

Kurzfassung

Empirische Wirtschaftsforschung

„Is it magic?“

[WS09/10 – Prof. Urban]

Die Zusammenfassung dient lediglich zur Klausurvorbereitung, kann jedoch nicht Vorlesung und Übung ersetzen. Es kann keine Garantie für Vollständigkeit und Richtigkeit der Angaben gegeben werden. Weiter kann nicht garantiert werden, dass der Stoff in späteren Semestern nicht erweitert, gekürzt oder anderweitig geändert wurde! Eine Veröffentlichung oder kommerzielle Nutzung ist ausdrücklich verboten!

1. OLS-Schätzer – Methode der kleinsten Quadrate

❖ OLS – Ordinary Least Squares

- Schätzer der kleinsten Quadrate → Die Summe der quadrierten Fehlerterme soll minimiert werden. Würde man nur die Summe der Fehlerterme minimieren, heben sich negative und positive Fehler auf und man kommt zu sehr schlechten Lösungen. Nimmt man nur die Summe der Beträge der Fehler, werden sehr große Fehler nicht relativ stärker bestraft als kleine Fehler.

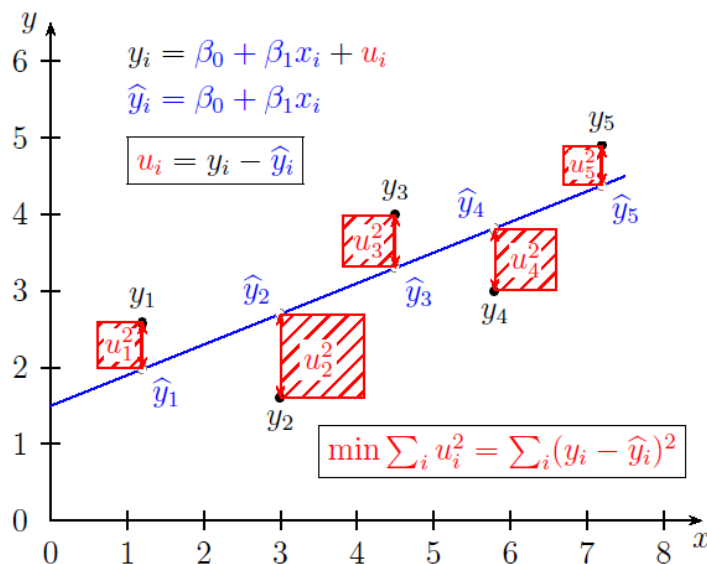


Abbildung 1

❖ OLS-Schätzer: $b_{OLS} = (X'X)^{-1} * X'y$

- β → „wahrer“ Parameter (Koeffizient)
- b_{OLS} → geschätzter Parameter

- geschätzte Werte (fitted value): $Xb_{OLS} = \hat{y}$

- geschätzte Fehler: $[I_n - X(X'X)^{-1}X'] * y = y - Xb_{OLS} = y - \hat{y} = e$
 - $y = \hat{y} + e$

- Fehlerquadratsumme:

$$y'[I_n - X(X'X)^{-1}X']y = y' * e = (e + \hat{y})' * e = e'e + \underbrace{\hat{y}' * e}_{=0} = e_1^2 + e_2^2 + \dots$$

da
orthogonal
zueinander

- sollte möglichst klein sein!

❖ **Wichtige Beweise:**

➤ **Herleitung des OLS-Schätzer:**

$$\begin{aligned} \argmin(e'e) &= \argmin[(y - Xb)' * (y - Xb)] \\ &= \argmin[y'y - (Xb)'y - Xb'y + (Xb)'Xb] \\ &= \argmin(y'y - 2Xb'y + b'X'Xb) = S(b) \end{aligned}$$

- Bedingung für Minimum: $\frac{\delta S(b)}{\delta b} = -2X'y + 2X'Xb = 0$
 - $b_{OLS} = (X'X)^{-1} * X'y$

➤ **Konsistenz**

- Ein Schätzer $\hat{\theta}$ eines Parameter θ ist konsistent, wenn gilt $\text{plim}(\hat{\theta}) = \theta$.
- Rechenregeln für „probability limits“:

➤ $\text{plim } c = c \text{ (} c=\text{const.)}$

➤ $\hat{\theta}_1$ und $\hat{\theta}_2$ = konsistente Schätzfunktion:

- $\text{plim}(\hat{\theta}_1 + \hat{\theta}_2) = \text{plim } \hat{\theta}_1 + \text{plim } \hat{\theta}_2$
- $\text{plim}(\hat{\theta}_1 \hat{\theta}_2) = \text{plim } \hat{\theta}_1 * \text{plim } \hat{\theta}_2$
- $\text{plim} \frac{\hat{\theta}_1}{\hat{\theta}_2} = \frac{\text{plim } \hat{\theta}_1}{\text{plim } \hat{\theta}_2}$

➤ Slutsky-Theorem: $\text{plim } h(\hat{\theta}) = h(\theta)$

- **Beweis:** $\text{plim}(b_{OLS}) = \beta$

$$\begin{aligned} \text{plim } b_{OLS} &= \text{plim } (X'X)^{-1} X'y = \text{plim } (X'X)^{-1} X' * (X\beta + \epsilon) \\ &= \text{plim } \beta + \text{plim } (X'X)^{-1} X' \epsilon = \beta + \underbrace{\text{plim} \left(\frac{X'X}{n} \right)^{-1}}_{\substack{<\infty \\ \text{Anna hme 5b}}} * \underbrace{\text{plim} \left(\frac{X'\epsilon}{n} \right)}_{\substack{=0 \\ \text{Anna hme 3 +5a}}} = \beta \end{aligned}$$

Die letzten beiden Annahmen müssen explizit gezeigt werden (siehe Skript)!

➤ Unverzerrtheit (Erwartungstreue):

- Ist ein Schätzer Erwartungstreu, so entspricht sein Erwartungswert dem wahren Wert des zu schätzenden Parameters!

- $E(b_{OLS}) = \beta$

- Beweis: $E[b_{OLS} - \beta | X] = E[(X'X)^{-1}X'y - \beta | X]$

$$= E[(X'X)^{-1}X' \overset{=y \text{ (Annahme 1)}}{(X\beta - \epsilon)} - \beta | X] = E[\beta - \beta - (X'X)^{-1}X'\epsilon | X]$$

$$= (X'X)^{-1}X' * \underset{=0}{E[\epsilon | X]} = \mathbf{0}$$

Annahme 3

2. Annahmen des OLS-Schätzers:

1. Annahme: **linearer Zusammenhang**

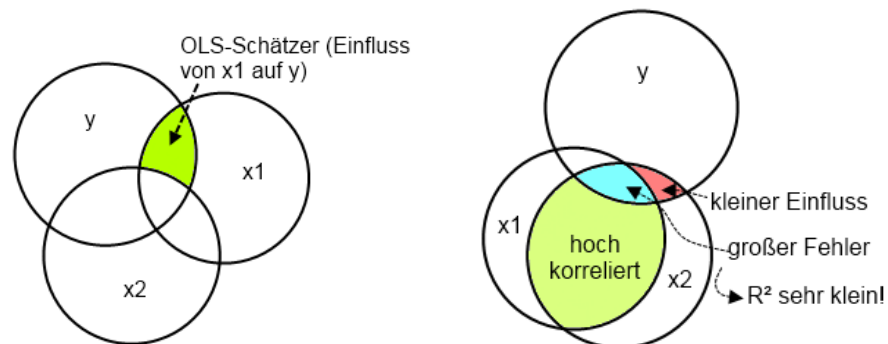
$$y = X\beta + \epsilon$$

- ❖ Kausalität von X auf y und nicht umgekehrt!
- ❖ keine *Heterogenität* im Zusammenhang zwischen X und y
 - z.B. kein **Strukturbruch**, durch den für eine Teilstichprobe ein Zusammenhang gilt & für eine andere nicht!
- ❖ Verletzt durch:
 - Strukturbruch
 - Kausalität von y auf X

2. Annahme: „Ausschluss perfekter Multikollinearität“

$$\text{Rang}(X) = k \quad (\text{implizit: } n \geq k)$$

- ❖ vollständiger Rang
- ❖ Variablen müssen linear unabhängig sein → sichert Invertierbarkeit von $(X'X)$
- ❖ Verletzt durch: linear abhängige Variablen
 - Problem der perfekter Multikollinearität ($\text{Korrelation} = 1$)
 - keine OLS-Schätzung mehr möglich!
- ❖ Lösung: Eine der anderen Variablen ausschließen, diese findet sich dann im Fehlerterm wieder!
 - Hohe Korrelation ($=0,95$) ist kein Problem!
 - Partial Effect:
 - Der Regressionskoeffizient b_i misst die gemeinsame Variation von y und x_i , die nicht gleichzeitig auch eine Variation mit einer anderen Variablen x_j ist!
 - *Beispiel:* b_2 misst die Veränderung des Erwartungswertes von y gegeben einer Veränderung von x_2 bei Konstanz von x_1 !
- ❖ Venn-Diagramm:



- Der türkise Bereich, ist der gleichzeitige Einfluss von x_1 und x_2 auf y und wird nicht im OLS-Schätzer erfasst (*Partial Effekt*). Dieser Einfluss findet sich im Fehlerterm wieder.

3. Annahme: Unkorreliertheit

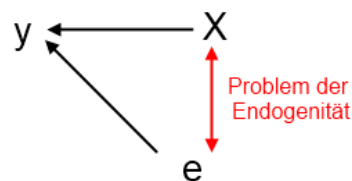
$$E[\varepsilon_i|X] = 0, \quad \text{Cov}(x, \varepsilon) = 0$$

❖ Verletzt durch: Endogenität

➤ $\text{Cov}(x, \varepsilon) \neq 0$

➤ b_{OLS} ist **verzerrt & inkonsistent!**

- Annahme 3 wird in beiden Beweisen benutzt!



❖ Auslöser:

➤ Simultanität (*simultane Kausalität*):

- Simultane Kausalität liegt vor, wenn die Kausalität in beide Richtungen geht! Wenn also X kausal auf y wirkt, aber auch y kausal auf X!

▪ Beispiel:

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

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

- ♦ Im Gleichgewicht gilt $L = L^d = L^s$ (daher ist **W endogen** gegeben)

- ♦ → unterschiedliche X und Z Werte, daher liegt b_{OLS} von $L = b_0 + b_{OLS} * W + \dots + e$ irgendwo zwischen β_1 & γ_1

- ♦ → also verzerrt und inkonsistent!

- ♦ positive Korrelation zwischen erklärender Variable **W** und dem wahren Fehlerterm $\varepsilon \rightarrow \text{Cov}(\varepsilon_i, W_i) > 0$

- ♦ Ursache: **Simultanität**, mit der die abhängige Variable **L** und die erklärende Variable **W** gleichzeitig bestimmt werden!

➤ Messfehler: $X^* = X - \underbrace{\mu}_{\text{Messfehler}} \rightarrow y = b_0 + b_1 X^* + \underbrace{b_1 \mu + \varepsilon}_{\text{Fehler } v}$

- $\text{Cov}(X^*, v) \neq 0 \rightarrow$ Annahme 3 verletzt!
- je größer τ_μ^2 (= Messfehlervariation), desto mehr wird b_1 gegen Null verzerrt! (*Attenuation Bias*)

- Konservativer Fehler: Ist b_1 trotzdem signifikant $\neq 0$, ist die Größe von b_1 zwar fraglich, aber der Effekt ist gesichert!

➤ **rationale Erwartungen**

- ...

➤ **Zeitverzögerung der abhängigen Variable**

- ...

➤ **Omitted Variable Bias** (*Ausschluss einer relevanten Variable*)

- Effekt auf OLS-Schätzer hängt davon ab, ob die ausgeschlossene Variable mit den verbleibenden erklärenden Variablen korreliert ist!

$$y = \beta_0 + \beta_1 X + \beta_2 Z + \varepsilon \quad \text{mit } \text{Cov}(X, Z) \neq 0$$

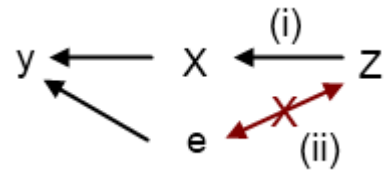
$$y = b_0 + b_1 X + u_i$$

- Wenn $\text{Cov}(X, Z) \neq 0$ ist, dann beinhaltet b_1 nicht nur den Einfluss von X auf y, sondern auch noch den Einfluss von Z auf y, da wir nicht explizit auf Z kontrollieren!
- **Folge:** Verzerrung von b_1
 - ♦ Wenn ausgeschlossene Variablen positiv mit verbleibender Variable korreliert sind, ändert sich der Koeffizient b_1 .

❖ **Lösung: Instrumentenvariablen-Schätzer** (*IV-Schätzer*)

➤ **Beheben mit „Instrument“ Z**

- (i) Z erklärt Einfluss von X auf y!
- (ii) Z ist **nicht** mit ε korreliert!



➤ **Anforderungen an ein gutes Instrument:**

- X & Z möglichst hoch korreliert (*Testbar*)
 - ♦ $\text{Cor}(X; Z) \neq 0 \rightarrow Z$ auf X regressieren und Signifikanz testen!
 - ♦ $X = \beta_0 + \beta_1 \dots + \beta_i * Z + \varepsilon \rightarrow$ Signifikanz β_i ?
- Z & ε nicht korreliert! (*muss theoretisch begründet werden*)

➤ **Schätzer in zwei Stufen** (*OLS in 2 Stufen $\rightarrow$ IV-Schätzer*)

- Two-Stage Least Squares Estimator
- **1. Stufe:** *OLS-Regression von X auf Z $\rightarrow \hat{X}$*
 - ♦ ε Teil (der die Verzerrung verursachte) wird aus X „rausgeschnitten“ übrig bleibt der gute Teil von X, mit dem in der zweiten Stufe normal regressiert wird.
- **2. Stufe:** *OLS-Regression von y auf $\hat{X}$*
 - ♦ $\hat{X}$ an Stelle von X-Regression mit y, ohne dass Annahme $\text{Cov}(\hat{X}; \varepsilon) = 0$ verletzt wird!

- in Stata:
 - ivreg y x1 (x2=z)
 - ivreg y x1 (x2=z), first robust
 - ♦ *first* → auch erste Stufe wird als Tabelle ausgegeben

4. Annahme: **Homoskedastizität & keine Autokorrelation**

$$E[\varepsilon * \varepsilon' | X] = I_n * \tau^2 < \infty$$

- ❖ $Cov(\varepsilon_i; \varepsilon_j) = 0 \quad \forall i \neq j$ → keine Autokorrelation
- ❖ $Var(\varepsilon_i) = Var(\varepsilon_j) = \tau^2 \quad \forall j, i$ → **Homoskedastizität!!**

- Fehlerterme besitzen die gleiche Varianz, d.h. sie folgen der gleichen Verteilung (*i.i.d.*)!

❖ Verletzt durch: **Heteroskedastizität**

- Fehlerterme haben unterschiedliche Varianzen!
- Kann schnell im Streudiagramm (*Scatterplot*) erkannt werden.

- $Var(\varepsilon_i)$ ist wichtig für die t-Statistik!

$$\text{♦ } t = \frac{\widehat{b_k} - \mu}{s / \sqrt{n}} \quad s = \sqrt{\frac{e' e * S_{kk}}{n-k}}$$

- ♦ s Standardabweichung (*Std. Err.*)
- ♦ $n = 1$
- ♦ $\mu = 0$

- Standard Error: wird unter der Annahme berechnet, dass die Varianzen der Residuen identisch sind: $e' e$: *Skalar*
 - Standard Error ist demnach falsch, falls in Wahrheit Heteroskedastizität vorliegt!
- Als anschauliches Beispiel dient die *Wettervorhersage*, bei welcher die Varianz der Fehler mit größerem Abstand vom heutigen Tag zunimmt. Die Vorhersagefehler also stärker streuen.

- ❖ **Folge:** Die Änderung der Varianzen hat somit Auswirkungen auf die t-Statistik!
→ Aussage über Signifikanzniveau wird verfälscht!
 - OLS-Schätzer **bleib unverzerrt & konsistent!** (*In den Herleitungen wird Annahme 4 nicht verwendet*)
 - OLS hatte die kleinste Varianz, da sich die Varianz bei Heteroskedastizität ändert, werden **t-Test & F-Test ungültig!** Die OLS-Schätzer für die Koeffizienten sind *nicht mehr effizient*. Die geschätzten Standardfehler der Koeffizienten sind verzerrt, sodass die Hypothesentests ungültig werden.
 - R^2 bleibt **gültig** als Maß der Güte der Regression!
- ❖ **Lösung:** konsistente Kovarianzmatrix → „**White**“-Matrix

5. Annahme:

- ❖ $[x_i; \varepsilon_i]; i = 1, \dots, n$ ist ein Vektor von **i.i.d. Zufallsvariablen**, die nach einer *Zufallsauswahl* aus einer Grundgesamtheit „gezogen“ wurden!
 - *i. i. d.* identically independently distributed
 - $Cov(x_i; \varepsilon_i) = 0$
 - Verletzt durch: nicht *i.i.d.* verteilte Stichproben!
- ❖ X folgt einer Multivariaten Verteilung mit
 - $E[X] = \mu_x < \infty$
 - $Var[X] = \sum_{xx} < \infty$ endliche Varianz
 - $E[x_i x_j x_k x_l] = \phi_{ijkl} < \infty$ endliches 4. Moment

6. Annahme:

- ❖ **kleine Stichprobe:** $\varepsilon | X \sim N[0, I_n \tau^2]$
 - Verletzt, wenn Störterme nicht Normalverteilt sind!
- ❖ **große Stichproben:** $n \rightarrow \infty$ (*gilt ab $n > 30$*)
 - Annahme der Normalverteilung ist hier nicht mehr nötig.

3. Tests

❖ **t-Test:** $H_0: \hat{\beta}_k = 0 \rightarrow X_k \text{ erklärt } y \text{ nicht!}$

- Test auf nur einen Koeffizienten und eine Restriktion!

$$t = \frac{\widehat{b}_k - \mu}{s/\sqrt{n}}$$

- Variablen: Standard Error (s), Erwartungswert ($\mu = 0$), n normal = 1
- Wenn $t < \text{als } t \text{ von } \phi(t) = 95\%$ (5% Signifikanzniveau), dann kann H_0 nicht abgelehnt werden! Der Koeffizient ist also insignifikant!

❖ **F-Test:** $F = \left(\frac{R^2}{1-R^2} \right) * \frac{n-k-1}{j}$

- $n-k-1$ nur, wenn man davon ausgeht, dass bei uns für gewöhnlich $k=j$ gilt, da wir alle Koeffizienten testen. -1 kommt hinzu, falls eine Konstante vorhanden ist, ansonsten fällt dies weg.
- Test für mehrere Koeffizienten (z.B. $\beta_1 = \beta_2 = \dots = 0$) oder mehrere Restriktionen!
- Variablen:
 - Gütemaß R^2 aus Stata Output
 - n Anzahl der Beobachtungen (*Number of obs*)
 - k Anzahl der geschätzten Koeffizienten
 - j Anzahl der Restriktionen

❖ **Bestimmtheitsmaß R^2** $R^2 = \frac{s_y^2}{s_y^2} = 1 - \frac{s_e^2}{s_y^2}$

- R^2 ist der Anteil der Variation der zu erklärenden Variable y , der durch das Regressmodell erklärt wird.
- Bei $R^2 = 0$ hat das Modell überhaupt keine Erklärungskraft.
- Die Höhe von R^2 verschiedener Schätzmodelle ist nicht untereinander vergleichbar.
- R^2 steigt immer, wenn zusätzliche erklärende Variablen hinzukommen. $\overline{R^2}$ sinkt hingegen, wenn die zusätzliche Variable nur eine geringe oder keine Erklärungskraft besitzt!

4. Interpretation der marginalen Effekte

- ❖ **lin-log-Modell** $y = \beta_0 + \beta_1 \ln(X) + \epsilon \rightarrow \frac{\Delta x}{x_0} = 1\% \rightarrow \Delta y = 0,01\beta$
 - Eine relative Änderung von X um **1%**, geht einher mit einer absoluten Änderung (des Erwartungswertes) von y um $\beta_1 * 0,01$!
- ❖ **log-lin-Modell** $\ln(y) = \beta_0 + \beta_1 X + \epsilon \rightarrow \Delta x = 1 \rightarrow \frac{\Delta y}{y_0} = \beta * 100\%$
 - Eine absolute Änderung von X um eine Einheit ($\Delta X = 1$), bedeutet eine relative Änderung von y um $\beta_1 * 100\%$!
- ❖ **log-log-Modell** $\ln(y) = \beta_0 + \beta_1 * \ln(X) + \epsilon \rightarrow \frac{\Delta x}{x_0} = 1\% \rightarrow \frac{\Delta y}{y_0} = \beta\%$
 - Eine relative Änderung von X um **1%**, geht einher mit einer relativen Änderung von y um $\beta_1\%$ → **Elastizität**!

5. STATA Output

```
. reg price sqrft bdrms bdrms_sqrft
```

Source	SS	df	MS
Model	600164.068	3	200054.689
Residual	317690.438	84	3782.02902
Total	917854.506	87	10550.0518

Number of obs =	88
F(3, 84) =	52.90
Prob > F =	0.0000
R-squared =	0.6539
Adj R-squared =	0.6415
Root MSE =	61.498

P Wert für die
gesamte
Regression
Gütemaß R^2

price	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
sqrft	.0326205	.0436415	0.75	0.457	-.0541655 .1194065
bdrms	-35.95534	24.01237	-1.50	0.138	-83.70657 11.79589
bdrms_sqrft	.023448	.0101573	2.31	0.023	.0032491 .043647
_cons	181.6908	92.18885	1.97	0.052	-1.636817 365.0184

Koeffizienten der
Schätzgleichung

Standard
Error

P Wert für die
einzelnen
Koeffizienten

- ❖ **Number of obs:** Anzahl der Beobachtungen
- ❖ **Coef.:** Koeffizienten der Schätzgleichung
 - Hier: $y = 181,69 + 0,03262 * sqrft - 35,955 * bdrms + 0,023448 * bdrms_sqrft + \epsilon$
- ❖ **t-Test Nullhypothese:** $H_0: b_k = 0$
 - Hypothese: x_k erklärt y **nicht**!
- ❖ **t:** Wert des t-Test! Hier einfach $t = \frac{Coef.}{Std.Err.}$

- ❖ $P > |t|$: Marginale Wahrscheinlichkeit mit der der Koeffizient b_k in der Grundgesamtheit = 0 ist!
 - $> 0,05 \rightarrow$ **nicht signifikant!** Der P Wert gibt das maximale Signifikanzniveau an, auf dem die Hypothese H_0 nicht abgelehnt werden kann!
 - Die Wahrscheinlichkeit das x_k keinen Einfluss auf y hat ist größer als 5%.
 - Unsere Nullhypothese kann auf dem 5% Niveau also nicht abgelehnt werden! Und b_k hat so wahrscheinlich keinen erklärenden Einfluss auf y !
 - $< 0,05 \rightarrow$ **signifikant!** Die Nullhypothese kann also auf dem 5% Signifikanzniveau zugunsten der Alternative abgelehnt werden.
 - Die Alternative ist, dass b_k in der Grundgesamtheit y erklärt!
- ❖ Prob > F: P-Wert für die gesamte Regression

6. Binäre abhängige Variable

- ❖ Neue Schätzer für binäre Variablen $y \in \{0,1\}$
- ❖ Problem bei OLS-Schätzer: Bei binären Variablen ist eine sinnvolle Interpretation nicht mehr möglich
 - $P(y = 1|X) = X_i\beta \rightarrow X_i\beta > 1$ oder $X_i\beta < 0$
 - führt zu Unsinnswahrscheinlichkeiten!
- ❖ Lösung: Funktion $F(X_i\beta)$ um OLS interpretierbar zu machen und zwischen 0 und 1 zu halten!
 - $P(y = 1|X) = F(X_i\beta)$
 - Definitionsbereich $[-\infty, \infty]$
 - Wertebereich: $[0,1]$
- ❖ Probit-Schätzmodell: $F(X_i\beta) = \Phi(X_i\beta)$ Standardnormalverteilung
- ❖ Logit-Schätzmodell: $P(y = 1|X) = \frac{e^{X_i\beta}}{1+e^{X_i\beta}} = \Lambda(X_i\beta)$ logistische Verteilung
- ❖ Schätzer werden mit der *Max-Likelihood-Methode* hergeleitet.