

**Einführung in die Ökonometrie für Doktoranden und Paneldatenverfahren**  
WS 2005/06 und SS 2006

Dozent: Dieter Urban, Ph.D.

---

**Teil I: Grundlagen der Ökonometrie**

A Einführung

- 1) Wissenschaftliche Kontribution
- 2) Aufbau einer empirischen Arbeit

B Ökonometrische Grundlagen

- 1) OLS-Schätzer
  - a) Annahmen
  - b) Herleitung
  - c) Eigenschaften in der kleinen Stichprobe (BLUE)
  - d) Partitioned Regression
    - d1) Interpretation des Partial-Effekts
    - d2) Omitted-Variable Problem
    - d3) Irrelevante Variablen
  - e) Eigenschaften in der großen Stichprobe (Konsistenz und asymptotische Normalverteilung des OLS-Schätzers)
- 2) GLS-Schätzer und Heteroskedastizität
  - a) GLS-Schätzer
  - b) Heteroskedastizität
- 3) IV-Schätzer, Hausman-Test und Weak-Instrument-Problem
  - a) IV-Schätzer
  - b) Hausman-Test
  - c) Weak-Instrument-Problem
    - c1) Grosse Varianz des Standardfehlers
    - c2) Rest-Korrelation der Instrumente mit dem Fehlerterm der 2. Schätzstufe
    - c3) Verzerrtheit des IV-Schätzers in der kleinen Stichprobe
    - c4) Test auf Vorliegen eines Weak-Instrument-Problems
- 4) GMM-Schätzer
- 5) Selection Bias und Heckman-Schätzer

- 6) Empirische Analysestrategie bei Cross-Section Daten mit Beispielen (Wie geht man mit Multikollinearität, Heteroskedastizität, Datenfehler, Strukturbruch, funktionale Form, Ausreißer, etc. um?)
  - a) Einige Faustregeln
  - b) Vorgehensweise

## **Teil II: Paneldatenverfahren**

### II.1 Einführung

- 1) Vorteile und Anwendungen von Paneldatenverfahren

### II.2 Statische Paneldatenverfahren

- 2) Fixed-Effekt (LSDV-) Schätzer
  - a) Schätzer
  - b) Kovarianzmatrix für Regressionskoeffizienten
  - c) „Pooled Regression“, „Between-“ und „Within-“ Schätzungen
  - d) Test auf Fixed-Effekt- vs. „Pooled Model“
  - e) Fixed-Effekt Falle
  - g) Zeitdifferenzschätzung
- 3) Random-Effekt-Schätzer
  - a) Schätzer
  - b) Test für „Random Effekt“ gegen „pooled OLS“
  - c) Heteroskedastizität und Random-Effekt-Modell
  - d) Test Random-Effekt- vs. Fixed-Effekt-Modell (Hausman-Test)
- 4) Cluster-Regression
- 5) „Pooling-“ Test
- 6) Test auf Autokorrelation des Störterms
- 7) Schätzung mit autoregressivem Fehlerterm
- 8) Hausman-Taylor (1981) Schätzer
- 9) „Hildreth-Houck Random Coefficient Model“

### II.3 Dynamische Paneldatenverfahren

- 1) Stationäre dynamische Paneldatenverfahren

#### II.3.1.1 Grundproblem

#### II.3.1.2 Arellano-Bond-Schätzer

##### II.3.1.2.1 Grundidee

##### II.3.1.2.2 Inkonsistenz bei Autokorrelation des Störterms

##### II.3.1.2.3 GMM-Schätzer bei Homoskedastizität

- II.3.1.2.4 GMM-Schätzer bei Heteroskedastizität
- II.3.1.2.5 Arellano-Bond-Schätzer mit exogenen erklärenden Variablen
- II.3.1.2.6 Tests
- II.3.1.2.7 Arellano-Bond-Schätzer in STATA

### II.3.1.3 Blundell-Bond-Schätzer

- II.3.1.3.1 Ausgangsproblem
- II.3.1.3.2 Lösungsvorschlag nach Blundell und Bond (1998)
- II.3.1.3.3 Blundell-Bond-Schätzer
- II.3.1.3.4 Der Blundell-Bond-Schätzer in STATA

## 2) Nicht-stationäre dynamische Paneldatenverfahren

### II.3.2.1 Einführung in die Nicht-stationäre Zeitreihenanalyse

- II.3.2.1.1 Unit Roots und Spurious Regression
- II.3.2.1.2 Kointegration

### II.3.2.2 Nicht-Stationäre Paneldatenanalyse

- II.3.2.2.1 Spurious Regression bei Paneldaten
- II.3.2.2.2 Panel Unit Root Tests
  - i. Test basierend auf OLS
  - ii. Test nach Breitung und Meyer (1994)
  - iii. Test nach Harris und Tzavalis (1999)
  - iv. Test nach Im, Pearsan und Shin (1997)
  - v. Test nach Maddala und Wu (1999)

## Literatur

### Allgemeine Ökonometrielehrbücher für Fortgeschrittene

Cameron, A. Colin, und Pravin K. Trivedi, 2005, Microeconometrics – Methods and Applications, New York, Cambridge University Press.

Greene, William H., 2001, Econometric Analysis, 4. Aufl., New York. \*\*\*

Greene, William H., 2003, Econometric Analysis, 5. Aufl., New York.

Davidson, James, 2000, Econometric Theory.

Wooldridge, Jeffrey, Econometric Analysis of Cross-Section and Panel Data, Cambridge MA: MIT Press, 2002. \*\*\*

Wooldridge, Jeffrey, Solutions Manual and Supplementary Materials for Econometric Analysis of Cross-Section and Panel Data, Cambridge MA: MIT Press, 2003.

## Paneldatenbücher für Fortgeschrittene

Arellano, Manuel, Panel Data Econometrics, Oxford University Press: Advanced Texts in Econometrics, 2003.

Arellano, Manuel, und Bo Honoré, Panel data models: some recent developments, 2000, mimeo. (Erscheint demnächst im Handbook of Econometrics)

Baltagi, Badi H., 2001, Econometric Analysis of Panel Data, 2. Aufl., Oxford.\*\*\*

Hsiao, Cheng, 2003, Analysis of Panel Data, 2. Aufl., Cambridge.

## **Artikel zum Problem schwacher Instrumente**

Bound, John, David A. Jaeger, und Regina M. Baker, 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak, Journal of the American Statistical Association, Bd. 90, Heft 430, S. 443-450.

Nelson, Charles R., und Richard Startz, 1990, The distribution of the instrumental variables estimator and its t-ratio when the instrument is a poor one, The Journal of Business, Bd. 63, Heft 1, S. S125-S140.

Staiger, Douglas, und James H. Stock, 1997, Instrumental variables regression with weak instruments, Econometrica, Bd. 65, Heft 3, S. 557-586.

## **Artikel über Stationäre Dynamische Paneldatenverfahren**

Arellano, Manuel, 1987, Computing robust standard errors for within-groups estimators, Oxford Bulletin of Economics and Statistics, Bd. 49, Heft 4, S. 431-434.

Arellano, Manuel, und Stephen R. Bond, 1991, Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, Review of Economic Studies, Vol. 58, p. 277-297.

Arellano, Manuel, und Stephen R. Bond, 1998, Dynamic panel data estimation using DPD98 for GAUSS: A guide for users, mimeo. (<http://www.ifs.org.uk/econometindex.shtml>)

Blundell, Richard, und Stephen Bond, 1998, Initial Conditions and moment restrictions in dynamic panel data models, Journal of Econometrics, Jg. 87, S. 115-143.

Bond, Stephen R., 2002, Dynamic panel data models: A guide to micro data methods and practice, CEMMAP Working Paper CWP09/02. (<http://cemmap.ifs.org.uk/docs/cwp0902.pdf>)

Bond, Stephen R., Anke Hoeffler und Jonathan Temple, GMM estimation of empirical growth models, CEPR Discussion Papers No. 3048.

Dahlberg, M., E. Johansson und P. Tovmo, 2002, Power properties of the Sargan test in the presence of measurement errors in dynamic panels, mimeograph.

Kiviet, Jan F., 1995, On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, *Journal of Econometrics*, Bd. 68, S. 53-78.

STATA-Handbuch, Version 7.0, Bd. 4, S. 291-308.

Trognon, Alain, 1978, Miscellaneous asymptotic properties of ordinary least squares and maximum likelihood estimators in dynamic error components models, *Annales de l'INSEE*, Bd. 30/31, S. 631-657.

Windmeijer, Frank, ExpEnd, A Gauss programme for non-linear GMM estimation of exponential models with endogenous regressors for cross section and panel data. CEMMAP Working Paper CWP 14/02, (<http://cemmap.ifs.org.uk/docs/cwp1402.pdf>)

### **Artikel über Nicht-stationäre dynamische Paneldatenverfahren**

Baltagi, Badi H., und Chihwa Kao, 2000, Nonstationary panels, cointegration in panels and dynamic panels: A survey, Center for Policy Research Working Paper No. 16, Syracuse University. (<http://www-cpr.maxwell.syr.edu/cprwps/wpslst.htm>)

Binder, Michael, Cheng Hsiao und M. Hshem Pesaran, 2003, Estimation and Inference in Short Panel Vector Autoregressions with Unit Roots and Cointegration, mimeo.

Bond, Stephen, Celine Nauges, und Frank Windmeijer, 2002, Unit roots and identification in autoregressive panel data models: A comparison of alternative tests, January, mimeo.

Phillips, Peter C.B., und Hyungsik R. Moon, 1999, Linear regression limit theory for nonstationary panel data, *Econometrica*, Vol. 67, No. 5, p. 1057-111.

### **Anwendungsorientierte Bücher**

Kennedy, Peter, 1992, *A Guide to Econometrics*, 3. Aufl., Oxford.

### **Sonstiges**

Arellano, Manuel, und Jinyong Hahn, 2006, Understanding bias in nonlinear panel models: Some recent developments, mimeograph. (Incidental parameter problem: Approximationen für Fixed-Effekt Schätzer bei nicht-linearen Panelschätzern (probit, tobit, etc.))

### **Web-Links zu ähnlichen Veranstaltungen mit Literaturliste**

Web-page von Banerjee am Europäischen Hochschulinstitut in Florenz:  
<http://www.iue.it/Personal/Banerjee/Welcome.html#Teaching>

## II Ökonometrische Grundlagen

### 1.) OLS-Schätzer

#### 1a) Annahmen

Annahme 1:  $y = X\beta + \varepsilon$

$$\text{wobei } y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nk} \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \text{ eine}$$

Datenstichprobe aus einer Grundgesamtheit darstellt, die den Kriterien der Zufallsstichprobe genügt (letzteres ist wichtig für die Verallgemeinerung der Ergebnisse von der Stichprobe auf die Grundgesamtheit  $\implies$  empirisches Gesetz);  $n$  ist die Größe der Stichproben;  $\beta$  ist der „wahre“ Zusammenhang in der Grundgesamtheit, der aber nicht bekannt ist und „bestmöglich“ geschätzt werden muss;  $k$  ist die Anzahl der unabhängigen Variablen  $x$  oft mit der Konstanten  $x_{i1} = \dots = x_{in} = 1$ ;  $i$  ist der Index der Beobachtungen und

$x_i = [x_{i1} \dots x_{ik}]'$   $\varepsilon$  ist ein stochastischer Störterm (siehe unten).

Bemerkung 1: Annahme 1 impliziert

- dass bekannt ist, welche X- Variablen die y-Variablen erklären (kein Suchen nach den „richtigen“ X-Variablen!);
- dass der Zusammenhang linear ist;
- dass eine Kausalität von X auf y besteht und nicht umgekehrt!
- dass keine Heterogenität in dem Zusammenhang zwischen X und y besteht (z.Bsp. kein Strukturbruch, durch den für eine Teilstichprobe ein Zusammenhang gilt und für eine andere Teilstichprobe ein anderer!).

Bemerkung 2: Implikationen a) und c) können nur durch die Wirtschaftstheorie begründet werden. Damit erfordert die Anwendung des OLS-Schätzers stets eine theoretische Fundierung.

Bemerkung 3: Üblicherweise werden alle Variablen in Logarithmen transformiert (außer Dummys- Variablen mit den Werten 0 oder 1 und der Konstanten)

$$\ln(y_i) = \sum_{l=2}^k \beta_l \ln(x_{il}) + \beta_1 + \varepsilon_i$$

Gründe: a)  $\beta_k = \frac{\partial \ln y_1}{\partial \ln x_{ik}} = \frac{\partial y_1 / \partial x_{ik}}{y_1 / x_{ik}}$  ist eine Elastizität

b) Ausreißer in den Daten machen sich weniger bemerkbar und der „Fit“ wird meistens besser  $\implies$  log-lineare funktionale Form ist vorzuziehen.

Annahme 2: Rang  $[X] = k$  (impliziert  $n \geq k$ )  
 $\Rightarrow$  X Variablen sind untereinander keine Linearkombinationen  
 $\Rightarrow$  Sichert Invertierbarkeit von  $X'X$

Bemerkung 1: Annahme 2 wird auch als Ausschluss perfekter Multikollinearität bezeichnet.

Gegenbeispiel: Dummy-Variablen Falle:  
*i* sei ein Jahresindex mit  $i=1960 \dots 2000$

$$x_{i1} = 1; \quad x_{i2} = \begin{cases} 1 & \text{wenn } i \geq 1990 \\ 0 & \text{sonst} \end{cases} \quad x_{i3} = \begin{cases} 1 & \text{wenn } i < 1990 \\ 0 & \text{sonst} \end{cases}$$

$x_1 = x_2 + x_3 \Rightarrow$  perfekte Multikollinearität  $\Rightarrow$  ein  $x$  weglassen:  
 $y = 0,5x_2 + 0,7x_3 = 0,7x_1 - 0,2x_2 = 0,5x_1 + 0,2x_3$

$\downarrow$  Konstante für Beobachtungen  $i < 1990$   
 $\downarrow$  Konstante für Beobachtungen  $i \geq 1990$

Annahme 3:  $E[\varepsilon_i | X] = 0 \quad \forall i=1 \dots n$   
 $E[\cdot | \cdot]$  ist bedingter Erwartungswert  
 Annahme 3 impliziert auch  $E[\varepsilon_i] = 0$   
 Annahme 1 und 3 implizieren  $E[y | X] = X\beta$

Annahme 4: Homoskedastizität und keine Autokorrelation  
 $E[\varepsilon\varepsilon' | X] = I_n \tau^2 < \infty$   
 $I_n$  ist  $n \times n$ -Einheitsmatrix  
 Dies impliziert auch  $E[\varepsilon\varepsilon'] = I_n \tau^2$   
 $\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \quad \forall i \neq j \iff$  keine Autokorrelation  
 $\text{Var}(\varepsilon_i) = \text{Var}(\varepsilon_j) = \tau^2 \quad \forall i, j \iff$  Homoskedastizität

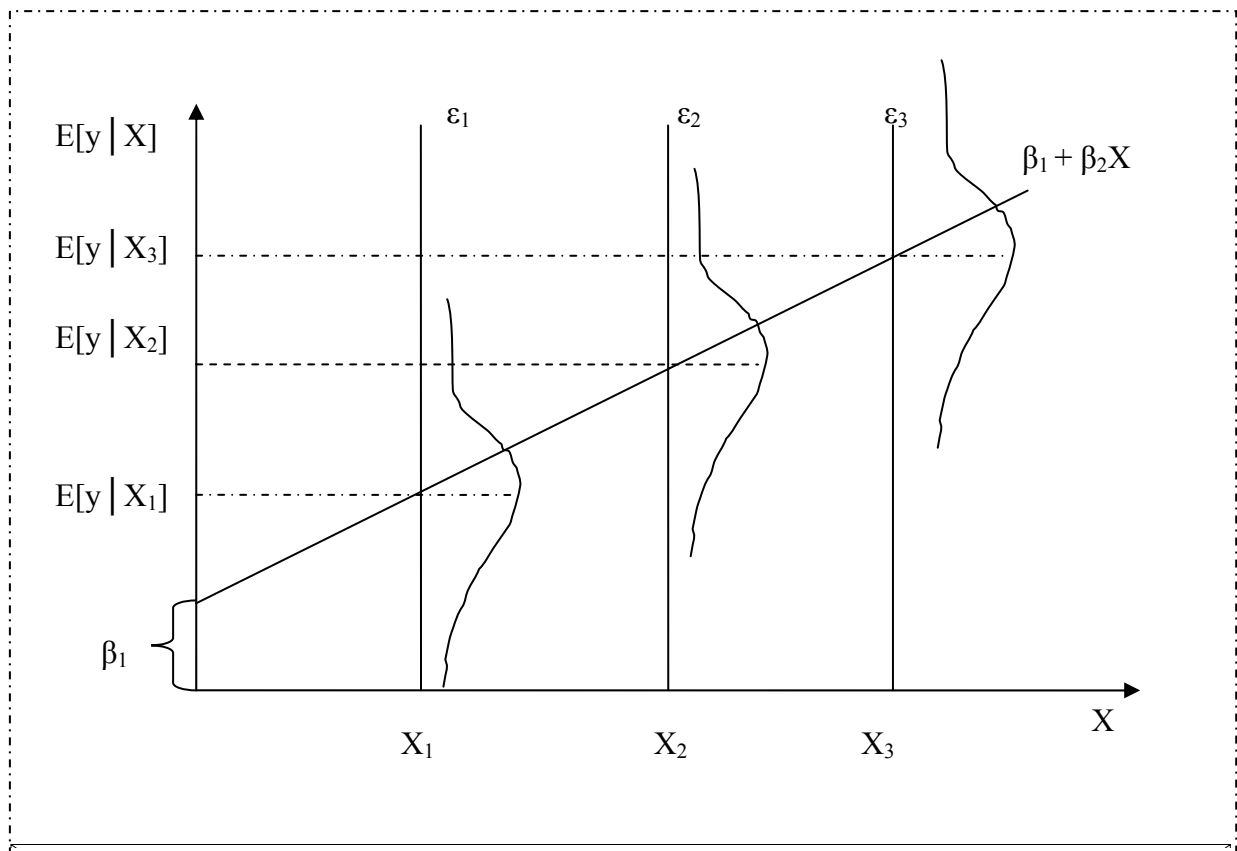
Annahme 5: a)  $[x_i, \varepsilon_i], i=1 \dots n$  ist eine i.i.d. Folge von Zufallsvariablen, die nach einer Zufallsauswahl aus einer Grundgesamtheit „gezogen“ wurden.  
 i.i.d. identically independently distributed  
 (Hier wird angenommen, daß die  $x_i$  Variablen über alle Beobachtungen hinweg identisch verteilt sind. Diese Annahme kann bei anderer Beweisführung auch fallen gelassen werden.)

Wegen  $E[\varepsilon_i | X] = 0 \quad \forall i=1 \dots n$  gilt dann auch  $\text{Cov}[x_i, \varepsilon_i] = 0$

b) X folgt einer multivariaten Verteilung mit  $E[X] = \mu_x < \infty$   
 $\text{Var}[X] = \Sigma_{xx} < \infty$  und  $E[x_i x_j x_k x_l] = \phi_{ijkl} < \infty$ .  
 (endliche Varianz und endliches 4. Moment von X)

Annahme 6: a) kleine Stichprobe:  $\varepsilon | X \sim N(0, \tau^2)$   
 oder b) große Stichprobe:  $n \rightarrow \infty$   
 Faustregel  $n > 30$  und  
 $E[\varepsilon_i^4] = \Phi_\varepsilon < \infty$

Zusammenfassung für  $k=2$  :  $y=\beta_1 + \beta_2 X+\varepsilon$



### 1 b) Herleitung des OLS-Schätzers

Problem:  $\beta$  ist nicht bekannt und muss geschätzt werden.

$\Rightarrow$  Stichprobenmodell  $y= X b_{ols} + e$

Definition1:  $b_{ols} = \underset{\text{argmin}}{e} (b_{ols})' e(b_{ols}) = \underset{\text{argmin}}{\sum_{i=1}^n e_i^2}$

$$= \underset{\text{argmin}}{(y - X b_{ols})' (y - X b_{ols})}$$

$$= \underset{\text{argmin}}{[y'y - (X b_{ols})' y - y' X b_{ols} + (X b_{ols})' X b_{ols}]}$$

$$= \underset{\text{argmin}}{[y'y - b'_{ols} X'y - y' X b_{ols} + b'_{ols} X'X b_{ols}]}$$

$$b'_{ols} X'y = \sum_i \sum_e b_e X_{ie} Y_i = (y' X b_{ols})' = y' X b_{ols}$$

$1 \times k$   $k \times n$   $n \times 1$   
 $\underbrace{1 \times 1}_{= \text{Skalar}}$

weil Skalar

$$X'X = \begin{bmatrix} x_{11} & \dots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{ik} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nk} \end{bmatrix} = \begin{bmatrix} n \\ \sum_{i=1}^n x_{il} x_{in} \end{bmatrix}_{l,n=1 \dots k} = \sum_{i=1}^n x_i x_i'$$

$k \times n$   $n \times k$   
 $\underbrace{k \times k}$

$k \times n$

$n \times k$

$k \times k$

$$\underbrace{b' X' X b}_{\substack{1 \times k \quad k \times n \quad n \times k \quad k \times 1 \\ 1 \times 1 \\ \Rightarrow \text{Skalar}}} = \sum_i \sum_l \sum_m b_l b_m x_{il} x_{im}$$

$$b_{ols} = \operatorname{argmin} \left[ \sum_{i=1}^n y_i^2 + \underbrace{\sum_{i=1}^n \sum_{l=1}^k \sum_{m=1}^k b_l b_m x_{il} x_{im}}_{\sum_{i=1}^n \begin{bmatrix} b_1 b_1 x_{i1} x_{i1} + \dots + b_1 b_l x_{i1} + \dots + b_1 b_k x_{ik} x_{i1} \\ \dots \\ + b_l b_1 x_{i1} x_{il} + \dots + b_l b_l x_{il}^2 + \dots + b_l b_k x_{ik} x_{il} \\ \dots \\ + b_k b_1 x_{i1} x_{ik} + \dots + b_k b_l x_{il} + \dots + b_k b_k x_{ik} x_{ik} \end{bmatrix}} - 2 \sum_{i=1}^n \sum_{l=1}^k b_l x_{il} y_i \right]$$

Bedingungen erster Ordnung:

$$\frac{\partial e'e}{\partial b_l} = 2 \sum_{i=1}^n \sum_{m=1}^k b_m x_{im} x_{il} - 2 \sum_{i=1}^n x_{il} y_i = 0 \quad \forall l = 1 \dots k$$

bzw.

$$\begin{bmatrix} 2 \sum_i \sum_m b_m x_{i1} x_{im} \\ \vdots \\ 2 \sum_i \sum_m b_m x_{ik} x_{im} \end{bmatrix} + \begin{bmatrix} -2 \sum_i x_{i1} y_i \\ \vdots \\ -2 \sum_i x_{ik} y_i \end{bmatrix} = 0$$

bzw.

$$2X'X b_{ols} - 2X'y = 0$$

$$b_{ols} = (X'X)^{-1} (X'y) \quad \text{wegen Annahme 2 (Rang}(X)=k)$$

### 1c) Eigenschaften des OLS-Schätzers

Ziel der empirischen Wirtschaftsforschung ist es, die „Signifikanz“ eines Zusammenhangs zwischen  $y$  und einer erklärenden Variablen  $x_k$  herzustellen. Also:

$$y = z\gamma + x_k \beta_k + \varepsilon \quad X = [z|x_k]; \quad \beta = \begin{bmatrix} \gamma \\ \beta_k \end{bmatrix}_{1 \times k}$$

wird geschätzt als

$$y = z c_{ols} + x_k b_{bols} + e \quad b_{ols} = [c_{ols} \ b_{kols}]'$$

Wenn  $b_{kols} \neq 0$ , liegt das daran, dass tatsächlich  $\beta_k \neq 0$  oder ist  $|b_{kols} - \beta_k|$  lediglich ein Schätzfehler? Ein Test ist erforderlich!

$$\left. \begin{array}{l} H_o : \beta_k = 0 \\ H_a : \beta_k \neq 0 \end{array} \right\} \begin{array}{l} \rightarrow \text{das wahre } \beta_k = 0 \\ \rightarrow \text{das wahre } \beta_k \neq 0 \end{array}$$

Dazu muss aber die Wahrscheinlichkeitsverteilung von  $b_{kols}$  als Schätzer für  $\beta_k$  bekannt sein. Dazu entwickeln wir Schritt für Schritt die Eigenschaften des OLS-Schätzers zunächst in der kleinen Stichprobe (Annahme 6a:  $\varepsilon | X \sim N(0, \tau^2)$ ); zur Vereinfachung sei  $X$  fixiert.

### i) unverzerrt

$$\begin{aligned} E[b_{ols} - \beta | X] &= E[\underbrace{(X'X)^{-1} X' (X\beta + \varepsilon)}_y - \beta | X] && \text{Annahme 1} \\ &= E[(X'X)^{-1} X' \varepsilon | X] \\ &= (X'X)^{-1} X' \underbrace{E[\varepsilon | X]}_0 = 0 && \text{Annahme 3} \end{aligned}$$

### ii) Varianz

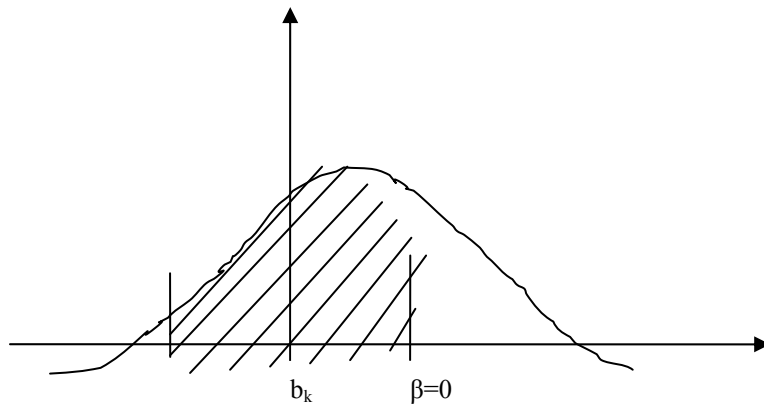
$$\begin{aligned} \text{Var}(b | X) &= E[(b_{ols} - \beta)(b_{ols} - \beta)' | X] \\ &= E\left\{ \left[ (X'X)^{-1} X' \varepsilon \right] \left[ (X'X)^{-1} X' \varepsilon \right]' \middle| X \right\} \\ &= E[(X'X)^{-1} X' \varepsilon \varepsilon' X (X'X)^{-1} | X] \\ &= (X'X)^{-1} X' \underbrace{E[\varepsilon \varepsilon' | X]}_{\tau^2 I_n \text{ Annahme 4}} X (X'X)^{-1} = (X'X)^{-1} \tau^2 \end{aligned}$$

### iii) Normalverteilung von $b_{ols}$

$$b_{ols} - \beta | X = (X'X)^{-1} X' \varepsilon \mid X \sim N(0, (X'X)^{-1} \tau^2)$$

weil  $(X'X)^{-1} X' \varepsilon | X$  aus einer Summe der nach Annahme 6a normalverteilten Zufallsvariablen besteht, die selbst normalverteilt ist.

Testidee:



Wenn ///-Fläche  $> 0,95 \Rightarrow b_{ols} \neq 0$  mit mehr als 95% Wahrscheinlichkeit = „signifikant auf dem 5% Niveau“.

Problem  $\tau^2$  ist nicht bekannt und muss daher geschätzt werden.

iv) Erwartungstreuer Schätzer für  $\tau^2$

$$e = y - Xb = y - X(X'X)^{-1}X'y = [I_n - X(X'X)^{-1}X']y \equiv My$$

mit  $M \equiv [I_n - X(X'X)^{-1}X']$  der „Residualmaker Matrix“

$$e = My = M[X\beta + \varepsilon] = [I_n - X(X'X)^{-1}X'] [X\beta + \varepsilon] = X\beta - X\beta + M\varepsilon = M\varepsilon$$

M ist idempotent:  $M'M = M$

$\Rightarrow$  Beweis:

$$\begin{aligned} M'M &= [I_n - X(X'X)^{-1}X']' [I_n - X(X'X)^{-1}X'] \\ &= I_n - \{X(X'X)^{-1}X'\}' - X(X'X)^{-1}X' + \{X(X'X)^{-1}X'\}' \{X(X'X)^{-1}X'\} \\ &= I_n - 2X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X' = I_n - X(X'X)^{-1}X' = M \end{aligned}$$

Dann gilt:  $e'e = \varepsilon'M'M\varepsilon = \varepsilon'M\varepsilon$

$$E[e'e | X] = E[\varepsilon'M\varepsilon | X] = E[\underbrace{\text{trace}(\varepsilon'M\varepsilon)}_{1 \times n \quad n \times n \quad n \times 1} | X]$$

$\underbrace{\hspace{10em}}_{1 \times 1}$

weil die „trace“ eines Skalars natürlich gleich dem Skalar selbst ist.

Nach der Permutationsregel ( $\text{trace}(ABC) = \text{trace}(BCA)$ ) gilt dann:

$$E[e'e | X] = E[\text{trace}(M\varepsilon\varepsilon') | X] = \text{trace}(M) E(\varepsilon\varepsilon' | X) = \text{trace}(M)\tau^2$$

$$\text{trace}(M) = \text{trace}(I_n) - \text{trace}(X(X'X)^{-1}X')$$

Annahme 4

und nach erneuter Anwendung der Permutationsregel

$$\text{trace}(M) = \text{trace}(I_n) - \text{trace}(\underbrace{(X'X)^{-1}}_{k \times k} \underbrace{X'X}_{k \times k})$$

$$= \text{trace}(I_n) - \text{trace}(I_k)$$

$$= n - k$$

Also:  $E[e'e | X] = (n-k)\tau^2$  bzw.

$$E \left[ \frac{\sum_{i=1}^n e_i^2}{n-k} \mid X \right] = \tau^2$$

### v) t-Test

Man definiere  $S_{kk}$  als das k-te Element von  $(X'X)^{-1}$ .

$z_k = \frac{b_k - \beta_k}{\sqrt{\tau^2 S_{kk}}}$  ist standard-normalverteilt (siehe iii)).

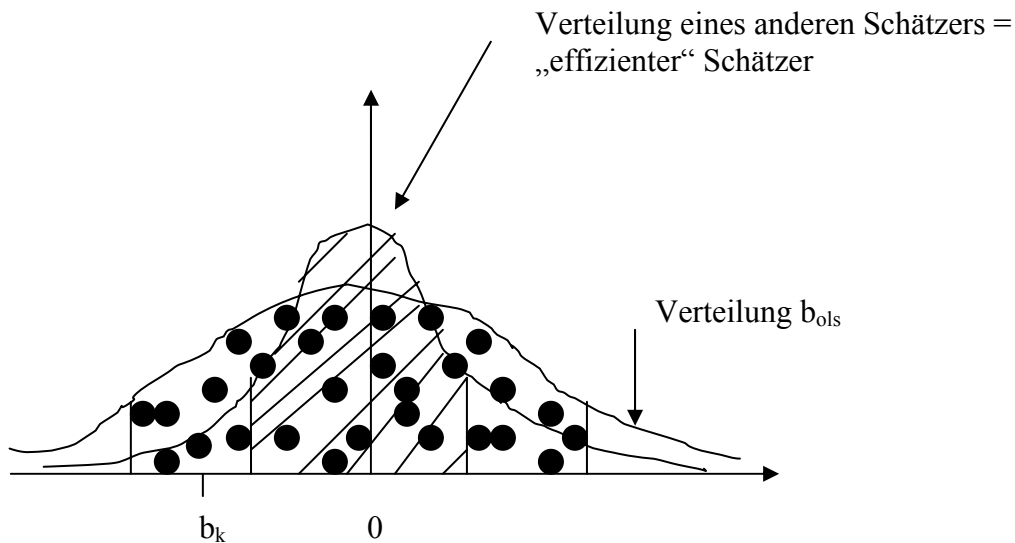
$\frac{e'e}{\tau^2} = \left(\frac{\varepsilon}{\tau}\right)' M \left(\frac{\varepsilon}{\tau}\right)$  ist  $\chi^2$  verteilt mit  $(n-k)$  – Freiheitsgraden, weil  $\left(\frac{\varepsilon}{\tau}\right)$  nach Annahme 6a standard-normalverteilt ist.

Die Test- Statistik

$$t_k = \frac{(b_k - \beta_k)}{\sqrt{\tau^2 S_{kk}}} / \sqrt{\frac{\tau^2}{e'e} (n-k)} = \frac{b_k - \beta_k}{\sqrt{\frac{e'e S_{kk}}{n-k}}}$$

Freiheitsgraden, weil  $z_k$  und  $\frac{\tau^2}{e'e}$  voneinander unabhängig sind (Beweis wird unterlassen).

Nächstes Problem: ist  $b_{ols}$  der beste Schätzer unter allen linearen unverzerrten Schätzern?



- ▨ Fläche = 0,95 für „effizienten“ Schätzer
- ▣ Fläche = 0,95 für  $b_{ols}$

Wenn  $b_{ols}$  ineffizient wäre, könnte  $b_k$  als „insignifikant“ angesehen werden, obwohl  $b_k$  tatsächlich „signifikant“ ist!  $\Rightarrow$  Der „Signifikanz- Test“ hätte geringe „power“, wenn  $b_{ols}$  nicht effizient wäre.

#### vi) Effizienz des OLS-Schätzers

Wir wollen zeigen,  $b_{ols}$  hat unter den linearen, unverzerrten Schätzern die geringste Varianz. Lineare Schätzer  $b_L = Ly$  mit  $L$  eine stochastische  $k \times n$ - Matrix.  $b_{ols}$  ist ein Spezialfall mit  $L = (X'X)^{-1}X'$

Zusätzliche Annahmen:

$$E[\varepsilon | X, L] = 0; E[\varepsilon \varepsilon' | X, L] = \tau^2 I_n$$

$$E[b_L] = \beta \quad (\text{Annahme der Unverzerrtheit})$$

$$E[b_L] = \beta = E[Ly] = E[LX\beta + L\varepsilon]$$

$$\Rightarrow LX = I_k \text{ und } E[b_L - \beta] = E[L\varepsilon]$$

$$\begin{aligned} Var [b_L | X, L] &= E[(b_L - \beta)(b_L - \beta)' | X, L] \\ &= E[L\varepsilon(L\varepsilon)' | X, L] \\ &= E[L\varepsilon\varepsilon'L' | X, L] \\ &= LE [\varepsilon\varepsilon' | X, L] L' \\ &= \tau^2 LL' \end{aligned}$$

Annahme 4

$$\begin{aligned} \text{Man definiere } D &\equiv L - (X'X)^{-1}X' \\ \Rightarrow DX &= LX - (X'X)^{-1}X'X = 0 \end{aligned}$$

$$\begin{aligned} L &= D + (X'X)^{-1}X' \\ LL' &= [D + (X'X)^{-1}X'] [D + (X'X)^{-1}X']' \\ &= [D + (X'X)^{-1}X'] [D' + X(X'X)^{-1}] \end{aligned}$$

$$\begin{aligned}
&= DD' + (X'X)^{-1} \underbrace{X'D'}_{=0} + \underbrace{DX(X'X)^{-1}}_{=0} + (X'X)^{-1} X'X(X'X)^{-1} \\
&= DD' + (X'X)^{-1}
\end{aligned}$$

$$\begin{aligned}
\text{Var}[b_L | X, L] &= \tau^2 LL' \\
&= \tau^2 [DD' + (X'X)^{-1}] \\
&= \tau^2 [DD'] + \tau^2 [(X'X)^{-1}] \\
&= \tau^2 [DD'] + \text{Var}[b_{ols} | X, L]
\end{aligned}$$

Zum besseren Verständnis:

$$\begin{aligned}
D &= [D_{mi}]_{k \times n} \\
\Rightarrow \underbrace{DD'}_{k \times k} &= \left[ \sum_{i=1}^n D_{mi} \cdot D_{ri} \right]_{k \times k}
\end{aligned}$$

Für jedes Element m in  $b_L$  gilt:

$$\begin{aligned}
\text{Var}[b_{Lm} | X, L] - \text{Var}[b_{olsm}] &= [0 \dots \underset{\substack{\downarrow \\ \text{m-tes Element}}}{1} \dots 0] [DD'] \tau^2 [0 \dots \underset{\substack{\downarrow \\ \text{m-tes Element}}}{1} \dots 0]' \\
&= \left[ \sum_{i=1}^n D_{mi}^2 \right] \tau^2 \geq 0 \\
\Rightarrow \text{Var}[b_L | X, L] &\geq \text{Var}[b_{OLS} | X, L]
\end{aligned}$$

### 1d) Partitioned Regression:

- Zweck:
- a) Interpretation von  $b_{ols}$
  - b) Auslassen relevanter Variablen (omitted variable bias)
  - c) Einfügen irrelevanter Variablen
  - d) Fixed-Effect Panel Schätzer (LSDV)

#### 1d1) Interpretation des Partial Effekts

$$y = X\beta + \varepsilon = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

Die Normalgleichungen des OLS- Schätzers lauteten (Kapitel 1b):

$$\begin{aligned}
&X'Xb_{ols} = X'y \\
\Leftrightarrow [X_1'X_2]' [X_1'X_2] \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &= [X_1'X_2]'y \quad | \quad k=k_1+k_2 \\
\Leftrightarrow \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} &= \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\Leftrightarrow X_1'X_1b_1 + X_1'X_2b_2 &= X_1'y & | \quad 1. \text{ Gleichungssystem} \\
\Leftrightarrow X_2'X_1b_1 + X_2'X_2b_2 &= X_2'y & | \quad 2. \text{ Gleichungssystem}
\end{aligned}$$

$$\begin{aligned}
& X_1'X_1b_1 = X_1'y - X_1'X_2b_2 = X_1'[y - X_2b_2] \\
& b_1 = (X_1'X_1)^{-1} X_1'[y - X_2b_2] \\
& X_2'X_1(X_1'X_1)^{-1}X_1'[y - X_2b_2] + X_2'X_2b_2 = X_2'y \quad | \text{ b}_1 \text{ in 2. Gleichung} \\
& [X_2'X_2 - X_2'X_1(X_1'X_1)^{-1}X_1'X_2]b_2 = [X_2' - X_2'X_1(X_1'X_1)^{-1}X_1']y \\
& X_2' \underbrace{[I_n - X_1(X_1'X_1)^{-1}X_1']}_{\equiv M_1} X_2b_2 = X_2' \underbrace{[I_n - X_1(X_1'X_1)^{-1}X_1']}_{\equiv M_1} y \quad | \text{ M ist «Residualmaker» Matrix (siehe 1c iv)} \\
& b_2 = [X_2'M_1X_2]^{-1} X_2'M_1y
\end{aligned}$$

Was bedeutet die „Residual- Maker“ Matrix  $M_1$ ?

$$\begin{aligned}
M_1X_2 &= X_2 - X_1(X_1'X_1)^{-1}X_1'X_2 = X_2 - X_1\gamma = \phi \\
\gamma: X_2 &= X_1\gamma + \phi \\
&\quad \begin{matrix} n+k_2 & nxk_1 & k_1xb_2 & nxk_2 \end{matrix}
\end{aligned}$$

$\implies$   $M_1$ - Matrix erzeugt Residuen einer Regression der Variablen  $X_2$  auf  $X_1$

Wir erinnern uns:  $M_1$  ist idempotent ( $M_1 = M_1'M_1$ ) (1c (iv))

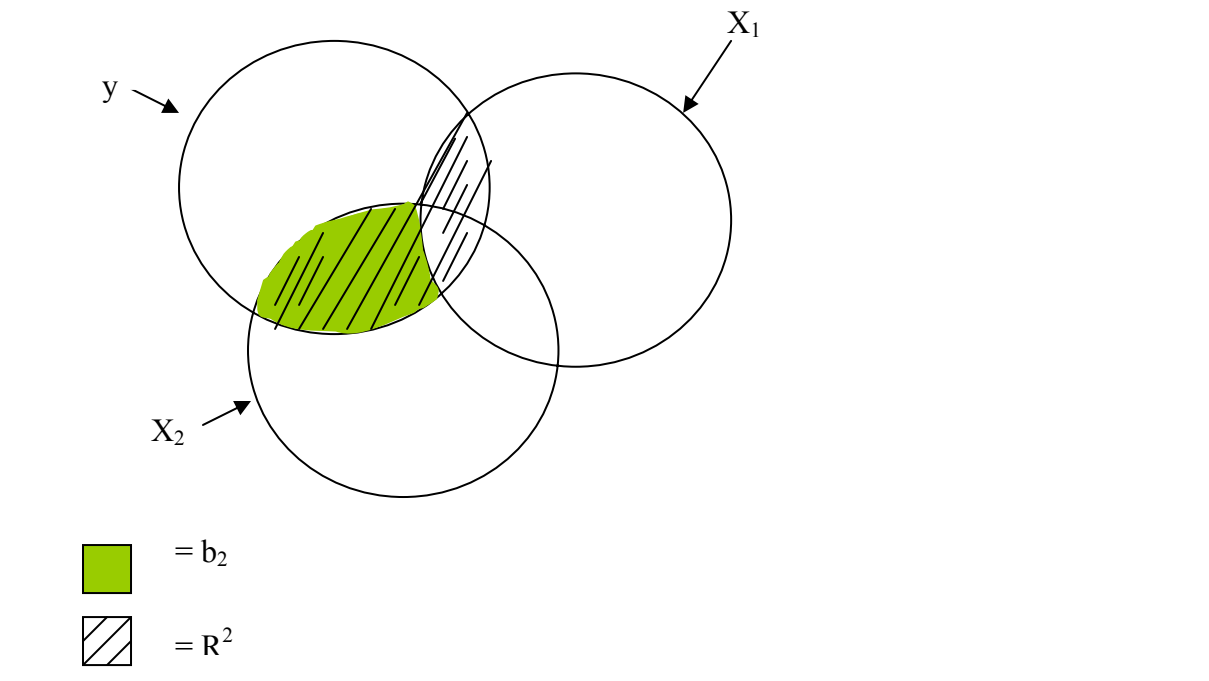
$$b_2 = [X_2' M_1' M_1 X_2]^{-1} X_2' M_1' M_1 y$$

Definition:  $X_2^* = M_1 X_2$  = Fehlerterm der Regression  $X_2$  auf  $X_1$   
 $y^* = M_1 y$  = Fehlerterm der Regression  $y$  auf  $X_1$

$$(**) \quad b_2 = [X_2^{*'} X_2^*]^{-1} X_2^{*'} y^* \quad \Rightarrow \text{Frisch- Waugh Theorem}$$

$b_2$  ist der Koeffizient, der den Teil des Einflusses von  $X_2$  auf  $y$  erklärt, der nicht gleichzeitig von  $X_1$  erklärt wird.

## Venn-Diagramm



Der Regressionskoeffizient misst nur die gemeinsame Variation von  $y$  und  $x_2$  Variablen, die nicht gleichzeitig auch gemeinsame Variation mit der  $x_1$  Variablen ist. Dies ist noch nicht hinreichend, um Kausalität zu behaupten. Aber es ist eine notwendige Bedingung für Kausalität. Daher misst der Regressionskoeffizient  $b_2$  die partielle Korrelation zwischen  $y$  und  $x_2$ .

Gleichzeitig ist dieser Koeffizient die partielle Ableitung des Erwartungswertes der  $y$ -Variablen nach der  $x_2$ -Variablen. Ein partieller Regressionskoeffizient korrespondiert damit mit einer partiellen Ableitung einer endogenen Variablen nach entweder einem exogenen Parameter (Kausalitätsfall) oder einer weiteren endogenen Variablen in einem theoretischen Modell (Korrelationsfall). Somit ist das Regressionsmodell ein geeignetes Instrument, um Modelle zu testen.

## Beispiel: STATA Log-Datei mit einem Zufallszahlenexperiment

```

. clear
. set memory 500m
. set obs 1000000
. gen cons=1
.
. * Define a number of standard errors
. gen s_x=2
. gen s_z2=1
. gen s_z3=4
. gen s_z4=3
.
. * Define several normally distributed random errors
.
. /* xi is the common variation of tariff and tax */
. gen xi=s_x*invnorm(uniform())
.
. * other random variables
. gen z2=s_z2*invnorm(uniform())
. gen z3=s_z3*invnorm(uniform())
. gen z4=s_z4*invnorm(uniform())
.
. * Generate explanatory variables
. gen tariff=1.2*xi+z2
. gen tax=6*xi+z3
.
. * Generate dependent variable
. gen newinv=(-4*tax)+(-1.5*tariff)+z4
.
. * "Full" regression
. * (1)
. reg newinv tariff tax, nocons

```

Source	SS	df	MS			
Model	2.9233e+09	2	1.4617e+09	Number of obs =	1000000	
Residual	8990924.15999998	8.99094213		F( 2,999998) =	.	
Total	2.9323e+091000000	2932.3048		Prob > F =	0.0000	
				R-squared =	0.9969	
				Adj R-squared =	0.9969	
				Root MSE =	2.9985	

```

-----+-----
newinv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
tariff | -1.500612   .0023869   -628.68  0.000   -1.50529   -1.495933
tax    | -3.999978   .0004909  -8147.78  0.000   -4.00094   -3.999015
-----+-----
.
. * regression using only the independent variation of the dependent variables
. * (2)
. reg newinv z2 z3, nocons

```

Source	SS	df	MS			
Model	257759368	2	128879684	Number of obs =	1000000	
Residual	2.6745e+099999998	2674.55078		F( 2,999998) =	48187.41	
Total	2.9323e+091000000	2932.3048		Prob > F =	0.0000	
				R-squared =	0.0879	
				Adj R-squared =	0.0879	
				Root MSE =	51.716	

```

-----+-----
newinv |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
z2     | -1.578723   .0517252   -30.52  0.000   -1.680103  -1.477344
z3     | -3.993107   .0129243   -308.96  0.000   -4.018438  -3.967776
-----+-----

```

```

. cor z2 z3
(obs=1000000)

-----+-----
          |          z2          z3
-----+-----
          |          1.0000          1.0000
z2        |          -0.0007          1.0000
z3        |
-----+-----

. test z2=-1.5

( 1)  z2 = -1.5

      F( 1,999998) =      2.32
      Prob > F    =      0.1280

. test z3=-4

( 1)  z3 = -4

      F( 1,999998) =      0.28
      Prob > F    =      0.5938

.
end of do-file

```

**Schlußfolgerung:** Die beiden Test-Statistiken zum Schluß der log-Datei zeigen, daß die Regressionskoeffizienten der erklärenden Variablen z2 und z3 nicht statistisch signifikant unterschiedlich von den theoretisch wahren Regressionskoeffizienten für die Variablen „tariff“ und „tax“ in der Grundgesamtheit sind. Der OLS-Schätzer berücksichtigt also nicht die gemeinsame Variation „xi“ in den Variablen „tariff“ und „tax“, sondern nur die Variation in den voneinander unabhängigen stochastischen Komponenten „z2“ und „z3“ aus den Variablen „tax“ und „tariff“, die wiederum selbst stochastische Komponente in der abhängigen Variablen „newinv“ sind.

### 1d b) Omitted Variable Bias

Angenommen

$$y = X_1\beta_1 + X_2\beta_2 + \varepsilon \quad \text{sei korrekt.}$$

Geschätzt werde aber

$$y = X_1b_1 + e_1$$

$$b_1 = (X_1'X_1)^{-1}X_1'y$$

$$\begin{aligned}
E[b_1 - \beta_1 | X] &= E[(X_1'X_1)^{-1}X_1'y | X] - \beta_1 \\
&= E[(X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + \varepsilon) | X] - \beta_1 \\
&= \beta_1 + E[(X_1'X_1)^{-1}X_1'X_2 | X] \beta_2 + \underbrace{E[(X_1'X_1)^{-1}X_1'\varepsilon | X]}_{=0} - \beta_1
\end{aligned}$$

$$= E[(X_1'X_1)^{-1}X_1'X_2 | X] \beta_2 \neq 0$$

$\implies$   $b_1$  ist ein verzerrter Schätzer, es sei denn  $X_1$  und  $X_2$  sind nicht korreliert ( $X_1'X_2=0$ ) (oder  $\beta_2=0$ ).

### 1d c) Einschluss irrelevanter Variablen

Wahr sei:

$$y = X_1\beta_1 + \varepsilon \quad (***)$$

Geschätzt werde:

$$y = X_1b_1 + X_2b_2 + e \quad (***)$$

Ergebnisse:

- $b_1$  muss unverzerrt sein, weil (\*\*\*) auch als  $y = X_1\beta_1 + X_2\theta + \varepsilon$  geschrieben werden kann und bereits angezeigt wurde, dass  $b_1$  dann unverzerrt ist.
- Effizienz:

$$\text{Var}[\beta_1 | X] = \tau^2 (X_1' X_1)^{-1} \quad | \beta_1 \text{ ist OLS- Schätzer der „wahren Regression“}$$

Dagegen gilt:

$$\text{Var}[b_1 | X] = \tau^2 (X_1' X_1^* X_1^*)^{-1} = \tau^2 (X_1' M_2 X_1)^{-1} \\ X_1' [I_n - X_2 (X_2' X_2)^{-1} X_2'] X_1$$

Nach dem Frisch-Waugh- Theorem mit  $M_2 \equiv I_n - X_2 (X_2' X_2)^{-1} X_2'$

$$\text{Var}[b_1 | X] = \tau^2 [X_1' X_1 - X_1' X_2 (X_2' X_2)^{-1} X_2' X_1]^{-1}$$

$$\text{Var}[\beta_1 | X]^{-1} - \text{Var}[b_1 | X]^{-1} = \frac{1}{\tau^2} \underbrace{[X_1' X_2 (X_2' X_2)^{-1} X_2' X_1]}_{\text{quadratische Form}}$$

⇒ Regel: eine quadratische Form ist nicht-negativ definit

⇒  $\boxed{\text{Var}[\beta_j | X] < \text{Var}[b_j | X]}$  für  $j=1 \dots k_1$  und  $\beta_j$  ist Element  $j$  des  $1 \times k_1$  Vektors  $\beta_1$  bzw.  $b_j$  Element  $j$  des  $1 \times k_1$  Vektors  $b_1$   
wobei die Regel  $A > B \iff A^{-1} < B^{-1}$  (Greene, 2000, S. 49) zur Anwendung kommt.

### **Schlussfolgerung:**

Einschluss von irrelevanten Variablen in einer Regression führt zu ineffizienten Schätzern  
 ⇒ Insignifikante Koeffizienten können in Wirklichkeit signifikant sein (konservativer Fehler).

## 1e) Eigenschaften des OLS- Schätzers in der großen Stichprobe

Problem mit der BLUE- Eigenschaft:

- Zwar ist die Varianz des OLS- Schätzers minimiert, aber das heißt nicht, dass diese Varianz „klein“ oder gar null ist. Je größer diese minimierte Varianz ist, umso schwieriger ist es, die Signifikanz von  $\beta_k$  nahe null nachzuweisen.
- Die Normalverteilungsannahme ist zu restriktiv. Daher wird ein Alternativkonzept vorgestellt: das des „konsistenten“ Schätzers an Stelle des „erwartungstreuen“ Schätzers.  
Hierzu sind neue statistische Konzepte notwendig.

$x_n$  sei eine Zufallsvariable basierend auf  $n$  i.i.d. Ziehungen aus einer Grundgesamtheit ( $X_n$  kann als Stichprobenmittelwert einer Stichprobe mit  $n$ - Beobachtungen gedacht werden).

### Definition 1: Convergence in probability

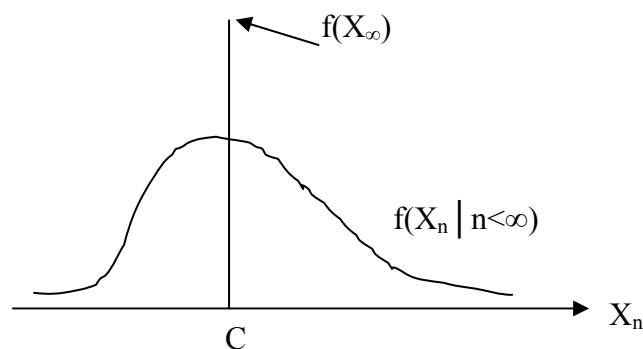
Eine Zufallsvariable  $x_n$  konvergiert „in probability“ gegen die Konstante  $C$ , wenn

$$\lim_{n \rightarrow \infty} \text{prob}(|x_n - C| > \varepsilon) = 0 \quad \text{für jedes } \varepsilon > 0 .$$

Kurzschreibweise:  $\text{plim } x_n = C$  oder  $x_n \xrightarrow{P} C$

Erläuterung:

$f(\cdot)$  sei Wahrscheinlichkeitsdichtefunktion

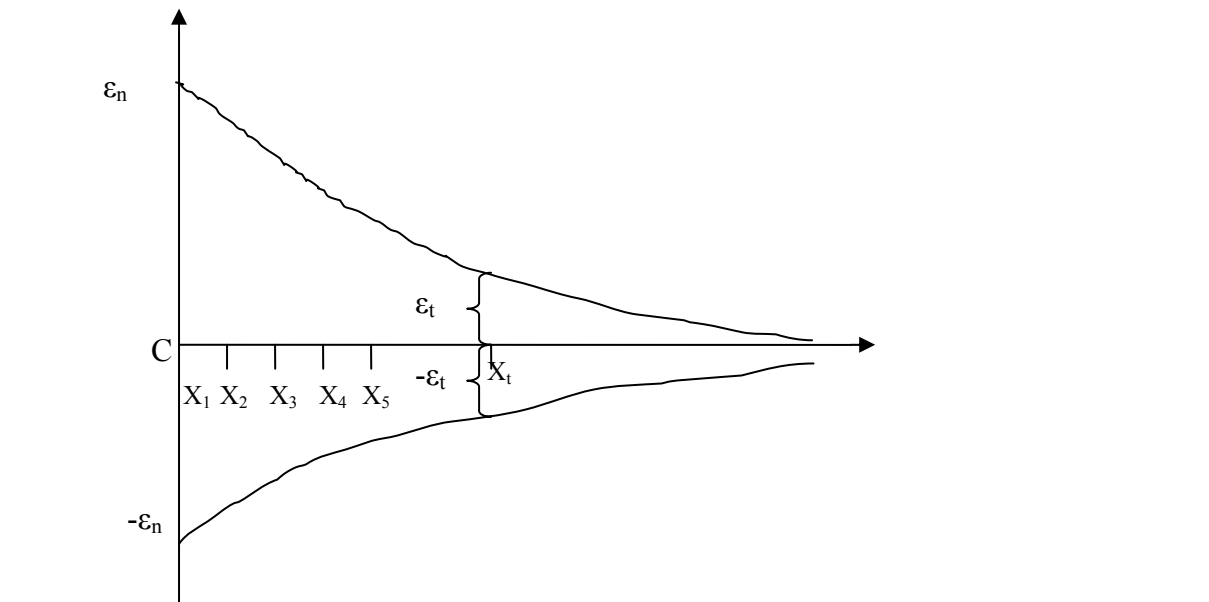


Bemerkung: „Ähnliche“ Konzepte sind „almost sure“, Konvergenz (strikt) und Konvergenz „in distribution“ (weniger strikt).  
„Convergence in probability“ ist die einfachere Textbuch-Variante.

Definition 2: Almost sure convergence (a.S.)

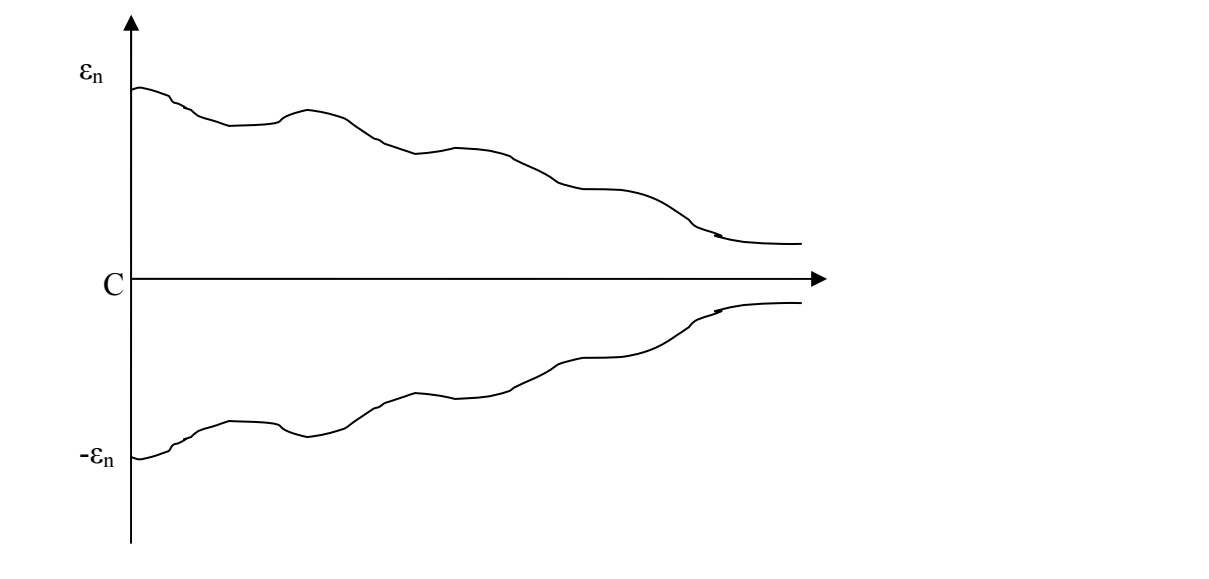
$$\text{prob}\left(\lim_{n \rightarrow \infty} x_n = C\right) = 1$$

Intuitive Unterscheidung von „convergence in probability“ und „almost sure convergence“:



Bemerkung: die  $(\varepsilon_n$  und  $-\varepsilon_n)$  – Funktionen geben die Grenzwerte an, außerhalb derer keine Werte der Zufallsvariable  $x_n$  mehr existieren.

Beispiel einer „convergence in probability“, die keine „almost sure convergence“ ist:



„Almost sure convergence“ stellt sicher, dass die Varianz von  $x_n$  bei jeder zusätzlichen Beobachtung in der Stichprobe kleiner wird. Damit gibt es eine endliche Stichprobengröße

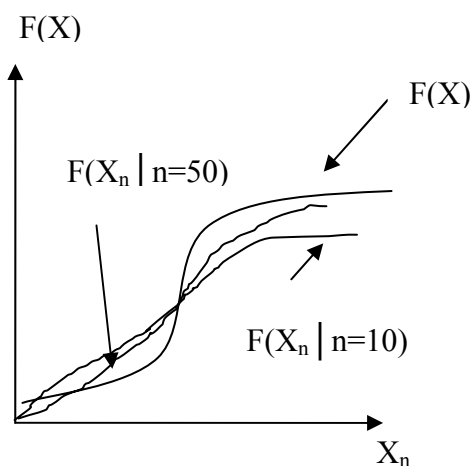
(Faustregel  $n \geq 30$ ) ab der die Varianz der Zufallsvariable so gering ist, wie man es sich wünscht.

### Definition 3: Convergence in Distribution

Die Zufallszahl  $X_n$  aus einer Stichprobe mit  $n$  Beobachtungen und einer stichprobengrößenabhängigen kumulativen Verteilungsfunktion  $F_n(X_n)$  konvergiert „in distribution“ zu einer Zufallsvariablen  $X$  mit kumulativer Verteilungsfunktion  $F(X)$ , wenn

$$\lim_{n \rightarrow \infty} |F_n(X_n) - F(X)| = 0 \quad \text{in allen Kontinuitätspunkten von } F(X).$$

Kurz:  $X_n \xrightarrow{d} X$



### Exkurs: Beispiel für die Konvergenz von Zufallsprozessen

```
. gen random=uniform()
. gen sample1=random<0.1
. gen sample2=random<0.01
. gen sample3=random<0.001
. gen sample4=random<0.0008
. gen sample5=random<0.0005
. gen sample6=random<0.0003
. gen sample7=random<0.0001
```

```
. reg lsales lemployment lcapital lmaterial
```

Source	SS	df	MS	Number of obs =	167229
Model	234770.211	3	78256.737	F( 3,167225) =	.
Residual	7230.3944167225		.043237521	Prob > F =	0.0000
Total	242000.605167228	1.4471297		R-squared =	0.9701
				Adj R-squared =	0.9701
				Root MSE =	.20794

lsales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lemployment	.1615397	.0006259	258.09	0.000	.1603129 .1627664
lcapital	.0318455	.000413	77.10	0.000	.0310359 .0326551
lmaterial	.7926037	.000573	1383.24	0.000	.7914806 .7937268
_cons	1.449894	.0039428	367.73	0.000	1.442166 1.457622

```
. reg lsales lemployment lcapital lmaterial if sample1==1
```

Source	SS	df	MS	Number of obs =
Model	23803.775	3	7934.59166	16758
Residual	681.349553	16754	.040667874	F( 3, 16754) = .
Total	24485.1245	16757	1.46118783	Prob > F = 0.0000
				R-squared = 0.9722
				Adj R-squared = 0.9722
				Root MSE = .20166

lsales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lemployment	.1581932	.0019027	83.14	0.000	.1544636 .1619227
lcapital	.0329588	.0012637	26.08	0.000	.0304819 .0354357
lmaterial	.7944815	.0017572	452.14	0.000	.7910373 .7979257
_cons	1.438326	.0120851	119.02	0.000	1.414638 1.462014

. reg lsales lemployment lcapital lmaterial if sample2==1

Source	SS	df	MS	Number of obs =
Model	2418.1434	3	806.047801	1648
Residual	45.9761461	1644	.027966026	F( 3, 1644) =28822.39
Total	2464.11955	1647	1.49612602	Prob > F = 0.0000
				R-squared = 0.9813
				Adj R-squared = 0.9813
				Root MSE = .16723

lsales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lemployment	.1702203	.0050808	33.50	0.000	.1602548 .1801858
lcapital	.0244708	.0034644	7.06	0.000	.0176758 .0312658
lmaterial	.7958627	.0047889	166.19	0.000	.7864698 .8052557
_cons	1.447786	.0329155	43.98	0.000	1.383226 1.512347

. reg lsales lemployment lcapital lmaterial if sample3==1

Source	SS	df	MS	Number of obs =
Model	230.999392	3	76.9997974	158
Residual	3.13871817	154	.020381287	F( 3, 154) = 3777.97
Total	234.13811	157	1.49132554	Prob > F = 0.0000
				R-squared = 0.9866
				Adj R-squared = 0.9863
				Root MSE = .14276

lsales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lemployment	.1721012	.0140831	12.22	0.000	.1442802 .1999222
lcapital	.0318865	.0093371	3.42	0.001	.0134412 .0503317
lmaterial	.7805357	.0137117	56.92	0.000	.7534484 .8076229
_cons	1.522881	.0931702	16.35	0.000	1.338824 1.706938

. reg lsales lemployment lcapital lmaterial if sample4==1

Source	SS	df	MS	Number of obs =
Model	178.135899	3	59.3786332	130
Residual	2.58435026	126	.020510716	F( 3, 126) = 2895.01
Total	180.72025	129	1.40093217	Prob > F = 0.0000
				R-squared = 0.9857
				Adj R-squared = 0.9854
				Root MSE = .14322

lsales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lemployment	.1798758	.0151124	11.90	0.000	.1499688 .2097828
lcapital	.0194149	.0106788	1.82	0.071	-.0017181 .0405478
lmaterial	.7875758	.0150436	52.35	0.000	.7578049 .8173467
_cons	1.521698	.1042035	14.60	0.000	1.315482 1.727913

```
. reg lsales lemployment lcapital lmaterial if sample5==1
```

Source	SS	df	MS	Number of obs =	80
Model	93.874217	3	31.2914057	F( 3, 76) =	1301.72
Residual	1.82692652	76	.024038507	Prob > F =	0.0000
				R-squared =	0.9809
				Adj R-squared =	0.9802
Total	95.7011435	79	1.21140688	Root MSE =	.15504

lsales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lemployment	.1687017	.0217382	7.76	0.000	.1254064 .211997
lcapital	.0316878	.0151949	2.09	0.040	.0014244 .0619511
lmaterial	.782881	.0209925	37.29	0.000	.7410708 .8246911
_cons	1.520622	.1500053	10.14	0.000	1.22186 1.819383

```
. reg lsales lemployment lcapital lmaterial if sample6==1
```

Source	SS	df	MS	Number of obs =	45
Model	42.5473238	3	14.1824413	F( 3, 41) =	447.59
Residual	1.29913313	41	.031686174	Prob > F =	0.0000
				R-squared =	0.9704
				Adj R-squared =	0.9682
Total	43.8464569	44	.996510385	Root MSE =	.17801

lsales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lemployment	.1853139	.0299393	6.19	0.000	.1248503 .2457775
lcapital	.0282158	.0236912	1.19	0.241	-.0196296 .0760612
lmaterial	.7676922	.0322098	23.83	0.000	.7026433 .8327412
_cons	1.651597	.238013	6.94	0.000	1.17092 2.132274

```
. reg lsales lemployment lcapital lmaterial if sample7==1
```

Source	SS	df	MS	Number of obs =	14
Model	12.8943052	3	4.29810174	F( 3, 10) =	99.36
Residual	.43258586	10	.043258586	Prob > F =	0.0000
				R-squared =	0.9675
				Adj R-squared =	0.9578
Total	13.3268911	13	1.02514547	Root MSE =	.20799

lsales	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
lemployment	.1701595	.0687259	2.48	0.033	.0170287 .3232904
lcapital	-.0342839	.0429966	-0.80	0.444	-.1300862 .0615184
lmaterial	.8688577	.0601479	14.45	0.000	.7348398 1.002876
_cons	1.232574	.4383036	2.81	0.018	.255973 2.209175

#### Definition 4: Konsistenzschätzer

Ein Schätzer  $\hat{\theta}$  eines Parameters  $\theta$  ist konsistent, genau dann, wenn  $p\lim \hat{\theta} = \theta$

Weitere Vorteile dieses Konzepts des konsistenten Schätzers relativ zum erwartungstreuen Schätzer:

- Die Annahme der Normalverteilung der Störvariablen ist oft nicht erfüllt und wir werden sehen, dass diese Annahme für die große Stichprobe überflüssig werden wird.
- Für zahlreiche Schätzverfahren (z.B. Kointegration, GMM-Schätzer, IV- Schätzer) existieren keine Ergebnisse für die kleine Stichprobe.

## 1e 1) Statistische Grundlagen für die Analyse von Schätzern in der großen Stichprobe

Theorem 1: Stichprobenverteilung des Stichprobenmittelwertes

$x_1, \dots, x_n$  sei eine i.i.d. Zufallsstichprobe mit Mittelwert  $\mu$  und Varianz  $\tau^2$  in der Grundgesamtheit, dann ist  $\bar{x}_n \equiv \frac{1}{n} \sum_{i=1}^n x_i$  eine Zufallsvariable mit Mittelwert  $\mu$  und Varianz  $\frac{\tau^2}{n}$

Beweis:

$$E[\bar{x}_n] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \frac{1}{n} \sum_{i=1}^n \mu = \mu$$
$$Var[\bar{x}_n] = Var\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n^2} Var\left[\sum_{i=1}^n x_i\right] = \frac{1}{n^2} \sum_{i=1}^n \tau^2 = \frac{\tau^2}{n},$$

wobei das vorletzte „=“ aus der Annahme folgt, dass  $X_i$  i.i.d. sind (Zufallsstichprobe).  
q.e.d.

(Siehe auch Greene, 2000, S. 101)

Theorem 2: Markov's Inequality

Wenn  $y_n$  eine nicht- negative Zufallsvariable und  $\delta$  eine positive Konstante ist, dann gilt:

$$prob[y_n \geq \delta] \leq E[y_n] / \delta$$

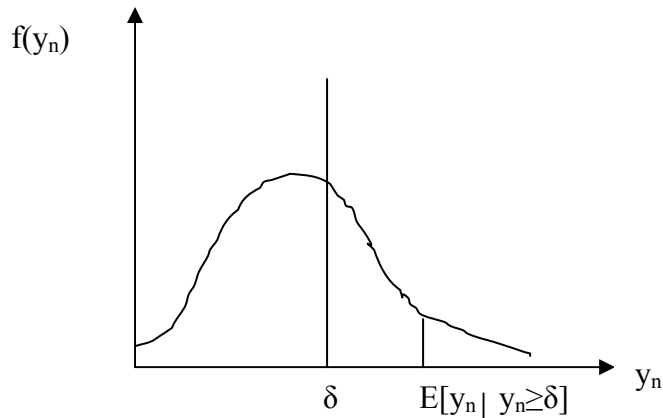
Beweis:

$$E[y_n] = \int_0^{\infty} y_n f(y_n) dy_n$$
$$= \int_0^{\delta} y_n f(y_n) dy_n + \int_{\delta}^{\infty} y_n f(y_n) dy_n$$
$$= F(\delta) \int_0^{\delta} y_n \frac{f(y_n)}{F(\delta)} dy_n + (1 - F(\delta)) \int_{\delta}^{\infty} y_n \frac{f(y_n)}{1 - F(\delta)} dy_n$$
$$= Prob[y_n \leq \delta] \cdot E[y_n | y_n < \delta] + Prob[y_n \geq \delta] \cdot E[y_n | y_n \geq \delta]$$

Wobei  $f(\cdot)$  eine beliebige Dichtefunktion und  $F(\cdot)$  die zugehörige kumulative Verteilungsfunktion darstellt.

Weil  $y_n > 0$ , gilt  $E[y_n | y_n < \delta] > 0$  und damit  
 $E[y_n] \geq prob[y_n \geq \delta] E[y_n | y_n \geq \delta]$

Aber  $E[y_n | y_n \geq \delta] > \delta$



$$E[y_n] \geq \delta \text{Prob}[y_n \geq \delta] \quad \text{q.e.d.}$$

### Theorem 3: Chebychev's Inequality

Wenn  $x_n$  eine Zufallsvariable ist, und  $C$  und  $\varepsilon$  Konstanten,  $\varepsilon > 0$ , dann gilt:

$$\text{Prob}(|x_n - C| > \varepsilon) \leq E[(x_n - C)^2] / \varepsilon^2$$

Beweis:

Man definiere  $y_n = (x_n - C)^2$  und  $\delta = \varepsilon^2$   
 $\text{Prob}[(x_n - C)^2 \geq \varepsilon^2] \leq E[(x_n - C)^2] / \varepsilon^2$

nach Markov's Ungleichheit

Aber  $(x_n)^2 \geq \varepsilon^2$  impliziert  $|x_n - C| \geq \varepsilon$  und daher folgt  
 $\text{Prob}[|x_n - C| \geq \varepsilon] \leq E[(x_n - C)^2] / \varepsilon^2 \quad \text{q.e.d.}$

### Theorem 4: Chebyshev's Theorem

$x_n$  sei eine Folge von Zufallsvariablen und  $\tau_n^2$  die Varianz dieser Zufallsvariablen. Wenn gilt

(1)  $\lim_{n \rightarrow \infty} E[x_n] = C$  und (2)  $\lim_{n \rightarrow \infty} \tau_n^2 = 0$  für eine Konstante  $C$ ,

dann  $\text{plim } x_n = C$ .

Beweis:

$$\lim_{n \rightarrow \infty} \text{Prob}[|x_n - C| \geq \varepsilon] \leq \lim_{n \rightarrow \infty} E[(x_n - C)^2] / \varepsilon^2$$

nach Markov's Ungleichheit

$$\lim_{n \rightarrow \infty} \frac{E[(x_n - C)^2]}{\varepsilon^2} = \lim_{n \rightarrow \infty} \frac{\int [(x_n - C)^2 f(x_n)] dx_n}{\varepsilon^2}$$

$$= \lim_{n \rightarrow \infty} \frac{\int x_n^2 f(x_n) dx_n + C^2 - 2C \int x_n f(x_n) dx_n}{\varepsilon^2} \quad | \quad (1)$$

$$= \lim_{n \rightarrow \infty} \frac{\int x_n^2 f(x_n) dx_n + E[x_n]^2 - 2E[x_n] \int x_n f(x_n) dx_n}{\varepsilon^2}$$

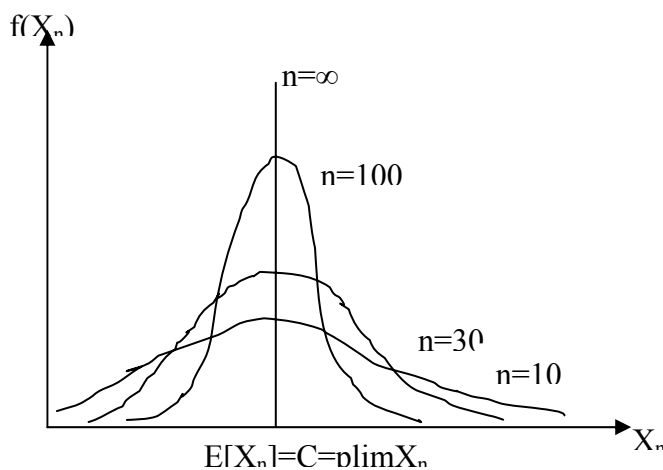
$$= \lim_{n \rightarrow \infty} \frac{E[(x_n - E[x_n])^2]}{\varepsilon^2}$$

$$= \lim_{n \rightarrow \infty} \frac{\tau_n^2}{\varepsilon^2} = 0 \quad | \quad (2)$$

$$\implies \lim_{n \rightarrow \infty} \text{Pr ob} [|x_n - C| \geq \varepsilon] = 0 \quad \text{q.e.d.}$$

Bemerkung 1: Wenn  $x_n$  i.i.d. verteilt sind, dann kann Bedingung (1) natürlich ersetzt werden durch (1a)  $E[x_n] = C$ .

Bemerkung 2: Intuition des Chebyshev Theorems:



### Theorem 5: Slutsky – Theorem

Für eine kontinuierliche Funktion  $g(x_n)$ , wobei  $g(\cdot)$  unabhängig von  $n$  ist, gilt:

$$\text{plim } g(x_n) = g(\text{plim } x_n)$$

Beweis: ausgelassen

z.Bsp.:  $W_n$  sei eine Zufallsmatrix mit  $\text{plim } W_n = \Omega$ ,  
dann gilt:

$$(3) \quad \text{plim } W_n^{-1} = (\text{plim } W_n)^{-1} = \Omega^{-1} \quad (\text{Matrix –Inversions-Regel})$$

(wenn  $\Omega$  invertierbar ist)

### Theorem 6: Lindberg-Levy- Central Limit Theorem

Wenn  $x_1, \dots, x_n$  eine Zufallsauswahl von einer multivariaten Verteilung mit Mittelwertvektor  $\mu < \infty$  und positiv definiter Kovarianz-Matrix  $Q < \infty$  ist, dann

$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N[0, Q]$ , wobei

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i .$$

Bemerkung: Mittelwert und Kovarianzmatrix sind für alle Beobachtungen  $i$  gleich!

### Theorem 7: Theorie des Stichprobenmittelwertes

$x_i, i=1, \dots, n$ , sei Zufallsstichprobe der Zufallsvariable  $X$ ;

$$\bar{x}_n = \frac{1}{n} \sum_{i=1}^n x_i ;$$

Wenn  $E(X) < \infty$  und  $\text{Var}(X) < \infty$ , dann  $p \lim \bar{X} = E(X)$

Beweis:

$$E(\bar{X}) = E(X) \quad \text{nach Theorem 1}$$

$$\text{Var}(\bar{X}) = \frac{\text{Var}(x)}{n} \quad \text{nach Theorem 1}$$

$$(1) \quad \lim_{n \rightarrow \infty} E\left(\frac{1}{n} \sum_{i=1}^n x_i\right) = \lim_{n \rightarrow \infty} E(X) = E(X) < \infty$$

$$(2) \quad \lim_{n \rightarrow \infty} \text{Var}\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \lim_{n \rightarrow \infty} \frac{\text{Var}(X)}{n} = 0$$

Da Bedingungen (1) und (2) gelten, kann auf  $\bar{X}$  das Chebychev- Theorem angewandt werden:

$$p \lim \bar{X} = E(X) \quad \text{q.e.d.}$$

### Corollary zu Theorem 7:

Wenn  $E(g(z)) < \infty$  und  $\text{Var}[g(z)] < \infty$ ,  $g(z)$  eine beliebige Funktion von  $z$  (unabhängig von  $n$ ),  $z$  eine i.i.d. Zufallsvariable ist, dann:

$$p \lim \frac{1}{n} \sum_{i=1}^n g(z_i) = E[g(z)].$$

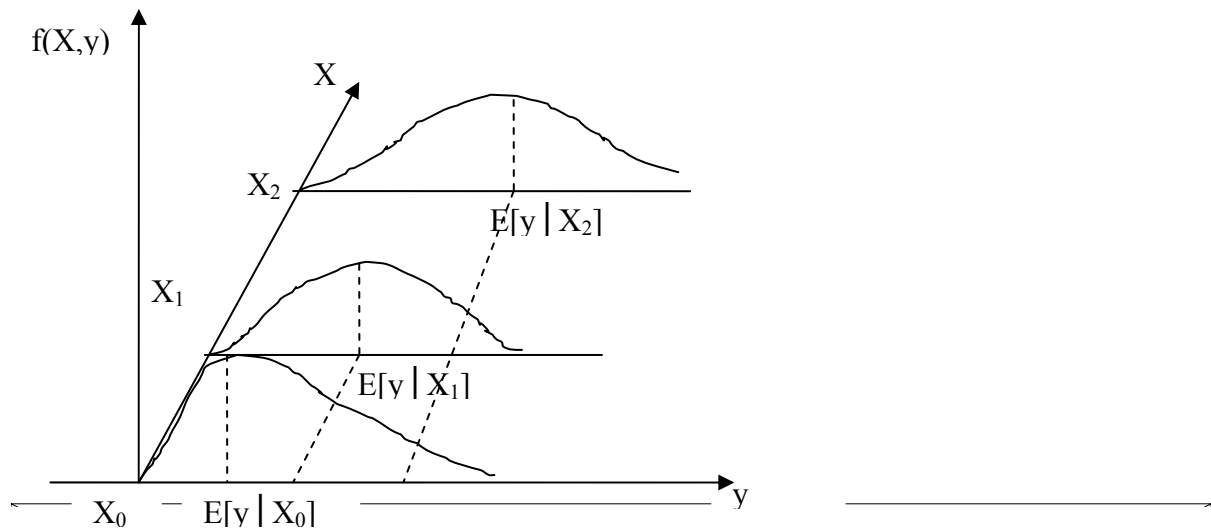
Beweis:

Man wende Theorem 7 auf  $X=g(z)$  an.      q.e.d.

Theorem 8: Gesetz von der wiederholten Erwartung (Law of iterated expectations)

$E[y] = E_X[E[y|X]]$ , wobei  $E_x[\cdot]$  der Erwartungswert über die X-Werte ist.

Beispiel:  $y$  kontinuierliche Zufallsvariable;  
 $X$  habe die Werte  $x_1, x_2, x_3$  jeweils mit der Wahrscheinlichkeit  $p_1, p_2, p_3$



$$E[Y] = \sum_{i=1}^3 p_i E[Y|X_i]$$

Theorem 9:

Für bivariate Verteilungen gilt:

$$Cov(X,y) = Cov(X, E[y|X])$$

(Greene, 2000, S. 80)

Theorem 10: Cramer's Theorem

(Davidson, 2000, S. 46)

Wenn für einen Zufallsvariablenvektor aus einer Zufallsstichprobe  $x_n$  gilt:

$$x_n \xrightarrow{d} N(\mu, \Sigma)$$

Und wenn  $Q_n$  eine Zufallsvariablen-Matrix aus derselben Stichprobe ist und mit  $x_n$  multiplizierbar ist und

$$p \lim Q_n = Q$$

dann folgt

$$Q_n x_n \xrightarrow{d} N(Q\mu, Q\Sigma Q')$$

Theorem 11: Konsistenz des OLS- Schätzers

Der OLS-Schätzer ist konsistent, i.e.  $\text{plim } b_{ols} = \beta$ .

Beweis:

$$\begin{aligned}
 (3a) \quad p \lim b_{ols} &= p \lim (X'X)^{-1} X'y \\
 &= p \lim (X'X)^{-1} X'(X\beta + \varepsilon) \\
 &= p \lim \beta + p \lim (X'X)^{-1} X'\varepsilon \quad (\text{Slutsky - Theorem}) \\
 &= \beta + p \lim \left( \frac{X'X}{n} \right)^{-1} p \lim \frac{X'\varepsilon}{n} \quad (\text{Slutsky - Theorem})
 \end{aligned}$$

Zu zeigen:

- a)  $p \lim (X'X)^{-1} n < \infty$       und
- b)  $p \lim \frac{1}{n} X'\varepsilon = 0$

Ad a) Beweisidee: Anwendung von Theorem 7. Hierzu ist zunächst zu zeigen, daß  $(X'X)^{-1}n$  ein Stichprobenmittelwert zu einer Zufallsmatrix ist. Zunächst schreiben wir den Ausdruck  $(X'X)$  aus:

$$\begin{matrix}
 X'X = \begin{bmatrix} x_{11} & \dots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1k} & \dots & x_{nk} \end{bmatrix} \begin{bmatrix} x_{11} & \dots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{n1} & \dots & x_{nk} \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 & \dots & \sum_{i=1}^n x_{i1}x_{ik} \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ik}x_{i1} & \dots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix} \\
 \text{kxn} \qquad \qquad \text{nxk} \qquad \qquad \qquad \qquad \qquad \qquad \text{kxk}
 \end{matrix}$$

Wenn wir einen Vektor  $x_i = (x_{i1}, \dots, x_{ik})'$  definieren, dann gilt

$$\begin{matrix}
 \frac{1}{n} X'X = \frac{1}{n} \sum_{i=1}^n x_i x_i' \\
 \text{kxk} \qquad \qquad \text{kx1} \quad \text{1xk}
 \end{matrix}$$

Damit ist also  $(X'X)^{-1}n$  der Stichprobenmittelwert zu  $x_i x_i'$ . Weiterhin gilt

$$\text{Var } [x_i] = \Sigma_{xx} \quad \text{per Annahme 5b)}$$

Aber per definitionem einer Kovarianzmatrix gilt auch:

$$\begin{matrix}
 \text{Var } [x_i] = E[x_i x_i'] - E[x_i] E[x_i'] & (\text{Multivariate Form von } \text{Var } [X] = E[X^2] - E[X]^2) \\
 \text{kxk} \qquad \qquad \underbrace{\text{kxk}}_{\mu_x \mu_x'} & \text{per Annahme 5b)}
 \end{matrix}$$

Nach Umformung erhalten wir dann:

$$(4) \quad \Rightarrow \quad E[x_i x_i'] = \Sigma_{xx} + \mu_x \mu_x' < \infty \quad (\text{Ungleichheit per Annahme 5b)}$$

Ergo gilt nach Theorem 1 und der Annahme, dass  $x_i$  i.i.d. sind, dass

$$E\left[\frac{1}{n} \sum_{i=1}^n x_i x_i'\right] = \Sigma_{XX} + \mu_X \mu_X' < \infty.$$

Desweiteren gilt:

$$(5) \quad \text{Var}[x_i x_i'] = \underbrace{E[x_i x_i' x_i x_i'] - E[x_i x_i'] E[(x_i x_i)']}_{\Phi} \quad \begin{matrix} \text{kxk} & \text{kxk} \\ & E[x_i x_i'] \\ & \text{kxk} \end{matrix}$$

per Annahme 5b

wobei  $\Phi$  die Matrix der 4. Momente von  $X$  mit Element  $\phi_{ijkl}$  ist.

$$= \Phi - [\Sigma_{XX} + \mu_X \mu_X'] [\Sigma_{XX} + \mu_X \mu_X']' < \infty$$

Wegen (4) und (5) sind die Bedingungen von Theorem 7 erfüllt und es gilt:

(6)

$$p \lim \frac{1}{n} X' X = p \lim \frac{1}{n} \sum_{i=1}^n x_i x_i' = E[x_i x_i'] = \Sigma_{XX} + \mu_X \mu_X' < \infty$$

$$p \lim \left( \frac{1}{n} X' X \right)^{-1} = (\Sigma_{XX} + \mu_X \mu_X')^{-1} < \infty,$$

wobei der letzte Schritt sich aus dem Slutsky Theorem (Theorem 5) ergibt.

Ad b) Analog erfolgt die Beweisführung zum zweiten Teil. Zunächst haben wir

$$(7) \quad X' \varepsilon = \begin{matrix} \begin{bmatrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1k} & \cdots & x_{nk} \end{bmatrix} & \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \\ \text{kxn} & \text{nx1} \end{matrix} = \begin{matrix} \begin{bmatrix} \sum_{i=1}^n x_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=1}^n x_{ik} \varepsilon_i \end{bmatrix} \\ \text{kx1} \end{matrix} = \sum_{i=1}^n x_i' \varepsilon_i \quad \text{kx1}$$

Damit ist  $n^{-1} X' \varepsilon$  der Stichprobenmittelwert des Zufallsvektors  $x_i \varepsilon_i$ . Es gilt nach Definition der Kovarianz:

$$\text{Cov}(x_i, \varepsilon_i) = E[x_i \varepsilon_i] - E[x_i] E[\varepsilon_i]$$

bzw.

$$(8) \quad E[x_i \varepsilon_i] = \text{Cov}(x_i, E[\varepsilon_i | x_i]) + \mu_X E_{X'}[E[\varepsilon_i | x_i]] = 0 \\ = \text{Cov}(x_i, 0) + \mu_X \cdot 0 = 0,$$

wobei beim ersten Gleichheitszeichen Theoreme 8 und 9 und beim zweiten Gleichheitszeichen Annahme 3 zur Anwendung kommen. Nach Definition einer Kovarianzmatrix finden wir:

$$\begin{aligned} \text{Var}[x_i \varepsilon_i] &= E[x_i \varepsilon_i \varepsilon_i' x_i'] - \underbrace{E[x_i \varepsilon_i] E[x_i' \varepsilon_i]}'_{=0 \text{ nach (8)}} \\ &= E[\varepsilon_i^2 x_i x_i'] \\ &= E_X[E[\varepsilon_i^2 x_i x_i' | x_i]] \\ &= E_X[x_i x_i' E[\varepsilon_i^2 | x_i]] \quad \text{nach law of iterated expectations} \end{aligned}$$

$$\begin{aligned}
&= \tau^2 E[x_i x_i'] \\
(9) \quad &= \tau^2 [\Sigma_{xx} + \mu_X \mu_X'] < \infty \quad \text{wegen Annahme (4)}
\end{aligned}$$

Wegen (8) und (9) sind die Annahmen für Theorem 7 erfüllt:

$$(10) \quad p \lim \frac{1}{n} \sum_{i=1}^n x_i \varepsilon_i = p \lim \frac{X' \varepsilon}{n} = E[x_i \varepsilon_i] = 0, \quad \text{was Aussage b) beweist}$$

q.e.d.

Theorem 12: Asymptotische Normalverteilung des OLS- Schätzers bei großer Stichprobe und stochastischen Regressoren

$$b_{ols} \stackrel{a}{\sim} N \left[ \beta, \frac{\tau^2}{n} [\Sigma_{xx} + \mu_X \mu_X'] \right]$$

Sketch eines Beweises:

$$\sqrt{n}(b_{ols} - \beta) = \left( \frac{X'X}{n} \right)^{-1} \left( \frac{1}{\sqrt{n}} \right) X' \varepsilon$$

$$X' \varepsilon = \sum_{i=1}^n x_i \varepsilon_i$$

Deshalb ist  $\frac{X' \varepsilon}{n}$  der Stichprobenmittelwert zu  $x_i \varepsilon_i$

(8) hat bewiesen:  $E[x_i \varepsilon_i] = 0$  und

(9) hat bewiesen:  $Var[x_i \varepsilon_i] = \tau^2 [\Sigma_{xx} + \mu_X \mu_X'] < \infty$

Da  $x_i \varepsilon_i$ ,  $i=1, \dots, n$ , die Voraussetzung einer Zufallsstichprobe erfüllt per Annahme 5 und (8) und (9) gelten, sind die Bedingungen für den zentralen Grenzwertsatz von Lindberg-Levy (Theorem 6) erfüllt und es gilt:

$$(11) \quad \sqrt{n} \left( \frac{X' \varepsilon}{n} \right) \xrightarrow{d} N[0, \tau^2 [\Sigma_{xx} + \mu_X \mu_X']]$$

Bedingungen (11) und (6),  $p \lim \frac{X'X}{n} = \Sigma_{xx} + \mu_X \mu_X'$ , erlauben die Anwendung von Cramer's Theorem (Theorem 10):

$$\left( \frac{X'X}{n} \right)^{-1} \left( \frac{1}{\sqrt{n}} \right) X' \varepsilon \xrightarrow{d} N \left[ 0, [\Sigma_{xx} + \mu_X \mu_X']^{-1} [\Sigma_{xx} + \mu_X \mu_X'] [\Sigma_{xx} + \mu_X \mu_X']^{-1} \tau^2 \right]$$

Und deshalb gilt:

$$\sqrt{n}(b_{ols} - \beta) \xrightarrow{d} N \left[ 0, [\Sigma_{xx} + \mu_X \mu_X']^{-1} \tau^2 \right] \quad \text{q.e.d}$$

Asymptotische Kovarianz Matrix des OLS-Schätzers:

Die Varianz der Fehlervariablen in der Grundgesamtheit ist nicht bekannt. Dann kann aber

stattdessen die Stichprobenvarianz  $S^2 \equiv \frac{1}{n-k} \sum_{i=1}^n e_i^2 = \frac{e'e}{n-k}$  verwendet werden und es gilt:

$$\sqrt{n}(b_{ols} - \beta) \xrightarrow{d} N(0, (X'X)^{-1} S^2)$$

Beweis:

Zunächst zeigen wir:  $p \lim S^2 = \tau^2$

$$e = My = M[X\beta + \varepsilon] = M\varepsilon$$

wobei

$$M = I - X(X'X)^{-1}X' \quad \text{die "Residualmaker-Matrix"}$$

also

$$My = y - X(X'X)^{-1}X'y = y - Xb_{ols} = e$$

und

$$MX = X - X(X'X)^{-1}X'X = X - X = 0$$

Deshalb:

$$S^2 = \frac{e'e}{n-k} = \frac{(M\varepsilon)'(M\varepsilon)}{n-k} = \frac{\varepsilon'M'M\varepsilon}{n-k} = \frac{\varepsilon'M\varepsilon}{n-k},$$

Weil M idempotent ist ( $M'M=M$ ) (siehe Kapitel über „partitioned regressions“)

$$S^2 = \frac{\varepsilon'(I - X(X'X)^{-1}X')\varepsilon}{n-k} = \frac{\varepsilon'(I\varepsilon - X(X'X)^{-1}X'\varepsilon)}{n-k}$$

$$= \left[ \frac{\varepsilon'\varepsilon}{n} - \left( \frac{\varepsilon'X}{n} \right) \left( \frac{X'X}{n} \right)^{-1} \left( \frac{X'\varepsilon}{n} \right) \right] \frac{n}{n-k}$$

$$p \lim S^2 = \left[ p \lim \frac{\varepsilon'\varepsilon}{n} - p \lim \left( \frac{\varepsilon'X}{n} \right) p \lim \left( \frac{X'X}{n} \right)^{-1} p \lim \left( \frac{X'\varepsilon}{n} \right) \right] p \lim \frac{n}{n-k}$$

$$\begin{array}{ccccccc} \text{nach Slutsky Theorem} & & =0 & & <\infty & & =0 & & =1 \\ & & \text{nach (10)} & & \text{nach (6)} & & \text{nach (10)} & & \end{array}$$

Also:

$$p \lim S^2 = p \lim \frac{\varepsilon'\varepsilon}{n} = p \lim \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$$

Wir haben die folgenden Eigenschaften:

- a)  $E[\varepsilon_i^2] = \tau^2$
- b)  $\text{Var}[\varepsilon_i^2] = E[\varepsilon_i^4] - E[\varepsilon_i^2]^2 = \Phi_\varepsilon - \tau^4 < \infty$  nach Annahme 6b
- c)  $\varepsilon_i^2$  ist Zufallsstichprobe (keine Autokorrelation)
- d)  $\frac{\varepsilon'\varepsilon}{n} = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2$ , ist Stichprobenmittelwert von  $\varepsilon_i^2$

Wegen a)-d) folgt nach Theorem 7:

$$(12) \quad p \lim S^2 = p \lim \frac{\varepsilon' \varepsilon}{n} = E[\varepsilon_i^2] = \tau^2$$

Nach dem Slutsky- Theorem folgt dann auch:

$$(13) \quad p \lim \frac{S^2}{\tau^2} = 1 \quad \text{aus (12)}$$

und

$$(14) \quad p \lim \frac{\left(\frac{1}{n} X' X\right)^{-1}}{Q^{-1}} = 1 \quad \text{aus (6)}$$

mit  $Q = \sum x x' + \mu_x \mu_x'$

Nach Theorem 12:

$$\sqrt{n} \frac{b_{kols} - \beta_k}{Q_{kk}^{-1} \tau^2} \xrightarrow{d} N(0,1), \quad \text{wobei k der Index für die Variable k ist.}$$

Wegen (13) und (14) und Cramer's Theorem (Theorem 10) gilt, dass aus der Verteilung

$$\sqrt{n} \frac{(b_{kols} - \beta_k)}{\frac{1}{n} (X' X)_{kk}^{-1} S^2} \left( \frac{\left(\frac{1}{n} X' X\right)^{-1}}{Q_{kk}^{-1}} \right) \left( \frac{S^2}{\tau^2} \right) \xrightarrow{d} N(0,1)$$

die Verteilung

$$\boxed{\sqrt{n} \frac{(b_{kols} - \beta_k)}{\frac{1}{n} (X' X)_{kk}^{-1} S^2} \xrightarrow{d} N(0,1)} \quad \text{folgt.} \quad \text{q.e.d.}$$

### Bemerkung 1:

Man kann aus der obigen Verteilung eine asymptotische Verteilung für  $b_{kols}$  folgern:

$$b_{kols} \xrightarrow{a} N(\beta_k, (X' X)_{kk}^{-1} S^2).$$

Hierbei nennt man  $(X' X)_{kk}^{-1} S^2$  die geschätzte asymptotische Kovarianzmatrix,

$$Est.Asy.Var. = (X' X)_{kk}^{-1} S^2$$

Durch die Anwendung des Cramer-Theorems reicht es also stets aus, einen konsistenten Schätzer für die Kovarianzmatrix in der Grundgesamtheit zu finden, sobald man weiß, daß ein Schätzer asymptotisch normalverteilt ist. Für jeden neuen Schätzer werden wir also diese drei Schritte wiederholen: a) zeigen, daß der Schätzer konsistent ist b) zeigen, daß der Schätzer

asymptotisch normalverteilt ist, c) einen konsistenten Schätzer für die Kovarianzmatrix des Schätzers finden. Aus diesen Informationen zusammen kann dann eine t-Test Statistik formuliert werden, die verwendet werden kann, um zu testen, ob der Schätzwert signifikant (also z. Bsp. Mit 95% Wahrscheinlichkeit) von Null verschieden ist.

**Wichtig:** Die geschätzte asymptotische Kovarianzmatrix, die für die Test-Statistik zum Test der Signifikanz des Regressionskoeffizienten verwendet wird, ist nicht gleich der Varianz des Wahrscheinlichkeitslimits des Schätzers! Letztere Varianz ist ja null per definitionem! Stattdessen ist die geschätzte asymptotische Kovarianzmatrix das Wahrscheinlichkeitslimit des Schätzers der wahren Varianz! Man geht also davon aus, daß die Stichprobe einerseits so groß ist, daß der konsistente Schätzer der Varianz hinreichend nahe an dem wahren Wert liegt, und andererseits nicht so groß, daß die Varianz praktisch Null wird.

#### Bemerkung 2:

In der großen Stichprobe sind die sogenannten „t-Werte“ also standard- normalverteilt. In der kleinen Stichprobe mit normalverteilter Fehlervariable sind die t-Werte dagegen student-(t-) verteilt.

#### Bemerkung 3:

Es kann gezeigt werden, daß der OLS-Schätzer auch in der großen Stichprobe (asymptotisch) effizient ist. (Siehe Greene, 2000, s. 387ff.)

## GLS – Schätzer

Erweiterung des Grundmodells:

$$(GLS\ 4) \quad E[\varepsilon\varepsilon' | X] = \tau^2 \Omega$$

Ersetzt nun Annahme 4 des klassischen Regressionsmodells, i.e.  $E[\varepsilon\varepsilon' | X] = \tau^2 I_n$

Die Matrix  $\Omega$  sei zunächst als bekannt angenommen. Es gelte weiter  $E[\varepsilon | X] = 0$ .

Wir definieren eine Matrix  $P$ , so dass  $P'P = \Omega^{-1}$ . Die Matrix  $P$  ist damit eindeutig bestimmt. Dann wird das Regressionsmodell  $y = X\beta + \varepsilon$  mit dem wahren Schätzer  $\beta$  multipliziert mit Matrix  $P$ :

$$Py = PX\beta + P\varepsilon$$

bzw. wenn  $\varepsilon$

$$\text{bzw. wenn } y^* \equiv Py; \quad X^* \equiv PX \quad \text{und } \varepsilon^* \equiv P\varepsilon$$

$$(1) \quad y^* = X^*\beta + \varepsilon^*$$

Dann gilt wegen

$$E[\varepsilon^*] = E[P\varepsilon] = PE[\varepsilon] = 0, \quad \varepsilon^* \text{ i.i.d.}$$

$$\text{Var}[\varepsilon^*] = E[\varepsilon^* \varepsilon^{*'}] = E[P\varepsilon(P\varepsilon)'] = E[P\varepsilon\varepsilon'P'] = PE[\varepsilon\varepsilon']P' = P\Omega P'\tau^2 = P(P'P)^{-1}P'\tau^2 = I\tau^2$$

Damit erfüllt (1) die klassischen Annahmen und der Schätzer

$$\begin{aligned} b_{\text{GLS}} &\equiv (X^{*'}X^*)^{-1}X^{*'}y^* \\ &= ((PX)'(PX))^{-1}(PX)'(Py) = ((X'P'PX)^{-1}(X'P'Py)) \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y \end{aligned}$$

Dies ist der GLS- oder Aitken- Schätzer.

Da (1) die klassischen Annahmen erfüllt, muss

i)  $b_{\text{GLS}}$  konsistent sein, wenn

$$p \lim \left( \frac{1}{n} X^{*'} X^* \right) = p \lim \left( \frac{1}{n} X' P' P X \right) = p \lim \left( \frac{1}{n} X' \Omega^{-1} X \right) < \infty$$

ii)  $b_{\text{GLS}}$  in der großen Stichprobe asymptotisch normalverteilt sein:

$$b_{\text{GLS}} \stackrel{\text{asy}}{\sim} N\left(\beta, \tau^2 [(X^*)'(X^*)]^{-1}\right)$$

$$\text{wobei } [(X^*)'X^*]^{-1} = [(PX)'PX]^{-1} = [X'P'PX]^{-1} = [X'\Omega^{-1}X]^{-1}$$

iii)  $b_{\text{GLS}}$  asymptotisch effizient sein.

Bemerkung: Auch die Eigenschaften des OLS-Schätzers für kleine Stichproben gelten für  $b_{\text{GLS}}$ .

Caveat:  $R^2$  ist bei der GLS-Schätzung bedeutungslos, da es negativ werden kann oder  $> 1$ .



$$b_{FGLS}^{asy} \sim b_{GLS} \sim N(\beta, Q^{-1}\tau^2)$$

Bemerkung:

$\hat{\theta}$  muss lediglich konsistent sein und nicht effizient, damit  $b_{FGLS}$  selbst effizient ist.

### Heteroskedastizität:

Heteroskedastizität liegt vor, wenn  $\text{Var}[\varepsilon_i | X_i] = \tau_i^2, i=1, \dots, n$ , wobei die Störterme untereinander weiter nicht korreliert sind.

Also:

$$E[\varepsilon\varepsilon' | X] = \tau^2 \Omega = \begin{bmatrix} \tau_1^2 & & 0 \\ & \ddots & \\ 0 & & \tau_n^2 \end{bmatrix} = \tau^2 \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{bmatrix}$$

Mit der Normierung  $\sum_{i=1}^n w_i = n$

Welche Eigenschaften hat der OLS- Schätzer, wenn Heteroskedastizität vorliegt?

i) Konsistenz:

$$b_{ols} = \beta + (X'X)^{-1} X' \varepsilon = \beta + \frac{\left( \sum_{i=1}^n x_i x_i' \right)^{-1}}{n} \frac{\left( \sum_{i=1}^n x_i \varepsilon_i \right)}{n}$$

Konsistenz ist gesichert, wenn nach dem Chebychev Theorem (Theorem 4)

$$1) \lim_{n \rightarrow \infty} E[b_{ols}] = \beta \quad \text{und} \quad 2) \lim_{n \rightarrow \infty} \text{Var}(b_{ols}) = 0$$

gezeigt werden kann.

$$\begin{aligned} \text{Ad (1a): } E[b_{ols}] &= E_X[E(b_{ols} | X)] && | \text{Theorem 8} \\ &= \beta + E_X[E[(X'X)^{-1} X' \varepsilon | X]] = \beta + E_X[(X'X)^{-1} X' E[\varepsilon | X]] = 0 \end{aligned}$$

$\underbrace{\hspace{10em}}_{=0 \text{ per Annahme 3}}$

(OLS-Schätzer ist unverzerrt bei Heteroskedastizität)

$$\begin{aligned} \text{Ad (2a) } \text{Var}[b_{ols} | X] &= E[(b_{ols} - \beta)(b_{ols} - \beta)' | X] \\ &= E[(X'X)^{-1} X' \varepsilon (X'X)^{-1} X' \varepsilon' | X] \\ &= E[(X'X)^{-1} X' \varepsilon \varepsilon' X (X'X)^{-1} | X] \\ &= (X'X)^{-1} X' \tau^2 \Omega X (X'X)^{-1} \end{aligned}$$

$$= \frac{\tau^2}{n} \left( \frac{X'X}{n} \right)^{-1} \left( \frac{X'\Omega X}{n} \right) \left( \frac{X'X}{n} \right)^{-1}$$

$$\lim_{n \rightarrow \infty} \text{Var}[b_{ols}] = \lim_{n \rightarrow \infty} E[\text{Var}[b_{ols}|X]] = \lim_{n \rightarrow \infty} E \left[ \left( \frac{X'X}{n} \right)^{-1} \left( \frac{X'\Omega X}{n} \right) \left( \frac{X'X}{n} \right)^{-1} \right] \frac{\tau^2}{n} = 0,$$

was wahr ist, wenn (a)  $p \lim \frac{1}{n} X'X = E[x_i x_i'] = \frac{\sum_i E[x_i x_i']}{n} = E \left[ \frac{X'X}{n} \right] < \infty$

und (b)  $p \lim \frac{1}{n} X'\Omega X < \infty$ ,

weil dann  $\lim_{n \rightarrow \infty} \frac{\tau^2}{n} = 0$

(a) ist bereits demonstriert worden beim Beweis der Konsistenz des OLS-Schätzers. (b) muss noch im Fall Heteroskedastizität gezeigt werden:

$$\begin{aligned} X'\Omega X &= \begin{bmatrix} x_{11} & \cdots & x_{n1} \\ \vdots & \ddots & \vdots \\ x_{1k} & \cdots & x_{nk} \end{bmatrix} \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{bmatrix} \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{1k} & \cdots & x_{nk} \end{bmatrix} \\ &= \begin{bmatrix} x_{11}w_1 & \cdots & x_{n1}w_n \\ \vdots & & \vdots \\ x_{1k}w_1 & \cdots & x_{nk}w_n \end{bmatrix} \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & \ddots & \vdots \\ x_{1k} & \cdots & x_{nk} \end{bmatrix} \\ &= \begin{bmatrix} \sum_{i=1}^n x_{i1}^2 w_i & \cdots & \sum_{i=1}^n x_{i1} x_{ik} w_i \\ \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ik} x_{i1} w_i & \cdots & \sum_{i=1}^n x_{ik}^2 w_i \end{bmatrix} \end{aligned}$$

Da  $w_1 < \infty$  per Annahme und  $x_{i1} x_{jm} < \infty$  wegen (a) muss gelten:

$$p \lim \frac{1}{n} X'\Omega X < \infty$$

deswegen ist der OLS-Schätzer bei Heteroskedastizität konsistent!

Anmerkung:

Dies gilt nicht für jede  $\Omega$ -Matrix. Gegenbeispiel:

$$\Omega = \begin{bmatrix} 1 & p & \cdots & p \\ p & \ddots & & \vdots \\ \vdots & & \ddots & p \\ p & \cdots & p & 1 \end{bmatrix} \quad \text{und Regression } y = \beta + \varepsilon$$

Der OLS-Schätzer ist dann der Mittelwert von  $y$ , i.e.  $\bar{y}$ , die X-Matrix ist nun reduziert zu einem  $n \times 1$  Vektor, bei dem alle Elemente den Wert 1 haben (Regressionskonstante). Dann erhalten wir aus der Gleichung in 2a)

$$\text{Var}[\bar{y}] = \frac{\tau^2}{n}(1 - p + np)$$

$$\lim_{n \rightarrow \infty} \text{Var}[\bar{y}] = \tau^2 p > 0 \quad \text{und damit ist } b \text{ nicht konsistent.}$$

Ist bei Heteroskedastizität  $b_{ols}$  noch asymptotisch normalverteilt? Was ist die Kovarianzmatrix?

Lindberg- Feller Central Limit Theorem:

Angenommen  $x_1, \dots, x_n$  ist ein Zufallsvariablen- Vektor aus einer Zufallsauswahl, so dass  $E[x_i] = \mu_i$ ,  $\text{Var}[x_i] = Q_i$  und alle dritten gemischten Momente sind endlich.

Man definiere:

$$\bar{\mu}_n = \frac{1}{n} \sum_{i=1}^n \mu_i \quad \text{und} \quad \bar{Q}_n = \frac{1}{n} \sum_{i=1}^n Q_i$$

und nehme an:

a)  $\lim_{n \rightarrow \infty} \bar{Q}_n = Q$  mit einer endlichen positiv definiten Matrix  $Q$  und

b)  $\lim_{n \rightarrow \infty} (\bar{Q}_n)^{-1} Q_i = \lim_{n \rightarrow \infty} \left( \sum_{i=1}^n Q_i \right)^{-1} Q_i = 0$  einzelne  $Q_i$  -Matrix darf  $Q$  nicht dominieren,

dann  $\sqrt{n}(\bar{X}_n - \bar{\mu}_n) \xrightarrow{d} N(0, Q)$

Nun kommen wir zu der asymptotischen Verteilung von  $b_{ols}$  bei Heteroskedastizität

$$V \equiv \sqrt{n}(b_{ols} - \beta) = \left( \frac{X'X}{n} \right)^{-1} \left( \frac{X'\varepsilon}{\sqrt{n}} \right) \quad \text{normalverteilt?}$$

$$p \lim \frac{1}{n} X'X = Q < \infty \quad \text{bleibt weiter gültig}$$

$$V = Q^{-1} \frac{\sum_{i=1}^n x_i \varepsilon_i}{\sqrt{n}}$$

Wir betrachten:

$$\begin{aligned} (15) \quad & \frac{1}{n} \sum_{i=1}^n \text{Var}[x_i \varepsilon_i] = \frac{1}{n} \sum_{i=1}^n E[x_i \varepsilon_i \varepsilon_i' x_i'] = \frac{1}{n} \sum_{i=1}^n E[x_i x_i' \varepsilon_i^2] = \frac{1}{n} \sum_{i=1}^n E_X [x_i x_i' E[\varepsilon_i^2 | x_i]] \\ & = \frac{1}{n} \sum_{i=1}^n E[x_i x_i'] \tau_i^2 = \frac{\tau^2}{n} \sum_{i=1}^n E[x_i x_i'] w_i, \end{aligned}$$

wobei beim dritten Gleichheitszeichen das Gesetz von der iterativen Erwartung (Theorem 8) angewendet wird und beim letzten Gleichheitszeichen die Definition der Heteroskedastizität.

Wenn

$$(16) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \text{Var}[x_i \varepsilon_i] = \lim_{n \rightarrow \infty} \frac{\tau^2}{n} \sum_{i=1}^n E[x_i w_i x_i'] = \lim_{n \rightarrow \infty} \frac{\tau^2}{n} E[X' \Omega X] < \infty,$$

wobei die Ungleichheit bereits oben diskutiert wurde, und die anderen Annahmen des Lindberg- Feller Theorems gelten, dann kann man schreiben:

$$16a) \quad \frac{X' \varepsilon}{\sqrt{n}} \xrightarrow{d} N\left(0, p \lim \frac{\tau^2}{n} X' \Omega X\right)$$

Wenn wiederum das Cramer Theorem angewendet wird, gilt:

$$\sqrt{n}(b_{ols} - \beta) = N\left(0, Q^{-1} p \lim \left(\frac{\tau^2}{n} X' \Omega X\right) Q^{-1}\right)$$

Bemerkung:

Bei Heteroskedastizität ist  $b_{ols}$  zwar konsistent und normalverteilt, aber die Varianz  $(X'X)^{-1} \tau^2$  ist nicht konsistent. Damit ist der t-Test auf Signifikanz von  $b_{ols}$  (und alle anderen Tests) inkonsistent (und auch verzerrt).

Da  $p \lim \frac{\tau^2}{n} \left(\frac{X'X}{n}\right)^{-1} \left(\frac{X' \Omega X}{n}\right) \left(\frac{X'X}{n}\right)^{-1}$  nicht bekannt ist, muss ein Stichprobenschätzer gefunden werden. Es kann gezeigt werden, dass

$$(16b) \quad p \lim \frac{1}{n} \sum_{i=1}^n e_i^2 x_i x_i' = p \lim \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 x_i x_i',$$

wenn  $e_i$  der Fehlerterm irgendeines konsistenten Schätzers ist, und dass

$$p \lim \frac{1}{n} \sum_{i=1}^n e_i^2 x_i x_i' = p \lim \frac{1}{n} \sum_{i=1}^n \tau_i^2 x_i x_i'$$

nach einem Gesetz der großen Zahl. Daher ist bei Heteroskedastizität

$$\text{Asy.Var}[b_{ols}] = n(X'X)^{-1} \left(\sum_{i=1}^n x_i x_i' e_i^2\right) (X'X)^{-1}$$

die White-heteroskedastizitätskonsistente Kovarianz- Matrix.

Bemerkung:

Bei Abwesenheit von Heteroskedastizität nimmt die White-konsistente Kovarianz- Matrix dieselben Werte an, wie der OLS Schätzer der Kovarianz- Matrix  $\tau^2(X'X)^{-1}$ . Daher wird grundsätzlich der White-Schätzer angegeben.

Ausnahme: Die Ursache der Heteroskedastizität ist genauer bekannt:

$$\tau_i^2 = \tau^2 z_i \quad \text{oder} \quad \tau_i^2 = \tau^2 z_i^\alpha.$$

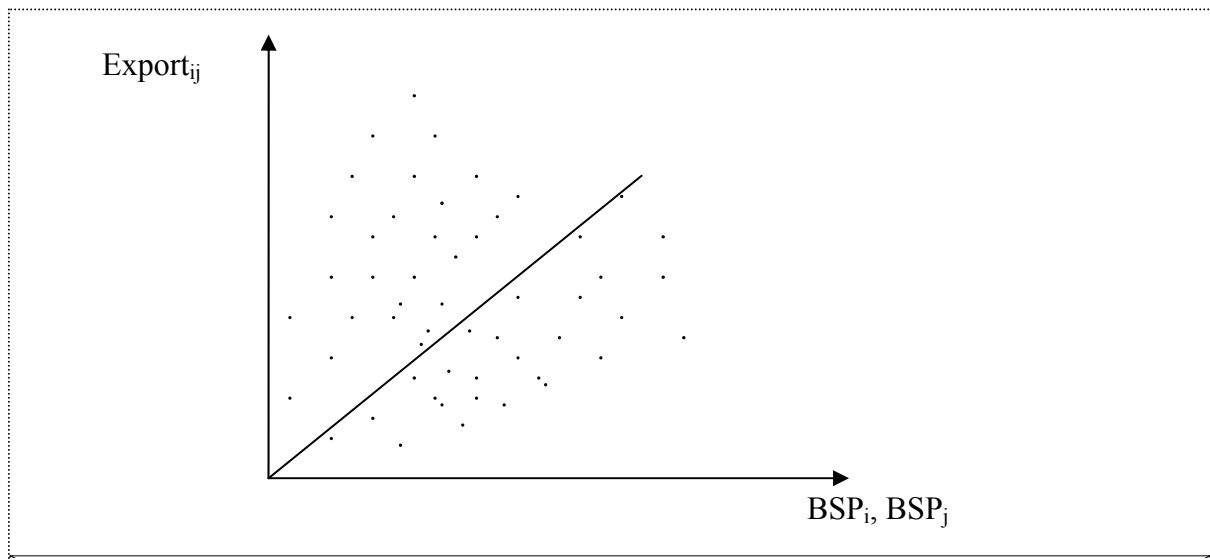
Beispiel:

Bilaterale Exporte<sub>ij</sub> =  $\beta_1 \text{BSP}_i + \beta_2 \text{BSP}_j + \beta_3 \text{Distanz}_{ij} + \varepsilon_{ij}$

i Exportland, j Importland,;

Dann  $\tau_{ij}^2 = \tau^2 \text{BSP}_i \text{BSP}_j$  wahrscheinlich.

⇒ Regressionen, die durch den Ursprung gehen



In einem solchen Fall wird ein Breusch- Pagan Test durchgeführt und eine FGLS- Schätzung, z.B. WLS- Schätzung, neben dem OLS- Schätzer berichtet, weil diese effizienter sind (siehe Greene, 2000, S. 521f), wenn die Form der Heteroskedastizität bekannt ist.

Kurz: WLS-Schätzung für den Fall  $\tau_i^2 = \tau^2 z_i^2$

Für GLS-Schätzer muss gelten:

$$P'P = \Omega^{-1} = \begin{bmatrix} z_1^{-2} & & 0 \\ & \ddots & \\ 0 & & z_n^{-2} \end{bmatrix} = \begin{bmatrix} z_1^{-1} & & 0 \\ & \ddots & \\ 0 & & z_n^{-1} \end{bmatrix} \begin{bmatrix} z_1^{-1} & & 0 \\ & \ddots & \\ 0 & & z_n^{-1} \end{bmatrix}$$

WLS- Schätzer ist dann gleich dem OLS- Schätzer von einer Regression

$y^* = Py$ ,  $X^* = PX$  und  $\varepsilon^* = P\varepsilon$  und damit der OLS- Schätzer von

$$\frac{y_i}{z_i} = \beta_1 \frac{1}{z_i} + \beta_2 \frac{x_{i2}}{z_i} + \dots + \beta_k \frac{x_{ik}}{z_i} + \frac{\varepsilon_i}{z_i}$$

### 3. IV- Schätzer (Instrumentenvariablen- Schätzer)

IV- Schätzer löst das Problem der Annahmenverletzung  $E[X_i \varepsilon_i] \neq 0$  (*Endogenitätsproblem*). Diese Annahmenverletzung hat dramatische Konsequenzen:  $b_{ols}$  ist verzerrt und inkonsistent. Dies folgt aus den jeweiligen Beweisen:

$$(3.1) \quad E[b - \beta | X] = (X'X)^{-1} E[X' \varepsilon | X] \neq 0 \quad (\text{Kapitel II1c1})$$

bzw.

$$(3.2) \quad p \lim [b_{ols} - \beta] = p \lim (X'X)^{-1} X' \varepsilon = [\Sigma_{XX} - \mu_X \mu_X']^{-1} E[X' \varepsilon] \neq 0,$$

weil nach (10) in Kapitel 1e, Beweis zu Theorem 11,  $p \lim \frac{X' \varepsilon}{n} = E[X' \varepsilon]$  gilt.

Wann treten Endogenitätsprobleme auf?

a) Simultaneität einer oder mehrerer X Variablen relativ zu y.

$$L^d = \beta_0 + \beta_1 W + X\beta + \varepsilon \quad \Longrightarrow \quad \text{Arbeitsnachfragekurve}$$

$$L^s = \gamma_0 + \gamma_1 W + Z\gamma + v \quad \Longrightarrow \quad \text{Arbeitsangebotskurve}$$

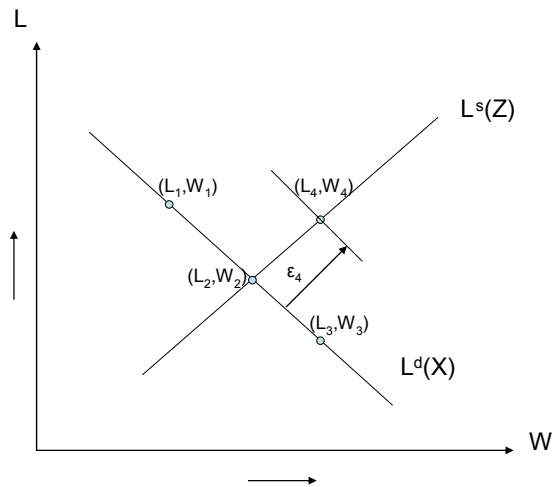
Im Gleichgewicht ist  $L = L^d = L^s$  und der Reallohn  $W$  ist endogen. Da die Beschäftigung  $L$  im Gleichgewicht bei unterschiedlichen  $X$  und  $Z$ -Werten beobachtet wird, wird  $b_{ols}$  von

$$(3.*) \quad L = b_0 + b_{ols}(W/P) + \dots + e$$

irgendwo zwischen  $\beta_0$  und  $\gamma_1$  liegen und damit verzerrt und inkonsistent sein.

Dies kann graphisch veranschaulicht werden wie folgt. Die Geraden  $L^d(X)$  und  $L^s(Z)$  seien die wahren Schätzgleichungen für Arbeitsangebot und Nachfrage in Abhängigkeit vom Reallohn. Die Beobachtungen  $(L_1, W_1)$ ,  $(L_2, W_2)$  und  $(L_3, W_3)$  liegen auf der  $L^d(X)$ -Kurven und der jeweilige Schätzfehler in der Grundgesamtheit ist  $\varepsilon_i = 0$  für  $i = 1..3$ . Dagegen ist der Schätzfehler  $\varepsilon_4 > 0$ . Es findet eine Verschiebung der  $L^d(X)$ -Kurve nach oben statt. Der neue Gleichgewichtspunkt impliziert aber, daß auch  $W_4$  gegenüber dem Ausgangspunkt  $W_3$  mit Fehler  $\varepsilon_3 = 0$  gestiegen ist. Damit liegt aber eine (positive) Korrelation zwischen der  $X$ -Variablen (also  $W$ ) und dem wahren Fehlerterm vor, i.e.  $\text{cov}(\varepsilon_i, W_i) > 0$ . Ursache der Korrelation ist die Simultaneität mit der die abhängige Variable in der Regressionsgleichung (3.\*) und die „unabhängige Variable“  $W$  gleichzeitig bestimmt werden. Mit anderen Worten die  $X$ -Variable verursacht die  $y$ -Variable, aber die  $y$ -Variable verursacht gleichzeitig auch die  $X$ -Variable. Damit die Schätzgleichung (3.\*) mit OLS konsistent geschätzt werden kann, muß die Arbeitsangebotskurve vertikal verlaufen. (In der Graphik würde dann die Korrelation zwischen  $\varepsilon_i$  und  $W$  verschwinden, weil  $W$  dann eine Konstante ist.) Dann ist aber der

Koeffizient  $\gamma_1=0$  und eine simultane Bestimmung von L und W liegt nicht mehr vor.



b) Messfehler einer oder mehrerer X-Variablen

Zur Illustration gebe es nur eine einzige X Variable und eine Konstante in der Regression. Das wahre Schätzmodell, das die OLS-Annahmen erfüllt, lautet:

$$(3.3) \quad y = \beta_0 + \beta_1 X + \varepsilon$$

Die Variable X werde mit einem Fehler gemessen, so dass

$$(3.4) \quad X^* = X - \mu, \quad \mu \sim N(0, \tau_\mu^2), \quad \text{cov}(X, \mu) = 0 \quad \text{cov}(\mu, \varepsilon) = 0$$

Damit gilt:

$$(3.5) \quad y = b_0 + b_1 X^* + b_1 \mu + \varepsilon = \beta_0 + \beta_1 X^* + W \quad \text{mit } W \equiv \varepsilon + b_1 \mu$$

Man beachte, dass

$$(3.5) \quad \text{cov}(X^*, W) = \text{cov}(X - \mu, \beta_1 \mu + \varepsilon) = -\beta_1 \tau_\mu^2 \neq 0.$$

Damit ist die Annahme  $\text{cov}(X^*, W) = 0$  verletzt und der Schätzer  $b_1$  des Stichprobenmodells

$$(3.6) \quad y = b_0 + b_1 X^* + e$$

ist inkonsistent.

Es kann sogar die Richtung der Verzerrung bestimmt werden. Mit Hilfe von Theorem 7 vom Stichprobenmittelwert erhalten wir

(3.7)

$$\begin{aligned} p \lim b_1 &= \frac{p \lim \sum_i x_i^* y_i}{p \lim \sum_i (x_i^*)^2} = \frac{p \lim \sum_i (x_i - \mu) [(x_i - \mu) \beta + \varepsilon]}{p \lim \sum_i (x_i - \mu)^2} = \frac{E[(x_i - \mu) [(x_i - \mu) \beta + \varepsilon]]}{E(x_i - \mu)^2} \\ &= \frac{\beta Q^*}{Q^* + \tau_\mu^2} = \frac{\beta}{1 + \left( \frac{\tau_\mu^2}{Q^*} \right)} \end{aligned}$$

mit  $Q^* = p \lim \frac{1}{n} \sum_{i=1}^n (x_i^*)^2 = E[x_i^2]$

⇒ je größer  $\tau_{\mu}^2$  (= Messfehlervarianz), umso mehr wird  $b_1$  gegen Null verzerrt (attenuation bias). Dieses Ergebnis kann auch für eine multivariate Regression mit einer fehlerhaft gemessenen erklärenden Variablen verallgemeinert werden.

⇒ Konservativer Fehler: wenn trotz Datenfehler  $b_1$  signifikant  $\neq 0$  ist, dann ist zwar die genaue Größe des Koeffizienten fraglich, aber der Effekt ist grundsätzlich gesichert. Umgekehrt kann ein nicht-signifikantes  $b_1$  in Wahrheit signifikant sein. (Randbemerkung: auch die Varianz des Schätzers ist größer mit als ohne Datenfehler.)

### c) Rationale Erwartungen

Beispiel: Investition = f(Nominalzins- erwartete Inflationsrate)  
= f( $i - \pi^e$ )

Da  $\pi^e$  nicht bekannt ist, wird  $E[\pi] = \pi + \varepsilon$  mit  $E[\varepsilon] = 0$  benutzt und statt  $\pi^e$  einfach  $\pi$  eingesetzt- ⇒ führt exakt zu den Problemen in b).

### d) Zeitverzögerung der abhängigen Variablen

Beispiel:  $y_t = \beta_0 + \beta_1 y_{t-1} + \varepsilon_t$  wobei t Zeitindex ist.

Auch hier gilt  $\text{cov}(y_{t-1}, \varepsilon_t) = \text{cov}(y_{t-1}, y_t - \beta_0 - \beta_1 y_{t-1}) \neq 0$

Mehr dazu im Kapitel über dynamische Paneldatenverfahren.

### e) Omitted variable bias

Das wahre Modell sei:

$$(3.8) \quad y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \varepsilon_i$$

Geschätzt werde aber

$$(3.9) \quad y_i = b_0 + b_1 x_i + u_i$$

Dann ist  $u_i = \beta_2 z_i + \varepsilon_i$  der wahre Fehlerterm des Schätzmodells und folglich  $\text{cov}(x_i, u_i) \neq 0$ , wenn  $\text{cov}(x_i, z_i) \neq 0$ .

### Lösung des Problems: Instrumenten- Variablen- Schätzer (IV)

Der Trick: Angenommen  $x_k$  ist endogen, aber es gebe eine Instrumentenvariable  $z_k$ , die in keinem ökonomischen Kausalitätsverhältnis zu  $x_k$  stehen muss, aber hoch korreliert mit  $x_k$  ist und weder notwendig zur Erklärung von  $y$  (exclusion criterium) ist noch mit dem Störterm  $\varepsilon$  korreliert ist. Man definiere  $z_i = [x_{i1} \dots x_{ik-1}, z_k]$

$$Z = \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix}$$

Formaler:  $E[Z'\varepsilon \mid X, Z] = 0$  wird angenommen.

Man pre-multipliziert die Regressionsgleichung  $y = X\beta + \varepsilon$  mit  $Z'$ :

$$(17) Z'y = Z'X\beta + Z'\varepsilon$$

$$(3.10) \quad b_{IV} = ((Z'X)'(Z'X))^{-1}(Z'X)'(Z'y) = (X'ZZ'X)^{-1}(X'ZZ'y) \\ = (Z'X)^{-1}(X'Z)^{-1}(X'Z)(Z'y) \\ = (Z'X)^{-1}Z'y$$

$$(3.11) \quad E[b_{IV} - \beta \mid XZ] = E[(Z'X)^{-1}Z'(X\beta + \varepsilon) \mid X, Z] - \beta = E[(Z'X)^{-1}Z'\varepsilon \mid X, Z] \\ = (Z'X)^{-1}E[Z'\varepsilon \mid X, Z] = 0$$

### 3a) Formal: Darstellung der IV- Schätzer- Annahmen

- Annahme 3 des klassischen Regressionsmodells wird ausgetauscht durch

$$(AI 3a) \quad E[\varepsilon_i \mid x_i] = \eta_i \quad \text{mit} \quad |\eta_i| < \infty;$$

$$(AI 3b) \quad E[\eta_i] = 0; \quad Var[\eta_i] = K^2 < \infty$$

Das impliziert zusammen mit Annahme 5:

$$Cov(\varepsilon_i, x_i) = \gamma = E[x_i, \varepsilon_i] \neq 0 \Rightarrow \text{Annahmenabschwächung}$$

- Eine 7. Annahme kommt zusätzlich hinzu:

(AI 7)  $[x_i; z_i; \varepsilon_i], i = 1 \dots n$ , ist eine i.i.d. Folge von Zufallsvariablen aus einer Zufallsauswahl

$$(AI 7a) \quad E[X^2_{ikm}] = Q_{xx,mm} < \infty \quad \forall m = 1 \dots k$$

$$(AI 7b) \quad E[z^2_{il}] = Q_{zz,ll} < \infty \quad \forall l = 1 \dots L$$

$L$  ist Dimension der IV – Variablen Matrix,  $L \geq k$  (Wir betrachten in diesem Unterabschnitt nur den Fall  $L=k$  der Einfachheit halber);

$$(AI 7c) \quad E[z_{il}x_{im}] = Q_{zx,lm} < \infty \quad \forall l = 1 \dots L, m = 1 \dots k$$

$$(AI 7d) \quad E[\varepsilon_i | z_i] = 0, \\ \text{was impliziert } E[z_i \varepsilon_i] = 0$$

### Konsistenz des IV- Schätzers

Wenn das Wahrscheinlichkeitslimit auf (17) angewendet wird, ergibt sich:

$$p \lim \frac{1}{n} Z' y = p \lim \left( \frac{Z' X}{n} \right) \cdot \beta + p \lim \left( \frac{Z' \varepsilon}{n} \right) \quad (18)$$

nach Anwendung des Slutsky- Theorems.

Nach Annahmen (AI 7a bis AI 7d) folgen:

$$(19) \quad p \lim \frac{1}{n} Z' X = Q_{ZX}$$

$$(20) \quad p \lim \frac{1}{n} Z' \varepsilon = 0, \quad (\text{was hier nicht gezeigt wird})$$

und daher kann (18) umgeformt werden unter Anwendung von (20)

$$(3.12) \quad \left( p \lim \left( \frac{1}{n} Z' X \right) \right)^{-1} \cdot p \lim \left( \frac{1}{n} Z' y \right) = \beta$$

Aber

$$(3.13) \quad p \lim b_{IV} = p \lim \left( \frac{1}{n} Z' X \right)^{-1} \cdot p \lim \left( \frac{1}{n} Z' y \right) = \beta$$

Der IV- Schätzer ist also konsistent.

### Asymptotische Normalverteilung des IV Schätzers

$$(3.13) \quad \sqrt{n}(b_{IV} - \beta) = \left( \frac{1}{n} Z' X \right)^{-1} \frac{Z' \varepsilon}{\sqrt{n}}$$

Es kann wie im Kapitel 1e, Theorem 12, gezeigt werden, dass

$$(21) \quad \frac{Z' \varepsilon}{\sqrt{n}} \xrightarrow{d} N(0, \tau^2 Q_{ZZ}), \text{ weil per Annahme AI7b) } p \lim \frac{1}{n} Z' Z = Q_{ZZ} \text{ gilt.}$$

Nach Anwendung des Cramer Theorems gilt weiter

$$(3.14) \quad \left( \frac{1}{n} Z' X \right)^{-1} \left( \frac{Z' \varepsilon}{\sqrt{n}} \right) \xrightarrow{d} N(0, \tau^2 Q_{ZX}^{-1} Q_{ZZ} Q_{XZ}^{-1})$$

unter Zuhilfenahme von (19), wobei  $Q_{XZ} \equiv Q_{ZX}'$

Entsprechend kann auch ein konsistenter Schätzer für die asymptotische Kovarianz- Matrix für den IV- Schätzer gefunden werden. (Wir erinnern uns dabei, daß nach dem Cramer- Theorem analog zu dem Beweis von Theorem 12 es hinreichend ist, daß die wahre Varianz des Schätzers ersetzt wird durch einen konsistenten Schätzer dieser Varianz, damit der t-Wert asymptotisch Normalverteilt ist.) Ein solcher Schätzer der Kovarianz ist die folgende geschätzte asymptotische Kovarianz-Matrix des IV-Schätzers:

$$(3.15) \quad Est.Asy.Var[b_{IV}] = \tilde{\tau}^2 (Z'X)^{-1} (Z'Z)(X'Z)^{-1}$$

$$\text{mit } \tilde{\tau}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - x_i b_{IV})^2$$

wobei (19) und AI7b)-d) zur Anwendung kommen.

---

### Praktische Berechnung des IV- Schätzers (Two – Stage Least Square Estimator)

#### 1. Stufe

Man berechne OLS- Regression von X auf Z:

$$(3.16) \quad X = Z\gamma_{OLS} + e_1 \quad \Rightarrow \gamma_{OLS} = (Z'Z)^{-1} Z'X$$

$$(3.17) \quad E[X | Z] = Z\gamma_{OLS} = Z(Z'Z)^{-1} Z'X \equiv \hat{X}$$

$\hat{X}$  wird als Projektion von X auf Z bezeichnet.

#### 2. Stufe

Man berechne OLS- Regression von y auf  $\hat{X}$

(3.18)

$$\begin{aligned} Y &= \hat{X}\hat{b}_{OLS} + e_2 \\ \hat{b}_{OLS} &= (\hat{X}'\hat{X})^{-1} \hat{X}'y \\ &= [(Z(Z'Z)^{-1}Z'X)'(Z(Z'Z)^{-1}Z'X)]^{-1} [Z(Z'Z)^{-1}Z'X]y \\ &= [X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X]^{-1} [X'Z(Z'Z)^{-1}Z'y] \\ &= [X'Z(Z'Z)^{-1}Z'X]^{-1} X'Z(Z'Z)^{-1}Z'y \\ &= (Z'X)^{-1} (Z'Z)(X'Z)^{-1} X'Z(Z'Z)^{-1}Z'y \\ &= (Z'X)^{-1} Z'y \\ &= b_{IV} \end{aligned}$$

In der Praxis wird dieses 2- stufige Schätzverfahren verwendet und auch in Tabellen aufgeführt. (Insbesondere wird auch die 1. Stufe berichtet). Allerdings müssen die Standardfehler korrigiert werden!

---

### 3b) Hausman-Test

#### **Wann ist nun der IV- Schätzer vorzuziehen?**

Wenn  $p \lim X'\varepsilon = 0$ , dann ist der IV- Schätzer nicht nur überflüssig, sondern ineffizient, was daraus folgt, dass der OLS- Schätzer effizient und die asymmetrischen Kovarianzmatrizen von OLS- und IV- Schätzer unterschiedlich sind. Daher ist ein Test notwendig, der bestimmt, ob der OLS- Schätzer inkonsistent und damit der IV Schätzer erforderlich ist

⇒ Hausman – Test

### Hausman- Test

$H_0$ :  $b_{OLS}$  und  $b_{IV}$  sind beide konsistent, aber  $b_{IV}$  ist ineffizient.

(3.19)

$$\Rightarrow d \equiv b_{IV} - b_{OLS} = 0$$

$H_a$ :  $b_{IV}$  ist konsistent, aber  $b_{OLS}$  ist es nicht

$$\Rightarrow d \equiv b_{IV} - b_{OLS} \neq 0$$

Test- Statistik:  $H = d' [\text{Est-Asy-Var}[d]]^{-1} d \Rightarrow$  (analog zu Wald- Test) (3.20)

$$(22) \quad \text{Asy. Var}[b_{IV} - b_{OLS}] = \text{Asy. Var}[b_{IV}] + \text{Asy. Var}[b_{OLS}] - 2 \text{Asy. Cov}[b_{IV} - b_{OLS}]$$

Hausman zeigt für jeden konsistenten und effizienten Schätzer  $b_E$  und jeden konsistenten, aber ineffizienten Schätzer  $b_I$  (also  $p \lim b_E = p \lim b_I$ ):

$$(23) \quad \text{Cov}[b_E, b_I] = \text{Var}[b_E]$$

Daher gilt mit (22) und (23) und  $b_{OLS}$  effizient und  $b_{IV}$  ineffizient

$$(3.21) \quad H = [b_{IV} - b_{OLS}]' [\text{Asy. Var}[b_{IV}] - \text{Asy. Var}[b_{OLS}]]^{-1} [b_{IV} - b_{OLS}]$$

Es kann gezeigt werden, dass  $H \xrightarrow{d} \chi^2[I]$ .

H ist also asymptotisch  $\chi^2$  - verteilt mit I Freiheitsgraden.

Bemerkung 1: Dieser Hausman- Test ist immer anwendbar, wenn die Bedingungen für  $H_0$  und  $H_a$  für beliebige Schätzer gelten ⇒ wird für Panel- Verfahren wichtig werden.

Bemerkung 2: Die Anzahl der Freiheitsgrade ist fallspezifisch.

Zunächst sei angedeutet, warum H in diesem Fall asymptotisch  $\chi^2$  - verteilt ist.

Theorem 13: Wenn  $X \sim N[\mu, \Sigma]$ , dann  $(X - \mu)' \Sigma^{-1} (X - \mu) \sim \chi^2[n]$

Es ist also zu zeigen, dass  $d = b_{IV} - b_{OLS}$  asymptotisch normalverteilt ist

$$\text{Mit } \hat{X} = Z(Z'Z)^{-1}Z'X \text{ und } \hat{X}'\hat{X} = X'Z(Z'Z)^{-1}Z'Z(Z'Z)^{-1}Z'X = X'Z(Z'Z)^{-1}Z'X'$$

$$= X'\hat{X} = \hat{X}X' \text{ ist}$$

(3.22)

$$d = (\hat{X}'\hat{X})^{-1}\hat{X}'y - (X'X)^{-1}X'y = (\hat{X}'\hat{X})^{-1}[\hat{X}'y - (\hat{X}'\hat{X})(X'X)^{-1}X'y]$$

$$= (\hat{X}'\hat{X})^{-1}[\hat{X}'y - \underbrace{(\hat{X}'X)(X'X)^{-1}X'y}_{b_{OLS}}] = (\hat{X}'\hat{X})^{-1}\hat{X}'[y - Xb_{OLS}] = (\hat{X}'\hat{X})^{-1}\hat{X}'e_{OLS}$$

damit wird offensichtlich, dass  $d$  asymptotisch normalverteilt mit Erwartungswert 0 ist, weil  $\hat{X}'e_{OLS}$  eine Summe aus der asymptotisch normalverteilten Zufallsvariablen  $e_{OLS}$  ist.

---

Nun bleibt noch zu klären, wie viele Freiheitsgrade  $I$  die  $\chi^2$ -Verteilung hat.

Nehmen wir an, dass  $\text{Cov}(x_k, \varepsilon) \neq 0 \quad \forall k = 1 \dots k_0$

Dann lautet die  $H_0$  auch nur:  $d_k = b_{IV_k} - b_{OLS_k} \quad \forall k = 1 \dots k_0$

Daher ist  $I$  in diesem Fall  $k_0$ .

---

Der Hausman-Test ist in STATA eine Standardprozedur.

Problem: die Varianz von  $b_{IV}$  kann in Stichprobe kleiner sein als die Varianz von  $b_{OLS}$

$\Rightarrow H$  wird negativ  $\Rightarrow$  keine  $\chi^2$ -Verteilung, da diese strikt positiv ist.

Daher ist eine alternative Vorgehensweise nach Wu (1973) üblich.

1. Schritt:  $X = Z\gamma_{OLS} + e_1$   
 $E[X^*\gamma | Z] \equiv \hat{X}^* = Z'(Z'Z)^{-1}Z'X^*$ ,  
wobei  $X^*$  die Variablen mit  $\text{Cov}(x_k, \varepsilon) \neq 0, \quad k=1 \dots k_0$

2. Schritt:  $y = X\beta_{IV} + \hat{X}^*\alpha + e_2$   
Man teste  $\alpha$  als F-Test mit  $k_0$  und  $n-k-k_0$  Freiheitsgraden.  
Dieser Test ist numerisch äquivalent zum Hausman-Test (Davidson and MacKinnon, 1993, S. 237ff)

---

### 3c) Weak-Instrument-Problem

Bleibt die Frage: Wo kommen die Instrumente  $Z$  her?

Im Zeitreihenzusammenhang oft Zeitverzögerungen, also z.B.  $x_{t-3}$  ist Instrument für  $x_t$ . Vereinfachend bei der Suche nach Instrumenten ist, dass sie in keinem ökonomischen (kausalen) Zusammenhang zu den instrumentierten  $x$ -Variablen stehen müssen.

In der Wachstumsforschung sind Variablen aus dem 19. Jahrhundert populär: z.Bsp. Dummy-Variablen über die Kolonialmacht eines Landes; oder geographische Variablen wie die Entfernung eines Landes zum Äquator oder ein Dummy dafür, ob ein Land einen Hafen hat.

Problem: Variablen, die exogen sind, sind meist auch kaum korreliert mit der endogenen Variablen. Dies führt zur Diskussion „schwacher Instrumente“. Selbst wenn eine minimale Korrelation zwischen Instrument und Störterm existiert kann der IV-Schätzer stärker verzerrt sein als der OLS-Schätzer. Außerdem hat der IV-Schätzer unvorteilhafte Eigenschaften in der

kleinen Stichprobe. Überraschenderweise kann „klein“ in diesem Kontext einige hunderttausend Beobachtungen bedeuten.

### 3c1) Grosse Varianz des Standardfehlers des IV-Schätzers

Der IV-Schätzer tendiert dazu, größere Standardfehler des Schätzers auszuweisen als der OLS-Schätzer. Damit werden häufig Zusammenhänge insignifikant, die mit OLS-Schätzern von Null signifikant verschieden sind. Dieses Problem wird um so größer, je geringer die Korrelation zwischen der x-Variablen und seinem Instrument z ist. Für den Fall einer x-Variablen und einer z-Variablen sieht man dies sofort, wenn man nochmals die geschätzte Varianz des IV-Schätzers aus (21) für diesen Fall aufschreibt:

(3.23)

$$Est.Asy.Var[b_{IV}] = \tilde{\tau}^2 \frac{\sum_i z_i^2}{\left[\sum_i z_i x_i\right]^2} = \frac{\tilde{\tau}^2}{\sum_i x_i^2} \frac{\left(\sum_i x_i^2\right)\left(\sum_i z_i^2\right)}{\left[\sum_i z_i x_i\right]^2} = \frac{\tilde{\tau}^2}{\sum_i x_i^2} \cdot \frac{1}{\rho_{zx}^2}$$

wobei  $\rho_{zx}$  der Korrelationskoeffizient zwischen der z- und x-Variablen ist. Wenn die Varianz des Störterms von IV-Schätzer und OLS-Schätzer gleich sind (was unter der  $H_0$ -Hypothese zutrifft, weil dann OLS und IV-Schätzer beide konsistent sind und damit konsistente Schätzer für die Varianz des Fehlerterms sowohl aus IV als auch aus OLS-Schätzungen gewonnen werden können und diese folglich asymptotisch identisch sein müssen), dann sieht man, daß der Term  $\frac{\tilde{\tau}^2}{\sum_i x_i^2}$  in der obigen Gleichung nichts anderes ist als ein Schätzer der Varianz des

OLS-Schätzkoeffizienten. Damit ist aber die Varianz des IV-Schätzers größer als die Varianz des OLS-Schätzers (mithin ist also auch das Konfidenz-Intervall um den IV-Schätzer größer). Wichtiger noch: Die Varianz des IV-Schätzers steigt mit geringerem Korrelationskoeffizienten  $\rho_{zx}$  zwischen Instrument und endogener x-Variable. Wenn die Korrelation z. Bsp. 0.1 beträgt, dann ist der IV-Standardfehler 10-mal größer als der OLS-Standardfehler!

Das obige Ergebnis kann verallgemeinert werden zu dem Fall **einer endogenen**  $x_k$ -Variablen mit **mehreren** Instrumenten  $z_k=[z_{k1} \dots z_{kL}]$  und **mehreren exogenen** X-Variablen  $x_1$  bis  $x_{k-1}$ . Dann ist der Standardfehler  $se[.]$  des IV-Schätzers gegeben durch (siehe Cameron und Trivedi, 2005, S. 107):

$$(3.24) \quad se[b_{IV}] = se[b_{OLS}] / R_p^2,$$

wobei  $R_p^2$  das Bestimmtheitsmaß ( $R^2$ ) der Regression

$$(3.25) \quad x_k^* = \beta_0 + \beta_1 z_k^* + \text{error}$$

und  $x_k^*$  und  $z_k^*$  sind analog zum Kapitel 1c1) über partitioned regression die Fehlerterme der Regression

$$(3.26) \quad x_k = a_{0OLS} + a_{1OLS} x_1 + \dots + a_{k-1OLS} x_{k-1} + x_k^*$$

bzw.

$$(3.27) \quad z_k = c_{0OLS} + c_{1OLS}x_1 + \dots + c_{k-1OLS}x_{k-1} + z_k^*,$$

wobei  $a_{iOLS}$  und  $c_{iOLS}$  OLS-Schätzer sind.

### 3c2) Rest-Korrelation der Instrumente mit dem Fehlerterm der 2. Schätzstufe

Es kann gezeigt werden (siehe Nelson und Startz, 1990), daß die Inkonsistenz des IV-Schätzer sehr schnell größer werden kann als die Inkonsistenz des OLS-Schätzers selbst, wenn nur eine minimale Rest-Korrelation zwischen Instrument und Fehlerterm der zweiten Schätzstufe besteht, also Annahme (AI 7d), i.e.  $E[\varepsilon_i|z_i]=0$ , nur minimal verletzt ist.

Wir beschränken uns der Einfachheit wieder auf eine Regression mit einer einzigen endogenen x-Variablen (ohne weitere exogene x-Variablen!) und einem einzigen Instrument z, wobei beide den Erwartungswert Null haben. Dann gilt nach Gleichung (3a) im Beweis zu Theorem 11, dass

$$(3.28) \quad p \lim b_{OLS} - \beta = p \lim (X'X)^{-1} X' \varepsilon = p \lim \frac{\sum_i x_i \varepsilon_i}{\sum_i x_i^2} = \frac{E[x_i \varepsilon_i]}{E[x_i^2]} = \frac{Cov(x_i, \varepsilon_i)}{Var(x_i)}$$

Für die Verzerrung des IV-Schätzers bei Verletzung von Annahme AI7d gilt dagegen

$$(3.29) \quad p \lim b_{IV} - \beta = p \lim (Z'X)^{-1} Z' \varepsilon = p \lim \frac{\sum_i z_i \varepsilon_i}{\sum_i z_i x_i} = \frac{E[z_i \varepsilon_i]}{E[z_i x_i]} = \frac{Cov(z_i, \varepsilon_i)}{Cov(x_i, z_i)}$$

Daraus erhalten wir das Verhältnis der Verzerrungen von IV- und OLS-Schätzer als:

$$(3.30) \quad p \lim \frac{b_{IV} - \beta}{b_{OLS} - \beta} = \frac{Cov(z_i, \varepsilon_i)}{\sqrt{Var(z_i) \cdot Var(\varepsilon_i)}} \cdot \frac{\sqrt{Var(x_i) \cdot Var(z_i)}}{Cov(x_i, z_i)} \cdot \frac{\sqrt{Var(x_i) \cdot Var(z_i)}}{Cov(x_i, \varepsilon_i)} \cdot \frac{Var(x_i)}{Var(x_i)}$$

$$= \frac{\rho_{z\varepsilon}}{\rho_{x\varepsilon}} \cdot \frac{1}{\rho_{zx}}$$

Wobei  $\rho_{qr}$  der Korrelationskoeffizient zwischen Variable q und r ist. Wenn  $\rho_{zx}$  gegen Null geht (also ein ganz schwaches Instrument im Sinne einer geringen Korrelation zwischen endogener Variablen und Instrument vorliegt), dann reicht eine Korrelation zwischen z und  $\varepsilon$  marginal größer als Null, damit der IV-Schätzer stärker verzerrt ist als der OLS-Schätzer. (Siehe Nelson und Starz, 1990)

Im verallgemeinerten Fall mit mehreren exogenen X-Variablen und einer endogenen  $x_k$  - Variablen gilt (Cameron und Trivedi, 2005, S. 107):

$$(3.31) \quad p \lim \frac{b_{IV} - \beta}{b_{OLS} - \beta} = \frac{\rho_{\hat{X}\varepsilon}}{\rho_{x\varepsilon}} \cdot \frac{1}{R_p^2}$$

wobei  $\hat{X}$  die Projektion der endogenen  $x_k$  -Variablen aus der ersten Schätzstufe und  $R_p^2$  wieder das in Kapitel 3d1 definierte Bestimmtheitsmaß ist. Es gilt also auch hier, daß bei geringer partieller Korrelation der Instrumente mit der endogenen Variablen, nachdem der

gemeinsame Einfluß der Instrumente und der exogenen Variablen auf die Korrelation mit der endogenen Variablen herausgerechnet worden ist, das Problem der Inkonsistenz des IV-Schätzers bei marginaler Rest-Korrelation zwischen Instrumenten und Fehlerterm verschärft wird.

### 3c3) Verzerrtheit des IV-Schätzers in der kleinen Stichprobe

Der IV-Schätzer ist grundsätzlich verzerrt. Warum dies so ist, kann sehr leicht gesehen werden:

$$(3.32) \quad E[b_{IV}] - \beta = E[(Z'X)^{-1}Z'\varepsilon] = E_{Z,X}[(Z'X)^{-1}Z'E[\varepsilon|X,Z]]$$

Der obige Ausdruck ist aber nur Null, wenn  $E[\varepsilon|X,Z]$  den Wert Null annimmt. Damit gilt nach dem Gesetz der iterativen Erwartung (Theorem 8) aber  $E[\varepsilon|X] = E_Z[E[\varepsilon|X,Z]] = 0$ . Aber wir betrachten ja gerade den Fall, daß  $E[\varepsilon|X] \neq 0$  (Annahme AI 3a). Im Umkehrschluß, muß also der IV-Schätzer verzerrt sein.

Die Verzerrung kann auch berechnet werden, wenn man eine multivariate Normalverteilung aller Zufallsvariablen annimmt. Es gebe nur eine endogene (und keine exogene) X-Variable, i.e.  $x_k$ , und  $k_0$  ( $k_0 > 1$ ) Instrumente. (Wenn nur ein einziges Instrument existiert, ist der Term der Verzerrung des IV-Schätzers nicht definiert.) Bound et al. (1995) zeigen, daß die Verzerrung des IV-Schätzers relativ zur Verzerrung des OLS-Schätzers (OLS-Schätzer ist verzerrt durch Korrelation von  $x_k$ -Variable mit Fehlerterm und IV-Schätzer ist grundsätzlich verzerrt) darstellbar ist bei großer Stichprobe und geringer Korrelation der Instrumente mit der  $x_k$ -Variablen durch die Funktion

$$(3.33) \quad \frac{E[b_{IV}] - \beta}{E[b_{OLS}] - \beta} \approx \frac{T^2}{k_0} {}_1F_1\left(1, \frac{k_0 + 2}{2}, -\frac{T^2}{k_0}\right) - 1$$

approximiert, wobei  ${}_1F_1(\dots)$  die konfluente hypergeometrische Funktion ist und

$$(3.34) \quad T^2 \equiv \frac{a'z_k'z_k a}{\sigma_v^2}$$

aus der zugrundeliegenden Schätzung der ersten Stufe

$$(3.35) \quad x_k = a_{k1OLS}z_{k1} + \dots + a_{kLOLS}z_{kL} + v = z_k a + v$$

mit Fehlerterm  $v$  und seiner Varianz  $\sigma_v^2$ .  $T^2$  ist hierbei der Wert der Grundgesamtheit, zu dem eine F-Test-Statistik auf gemeinsame Insignifikanz aller a-Koeffizienten der Instrumente  $z_k$  in der 1. Schätzstufe konvergiert. Je größer der  $T^2$ -Wert, um so kleiner ist die relative Verzerrung.

### 3c4) Test auf Vorliegen eines Weak-Instrument-Problems

Wir haben in Kapitel 3c3) gesehen, daß die Verzerrung des IV-Schätzers durch den Term in (3.33) approximiert werden kann. Diese Funktion hängt offenbar von zwei Faktoren ab: die

Anzahl der Instrumente  $k_0$  und den Term  $T^2$ , der selbst wiederum in der Stichprobe mit Hilfe eines F-Tests (konsistent aber leider nicht unverzerrt) geschätzt werden kann. Daraus haben Bound et al. (1995) eine Tabelle mit „kritischen Werten“ zusammengestellt, wobei die Kennzahlen der folgenden Tabelle die approximativen Werte von  $\frac{E[b_{IV}] - \beta}{E[b_{OLS}] - \beta}$  aus (3.33) darstellen, also die relative approximative Verzerrung des IV- zum OLS-Schätzer.

	$T^2=4$	$T^2=10$
$k_0=2$	0.02	0.00
$k_0=3$	0.09	0.03
$k_0=10$	0.17	0.08
$k_0=20$	0.19	0.08

Quelle: Bound et al. (1995), S. 450.

Als Konvention gilt zur Zeit, daß die relative Verzerrung des IV-Schätzers zum OLS-Schätzer nicht mehr als 0.1 betragen sollte. Daraus ergibt sich, daß eine Test-Statistik von  $T^2=10$  oder größer auf alle Fälle hinreichend ist.

Aus den zahlreichen Approximationen ergibt sich, daß eine eindeutige Testprozedur noch nicht existiert. Allerdings kommen Staiger und Stock (1997) zu einem ähnlichen Ergebnis für einen Zusammenhang von  $T^2 > 10$  mit einem völlig anderen Approximationsverfahren (ohne Normalverteilungsannahme).

Auch die beiden anderen Probleme der Weak-Instruments – große Varianz des Standardfehlers und Restkorrelation zwischen Instrumenten und Fehlerterm – stehen in Relation zu der Kennzahl  $T^2$ . Das Bestimmtheitsmaß  $R_p^2$ , daß unter den gegebenen Annahmen der Kapitel 3c1) und 3c2) eine Meßzahl für diese beiden Probleme ist, ist nämlich gleichzeitig eine Funktion der F-Statistik  $T^2$ . Wir definieren genauer eine F-Statistik der First-Stage-IV-Regression (in den Kapiteln 3c1 und 3c2 wurden ja exogene X-Variablen,  $x_1 \dots x_{k-1}$ , angenommen in Kapitel 3c3 existierten diese aber nicht!):

$$(3.36) \quad x_k = a_{0OLS} + a_{1OLS}x_1 + \dots + a_{k-1OLS}x_{k-1} + \alpha_{1OLS}z_{k1} + \dots + \alpha_{k0OLS}z_{k0} + \text{error},$$

wobei die F-Statistik ein Test auf die gemeinsame Insignifikanz der Koeffizienten  $\alpha_{1OLS}=0, \dots, \alpha_{k0OLS}=0$  darstellt.

Es kommt also darauf an, wie hoch die zusätzliche Erklärungskraft der Instrumente auf die endogene Variable ist, nachdem die gemeinsame Erklärungskraft der Instrumente und der anderen exogenen X-Variablen auf die endogene X-Variable herausgerechnet wurde.

In STATA kann die F-Statistik z.Bsp. mit dem Befehl „test  $\alpha_{1OLS} \dots \alpha_{k0OLS}$ “ nach berechnen der First-Stage-IV-Regression berechnet werden. Diese sollte nicht kleiner als 10 sein.

#### Literatur:

Eine exzellente Übersicht über IV-Schätzung gibt Cameron und Trivedi, 2005, S. 95-112. Die Probleme „schwacher Instrumente“ werden insbesondere auf den S. 104-111 behandelt.

## 4) GMM- Schätzer

### a) MM - Schätzer

Damit ein Schätzer bestimmte Eigenschaften hat, müssen bestimmte „Moment-Restriktionen“ erfüllt sein, z.B. beim OLS- Schätzer muss gelten  $E[X_i \varepsilon_i] = 0$ , damit  $E[b_{OLS}|X] = \beta$  gilt (siehe Kapitel II. 1c).

In diesen Moment – Restriktionen werden dann die Stichprobenschätzer dieser Momente eingesetzt:

$$E[x_i' \varepsilon_i] = \frac{1}{n} \sum_{i=1}^n x_i' e_i = \frac{1}{n} \sum_{i=1}^n x_i' (y_i - x_i' b_{MM}) = \frac{1}{n} X' e = 0$$

$k \times 1$

$$\Leftrightarrow \frac{1}{n} \sum_{i=1}^n x_i' y_i - \frac{1}{n} \sum_{i=1}^n x_i' y_i' b_{MM} = 0$$

$$\Leftrightarrow X' Y = X' X b_{MM}$$

$$b_{MM} = (X' X)^{-1} X' y = b_{OLS}$$

Der OLS- Schätzer ist also auch ein spezieller MM- Schätzer. Die Herleitung klappt so gut, weil es  $k$ - Momentrestriktionen gibt und ein  $k$ - dimensionaler Regressionskoeffizienten gesucht wird  $\rightarrow$  eindeutige Lösung

Was passiert, wenn es  $L > k$  Moment- Restriktionen gibt?  
 $\rightarrow$  GMM- Schätzer

### b) GMM Schätzer

$\theta$  sei ein  $K$ - dimensionaler Vektor von zu schätzenden Parametern. Das Moment  $l, l=1 \dots L$ , sei eine Funktion von  $\theta$  und den Variablen(-Vektoren)  $X_i, y_i, Z_i$ , kurz:  $m_l(\theta, X_i, Z_i, y_i) \equiv m_{li}(\theta)$ .

Die Moment- Restriktionen können dann geschrieben werden als

$$E[m_{li}(\theta)], \quad l=1 \dots L$$

Das Stichprobenäquivalent zu den Momenten ist:

$$\bar{m}_l(\theta) \equiv \frac{1}{n} \sum_{i=1}^n m_{li}(\theta) = 0, \quad l = 1 \dots L$$

Wenn man

$\bar{m}_l(\theta) = 0, \quad l = 1 \dots K$  nach  $\theta$  gelöst wird, gibt es eine eindeutige Lösung  $\hat{\theta}$ . Wenn diese nun in

$\bar{m}_l(\theta), \quad l = K + 1 \dots L$  eingesetzt wird, gilt natürlich nicht die Moment- Restriktion, also

$$\bar{m}_l(\theta) \neq 0 \quad l = K + 1 \dots L$$

Schlimmer noch: Auch wenn die Reihenfolge der ersten K- Restriktionen ausgetauscht wird, gibt es andere Lösungen für  $\hat{\theta}$ .

Der Ausweg:

$$\bar{m}_l(\theta) = u_l \text{ mit } E[u_l] = 0 \text{ für } l = 1 \dots L$$

$$|u - 0|, \quad u = (u_1 \dots u_L)'$$

soll minimiert werden wie beim OLS- Schätzer.

Dazu wird die euklidische Distanz von  $|u-0|$  minimiert:

$$\theta_{\min} = \arg \min \sum_{l=1}^L u_l^2 = \arg \min \sum_{l=1}^L \bar{m}_l(\theta)' \bar{m}_l(\theta)$$

Dieses Problem kann noch weiter verallgemeinert werden, indem eine beliebige positiv-definite Matrix A hinzugefügt wird zu

$$\theta_{GMM} = \arg \min \sum_{l=1}^L \bar{m}_l(\theta)' A^{-1} \bar{m}_l(\theta) = \bar{m}(\theta)' A^{-1} \bar{m}(\theta)$$

Dann kann gezeigt werden (Hansen, 1982):

#### Theorem 14:

Wenn  $A = \text{Asy. Var}[\bar{m}]$ , also A ein konsistenter Schätzer der Momentmittelwertvarianz und  $p \lim \bar{m}(\theta) = 0$ , dann

a)  $p \lim \theta_{GMM} = \theta$

b)  $\text{Asy. Var}[\theta_{GMM}] = [G' A^{-1} G]^{-1}$  mit G Matrix mit Reihe j

$$G^j = \frac{\partial \bar{m}_j(\theta)}{\partial \theta'}$$

c)  $\theta_{GMM} \overset{a}{\sim} N[\theta, [G' A^{-1} G]^{-1}]$

d)  $\theta$  ist asymptotisch effizient

e)  $\theta = \arg \min \bar{m}(\theta)' B \bar{m}(\theta)$ ,  $B \neq A$ , ist konsistent, aber nicht asymptotisch effizient

#### 4c) Beispiel für GMM- Schätzung

(i)  $y = X\beta + \varepsilon$ ;  $[z_i, x_i, \varepsilon_i]$  ist n-fache Zufallsauswahl, wobei  $z_i, x_i$  die i.i.d. Eigenschaft haben.

(ii)  $E[\varepsilon \varepsilon' | X, Z] = \begin{bmatrix} w_1 & & 0 \\ & \ddots & \\ 0 & & w_n \end{bmatrix} r^2 = \begin{bmatrix} r_1^2 & & 0 \\ & \ddots & \\ 0 & & r_n^2 \end{bmatrix} = \Omega$

(iii)  $\text{cov}(x_i, \varepsilon_i) \neq 0$

(iv)  $\text{cov}(z_i, \varepsilon_i) = 0 \Leftrightarrow E[z_i \varepsilon_i] = 0; \quad E[z_i z_i'] = \sum_{zz_i} < \infty$

Momentbedingungen:

$$E[z_i \varepsilon_i] = 0 = \frac{1}{n} \sum_{i=1}^n z_i e_i = \frac{1}{n} \sum_{i=1}^n z_i \overbrace{(y_i - x_i b_{GMM})}^{\equiv \varepsilon_i} = \frac{Z' e(b_{GMM}, X)}{n}$$

Der GMM- Schätzer ist für diesen Fall definiert als

$$(24) \quad b_{GMM} = \arg \min_b \frac{1}{n^2} e(b, X)' Z A^{-1} Z' e(b, X)$$

Weil  $E[Z_i \varepsilon_i] = 0$  nach Annahme (iv) und

$\text{Var}[z_i \varepsilon_i] = E[z_i z_i' \varepsilon_i^2] = E[z_i z_i' E[\varepsilon_i^2 | x_i, z_i]] = \tau_i^2 \sum_{zz_i} < \infty$  und  $[Z_i, X_i, \varepsilon_i]$  eine Zufallsstichprobe ist, gilt nach dem Lindberg- Feller Central Limit Theorem (siehe Kapitel über Heteroskedastizität)

$$\sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i \right) \xrightarrow{d} N(0, \text{Asy.Var}[\frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i]),$$

wobei

$$\begin{aligned} \text{Asy.Var}[\frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i] &= \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \text{Var}[z_i \varepsilon_i] = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \tau_i^2 \sum_{zz_i} \\ &= p \lim \frac{\tau^2}{n} Z' \Omega Z \quad (\text{siehe 16a}) \end{aligned}$$

nach Theorem 14 (Hansen, 1982) gilt aber für A in (24):

$$A = \text{Asy.Var}[\frac{1}{n} \sum_{i=1}^n z_i \varepsilon_i] = p \lim \frac{\tau^2}{n} Z' \Omega Z = p \lim \frac{1}{n} \sum_{i=1}^n z_i z_i' \varepsilon_i^2 \quad \text{damit } b_{GMM} \text{ effizient ist.}$$

Nach (16b) (White, 1980) kann A konsistent geschätzt werden als:

$$\hat{A} = \frac{1}{n} \sum_{i=1}^n z_i z_i' \varepsilon_i^2$$

Das Optimierungsproblem (24) kann zunächst wie folgt gelöst werden:

$$\begin{aligned} q &\equiv \frac{1}{n^2} e' Z A^{-1} Z' e = \frac{1}{n^2} [y - X b_{GMM}]' Z A^{-1} Z' [y - X b_{GMM}] \\ &= \frac{1}{n^2} [y - X b_{GMM}]' Z [A^{-1} Z' y - A^{-1} Z' X b_{GMM}] \\ &= \frac{1}{n^2} [y' Z A^{-1} Z' y - b_{GMM}' X' Z A^{-1} Z' y - y' Z A^{-1} Z' X b_{GMM} + b_{GMM}' X' Z A^{-1} Z' X b_{GMM}] \end{aligned}$$

die Optimierungsbedingung 1. Ordnung lautet dann:

$$\frac{\partial q}{\partial b_{GMM}} = 0 = -\frac{2}{n^2} (X'ZA^{-1}Z'y) + \frac{2}{n^2} X'ZA^{-1}Z'Xb_{GMM}$$

Also:

$$(25) b_{GMM} = [X'ZA^{-1}Z'X]^{-1} X'ZA^{-1}Z'y$$

Nun besteht das Problem, dass A nicht bekannt ist. Nach (16b) (White, 1980) kann A aber konsistent geschätzt werden als:

$$(26) \hat{A} = \frac{1}{n} \sum_{i=1}^n z_i z_i' e_i^2$$

Das nächste Problem besteht darin, dass  $e_i = y_i - x_i b$  schon die Kenntnis von b erfordert. Aber wenn wie beim FGL- Schätzer irgendein konsistenter Schätzer b existiert, mit dem  $e_i$  bzw.  $\hat{A}$  berechnet werden können, dann bleibt nach dem Theorem von Cramér (Theorem 12)  $\hat{b}_{GMM} = [X'Z\hat{A}^{-1}Z'X]^{-1} X'Z\hat{A}^{-1}Z'y$  asymptotisch effizient, selbst wenn b nicht asymptotisch effizient ist. Nach Theorem 14(e) (Hansen, 1982) kann für A eine beliebige Matrix, Z.B. die Einheitsmatrix I gewählt werden und  $b = [X'ZZ'X]^{-1} X'ZZ'y = (Z'X)^{-1} Z'y = b_{IV}$  (nach (17) im Kapitel IV- Schätzer) ist konsistent, aber nicht effizient. Daraus ergibt sich der

#### Zweistufige GMM- Schätzer als

$$b = (Z'X)^{-1} Z'y,$$

$$e_i = y_i - x_i b,$$

$$\hat{A} = \frac{1}{n} \sum_{i=1}^n z_i z_i' e_i^2$$

und

$$(27) b_{GMM} = [X'Z\hat{A}^{-1}Z'X]^{-1} X'Z\hat{A}^{-1}Z'y$$

Die asymptotische Kovarianz muss als letztes noch ermittelt werden. Nach Theorem 14 (b) gilt (j ist Reihe):

$$\sigma^j = \frac{\partial \bar{m}_j(\theta)}{\partial \theta'} = \frac{\partial \frac{1}{n} \sum_{i=1}^n z_{ij} \varepsilon_i}{\partial \beta'} = \frac{\partial Z'(y - X\beta)}{\partial \beta'} = -Z'X$$

$$(28) \text{Est.Asy.Var}[b_{GMM}] = [\sigma' \hat{A}^{-1} \sigma]^{-1} = [X'Z\hat{A}^{-1}Z'X]^{-1}$$

Ende des Beispiels

#### 4d) Sargan- Test auf „Overidentifying Restrictions“

Wir kehren nochmals zum Ausgangspunkt des GMM- Schätzers zurück. Die wesentliche Annahme  $E[u_i]=0, i=1...L$  mit  $u_i$  definiert durch  $\bar{m}_i(\theta_{GMM}) \equiv u_i$ , muss getestet werden durch den Sargan- Test. Intuitiv:  $\sum_{i=1}^L u_i^2$  soll nicht signifikant unterschiedlich von 0 sein.

Genauer:

$$H_0: \quad \bar{m}(\theta)' A^{-1} \bar{m}(\theta) = 0$$

$$H_1: \quad \bar{m}(\theta)' A^{-1} \bar{m}(\theta) > 0$$

Die Teststatistik  $q \equiv \bar{m}(\theta_{GMM})' \hat{A}^{-1} \bar{m}(\theta_{GMM})$  konvergiert asymptotisch zu dem Term in  $H_0$  und ist  $\chi^2$ - verteilt mit  $L-k$  Freiheitsgraden. Begründung: Bei Anwendung eines Central Limit Theorems ist  $u_i$  asymptotisch normalverteilt und  $q$  ist dann eine quadratische Form einer asymptotisch- normalverteilten Zufallsvariablen und damit  $\chi^2$ - verteilt.

Wenn die Anzahl der Moment- Restriktionen  $L$  gleich der Anzahl  $k$  der zu schätzenden Parameter in  $\theta$  ist, dann ist  $q$  per Konstruktion immer 0. Daher bleiben  $L-k$  Freiheitsgrade.

## 5) Selection Bias und Heckman- Schätzer

Dieses Kapitel behandelt einen Schätzer für einen Fall, bei dem die Stichprobe keine reine Zufallsauswahl ist (Annahme 5 ist verletzt).

### a) Nicht repräsentative Daten

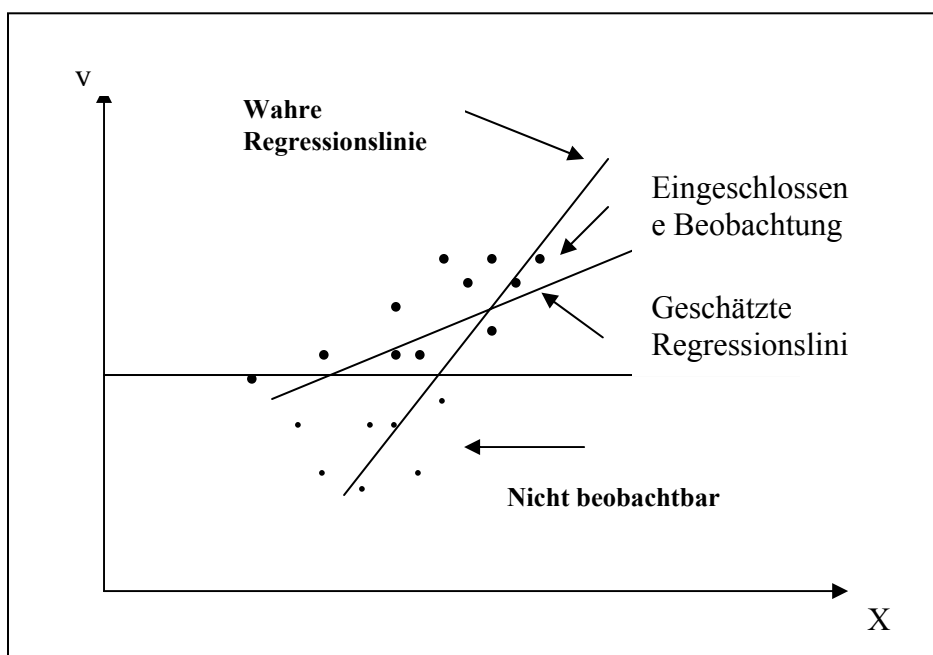
- Oft gibt es explizite oder implizite Ausschlusskriterien bei der Datensammlung

$y_i = x_i\beta + \varepsilon_i > L$  in Datenbasis

$y_i = x_i\beta + \varepsilon_i \leq L$  nicht in Datenbasis (L ist Ausschlusskriterium)

Bsp.:

„Amadeus 200000“ Firmendatenbank mit Bilanzdaten von europäischen Firmen enthält nur Firmen ab 100 Beschäftigten (oder einem bestimmten Umsatz oder einer bestimmten Bilanzsumme)



- Lösung des Problems  $\longrightarrow$  Heckman-Schätzer

### b) Exkurs: Binary Choice Models:

Binary Choice Models sind Modelle, bei denen die  $y$ -Variable (abhängige Variable) eine Wahlentscheidung widerspiegelt und daher nur 2 Werte annehmen kann:  $y_i \in \{0,1\}$

z.B.  $y_i$  könnte die Antworten „Ja“ und „Nein“ auf irgendeine Frage sein

Dann wird angenommen, dass

$$(29) \quad \begin{aligned} \text{Prob}(y=1) &= F(X, \beta) \\ \text{Prob}(y=0) &= 1 - F(X, \beta) \end{aligned}$$

wobei

$\text{Prob}(\cdot)$  für die Wahrscheinlichkeit des Eintretens eines Ereignisses „ $\cdot$ “ steht.

und

$F(X, \beta)$  diese Wahrscheinlichkeit von unabhängigen Variablen  $X$  und einem zu schätzenden Parametervektor  $\beta$  abhängt.

Dann gilt:

$$\begin{aligned} E[y|X] &= 1 \cdot F(X, \beta) + 0 \cdot (1 - F(X, \beta)) \\ &= F(X, \beta) \end{aligned}$$

Dann gilt die Tautologie:

$$(30) \quad y = E[y|X] + (y - E[y|X]) = F(X, \beta) + \varepsilon$$

mit dem Störterm  $\varepsilon$ .

Wenn  $F(X, \beta) = \beta'X$ ,  $\beta$  ein  $k \times 1$  Vektor,  $X$  ein  $k \times 1$  Zufallsvariablenvektor ist, dann wird (30) zum linearen Regressionsmodell.

Das Problem liegt darin, dass  $F(X, \beta) = X'\beta$  dem Wertebereich  $(-\infty, \infty)$  annehmen kann, was nach (29) nicht erlaubt ist, weil Wahrscheinlichkeiten mindestens 0 und höchstens 1 sein dürfen:

$$(31) \quad F(-\infty, \beta) = 0 \quad \text{und} \quad F(\infty, \beta) = 1.$$

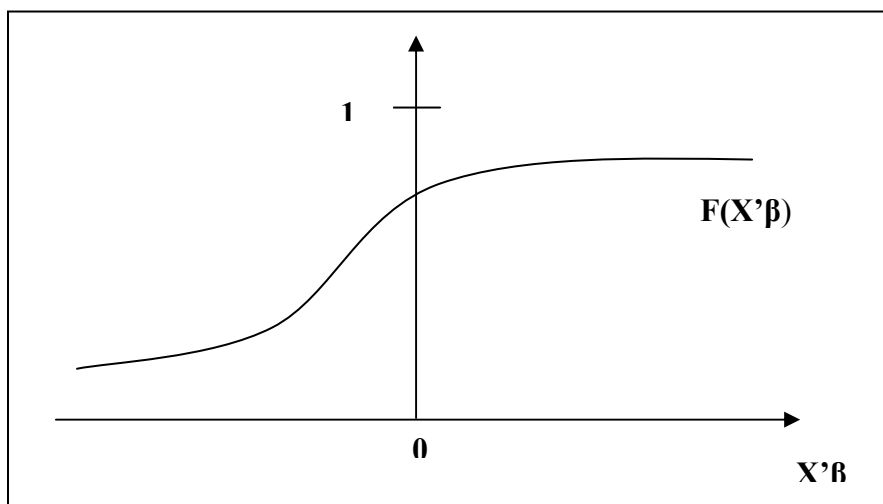
Daher wird für  $F(X, \beta)$  eine Wahrscheinlichkeitsverteilungsfunktion angenommen.

Wenn  $F(X, \beta) = \Phi(X'\beta)$  mit  $\Phi(\cdot)$  Verteilungsfunktion der Normalverteilung, dann liegt ein

Probit-Modell vor. Wenn  $F(X, \beta) = \frac{e^{\beta'X}}{1 + e^{\beta'X}}$  (logistische Verteilungsfunktion), dann liegt ein

Logit-Modell vor.

Graphisch



Bemerkung: die Schätzer der Probit- und Logit- Modelle werden nach der Maximum-Likelihood Methode berechnet. Wir werden uns nicht weiter mit den mathematischen Herleitungen auseinandersetzen, da wir uns generell auf metrische y- Variablen beschränken wollen. Stattdessen sei auf Green (2000), S. 812- 857), verwiesen.

## (b2) Heckmann- Schätzer

Das Problem, dass der Heckman- Schätzer löst, ist eine bestimmte Form einer nicht repräsentativen Stichprobe (Verletzung der Annahme 5).

Es gebe in der Stichprobe die Beobachtungen  $z_i, i=1, \dots, n$  und  $y_i$  oder  $x_i, i=1, \dots, n_1, n_1 < n$ . D.h. z.B. die Werte  $y_{n_1+1}, \dots, y_n$  der y- Variablen fehlen (missing observations) und diese fehlenden Beobachtungen sind überproportional häufig, wenn ein bestimmtes beobachtbares Kriterium  $\gamma'w$  besonders ausgeprägt ist.

---

Beispiel:

In einem Firmendatensatz fehlen oft die Bilanzdaten von „kleinen“ Firmen, weil die Datensammlungskosten vielleicht höher sind, größere Unternehmen oft strikteren Publikationspflichten unterliegen oder einfach ein größeres Interesse für Großunternehmen besteht und daher bei der Datensammlung größere Sorgfalt auf Großunternehmen angewendet wurde. Dann könnte  $\gamma'w$  aus dem Umsatz, den Beschäftigten und dem Bilanzkapital bestehen, wobei die Gewichte  $\gamma$  geschätzt werden müssen.

---

Dann wird für den Selektionsmechanismus, der zur nicht repräsentativen Stichprobe führt, angenommen:

$$z_i^* = \gamma'w_i + u_i$$

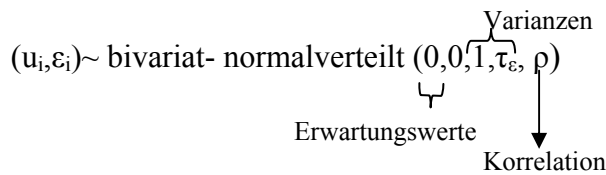
und eine Beobachtung  $y_i$  ist im Datensatz enthalten, wenn  $z_i^* > 0$  (z.B. Firmengrößenschwellenwert). Aber  $z_i^*$  selbst ist nicht beobachtbar bzw. nicht bekannt. Beobachtbar ist aber  $z_i \in (0, 1)$ ; der Wert 1 bedeutet, dass die Beobachtung  $i$  vollständig in der Stichprobe ist; der Wert 0 bedeutet, dass die Beobachtung  $i$  „systematisch“ fehlt. Es wird weiterhin unterstellt, dass

$$\begin{aligned} \text{Prob}(z_i=1) &= \Phi(\gamma'w) \\ \text{Prob}(z_i=0) &= 1 - \Phi(\gamma'w), \end{aligned}$$

wobei  $\Phi(\cdot)$  wieder für die kumulative Verteilungsfunktion der Normalverteilung steht. Das eigentliche Regressionsmodell sei linear und laute:

$$y_i = \beta'x_i + \varepsilon_i, \text{ wobei } i \text{ beobachtet wird, wenn } z_i=1$$

und für die beiden Fehlerterme gelte:



Um einen unverzerrten Schätzer zu berechnen, brauchen wir ein neues mathematisches Theorem:

**Theorem 15:** Momente einer trunkierten bivariaten Normalverteilung.

Wenn die Zufallsvariablen  $y$  und  $Z$  eine bivariate Normalverteilung mit Mittelwerten  $\mu_y$  und  $\mu_z$  und Standardabweichung  $\tau_y$  und  $\tau_z$  und Korrelationskoeffizient  $\rho$  haben und  $a$  eine beliebige Konstante ist, dann:

$$E[y | Z > a] = \mu_y + \rho \tau_y \lambda\left(\frac{a - \mu_z}{\tau_z}\right),$$

$$Var[y | Z > a] = \tau_y^2 [1 - \rho^2 \delta\left(\frac{a - \mu_z}{\tau_z}\right)],$$

mit

$$\lambda(\cdot) \equiv \frac{\Phi(\cdot)}{1 - \Phi(\cdot)}, \quad \Phi(\cdot) \text{ kumulative Verteilungsfunktion der Normalverteilung}$$

$$\delta(X) \equiv \lambda(X) / (\lambda(X) - X)$$

Mit Hilfe dieses Theorems gilt ( $x_i, w_i$  zur Vereinfachung fixiert)

$$\begin{aligned} E[y_i | \underbrace{y_i \text{ beobachtet}}_{Z_i=1}] &= E[y_i | z_i^* > 0] \\ &= E[y_i | \mu_i > -\gamma' w_i] \\ &= \beta' x_i + E[\varepsilon_i | \mu_i > -\gamma' w_i] \\ &= \beta' x_i + \rho \tau_\varepsilon \lambda_i\left(\frac{\gamma' w_i}{\tau_u}\right) \quad | \tau_u = 1 \end{aligned}$$

$$Var[y_i | z_i] = (X' X)^{-1} \tau_\varepsilon^2 [1 - \rho^2 \delta(\gamma' w_i)]$$

Problem:  $\gamma$  ist nicht bekannt!

Heckman (1979) zeigt, dass dieses Wahrscheinlichkeitsmodell in folgender Weise 2-stufig geschätzt werden kann:

1) Man schätze  $\hat{\gamma}$  mit einer Probit-Schätzung von

$$E[Z_i | Z_i=1] = \Phi(\gamma' w_i) + u_i$$

Daraus berechnet man für jede vorhandene Beobachtung  $i$ :

$$\hat{\lambda}_i = \frac{\Phi(\hat{\gamma}' w_i)}{1 - \Phi(\hat{\gamma}' w_i)}$$

$$\hat{\delta}_i = \hat{\lambda}_i / [\hat{\lambda}_i - \hat{\gamma}' w_i]$$

2) Man schätze mit OLS die Regression

$$(32) \quad Y_i = X_i' \beta + \beta_\lambda \hat{\lambda}_i + v_i$$

wobei  $\beta_\lambda = \rho \tau_\varepsilon$

Wie erhält man nun den Schätzer für  $\hat{\rho}$  und  $\tau_\varepsilon$  aus  $\beta_\lambda$ ?

Nach dem Theorem 15 gilt:

$$(33) \quad \text{Var}(v_i) = \tau_\varepsilon^2 (1 - \rho^2 \delta_i) \equiv \tau_i^2$$

Wenn als  $\bar{\delta}$  der wahre Stichprobenmittelwert von  $\delta_i$  in der Stichprobe bezeichnet wird, dann konvergiert die Fehlervarianz  $\tau_i^2$  im Stichprobenmittel zu

$$(34) \quad p \lim \frac{1}{n} \sum_{i=1}^n \tau_i^2 = \tau_\varepsilon^2 (1 - \rho^2 \bar{\delta}) = \tau_\varepsilon^2 - \tau_\varepsilon^2 \rho^2 \bar{\delta}$$

Dieser Term kann geschätzt werden durch die Stichprobenfehlervarianz  $\frac{e'e}{n}$  aus der OLS-Regression der 2. Stufe.

Aus der Tatsache, dass  $b_\lambda$  der OLS-Schätzer für  $\beta_\lambda = \rho \tau_\varepsilon$  ist, folgt unter Zuhilfenahme des Slutsky Theorems:

$$(35) \quad p \lim b_\lambda^2 = \rho^2 \tau_\varepsilon^2$$

Außerdem folgt aus der konsistenten Schätzung der ersten Stufe:

$$p \lim \frac{1}{n} \sum_{i=1}^n \hat{\delta}_i = \bar{\delta}$$

Mit Hilfe der letzten beiden Gleichungen kann (34) in geschätzten Termen geschrieben werden als

$$\frac{e'e}{n} = \hat{\tau}_\varepsilon^2 - b_\lambda^2 \hat{\delta}$$

bzw

$$\hat{\tau}_\varepsilon^2 = \frac{e'e}{n} + b_\lambda^2 \hat{\delta}$$

und nach (35) ergibt sich

$$\hat{\rho}^2 = b_\lambda^2 / \hat{\tau}_\varepsilon^2$$

### Bemerkung 1:

Wenn  $\rho=0$ , dann ist  $b_\lambda=0$  und die Regression in (32) kollabiert zu einer einfachen OLS-Regression. Dann liegt keine Verzerrung des OLS- Schätzers vor, obwohl die Stichprobe nicht repräsentativ ist!

Es gibt einen Wald- oder LR- Test, der testet, ob  $\rho$  signifikant unterschiedlich von 0 ist ( $H_0$  ist:  $\rho=0$ )

---

### Bemerkung 2:

Die Varianz von  $v_i$  in (34) impliziert Heteroskedastizität. Die Kovarianzmatrix von  $X_* \equiv (X, \lambda)$  für die Regressionskoeffizienten  $b_* \equiv (b, b_\lambda)$  ist dann nach White (1980) (siehe Kapitel II.2)

$$\text{Var}[b_*] = \tau_\varepsilon^2 [X_*' X_*]^{-1} \left[ \sum_{i=1}^n (1 - \rho^2 \delta_i) X_i^* X_i^{*'} \right] [X_*' X_*]^{-1}$$

Da  $\delta_i$  nicht mit Sicherheit bekannt ist, weil  $\delta_i = \delta_i(\hat{\gamma}' w_i)$  gilt und  $\hat{\gamma}$  in der ersten Stufe mit einem Fehler geschätzt wird, muss die aus der Stichprobe geschätzte asymptotische Kovarianzmatrix durch einen Korrekturfaktor versehen werden, der die Varianz von  $\hat{\gamma}$  mitberücksichtigt (Murphy- Topel Korrektur). Näheres dazu findet sich in Green (2000), S. 932f.

---

### Bemerkung 3:

Anstelle der 2- Stufen – Schätzung gibt es alternativ auch eine 1- stufige Maximum-Likelihood- Schätzung, die als „effizienter“ gilt und daher in der Praxis vorgezogen wird. Sie hat aber den Nachteil, dass sie nur numerisch gelöst werden kann und manchmal nicht „konvergiert“, i.e. keine Lösung gefunden wird.

## 6) Zusammenfassung von Teil 1 der Vorlesung

Bei den bisherigen Kapiteln gilt zu beachten, dass wichtige Teilbereiche der Ökonometrie ausgelassen wurden, weil nur die Grundlage für Paneldatenanalyse gelegt werden sollte. Zu den ausgelassenen Kapiteln gehören insbesondere:

Maximum- Likelihood Schätzer, LR-/Wald-/LM- Test, F-Test, Nicht- lineare Regressionen, Nicht- bzw. Semiparametrische Regressionen, Heteroskedastizitäts- und Autokorrelationstests, Zeitreihenanalyse, Nicht- metrische y- Variablen, Hypothesen- Tests von „non- nested hypothesis“ (eine Hypothese ist nicht ein Spezialfall einer anderen Hypothese), Chow- Test auf Strukturbruch, etc.

Im folgenden soll (ohne Gewähr) eine kurze Check- Liste für die Praxis der Datenanalyse gegeben werden.

Ausgangspunkt sei ein Problem der Art:

$$y = X\beta + z\gamma + \varepsilon,$$

wobei  $z$  die Variable des Interesses ist, i.e. eine neue Erklärung für die Variable  $y$  liefern soll.

Wenn bei einer einfachen OLS- Regression bei Benutzung der White- konsistenten Kovarianzmatrix der Regressionskoeffizient  $\hat{\gamma}$  signifikant unterschiedlich von 0 ist und das nach der Theorie erwartete Vorzeichen hat, dann gibt es einen „case“.

Dann ist zu überprüfen, ob dieses Ergebnis einer kritischen Hinterfragung standhält. Dazu geht man die Annahmen für die OLS- Regression durch und versucht zu zeigen, dass  $\gamma$  immer noch signifikant ist, wenn eine Annahmeverletzung unterstellt wird und angemessene Korrekturen bei den Schätzern vorgenommen werden.

} **Baseline Spezifikation**

### Check- Liste

- 1) In einer Regressionstabelle werden die geschätzten Koeffizienten, Standardfehler oder t-Werte, die Anzahl der Beobachtungen und das  $R^2$  (oder Pseudo- $R^2$ , Adjusted  $R^2$ , Likelihood- Wert angegeben). Signifikanz- Niveaus werden durch „\*“ gekennzeichnet. Gegebenenfalls gibt es Test- Ergebnisse.
- 2.) Omitted variable bias: hinzufügen zusätzlicher Kontrollvariablen zur Baseline- Spezifikation, die mit der Variable des Interesses  $z$  hoch korreliert sein könnten.
- 3.) Strukturbruch (Verletzung der Annahme 1):  
Es könnte sein, dass die Daten der Stichprobe aus unterschiedlichen Grundgesamtheiten stammen. Wenn man einen Verdacht hat (z.B. vor oder nach der Wiedervereinigung/ Low- oder High- Tech- Industrien /Stadt- oder Land/ Entwicklungsländer, Industrieländer oder Emerging Market Economies/ vor oder nach einer Steuerreform etc.) werden Regressionsergebnisse für beide Stichprobenhälften berichtet und ein sogenannter Chow- Test auf Strukturbruch durchgeführt (wenn der Test- Befehl aus STATA angewendet wird, ist dieser Test heteroskedastizitätskonsistent).  
Wenn die Bruchstelle nicht genau bekannt ist  $\Rightarrow$  Threshold Regression (Hansen, Econometrica, 2000)

4) Nicht- lineare funktionale Form (Annahme 1)

Man definiere

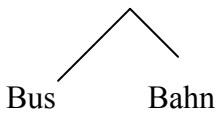
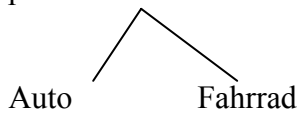
$$y' \equiv y - Xb_{OLS}$$

Und „plotted“ die Daten in einem Diagramm mit  $y'$  auf der vertikalen und  $z$  auf der horizontalen Achse. Wenn augenscheinlich kein linearer bzw. log- linearer Zusammenhang besteht, dann

- a) quadratische, kubische etc. Terme der  $Z$ - Variablen in die Regression aufnehmen
- b) nicht- lineare Regressionstechniken verwenden (Greene, 2000, Kapitel 10)
- c) nicht- parametrische / semi- parametrische Schätzverfahren verwenden  
⇒ Lehrstuhl Kneip  
(Ein formeller Test für Linearität einer funktionalen Form ist z.B. Zheng, Journal of Econometrics, 1996)

5) Nicht- metrische  $y$ - Variablen

$y$  sollte zumindest theoretisch eine annähernd kontinuierliche Variable im Wertebereich  $(-\infty, +\infty)$  sein, weil sonst  $\varepsilon$  nicht /asymptotisch) normalverteilt ist:

- a)  $y$  ist trunziert ( $\underline{a} \leq Y \leq \bar{a}$ ) mit Häufungen der Beobachtungen bei  $\underline{a}$  bzw.  $\bar{a}$   
z.B.:  $y$  seien Wahrscheinlichkeiten oder Prozentangaben  
Lösung ⇒ Tobit- Regression
- b)  $y$  ist zensiert: Häufung der Beobachtungen in einem  $y$ - Wert z.B.  $y =$  bilaterale Exporte; bei zahlreichen Länderpaaren ist der Wert von  $y$  null. Die Distanz ist zu groß, die Märkte zu klein, etc. ⇒ Censored Regression
- c)  $y$  ist  $\{0,1\}$ - Variable ⇒ Probit- oder Logit- Schätzung
- d)  $y$  besteht aus mehreren Zuständen (nicht nur 2)  
z.B. Transport mit Bus, Auto, zu Fuß, Fahrrad  
⇒ Multinomial Logit- Schätzung
- e)  $y$  besteht aus mehreren Zuständen, die einer Ordnung unterliegen. Z.B. sehr gut, gut, ..., mangelhaft  
⇒ ordered logit / probit- Schätzung
- f) Zustände sind „hierarchisch“  
z.B. öffentliche Verkehrsmittel  
  
private Verkehrsmittel  
  
⇒ nested probit-logit Modelle
- g)  $y$  ist  $\in \{1,2,\dots,N\}$   
z.B. Anzahl der Punkte, die ein Autofahrer in der Flensburger Verkehrskartei hat ⇒ Count- Data- Modelle (z.B.: Poisson- Regressionsmodell)

- h) y ist die „Überlebenszeit“  
z.B.: ein Datensatz mit Arbeitslosen, wobei die y- Variable, die bisherige Länge der Arbeitslosigkeit angibt  $\Rightarrow$  Duration Modelle
- i) y ist nur in Bandbreiten bekannt.  
z.B. y sei Industriebeschäftigung in Betrieben mit 0...10, 11...20, 21...50, 51...100, 101...500, 501...1000, 1001... Beschäftigten  $\Rightarrow$  Quantile-Regression
- j) Treatment Probleme (aus der Medizin)

Es gibt eine Gruppe in der Stichprobe, der ein bestimmtes Medikament verabreicht wurde. Der anderen Gruppe, die ansonsten der ersten möglichst gleich ist, wurde lediglich ein „Placebo“ verabreicht.

Zahlreiche ökonomische Probleme insbesondere in der experimentellen Ökonomie können als Treatment- Problem angesehen werden

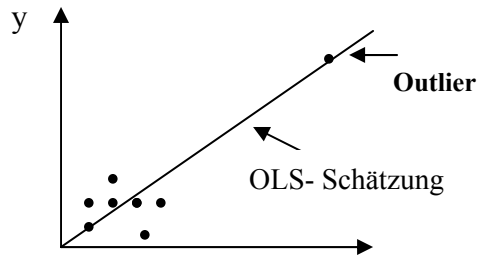
$\Rightarrow$  Diffrence- in- Diffrence Verfahren + treatment regression model (treat reg in STATA); psmatch2 im Internet

Bemerkung: für die unter 5a)-5j) behandelten Verfahren gibt es natürlich GLS- und IV-Schätzer und entsprechend modifizierte Testverfahren. Insbesondere gibt es für diese speziellen Verfahren auch statistische und/ oder dynamische Paneldaten-Variationen. Damit werden wir uns jedoch nicht beschäftigen. Stattdessen sei als umfassende Referenz auf Hsiao (2003) verwiesen. Eine Einführung in die Verfahren 5a)-5j) gibt Greene (2000), Kapitel 19 und 20.

## 6) Messprobleme

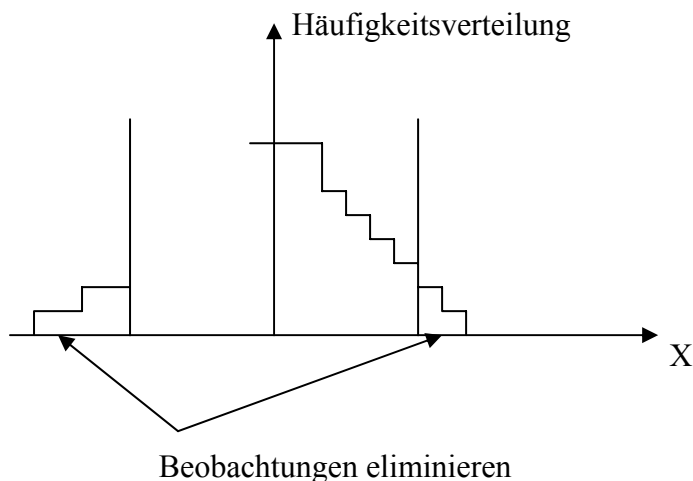
- a) y- Variable wird nur mit „Rauschen“ gemessen:  
 $y = y^* + \varepsilon$ ,  $E[y] = E[y^*]$ ,  $y^*$  und  $\varepsilon$  nicht beobachtbar  
 $\Rightarrow$  kein Problem, weil Rauschen im Störterm aufgeht
- b) X- Variable wird nur mit „Rauschen“ gemessen:  
 $\Rightarrow$  attenuation- Problem (siehe S. 33f)  
 $\Rightarrow$  konservativer Fehler  $\Rightarrow$  insignifikante Koeffizienten können in Wahrheit signifikant sein, aber nicht umgekehrt.
- c) Nachweis, dass eine X- Variable, bei der man „weißes Rauschen“ bei der Messung vermutet, nicht zu verzerrten Schätzern geführt hat, erfolgt durch eine IV- Schätzung (siehe Kapitel 3, S. 33ff)
- d) Outlier. Wenn die Datenmenge nicht sehr groß ist, können wenige extreme Ausreißer-Beobachtungen das Schätzergebnis völlig verzerren.

Bsp.:



Immer „summary“ Statistiken aller Variablen anschauen (min, max, etc). Am einfachsten ist ein Daten- Plot wie bei 4), um ein solches Problem zu entdecken. Es gibt auch formelle Tests. Outlier sollten Anlass zu einer Überprüfung der Richtigkeit einer Beobachtung geben. Outlier können aber nicht einfach weggelassen werden! Bei sehr großen Datensätzen (z.B.

100.000 Beobachtungen) sind Datenfehler unvermeidlich, haben aber auch nur einen geringen Einfluss auf das Schätzergebnis. Oft eliminiert man dann die Beobachtungen, deren X- Werte zu dem 1% höchsten oder niedrigsten Werten gehören.



e) Proxy- Variablen. Für ein theoretisches Konzept gibt es keine beobachtbare Variable, sondern nur mehrere „Proxy“-Variablen, die die „wahre“ Variable nur annähernd widerspiegeln.

---

z.B. Zusammenhang zwischen Investitionen und Lohnkosten eines Unternehmens. Lohnkosten eines Unternehmens seien nicht bekannt, sondern nur die durchschnittlichen Lohnkosten in der Industrie und die durchschnittlichen Lohnkosten in der Region.

---

⇒ Man zeigt, dass der Zusammenhang zwischen X und y für jede einzelne Proxy- Variable gilt. (Proxy- Variablen nur für abhängige Variable und unabhängige Variable des Interesses)

---

- 7) Nicht repräsentative Daten
- a) Heckmann Schätzer (siehe Kapitel 5)
- b) Weighted Least Squares (WLS-) Schätzer (Kapitel 2)

$$y_i = \alpha_0 + \alpha_1 K_i + \alpha_2 L_i + \alpha_3 D_g + \varepsilon_i$$

$y_i$  sei Value Added der Firma  $i$

$L_i$  Anzahl der Beschäftigten der Firma  $i$

$K_i$  Kapitalstock der Firma  $i$

$D_g$  sei eine Industrie- Dummy- Variablen Matrix und  $g$  sei ein Industrieindex

$\bar{L}_g$  sei die aggregierte Industriebeschäftigung im Sektor  $g$  nach einer amtlichen Statistik (Grundgesamtheit)

$$\bar{L}_g = \sum_{i \in g} L_i$$

dann kann ein Gewicht der Über- / Unterrepräsentation wie folgt berechnet werden:

$$weight_g = \frac{\bar{L}_g}{\sum_g \bar{L}_g} / \frac{\hat{L}_g}{\sum_g \hat{L}_g}$$

Wenn ein Sektor im Datensatz unterrepräsentiert ist, dann ist  $weight_g > 1$ . Wenn alle Beobachtungen von untergewichteten Sektoren einen größeren Einfluss auf die Berechnung der Regressionskoeffizienten ausüben,

z.B.  $weight_g = 2$ , dann werden alle Beobachtungen von Sektor  $g$  doppelt „gezählt“.

In STATA:

Regress yX [a weight=weight<sub>g</sub>], robust

## 8) Multikollinearität

Beinahe- Multikollinearität: die X- Variablen sind untereinander so stark korreliert (beinahe Linearkombinationen), dass der partielle Effekt von einer X- Variablen allein auf die y- Variablen nicht mehr signifikant ist (z.B.: Problem der empirischen Wachstumsforschung). Wenn die Anzahl der Variablen groß ist, relativ zu der Anzahl der Beobachtungen, die in einer Regression verwendet werden, ist die Wahrscheinlichkeit groß, dass Beinahe-Multikollinearität entsteht.

Lösung: mehr Beobachtungen zusammenbekommen.

Wenn es viele Messvariablen X gibt, die dasselbe theoretische Konzept messen, dann kann man das Multikollinearitätsproblem dadurch umgehen, dass man eine Indexzahl aus ihnen formt. Die statistisch optimalen Gewichte in diesem Index können mittels der Principal Component Analysis ermittelt werden.

Z.B.: „Freiheit“ als Index aus „Freiheit der Presse“, „freier Außenhandel“, „Grad der Demokratie“, „unabhängige Gerichte“, etc.

## 9) Heteroskedastizität

Siehe Kapitel II.2

10) Autokorrelation

Siehe Paneldatenverfahren

---

11) Simultaneität der „unabhängigen Variablen“

- a) IV- Schätzer (Kapitel II.3)
  - b) GMM- Schätzer (Kapitel II.4)
  - c) Treatment – Modell (STATA: treatreg)
  - d) Gleichungssystem- Schätzung: Two- Stage- Least Squares, Three- Stage Least Squares, System- GMM
  - e) Matching- Techniken
  - f) event studies/ natürliche Experimente
- 

12) Diskriminierung von Modellen

Hierzu gibt es lineare oder nicht- lineare LR-, Wald-, LM- oder F- Tests. Üblicherweise werden Wald- Tests verwendet, weil sie heteroskedastizitätskonsistent sind. (STATA: test für lineare Hypothesen und nltest für nicht-lineare Hypothesen.)