ÖKONOMETRIE I

Lehrveranstaltung Empirische Forschung zu Arbeit und Personal

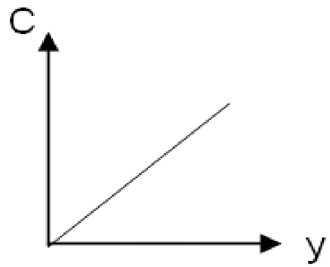
Universität Basel,
Herbstsemester 2019

Lutz Bellmann

INHALT

1. ÖKONOMISCHE MESSUNG ODER BESSER: QUANTIFIZIERUNG ÖKONOMISCHER SACHVERHALTE

z. B. Konsumfunktion



$$C = a + b\,Y^v + u$$

konstant = Koefizienten

veränderlich = Variablen

(a-Konsumniveau, wenn $Y^v = 0$, b ist die marginale KQ)

2. AUFGABEN DER ÖKONOMETRIE

Modellspezifikation:

Damit soll der für das jeweils zu untersuchende ökonomische Phänomen unterstellte Bedingungsrahmen explizit gemacht werden. Durch das Modell soll eine auf wesentliche Determinanten reduzierte Realität wiedergegeben werden.

Bestimmung numerischer Zusammenhänge (Punktschätzung):

Mit Hilfe statistischer Verfahren werden auf Basis von Daten numerische Angaben über die als relativ rigide erachteten Beziehungen gemacht:

$$\hat{b} = 0.8$$

Intervallschätzung:

Angabe des Bereichs der zulässigen, durch unsystematische Bewegungen hervorgerufenen Abweichungen (Intervallschätzung).

$$\hat{b} = [0.78, 0.82] \text{ bei } \propto = 0.05$$

2. AUFGABEN DER ÖKONOMETRIE (FORTS.)

Sind diese drei Schritte (scheinbar erfolgreich) abgeschlossen, kann der anwendungsorientierte Teil der Ökonometrie-Arbeit folgen:

- ökonomische Regelmäßigkeiten ableiten,
- ökonomische Theorien falsifizieren,
- zukünftige Entwicklungen prognostizieren,
- wirtschaftspolitische Entscheidungshilfen liefern.

3. GRUNDSTRUKTUR ÖKONOMETRISCHER MODELLE

Endogene Variable	Vom Modell erklärte Variablen (durch Innenwelt erklärt) (C)
Exogene Variable	Vom Modell nicht erklärte Variablen Übernommene Variablen (übernommen aus der Außenwelt) (Y^v)
Störvariable	Restgröße oder „catch-all-Größe“ (u)

3. GRUNDSTRUKTUR ÖKONOMETRISCHER MODELLE

Allen Begründungen, die die Einführung einer vom ökonomischen Modell abweichenden, zusätzlichen „catch-all-Größe“ nahelegen, ist gemeinsam, dass sie auf latent vorhandene, aber nicht explizit erfasste Einflüsse hinweisen.

4. FUNKTIONSTYPEN

Lineare Regressionsmodelle sind leichter handhabbar als nichtlineare. Sind Gleichungen nichtlinear in den Variablen, so lassen sich durch **Variablentransformation** die Gleichungen als linear behandeln.

4. FUNKTIONSTYPEN

$$y = a + \frac{b}{x} \longrightarrow y = a + b\,z$$

wobei $z = \frac{1}{x}$

4. FUNKTIONSTYPEN

Bei einer Vielzahl nichtlinearer Funktionen lässt sich eine lineare Transformation durchführen

$y = a\,x^{b} \longrightarrow w = \alpha + \beta\,z$

$\ln y = b \ln ax = \ln a + b \ln x$

wobei $w = \ln y$

$\alpha = \ln a$

$\beta = b$ (ist eine Elastizität)

$z = \ln x$

$y = a\,e^{b\,x} \longrightarrow w = \alpha + b\,x$

wobei $w = \ln y$; $\alpha = \ln a$; $b = WR$

Dabei ist der Koeffizient b als Wachstumsrate interpretierbar.
Wenn es sich beim Regressor x um eine Dummyvariable handelt,
lässt sich die Wachstumsrate aus dem Koeffizienten b wie folgt berechnen:
$(\exp(b)-1) \cdot 100$

4. PRODUKTIONSFUNKTIONEN

Die **Cobb-Douglas Produktionsfunktion (CD)** hat folgende Grundform:

(1) $Y=A*L^{\alpha}K^{1-\alpha}$

wobei A=Effizienzparameter, L= Produktionsfaktor Arbeit und K= Produktionsfaktor Kapital.

Nach Logarithmieren und Anpassung an das OLS-Schätzumodell ergibt sich:

(2) $lnY = \beta_1 + \beta_2 lnL + \beta_3 lnK + \varepsilon_1$

wobei Y=Umsatz, L=Beschäftigtenzahlt, K= Kapital und ε=Störterm.

β_i (mit i=1,2,3) bilden die Regressionskoeffizienten ab.

4. PRODUKTIONSFUNKTIONEN

Die Cobb-Douglas-Produktionsfunktion ist ein Spezialfall der **CES-Produktionsfunktion**, bei der die restriktive Annahme einer Substitutionselastizität von eins aufgehoben ist. Die CES-Funktion in ihrer Grundform lautet wie folgt:

(3) $\ln Y = \ln\gamma - v/\rho \ln[\delta K^{-\sigma} + (1-\delta)L^{-\sigma}] + \varepsilon_2$

wobei δ= Distributionsparameter, ρ= Substitutionsparameter, γ= Niveauparameter und v= Homogenitätsparameter.

Mittels Taylorreihenapproximation erhält man folgende Schätzgleichung:

$$\ln Y = \beta_1 + \beta_2 \ln L + \beta_3 \ln K + \beta_4(\ln K - \ln L)^2 + \varepsilon_2$$

4. PRODUKTIONSFUNKTIONEN

Die Annahme einer konstanten Substitutionselastizität wird bei der **Translog-Produktionsfunktion** aufgehoben. Die Produktionsfunktion hat folgende unspezifizierte Form:

(5) $\ln Y = f(\ln L, \ln K)$

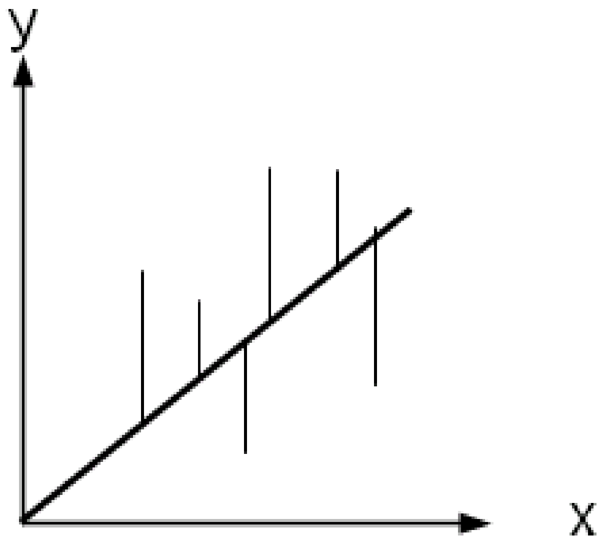
Für die Koeffizientenschätzungen wird die Funktion mittels einer Taylor-Reihe approximiert und logarithmiert, sodass folgendes Gleichungssystem entsteht (vgl. Greene 1993, S. 209):

(6) $\ln Y = \beta_1 + \beta_2 \ln L + \beta_3 \ln K + \beta_4 (\ln^2 L/2) + \beta_5 (\ln^2 K/2) + \beta_6 (\ln L * \ln K) + \varepsilon_4$

5. DAS EINFACHE LINEARE MODELL

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

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



$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Scatter-Diagramm

Wie verschaffe ich mir Werte für $\hat{\beta}_0$ und $\hat{\beta}_1$?

5. DAS EINFACHE LINEARE MODELL

Wahre Regressionsgerade: $y = \beta_0 + \beta_1 x$ unbekannt

$\hat{u}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$ senkrechter Abstand

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} \left[y_i - (\hat{\beta}_0 - \hat{\beta}_1 x_i) \right]^2 \longrightarrow \min_{\hat{\beta}_0, \hat{\beta}_1}$$

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

KQ-Residuen $\hat{u}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)$

$y_i = \hat{y}_i + \hat{u}_i$

6. DAS MULTIPLE LINEARE MODELL

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots + \beta_k x_{ki} + u_i$$

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

$$y = X\beta + u$$

$$y = X\beta + u^0$$

$$y_i = x_i{}'\beta + u_i$$

6. MULTIPLES BESTIMMTHEITSMAß

$$R^2 = 1 - \frac{\sum (\hat{y}_i - y)^2}{\sum (y_i - \bar{y})^2}$$

(1) Je größer, umso besser ist die Projektion

(2) Vergleichsmaß für Anpassungsgüte (goodness of fit): je mehr R^2 in der Nähe von 1 liegt, für umso geeigneter wird eine Modellspezifikation gehalten. Aber Probleme, wenn

- verschiedene endogene Variable
- Trend in exogenen Variablen
- Zusätzlich erklärende Variable (→korrigiertes Bestimmtheitsmaß)

6. KORRIGIERTES MULTIPLES BESTIMMTHEITSMAß

$$\overline{R}^2 = R^2 - \frac{K-1}{n-K}(1-R^2)$$

mit K als Zahl der exogenen Variablen und n als Fallzahl.

6. HYPOTHESENTEST

Für einzelne Regressionskoeffizienten oder für Linearkombination

H_0: $\beta_i = 0$

H_1: $\beta_i \neq 0$

Lehne H_0 ab, wenn $\frac{\hat{\beta}_i}{\sigma(\hat{\beta}_i)} > t(n - K - 1) = t(n - \nu)$

bei $\alpha = 0.01$ oder 0.05

6. TEST ÜBER DEN EINFLUSS VON VARIABLENGRUPPEN

Test von Kmenta, 1971, 371

$$F = \frac{(R_K^2 - R_{K1}^2)(n-K)}{(1-R_K^2)(K-K_1)} \sim F_{n-K,1-\alpha}^{K-K1}$$

7. STATISTISCHE PROBLEME UND IHRE KONSEQUENZEN

Kriterien

- Verzerrte Schätzung der Regressionskoeffizienten (Bias)
  Tritt ein bestimmtes Problem auf , so produziert OLS einen Schätzer für die β_i, von dem nicht erwartet werden kann, das er mit dem wahren β_i übereinstimmt.

- Verzerrte Schätzung bei der Standardabweichung der Regressionskoeffizienten
  Dasselbe für $[\![\sigma\left(\hat{\beta}_i\right)]\!]$

- t-Statistik
  wird ungültig, wen Regressionskoeffizient und/oder seine Standardabweichung verzerrt geschätzt wird.

- Effizienz
  Konfidenzintervall für Regressionskoeffizienten so klein wie möglich,
  z.B. $\left[\hat{\beta}_i - t(n-v)\sigma(\beta_i), \hat{\beta}_i + t(n-v)\sigma(\beta_i)\right]$
  gilt nur für unverzerrte Schätzungen(sonst mean square error als Kriterium)

Problem Nr. 1: Fehlspezifikation

- wichtigstes Problem: Problem Nr. 1
- falsche funktionale Form
- falsche Variable (zu wenig ist schlimmer als zu viel)

- bei Verdacht: testen

Heteroskedastie (HES)

- nicht so tragisch

Fälle:

- Varianz von u als Funktion der exogenen Variablen
- Heterogene Gruppenbildung
- Gruppierte Beobachtungsdaten
- Stochastische Regressionskoeffizienten

- bei Verdacht: testen und gewichtete Schätzung vornehmen

Autokorrelation (AK)

Bei Verwendung von Zeitreihendaten ist die Störgröße nicht mehr zeitinvariant, sondern der Störeffekt überträgt sich von Periode zu Periode. Idealtypisch besteht die Störgröße aus einer Vielzahl von weniger bedeutsamen, nicht messbaren Einflussgrößen auf die endogene Variable, deren Effekte sich im Durchschnitt kompensieren.

Je länger die jeweiligen Perioden sind, in denen jeweils eine Beobachtung gesammelt wird, umso stärker ist der Ausgleichsprozess.

Wahrscheinlichkeit für das Auftreten von AK umso größer, je mehr man von Jahres- zu Quartals- und schließlich zu Monatswerten übergeht.

- bei Verdacht: testen und gewichtete Schätzung vornehmen

7. STATISTISCHE PROBLEME UND IHRE KONSEQUENZEN

Fehler in den Variablen

Schlimm, aber selten feststellbar.

7. STATISTISCHE PROBLEME UND IHRE KONSEQUENZEN

mit exogenen Variablen korrelierte Störgröße

Schlimm, tritt immer dann (aber nicht nur dann) auf, wenn sich unter den Variablen auf der rechten Seite endogene Variablen befinden. Dann ist eine Instrumentalvariablenschätzung erforderlich. Im Prinzip wird 2. Gleichung mit x_i als abhängiger Variable geschätzt und $\hat{x}_i$ als Regressor an Stelle von x_i verwendet.

- bei Verdacht: testen

7. STATISTISCHE PROBLEME UND IHRE KONSEQUENZEN

Multikollinearität

- Stichprobenproblem
- Varianzinflationsfaktor berechnen.
- Damit muss man leben.

7. STATISTISCHE PROBLEME UND IHRE KONSEQUENZEN

Übersicht Verzerrungen

	Regressions-koeffizienten	Standard-abweichung der Regressions-koeffizienten	t-Statistik
Fehlspezifikation	ja	ja	ja
Heteroskedastie	nein	ja	ja
Autokorrelation	nein	ja	ja
Fehler in den Variablen	ja	ja	ja
Mit exogenen Variablen korrelierte Störgröße	ja	ja	ja

8.1 QUALITATIVE EXOGENE VARIABLE

$$D_1 = \begin{cases} 1 & \text{wenn in Betrieb weniger als 11 Beschäftigte} \\ 0 & \text{sonst} \end{cases}$$

$$D_2 = \begin{cases} 1 & \text{wenn in Betrieb 11 – 50 Beschäftigte} \\ 0 & \text{sonst} \end{cases}$$

$$D_3 = \begin{cases} 1 & \text{wenn in Betrieb > 50 Beschäftigte} \\ 0 & \text{sonst} \end{cases}$$

In Regression eine Variable weglassen:

$$\ln W = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + u$$

ACHTUNG: In diesem Fall müssen β_1 und β_2 in Bezug auf die weggelassene Kategorie 3 interpretiert werden.

DUMMY-VARIABLEN-FALLE

Multikollinearität trifft auf, wenn alle Dummies enthalten sind.

F-Test für Einfluss einer Dummy-Variable-Gruppe

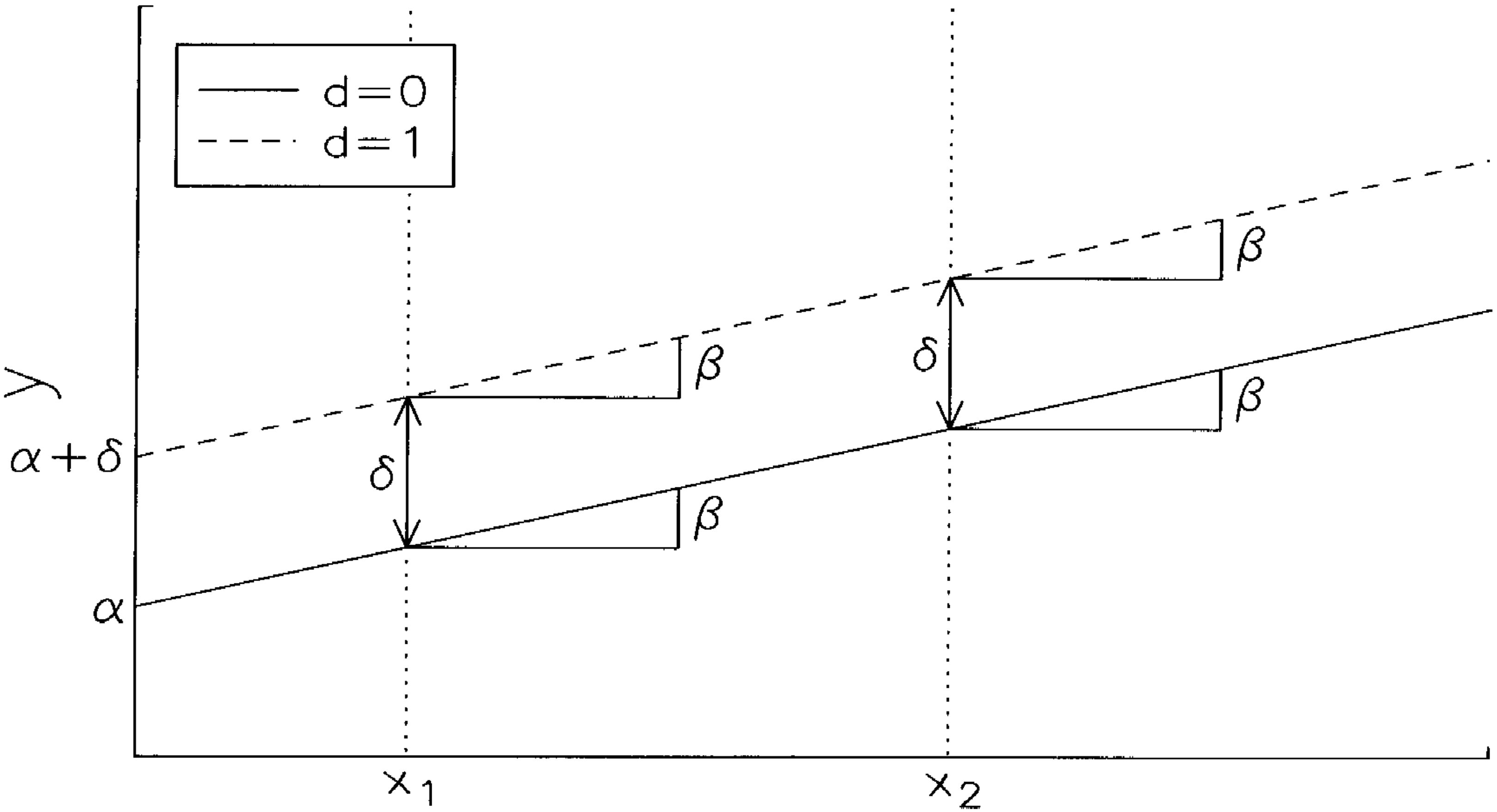
$\ln W = \beta_0 + \beta_1 S_1 + \beta_2 S_2 + \beta_3 D_1 + \beta_4 D_2 + u$

wobei S_1 und S_2 Sektoren (S_3 weglassen) und D_1 und D_2 Betriebsgröße (D_3 weglassen)

vgl. mit $\ln W = \beta_0 + \beta_1 S_1 + \beta_2 S_2$ um Einfluss der Betriebsgröße zu testen.

Effekt von D1 und D2 müssen in Bezug auf D3 interpretiert werden!

INTERPRETATION DER KOEFFIZIENTEN IM LINEAREN MODELL



INTERPRETATION DER KOEFFIZIENTEN IM LINEAREN MODELL

$$y = \alpha + \beta x + \delta d$$

$$\frac{\partial y}{\partial x} = \frac{\partial\left(\alpha + \beta x + \delta d\right)}{\partial x} = \beta$$

$$\frac{\Delta y}{\Delta d} = (\alpha + \beta x + \delta\, 1) - (\alpha + \beta x + \delta\, 0) = \delta$$

INTERPRETATION DER KOEFFIZIENTEN IM NICHT-LINEAREN MODELL