathbf{x}\boldsymbol{\beta} + \rho v \end{aligned} \tag{7.23}$$

    die bestmögliche Prognose für y gegeben $\mathbf{z}$ und v, d.h. **wenn v bekannt wäre**.

  - Da v nicht beobachtbar ist, sondern lediglich s und $\mathbf{z}$, lässt sich (7.23) nicht berechnen. Stattdessen muss man

$$E[y|\mathbf{z}, s = 1] = \mathbf{x}\boldsymbol{\beta} + \rho E[v|\mathbf{z}, s = 1]$$

    betrachten. Für $E[v|\mathbf{z}, s = 1]$ ergibt sich aufgrund der Normalverteilungsannahme (vgl. Abschnitt 7.4)

$$\begin{aligned} E[v|\mathbf{z}, s = 1] &= E[v|\mathbf{z}\boldsymbol{\gamma} + v > 0] \\ &= E[v|v > -\mathbf{z}\boldsymbol{\gamma}] \\ &= \lambda(\mathbf{z}\boldsymbol{\gamma}) \end{aligned}$$

und somit

$$E[y|\mathbf{z}, s = 1] = \mathbf{x}\boldsymbol{\beta} + \rho\lambda(\mathbf{z}\boldsymbol{\gamma}).$$

Wird der Term $\rho\lambda(\mathbf{z}\boldsymbol{\gamma})$ in der Regression zur Schätzung von $\boldsymbol{\beta}$ weggelassen, kommt es zum 'omitted variable bias', da man dann vernachlässigt, dass man y_i nur selektiv beobachtet und deshalb mit den vorliegenden Beobachtungen nur die Regression für $E[y|\mathbf{z}, s = 1]$, nicht jedoch für $E[y|\mathbf{z}]$ geschätzt werden kann.

- **Interpretation**

  Ist man im Beispiel des Lohnangebots an dem Einfluss von $\mathbf{x}$ auf das Lohnangebot

  – *unabhängig davon, ob es beobachtbar ist oder nicht* — interessiert, betrachtet man das Modell der Grundgesamtheit (7.16)

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

  und damit $\partial E[y|\mathbf{x}]/\partial x_j$. Damit hat $\boldsymbol{\beta}$ die übliche Interpretation.

  – *wenn es beobachtbar ist* — interessiert, muss zur Berechnung der marginale Effekt $\partial E[y|\mathbf{x}, s = 1]/\partial x_j$ betrachtet werden. Zur Berechnung siehe Abschnitt 7.4.

- **Schätzung**

  – **Heckit**-Methode (Heckman, 1976)

  1. Man schätzt $\boldsymbol{\gamma}$ mit einem Probit Modell.
  2. Man schätzt die inverse Mills-Ratio $\lambda(\mathbf{z}\boldsymbol{\gamma})$ durch Einsetzen von $\hat{\boldsymbol{\gamma}}$.
  3. Man verwendet OLS für

$$y = \mathbf{x}\boldsymbol{\beta} + \rho\lambda(\mathbf{z}\hat{\boldsymbol{\gamma}}) + Fehler.$$

  Beachte:

  ∗ Asymptotischer t-Test für $H_0 : \rho = 0$ möglich.

  ∗ Traditionelle Standardfehler im Fall $\rho \neq 0$ falsch, da Schätzunsicherheit von $\hat{\boldsymbol{\gamma}}$ vernachlässigt wird.

  ∗ Um starke Multikollinearität zwischen $\mathbf{x}$ und $\lambda(\mathbf{z}\boldsymbol{\gamma})$ zu vermeiden, sollte $\mathbf{x}$ mindestens eine Ausschlussrestriktion bezüglich $\mathbf{z}$, also mindestens eine Variable weniger als $\mathbf{z}$, enthalten.

– **Maximum-Likelihood-Schätzung** (siehe z.B. Wooldridge (2002, 2010, Section 17.4)).

- **Probleme**:
  Sind die Regressoren nicht exogen oder gilt die Standardnormalverteilungsannahme für v nicht, sind allgemeinere Verfahren zu verwenden, siehe z.B. Wooldridge (2002, 2010, Chapter 17).

8 Anhang

8.1 Theorie zu gestutzten Zufallsvariablen

- Die Zufallsvariable Z habe unendlichen Träger (=Z kann jeden Wert auf der reellen Achse annehmen) und Dichte $f(z)$.
- Beachte, dass $\int_{-\infty}^{\infty} f(z)dz = 1$.
- **Dichte einer gestutzten Zufallsvariable**: Die Dichte der gestutzten Zufallsvariable $Z > a$ kann nicht $f(z)$ selbst sein, da

$$\int_{a}^{\infty} f(z)dz < 1.$$

  Um 1 zu erhalten, muss auf der linken Seite der fehlende Teil addiert werden:

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

Damit erhält man

$$\int_a^\infty f(z)dz = 1 - F(a)$$

$$\int_a^\infty \underbrace{\frac{f(z)}{1 - F(a)}}_{f(z|Z>a)} dz = 1.$$

Somit ist

$$f(z|Z > a) = \frac{f(z)}{1 - F(a)}.$$

- **Erwartungswert einer gestutzten Zufallsvariable**:

$$E\left[Z|Z > a\right] = \int_a^\infty z f(z|Z > a)dz = \frac{1}{1 - F(a)} \int_a^\infty z f(z)dz.$$

Beachte, dass

$$E\left[Z|Z > a\right] > E[Z]$$

da wegen der Stutzung die kleinsten Werte nicht in die Berechnung des Erwartungswertes einbezogen werden.

– **Standardnormalverteilte Zufallsvariable** $Z \sim N(0,1)$

$$
\begin{aligned}
f(z) = \phi(z) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \\
F(z) = \Phi(z) &\equiv \int_{-\infty}^{z} \phi(u)du \\
f(z|Z>a) &= \frac{\phi(z)}{1-\Phi(a)} \\
E\left[Z|Z>a\right] &= \int_a^\infty z f(z|Z>a)dz \\
&= \int_a^\infty z \frac{\phi(z)}{1-\Phi(a)} dz \\
&= \frac{1}{1-\Phi(a)} \int_a^\infty z \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} dz.
\end{aligned}
$$

Es gilt der **Hauptsatz der Differential- und Integralrechnung**

$$
\frac{d}{dx}\int_a^x g(u)du = \frac{d}{dx}\left[G(x) - G(a)\right] = g(x)
$$

wobei $dG(x)/dx \equiv g(x)$.

Beachte, dass

$$\frac{d}{dz}\underbrace{\phi(z)}_{G(z)} = \frac{d}{dz}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}$$
$$= \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}z^2}\left(-\frac{2}{2}z\right) = \underbrace{-z\phi(z)}_{g(z)}$$

und somit

$$\int_a^b (-z\phi(z))dz = \phi(b) - \phi(a)$$

und

$$E\left[Z|Z > a\right] = \frac{1}{1-\Phi(a)}(-1)\left(\phi(\infty) - \phi(a)\right)$$
$$= \frac{\phi(a)}{1-\Phi(a)} \equiv \lambda(a).$$

Der Term $\lambda(a) = \frac{\phi(a)}{1-\Phi(a)}$ wird als **Hazard-Rate** oder **inverse Mills-Ratio** bezeichnet.

Die Varianz

$$Var(Z|Z > a) = 1 - \underbrace{\lambda(a)(\lambda(a) - a)}_{0 \leq \cdot \leq 1}$$

liegt zwischen 0 und 1, da

$$\lambda(a) = E[Z|Z > a] > a.$$

– Normalverteilte Zufallsvariable $W \sim N(\mu, \sigma^2)$

$$Z = \frac{W - \mu}{\sigma}, \quad a = \frac{b - \mu}{\sigma}$$

$$f(w) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\frac{(w-\mu)^2}{\sigma^2}} = \frac{1}{\sigma}\phi(z)$$

$$f(w|W > b) = \frac{1}{\sigma} \frac{\phi\left(\frac{w-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{b-\mu}{\sigma}\right)} = \frac{1}{\sigma} \frac{\phi(z)}{1 - \Phi(a)}$$

$$\begin{aligned} E[W|W > b] &= E[\sigma Z + \mu | \sigma Z + \mu > \sigma a + \mu] \\ &= \mu + \sigma E[Z|Z > a] \\ &= \mu + \sigma\lambda(a) \\ Var(W|W > b) &= \sigma^2 Var(Z|Z > a). \end{aligned}$$

8.2 Übersicht über wichtige R Befehle

Für eine sehr gute Einführung in R, siehe Kleiber und Zeileis (2008).

Benötigte R-Pakete in diesem Kurs:

- `wooldridge` für alle Daten, da Beispiele aus Wooldridge (2009),
- `plm` zum Durchführen von Panelregressionen,
- `AER` für die Instrumentvariablenschätzung, Tobit-Schätzung,
- `crch` für die Tobit-Schätzung von gestutzten, bedingt heteroskedastischen Daten.

Weitere wichtige R-Pakete finden sich in folgender Übersicht zu wichtigen R-Befehlen.

R-Befehl	Funktionsbeschreibung	R-Paket
	Schätz- und Prognoseoutput	
`print()`	einfaches gedrucktes Display	
`summary()`	Standard Regressionsoutput	
`coef()`	(oder `coefficients()`) extrahiert geschätzte Regressionsparameter	
`residuals()`	(oder `resid()`) extrahiert Residuen	
`fitted()`	(oder `fitted.values()`) extrahiert angepasste/gefittete Werte	
`anova()`	Vergleich von geschachtelten Modellen (nested models)	
`predict()`	Vorhersagen für neue Regressionswerte	
`confint()`	Konfidenzintervalle für Regressionskoeffizienten	
`confidenceEllipse()`	Konfidenzintervalle für Regressionskoeffizienten	`car`
`deviance()`	Residuenquadratsumme (RSS)	
`vcov()`	(geschätzte) Varianz-Kovarianzmatrix der Parameterschätzer	
`logLik()`	Log-Likelihood (unter der Annahme normalverteilter Fehler)	
`fixef()`	Ausgabe der fixen Effekte eines plm()-Objekts	`plm`

Testen		
`hccm()`	Heteroskedastie-korrigierte Varianz-Kovarianzmatrix der Parameterschätzer; mit `type=''hc0''` White-Varianz-Kovarianzmatrix, siehe **Methoden der Ökonometrie**	`car`
`coeftest()`	Standard-Regressionsoutput, ggf. mit heteroskedastie-robusten Standardfehlern	`lmtest`
`linearHypothesis()`	F-Test `test=c(''F'')` oder (asymptotischer) χ^2-Test `test=c(''Chisq'')`; mit `white.adjust=c(FALSE, TRUE, ''hc0'')` White-heteroskedastierobuste-Varianz-Kovarianzmatrix	`car`
`lrtest()`	Likelihoodratio-Test, siehe Abschnitt 7.2 oder **Advanced Econometrics**	`lmtest`
`waldtest()`	Wald-Test, siehe Abschnitt 7.2 oder **Advanced Econometrics**	`lmtest`

Modellspezifikation		
`AIC()`	Informationskriterien einschließlich AIC, BIC/SC (unter der Annahme normalverteilter Fehler) - Beachte: Im Gegensatz zu EViews wird die geschätzte Parametervarianz als Parameter mitgezählt und nicht durch die Zahl der Beobachtungen dividiert, siehe **Methoden der Ökonometrie**	
`SelectCritEViews()`	Informationskriterien à la EViews, siehe **Methoden der Ökonometrie**	eigenes R-Programm, siehe Abschnitt 8.3
`encomptest()`	Encompassing-Test zum Testen nicht geschachtelter Regressionsmodelle, siehe **Methoden der Ökonometrie**	`lmtest`
`jtest()`	J-Test zum Testen nicht geschachtelter Regressionsmodelle, siehe **Methoden der Ökonometrie**	`lmtest`
`acf()`	(graphische) Ausgabe der geschätzten Autokorrelationsfunktion einer Zeitreihe, verwende `acf(resid(model))` zur Betrachtung der Autokorrelationsfunktion in den Fehlern, siehe **Zeitreihenökonometrie**	
`pacf()`	analog zu `acf()`, jedoch Ausgabe der geschätzten partiellen Autokorrelationsfunktion	
`uc.df()`	**A**ugmented **D**ickey-**F**uller-Einheitswurzeltest zum Testen auf Stationarität einer Zeitreihe, siehe siehe **Zeitreihenökonometrie**	`urca`

Modelldiagnose

`plot()`	Graphiken zur Modellüberprüfung	
`resettest()`	RESET-Test zum Testen der funktionalen Form, siehe **Methoden der Ökonometrie**	`lmtest`
`jarque.test()`	Lomnicki-Jarque-Bera-Test zum Überprüfen normalverteilter Fehler, siehe **Methoden der Ökonometrie**	`moments`
`bptest()`	Breusch-Pagan-Test zum Testen auf Vorliegen von heteroskedastischen Fehlern, siehe **Methoden der Ökonometrie**	`lmtest`
`bgtest()`	Breusch-Godfrey-Test zum Testen auf Vorliegen von serieller Autokorrelation beliebiger Ordnung in den Fehlern, siehe **Zeitreihenökonometrie**	`lmtest`
`dwtest()`	Durbin-Watson-Test zum Testen auf Vorliegen von serieller Autokorrelation erster Ordnung in den Fehlern, siehe **Zeitreihenökonometrie**	`lmtest`
`whitetest()`	White-Test zum Testen auf Vorliegen von heteroskedastischen Fehlern, siehe **Methoden der Ökonometrie**	eigenes R-Programm, siehe **Methoden**

weitere Modelle und Schätzverfahren		
`plm()`	Schätzung Paneldatenmodellen, z.B. Differenzen-, Fixed-Effects- und Random-Effects-Schätzungen, siehe Abschnitt 4.2.1, 4.2.2 und 4.3	`plm`
`ivreg()`	Instrumentvariablenschätzung, siehe Abschnitt 5	`AER`
`glm()`	Schätzung einiger bekannter generalisierter linearer Modelle, z.B. Schätzung von Logit- und Probitmodellen, siehe Abschnitt 7.3	`AER`
`tobit()`	Schätzung von Tobit-Modellen mit zensierten Daten, siehe Abschnitt 7.5	`AER`
`truncreg()`	Schätzung von Tobit-Modellen mit gestutzten Daten, siehe Abschnitt 7.6	`truncreg`

Eine umfangreichere Übersicht weiterer Pakete für ökonometrische Analysen mit R befindet sich auf `http://cran.r-project.org/web/views/Econometrics.html`.

8.3 R-Programm für die empirischen Beispiele

```
################### WOE_ws21_Emp_Beispiele_final.R #############################
#
################################################################################
################################################################################
# R-Programm zum Reproduzieren der empirischen Beispiele in den
# Folien Weiterführende Fragen der Ökonometrie, Universität Regensburg
# erstellt von Patrick Kratzer und Rolf Tschernig
# Stand: 06.10.2021, RT
#
# Das Programm erfordert Daten aus Wooldridge (2016) oder frühere Editionen
# fertil1.txt, kielmc.txt, jtrain.txt, wagepan.txt, openness.txt, mroz.txt,
# Universität Regensburg
# Weiterführende Fragen der Ökonometrie, Wintersemester 2019/2020
# Stand: 24.01.2020, RT, PK
# erfordert Textdatendateien  aus Wooldridge (2016) oder frühere Editionen:
# fertil1, kielmc, jtrain, wagepan, openness, mroz.
# Aktuell werden die Daten via dem Paket wooldridge geladen.
# Das Laden via Textdateien ist im Code auskommentiert.


################################################################################
#                   Beginn Definition Funktionen
################################################################################

####################### Funktion SelectCritEviews ##############################
# Funktion zur Berechnung von Modellselektionskriterien wie in EViews
# RT, 2011_01_26

SelectCritEviews <- function(model)
{
```

```
  n            <- length(model$residuals)
  k            <- length(model$coefficients)
  fitmeasure   <- -2*logLik(model)/n

  aic          <- fitmeasure + k * 2/n
  hq           <- fitmeasure + k * 2*log(log(n))/n
  sc           <- fitmeasure + k * log(n)/n
  sellist      <- list(aic=aic[1],hq=hq[1],sc=sc[1])
  return(t(sellist))
}
############################### Ende ###########################################

############################ Funktion IC.loglik ################################
# Definiere Funktion zur Berechnung der Selektionskriterien auf Basis der
# Loglikelihoodfunktion (wird bei Tobitmodellen benötigt)
IC.loglik <- function(loglik, k, n){
    AIC <- (-2 * loglik + 2 * k) / n
    BIC <- (-2 * loglik + log(n) * k) / n
    HQ  <- (-2 * loglik + 2* log(log(n)) * k) / n
    return(list(AIC = AIC, BIC = BIC, HQ = HQ))
}
############################### Ende ###########################################

################################################################################
#                   Ende Definition Funktionen
################################################################################

################################################################################
#                   Beginn Hauptprogramm
################################################################################

############ Bestimme Parameter für das R-Programm #############################
```

```
print_pdf <- 0    # 1: Speichern von Graphiken als PDF, 0: Nichtspeichern
# setwd("C:/temp_rolf/WS2122/WOE/Vorlesung/R_code")
                  # Setze Working Directory,
                  # falls Daten in Textdateien,
                  # falls Abspeichern von Graphiken

########### Installieren und Laden von R-Paketen################################

# Paket zum Bereitstellen der Wooldridge-Daten
if (!require(wooldridge)){
  install.packages("wooldridge")
}
library("wooldridge")

# Paket zur Paneldatenanalyse
if (!require(plm)){
    install.packages("plm")
}
library("plm")

# für die Instrumentvariablenschätzung, Tobit-Schätzung
if (!require(AER)){
    install.packages("AER")
}
library("AER")

# für die Tobit-Schätzung von gestutzten, bedingt heteroskedastischen Daten
if (!require(crch)){
    install.packages("crch")
}
library("crch")
```

```
# # für die Tobit-Schätzung von gestutzten normalverteilten Daten
# if (!require(truncreg)){
#     install.packages("truncreg")
# }
# library("truncreg")

################################################################################
# Lade Daten
# Daten für  gepoolte Querschnittsregression: Fertilitätsbeispiel
# fertil1 <- read.table("fertil1.txt", header = TRUE)
data('fertil1')

# Daten für kausale Effekte: Müllverbrennungsanlage (Folien 13, 51ff)
# kielmc <- read.table("kielmc.txt", header = TRUE)
data('kielmc')

# Daten für Panelbeispiel: Jobtraining (ab Folie 97)
# jtrain <- pdata.frame(read.table("jtrain.txt", header = TRUE),
#                       index = c("fcode", "year"))
data('jtrain')

# Daten für Panelbeispiel: mit Stundenlöhnen (Table 14.2 in Wooldridge (2016))
# wagepan <- pdata.frame(read.table("wagepan.txt", header = TRUE),
#                         index = c("nr", "year"))
data('wagepan')

# Daten für IV-Beispiel: mit Inflation and Openess
#(Example 16.6 in Wooldridge (2016))
# openness <- read.table("openness.txt", header = TRUE)
data('openness')
```

```
# Daten für Probit- und Logitbeispiel: Arbeitsmarkteilnahme von Frauen
#(Example 17.1 in Wooldridge (2016)), Folien 243 folgende
# mroz <- read.table("mroz.txt", header = T)
data('mroz')

################################################################################
# Gepoolte Querschnittsregression: Fertilitätsbeispiel, ab Folie 9
# Datensatz: fertil1
# Folie 10
fertil_72_ols <- lm(kids ~ educ + age + I(age^2) + black + east + northcen +
            west + farm + othrural + town + smcity, data = fertil1,
            subset = year == 72)
summary(fertil_72_ols)
SelectCritEviews(fertil_72_ols)

# Folie 11
fertil_74_ols <- lm(kids ~ educ + age + I(age^2) + black + east + northcen +
            west + farm + othrural + town + smcity, data = fertil1,
            subset = year == 74)
summary(fertil_74_ols)
SelectCritEviews(fertil_74_ols)

################################################################################
# Querschnittsregression: Beispiel MüllverbrennungsanlageFertilitätsbeispiel
# Datensatz: kielmc
# Folie 13
kielmc_81_ols <- lm(rprice ~ nearinc, data = kielmc, subset = year == 1981)
summary(kielmc_81_ols)
SelectCritEviews(kielmc_81_ols)

# Folie 41 zu Fertilitätsbeispiel
y74 <- as.numeric(fertil1$year == 74)
```

```
fertil_72_74_ols <- (lm(kids ~ educ + age + I(age^2) + black + east + northcen +
           west + farm + othrural + town + smcity + y74, data = fertil1,
           subset = year == 72 | year == 74))
summary(fertil_72_74_ols)
SelectCritEviews(fertil_72_74_ols)

# Folie 42 (Ergänzung: Chow-Test )
summary(lm( kids ~ y74 + educ + I(y74*educ) + age + I(age^2) + black +
                east + northcen + west + farm + othrural + town + smcity,
            data = fertil1, subset = year == 72 | year == 74))

################################################################################
# Kausale Effekte: Beispiel Müllverbrennungsanlage, Folie 51
# Datensatz: kielmc

y81 <- as.numeric(kielmc$year == 1981)

# Keine Kontrollvariablen
kielmc_ols <- lm(rprice ~ nearinc + y81 + I(y81*nearinc), data = kielmc)
summary(kielmc_ols)
SelectCritEviews(kielmc_ols)

# Kontrollvariablen age und age^2
kielmc_age_age2_ols <- lm(rprice ~ y81 + nearinc + I(y81*nearinc) + age + agesq,
           data = kielmc)
summary(kielmc_age_age2_ols)
SelectCritEviews(kielmc_age_age2_ols)

# Alle Kontrollvariablen
kielmc_all_ols <-lm(rprice ~ y81 + nearinc + I(y81*nearinc) + age + agesq +
               intst + land + area + rooms + baths,
           data = kielmc)
```

```
summary(kielmc_all_ols)
SelectCritEviews(kielmc_all_ols)

################################################################################
# Paneldaten: Jobtraining und Ausschussrate, ab Folie 97
# Datensatz: jtrain
# Folien 97 und folgende: Fixed-Effects-Schätzer
    # Erstellen eines panel-dataframes
jtrain_pan <- pdata.frame(jtrain, index = c("fcode", "year"))

## 2. Berechnen des Wihin-Group-Schätzers ohne plm-Paket
#  eigene Berechnung des fixed-effects-Schätzer
    # Umwandlung der Variablen fcode von "factor" zu "character".
    # Letzteres ist notwendig, um Mittelwerte über die Zeit mit
    # tapply() zu berechnen
fcode <- as.character(jtrain_pan$fcode)
    # tapply: wendet Funktion "mean" (drittes Argument) auf
    # Variable (erstes Argument) an, wobei die Einheiten über den gleichlangen
    # Charakter-Vektor (zweites Argument) identifiziert werden.
    # Das Ergebnis wird dann angewendet auf den Charakter-Vektor in [].
lscrap_2p  <- jtrain$lscrap  - tapply(jtrain_pan$lscrap,  fcode, mean)[fcode]
grant_2p   <- jtrain$grant   - tapply(jtrain_pan$grant,   fcode, mean)[fcode]
grant_1_2p <- jtrain$grant_1 - tapply(jtrain_pan$grant_1, fcode, mean)[fcode]
d88_2p     <- jtrain$d88     - tapply(jtrain_pan$d88,     fcode, mean)[fcode]
d89_2p     <- jtrain$d89     - tapply(jtrain_pan$d89,     fcode, mean)[fcode]

    # selbsterstellter Within-group-Schätzer
eq_table14_1 <- lm(lscrap_2p ~ 0 + grant_2p + grant_1_2p + d88_2p + d89_2p)
summary(eq_table14_1)
  # BEACHTE: Std. Error, t value und Pr() sind im Vergleich zum plm()-Befehl
  # falsch, da die Freiheitsgrade nicht richtig berechnet werden. Es werden
  # zu wenig verwendet.
```

```
## 3. Direkte Berechnung des FE-Schätzers und expliziten Zeitdummies mit plm-Paket

eq_table14_1_fullR <- plm(lscrap ~ grant + grant_1 + d88 + d89,
                          data = jtrain_pan, model = "within", effect = "individual")
summary(eq_table14_1_fullR)
# Ausgabe der geschätzten fixen Effekte
fixef(eq_table14_1_fullR, effect="individual")
# Umrechnung in Cross-section Effects a la EViews
# Berechnung Konstante als Mittelwert der fixen Effekte
hat_beta_0 <- mean(fixef(eq_table14_1_fullR))
# Cross-section Fixed Effects a la EViews
fixef(eq_table14_1_fullR) - hat_beta_0

## 4. Direkte Berechnung des FE-Schätzers mit Cross-Section und Period-Effekts
#  mit plm-Paket

# Fixe Effekte und Period Effects mit plm()
eq_table14_1_cs_per <- plm(lscrap ~ grant + grant_1,
                           data = jtrain_pan, model = "within", effect = "twoways")
summary(eq_table14_1_cs_per)

# Durchscnitt der fixen Effekte
hat_beta_0 <- mean(fixef(eq_table14_1_cs_per, effect="twoways"))

# Cross-section Effects a la EViews
fixef(eq_table14_1_cs_per, effect="individual") - mean(fixef(eq_table14_1_cs_per, effect="individual"))
# Period Effects a la EViews
fixef(eq_table14_1_cs_per, effect="time") - mean(fixef(eq_table14_1_cs_per, effect="time"))

## 5. Fixed effects by hand and period effects a la EViews (Gleichung (4.19')
d87 <- as.numeric(jtrain_pan$year == 1987)
```

```
eq_fixed_time_effects <- lm(lscrap_2p ~ grant_2p + grant_1_2p + I(d87-d89) +
                                I(d88-d89), data = jtrain_pan)
summary(eq_fixed_time_effects)



################################################################################
# Paneldaten: Lohnbeispiel in Wooldridge (2016), Table 14.2, Folie 114
# Datensatz: wagepan

    # Erstellen eines Panel-Dataframes
wagepan_pan <- pdata.frame(wagepan, index = c("nr", "year"))

    # Pooled-OLS-Schätzer
    # estimated directly via the lm command
eq_table14_2_Pooled <- lm(lwage ~ educ + black + hisp + exper +
                           expersq + married + union + year,
                       data = wagepan_pan)
    # or using the plm command which produces the same estimates but contains
    # different quantities in the summary output
eq_table14_2_Pooled <- plm(lwage ~ educ + black + hisp + exper +
                           expersq + married + union + year,
                       data = wagepan_pan, model = "pooling", effect = "individual")

summary(eq_table14_2_Pooled)

    # Random-Effects-Schätzer
    # (in EViews. Cross-section: Random, Period: Fixed)
eq_table14_2_RE <- plm(lwage ~ educ + black + hisp + exper +
                           expersq + married + union + year,
                           data = wagepan_pan, model = "random", effect = "individual")
summary(eq_table14_2_RE) # enthält Output für Folie 114
```

```
    # Fixed-Effects-Schätzer mit Zeitdummies explizit
    # Beachte: das R^2 wird nach Bereinigung von Cross-Section Effects
    # ausgewiesen
eq_table14_2_FE <- plm(lwage ~ expersq + married + union + year,
                       data = wagepan_pan, model = "within", effect = "individual")
summary(eq_table14_2_FE)

    # Fixed-Effects-Schätzer mit automatisch generierten Period Effects
    # Beachte: dass R^2 wird nach Bereinigung von Cross-Section und
    # Period Effects ausgewiesen
eq_table14_2_FE_two <- plm(lwage ~ expersq + married + union,
                          data = wagepan_pan, model = "within", effect = "twoways")
summary(eq_table14_2_FE_two)

# Hausman-Test
# Beachte: die Teststatistik hängt davon ab, auf welche Weise Cross-Section
# und Period Effects bei der Berechnung berücksichtigt werden. Wichtig ist,
# dass dies in beiden Modellen auf die gleiche Weise passiert.

phtest(eq_table14_2_RE, eq_table14_2_FE)
    # berechnet zwar die gleiche Teststatistik wie EViews for Table 14.2, aber
    # verwendet 10 Freiheitsgraden (also testet auch die Äquivalenz der Zeitdummies)
    # im Gegensatz zu Eviews, was 3 Freiheitsgrade (also nur für die "normalen"
    # Variablen) verwendet. Dadurch unterscheidet sich der p-Wert.
phtest(plm(lwage ~ educ + black + hisp + exper +
               expersq + married + union + year,
           data = wagepan_pan, model = "random", effect = "individual"),
       plm(lwage ~ expersq + married + union + year,
           data = wagepan_pan, model = "within", effect = "individual")
       )
```

```
    # Eine andere Teststatistik ergibt sich, wenn sowohl Cross-Section als auch
    # Period Effects als Random modelliert werden. Dann erhält man
phtest(lwage ~ educ + black + hisp + exper +
               expersq + married + union,
       data=wagepan_pan, model=c("within","random"),
       method="chisq", effect="twoways")
    # bzw. äquivalent dazu
phtest(plm(lwage ~ educ + black + hisp + exper +
               expersq + married + union,
           data = wagepan_pan, model = "random", effect = "twoways"),
       plm(lwage ~ expersq + married + union,
           data = wagepan_pan, model = "within", effect = "twoways")
       )


################################################################################
# IV-Schätzung: Inflation and Openness (Example 16.6 in Wooldridge (2016))
# Datensatz: openness

openness_ols <- lm(open ~ log(land), data=openness)
summary(openness_ols)

inf_ols <- lm(inf ~open + log(pcinc), data=openness)
summary(inf_ols)

inf_iv <- ivreg(inf ~ 1+ open + log(pcinc)| log(land) + log(pcinc), data=openness)
summary(inf_iv)

################################################################################
# Probit- und Logit-Schätzung, ab Folie 205

# Simulation des latenten Modells
```

```
    # Folie 208
set.seed(12)
n <- 30
beta_0 <- 1
beta_1 <- 2

u <- rnorm(n,0,2)
x <- rnorm(n,0,1.5)

y_star <- beta_0 + beta_1 * x + u   # unbeobachtbar
y <- as.numeric(y_star > 0)

if (print_pdf) pdf("graph_r_binary_1.pdf", 6, 6)
plot(x,y, ylim = c(-0.1, 1.1))
if (print_pdf) dev.off()

    # Folie 211
if (print_pdf) pdf("graph_r_binary_2.pdf", 6, 6)
lin_prob <- lm(y ~  x)
plot(x,y, ylim = range(fitted(lin_prob)))
points(x,fitted(lin_prob), col = "red")
legend("bottomright", c("y", "linear"),
       bty = "n", lty = 1, col = c("blue", "red"))
if (print_pdf) dev.off()

# ##################################################################################
# Darstellung Dichte und Verteilungsfunktionen für Probit und Logitmodell,
# Folie 214

if (print_pdf) pdf("graph_r_logitprobit.pdf", 12, 6)
par(mfrow = c(1,2))
```

```
curve(dnorm(x), xlim = c(-4, 4), col = "blue", xlab = "", ylab = "", main = "Dichtefunktionen")
curve(dlogis(x), xlim = c(-4, 4), col = "red", add = T)
grid()
legend("topleft", c(expression(phi(x)), expression(lambda(x))),
       col = c("blue", "red"), lty = 1, bty = "n")

curve(pnorm(x), xlim = c(-4, 4), col = "blue", xlab = "", ylab = "", main = "Verteilungsfunktionen")
curve(plogis(x), xlim = c(-4, 4), col = "red", add = T)
grid()
legend("topleft", c(expression(Phi(x)), expression(Lambda(x))),
       col = c("blue", "red"), lty = 1, bty = "n")
if (print_pdf) dev.off()

# ################################################################################
# Fortsetzung Simulationsbeispiel, Folie 241

if (print_pdf) pdf("graph_r_binary_3.pdf", 6, 6)
probit_model <- glm(y ~ x, family = binomial(link = "probit"))
f.probit <- function(x){pnorm(probit_model$coefficients[1] + probit_model$coefficients[2]*x)}

logit_model <- glm(y ~ x, family = binomial(link = "logit"))
f.logit <- function(x){plogis(logit_model$coefficients[1] + logit_model$coefficients[2]*x)}

plot(x,y, ylim = c(-0.1, 1.1))
abline(lm(y~x), col = "red")
points(x,fitted(lm(y~x)), col = "red")
curve(f.probit, from = 1.2*min(x), to = 1.2*max(x), add = T, col = "blue")
points(x, f.probit(x), col = "blue")
curve(f.logit,  from = 1.2*min(x), to = 1.2*max(x), add = T, col = "green")
points(x, f.logit(x), col = "green")
legend("bottomright", c("y","linear", "probit", "logit"),
       bty = "n", lty = 1, col = c("blue","red", "blue", "green"))
```

```
if (print_pdf) dev.off()


#################################################################################
# Probit- und Logitschätzung: Arbeitsmarkteilnahme von Frauen (Example 17.1 in
# Wooldridge (2016)), Folien 244 folgende
# Datensatz: mroz.txt

# Schätzung eines linearen Wahrscheinlichkeitsmodells
linear_model <- lm(inlf ~ nwifeinc + educ + exper + expersq + age + kidslt6 + kidsge6, data = mroz)
summary(linear_model)

# Schätzung eines Probitmodells, Folie 245
probit_model <- glm(inlf ~ nwifeinc + educ + exper + expersq + age + kidslt6 + kidsge6,
                    family = binomial(link = "probit"), data = mroz)
summary(probit_model)
logLik(probit_model) # computes log likelihood for estimated model
    # nur mit Konstante
probit_model_kon <- glm(inlf ~ 1, family=binomial(link = "probit"), data=mroz)
    # Maße
AIC(probit_model) / nobs(probit_model) # computes AIC a la EViews
BIC(probit_model) / nobs(probit_model) # computes BIC a la EViews
AIC(probit_model, k=(log(nobs(probit_model))) ) / nobs(probit_model) # also BIC a la EViews
AIC(probit_model, k=(2*log(log(nobs(probit_model)))) ) / nobs(probit_model) # HQ a la EViews

1 - logLik(probit_model)/logLik(probit_model_kon) # McFadden Pseude-R^2
(LR_statistic_probit <- 2*(logLik(probit_model) - logLik(probit_model_kon))) # LR-Statistik
1 - pchisq(LR_statistic_probit, df=probit_model$df.null-probit_model$df.residual) # p-Wert

# Schätzung eines Logitmodells, Folie 246
logit_model <- glm(inlf ~ nwifeinc + educ + exper + expersq + age + kidslt6 + kidsge6,
                   family = binomial(link = "logit"), data = mroz)
```

```
summary(logit_model)
logLik(logit_model) # computes log likelihood for estimated model
    # nur mit Konstante
logit_model_kon <- glm(inlf ~ 1, family=binomial(link = "logit"), data=mroz)

AIC(logit_model) / nobs(logit_model) # computes AIC a la EViews
BIC(logit_model) / nobs(logit_model) # computes BIC a la EViews
AIC(logit_model, k=(2*log(log(nobs(logit_model)))) ) / nobs(logit_model) # HQ a la EViews
1 - logLik(logit_model)/logLik(logit_model_kon) # McFadden Pseude-R^2
(LR_statistic_logit <- 2*(logLik(logit_model) - logLik(logit_model_kon))) # LR-Statistik
1 - pchisq(LR_statistic_logit, df=logit_model$df.null-logit_model$df.residual) # p-Wert

# Berechnung von Evaluationskriterien
    # Tabelle mit korrekten und falschen Prognosen für lineare W'Modell
table(true = mroz$inlf, pred = round(fitted(linear_model)))
mean(mroz$inlf == round(fitted(linear_model))) # Anteil korrekter Prognosen

# Goodness-of-fit für Probit-Modell, Folie 247
y_tilde <- (fitted(probit_model) > 0.5)
mean(y_tilde == mroz$inlf)

# Tabelle mit korrekten und falschen Prognosen für Probitmodell für y = 1 und y = 0
# und entsrechenden Anteilen, Folie 248
table(true = mroz$inlf, pred = round(fitted(probit_model)))
table(true = mroz$inlf, pred = round(fitted(probit_model)))[1,1] / sum(mroz$inlf == 0)
table(true = mroz$inlf, pred = round(fitted(probit_model)))[2,2] / sum(mroz$inlf == 1)

    # Tabelle mit korrekten und falschen Prognosen für Logitmodell
table(true = mroz$inlf, pred = round(fitted(logit_model)))
mean(mroz$inlf == round(fitted(logit_model)))

# durchschnittliche partielle Effekte (APE), Folie 248
```

```
coef(linear_model)
(ape_probit <- mean(dnorm(predict(probit_model, type = "link"))) * coef(probit_model))
(ape_logit <- mean(dlogis(predict(logit_model, type = "link"))) * coef(logit_model))

# marginale Effekte auf Basis der Durchschnitte der Regressoren (PEA)
coef(linear_model)
dnorm(colMeans(model.matrix(probit_model)%*%coef(probit_model))) * coef(probit_model)
dlogis(colMeans(model.matrix(logit_model)%*%coef(logit_model))) * coef(logit_model)

################################################################################
# Tobit-Schätzung mit gestutzten Daten, ab Folie 272
# Datensatz: wagepan

# OLS mit gestutzten Daten, Folie 272
ols_subset_gestutzt <- lm(lwage ~ educ,
                     subset = (year == 1980 & black == 0 & hisp == 0 & married == 0
                               & exper == 3 & lwage<=log(8) ),
                     data = wagepan)
summary(ols_subset_gestutzt)
SelectCritEviews(ols_subset_gestutzt)

# OLS mit allen Daten, Folie 273
ols_subset_all <- lm(lwage ~ educ,
                subset = (year == 1980 & black == 0 & hisp == 0 & married == 0
                          & exper == 3),
                data = wagepan)
summary(ols_subset_all)
SelectCritEviews(ols_subset_all)

# Scatterplot mit Regressiongeraden der verschiedenen Modelle, Folie 275
    # Bestimme Subset für OLS mit allen Daten
subset_wagepan <- (wagepan$year == 1980 & wagepan$black == 0 &
```

```
                       wagepan$hisp == 0 & wagepan$married == 0 &
                       wagepan$exper == 3 )
if (print_pdf) pdf("graph_scatter_Daten_gestutzt.pdf", 6, 6)
    # Scatterplot mit allen Daten
plot(x = subset(wagepan$educ, subset = subset_wagepan),
     y = subset(wagepan$lwage, subset = subset_wagepan),
     ylab = "LWAGE", xlab = "EDUC")
    # Regressionsgerade von OLS mit allen Daten
points(x = subset(wagepan$educ, subset = subset_wagepan),
      y = fitted(ols_subset_all), pch = 19, type = "p", col = "red")
    # Regressionsgerade von OLS mit gestutzten Daten
points(x = subset(wagepan$educ, subset = subset_wagepan & wagepan$lwage <= log(8)),
       y = fitted(ols_subset_gestutzt), pch = 17, type = "p", col = "green")
legend("bottomright", c("LWAGE", "LWAGEF", "LWAGE_TRUNC"),
       bty = "n", lty = 1, col = c("black","red", "blue"))
if (print_pdf) dev.off()


# # Tobit mit gestutzten normalverteilten Daten, Folie 275
# # mit R package truncreg
# tobit_gestutzt <- truncreg(lwage ~ educ, data = wagepan,
#                           subset = (year == 1980 & black == 0 & hisp == 0 & married == 0
#                                     & exper == 3 & lwage<=log(8) ),
#                           point = log(8), direction = "right", scaled = TRUE)
# summary(tobit_gestutzt)
# (AIC <- (-2*logLik(tobit_gestutzt) + 2*length(coefficients(tobit_gestutzt)))/nobs(tobit_gestutzt))
# (BIC <- (-2*logLik(tobit_gestutzt) + log(nobs(tobit_gestutzt))*length(coefficients(tobit_gestutzt)))/
#         nobs(tobit_gestutzt))

# Tobit-Schätzung mit gestutzten normalverteilten Daten, Folie 275
# mit R package crch
tobit_gestutzt_norm <- trch(lwage ~ educ,
```

```
                       data = wagepan,
                       subset = (year == 1980 & black == 0 & hisp == 0 & married == 0
                       & exper == 3 & lwage<=log(8) ),
                       link.scale = "identity",
                       dist = "gaussian",
                       right = log(8),
                       type = "ml"
                       )
summary(tobit_gestutzt_norm)
    # Berechne Informationskriterien mit ganz oben definierter Funktion
IC.loglik(tobit_gestutzt_norm$loglik, length(tobit_gestutzt_norm$coefficients),
        tobit_gestutzt_norm$n)

# Tobit-Schätzung mit gestutzten logistisch verteilten Daten, Folie 276
# mit R package crch
tobit_gestutzt_logit <- trch(lwage ~ educ,
                            data = wagepan,
                            subset = (year == 1980 & black == 0 & hisp == 0 & married == 0
                                      & exper == 3 & lwage<=log(8) ),
                            link.scale = "identity",
                            dist = "logistic",
                            right = log(8),
                            type = "ml",
                            x = TRUE,
                            y = TRUE
)
summary(tobit_gestutzt_logit)

IC.loglik(tobit_gestutzt_logit$loglik, length(tobit_gestutzt_logit$coefficients),
          tobit_gestutzt_logit$n)
```

```
# zum Plotten der Regressionsgerade aus Tobitmodell mit logistisch verteilten Daten
# plot(x = tobit_gestutzt_logit$x$location[,"educ"], y= tobit_gestutzt_logit$fitted.values$location, col = "green")

################################################################################
# Heckit-Schätzung: Lohnangebotsgleichung von verheirateten Frauen
# Example 17.5 in Wooldridge (2016)
# Datensatz: mroz.txt

# Schätze OLS (Spalte 2 in Table 17.5 in Wooldridge (2016))
eq_table17_5_ols <- lm( lwage ~ educ + exper + expersq, data = mroz)
summary(eq_table17_5_ols)

# Berechne Heckit-Schätzer (Spalte 3 in Table 17.5)
    # Berechne Probit-Schätzung für lambda
probit_model <- glm(inlf ~ nwifeinc + educ + exper + expersq + age + kidslt6 + kidsge6,
                    family = binomial(link = "probit"), data = mroz, x = TRUE)
    # Berechne Index des Probitmodells
index_probit <- probit_model$x %*% coefficients(probit_model)
    # Berechne Lambda
lambda <- dnorm(index_probit) / pnorm(index_probit)
    # Heckit-Schätzung
eq_table17_5_heckit <- lm( lwage ~ educ + exper + expersq + lambda, data = mroz)
summary(eq_table17_5_heckit)

################################################################################
#                 Ende
################################################################################
```

Listing 8.1: ../../R_code/WOE_ws24_Emp_Beispiele_final.R

8.4 Beweis zum Testen der 'overidentifying restrictions' auf Folie 176

$$\mathbf{y}_1 = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}_1$$

IV-Schätzer, vgl. Folie 161:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{2SLS} &= (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\mathbf{y}_1 \\ &= \left(\mathbf{X}'\mathbf{P}_Z\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_Z\mathbf{y}_1,\end{aligned}$$

so dass

$$\begin{aligned}\hat{\mathbf{u}}_{1,2SLS} &= \mathbf{y}_1 - \mathbf{X}\hat{\boldsymbol{\beta}}_{2SLS} \\ &= \left(\mathbf{I} - \mathbf{X}\left(\mathbf{X}'\mathbf{P}_Z\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_Z\right)\mathbf{y}_1 \qquad (8.1)\end{aligned}$$

Testen, ob $\mathbf{u}_1$ unkorreliert mit $\mathbf{Z}$

Test mit $\hat{\mathbf{X}}$ ist nicht möglich. Bei Vorliegen von Korrelation muss für

$$\hat{\mathbf{u}}_{1,2SLS} = \hat{\mathbf{X}}\mathbf{b} + \eta$$

$\mathbf{b} \neq 0$ gelten. Aber für den KQ-Schätzer gilt

$$\hat{\mathbf{b}} = \left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)^{-1} \underbrace{\hat{\mathbf{X}}'\hat{\mathbf{u}}_{1,2SLS}}_{=0} = 0,$$

da für die Normalenlgleichungen für die zweite Stufe des IV-Schätzers gilt

$$\begin{aligned}
\left(\hat{\mathbf{X}}'\hat{\mathbf{X}}\right)\hat{\boldsymbol{\beta}}_{2SLS} &= \hat{\mathbf{X}}'\mathbf{y}_1 \\
&= \hat{\mathbf{X}}'\left(\mathbf{X}\hat{\boldsymbol{\beta}}_{2SLS} + \hat{\mathbf{u}}_{1,2SLS}\right) \\
0 &= \hat{\mathbf{X}}'\hat{\mathbf{u}}_{1,2SLS}.
\end{aligned}$$

Schätzen mit $\mathbf{Z}$ bei m Instrumenten

$$\hat{\mathbf{u}}_{1,2SLS} = \mathbf{Z}d + Fehler$$

Falls $m = k$, gilt wegen (8.1)

$$\begin{aligned}
\hat{d} &= \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\hat{\mathbf{u}}_{1,2SLS} \\
&= \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\left(\mathbf{I} - \mathbf{X}\underbrace{\left(\mathbf{X}'\mathbf{P}_Z\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{P}_Z}_{=\left(\mathbf{Z}'\mathbf{X}\right)^{-1}\mathbf{Z}'}\right)\mathbf{y}_1 \\
&= \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y}_1 \\
&- \left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\underbrace{\mathbf{Z}'\mathbf{X}\left(\mathbf{Z}'\mathbf{X}\right)^{-1}}_{=\mathbf{I}}\mathbf{Z}'\mathbf{y}_1 \\
&= 0.
\end{aligned}$$

wobei die erste Unterklammer für $m = k$ wegen Folie 171 folgt.

Literaturverzeichnis

Angrist, J., und J. Pischke (2009), *Mostly harmless econometrics. An Empiricist's Companion*, Princeton University Press. 15, 24, 32, 52, 66

Angrist, J., und J. Pischke (2014), *Mastering 'Metrics. The Path from Cause to Effect*, Princeton University Press.

Cameron, A., und P. Trivedi (2005), *Microeconometrics*, Cambridge University Press, doi:10.2277/0521848059. 249, 250

Davidson, R., und J. G. MacKinnon (2004), *Econometric Theory and Methods*, Oxford University Press, Oxford. 231, 265

Guggenberger, P. (2009), "The Impact of a Hausman Pretest on the Asymptotic Size of a Hypothesis Test," *Econometric Theory*, 1–14. 175

Hausman, J. (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271. 174

Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models," *Annals of Economic and Social Measurement*, 5, 475–492. 289

Kleiber, C., und A. Zeileis (2008), *Applied Econometrics with R*, Springer, doi:10.1007/978-0-387-77318-6. VII

McFadden, D. (1974), "Conditional Logit Analysis of Qualtitative Choice Behavior," in: P. Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, New York, pp. 105–142. 244

Wooldridge, J. M. (2002), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press. 101, 231, 249, 250, 278, 290

Wooldridge, J. M. (2006), *Introductory Econometrics. A Modern Approach*, 3rd edn., Thomson South-Western, Mason. 9, 12, 42, 93, 97, 115, 116, 124, 150, 166, 167, 190, 199, 244

Wooldridge, J. M. (2009), *Introductory Econometrics. A Modern Approach*, 4th edn., Thomson South-Western, Mason. 9, 12, 13, 42, 51, 93, 97, 114, 115, 116, 124, 150, 166, 167, 190, 199, 244, VII

Wooldridge, J. M. (2010), *Econometric Analysis of Cross Section and Panel Data*, The MIT Press. 101, 240, 249, 250, 278, 290

Wooldridge, J. M. (2015), *Introductory Econometrics. A Modern Approach*, 6th edn., Thomson South-Western, Mason. 87