



Übung zur Empirischen Wirtschaftsforschung

V. Lineares Regressionsmodell

- 1 Das lineare Regressionsmodell
- 2 Die Methode der Kleinsten Quadrate
- 3 Güteeigenschaften des KQ-Schätzers
- 4 Bestimmtheitsmaß
- 5 Standardfehler
- 6 Testverfahren

1 Das lineare Regressionsmodell

- Im linearen Regressionsmodell besteht zwischen einer endogenen Variable y (Regressand, erklärte bzw. abhängige Variable) und exogenen Variablen x_i (Regressor, erklärende bzw. unabhängige Variable) ein **funktionaler Zusammenhang**:

$$y_t = \beta_1 + \beta_2 \cdot x_t + \varepsilon_t \text{ (bei einem Regressor)}$$

$$y_t = \beta_1 + \beta_2 \cdot x_{2,t} + \beta_3 \cdot x_{3,t} + \dots + \beta_k \cdot x_{k,t} + \varepsilon_t \text{ (bei } k - 1 \text{ Regressoren)}$$

- Abkürzungen:

β_1 : Absolutglied

β_2 : Koeffizient der exogenen Variable

t : Zeitindex mit $t = 1, 2, \dots, T$

ε : Störterm (Residuum, Fehler)

- Im Normalfall existiert ein **Störterm ε** , da der funktionale Zusammenhang zwischen y und x nicht zu einer 100%igen Erklärung führt.

- Bei der **Modellierung** unterscheidet man:

- ökonomisches Modell:

$$y_t = \beta_1 + \beta_2 \cdot x_t$$

- empirisches Modell:

$$y_t = \beta_1 + \beta_2 \cdot x_t + \varepsilon_t$$

- ökonometrisches Modell:

$$y_t = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_t + \hat{\varepsilon}_t$$

bzw. $\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_t$

- **Ziel:** Möglichst kleine Störterme.

- Es gibt viele Möglichkeiten, um die Störterme zu minimieren:

- Minimierung der Summe der Beträge der Störterme ($\sum_{t=1}^T |\hat{\varepsilon}_t|$)

- Minimierung des maximalen Störterms ($\max_{t \in \{1, \dots, T\}} |\hat{\varepsilon}_t|$)

– ...

- Die Methode der Kleinsten Quadrate minimiert die **Summe der Residuenquadrate** ($\sum_{t=1}^T \hat{\varepsilon}_t^2$)

- Die Quadrierung verhindert, dass sich Residuen mit unterschiedlichen Vorzeichen aufheben.

2 Die Methode der Kleinsten Quadrate

- Im linearen Regressionsmodell mit einem Regressor ergibt sich für die Summe der Residuenquadrate:

$$\sum_{t=1}^T \hat{\varepsilon}_t^2 = \sum_{t=1}^T (y_t - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_t)^2$$

- Die Minimierung der Summe der Residuenquadrate kann somit über die Minimierung des Ausdrucks auf der rechten Seite der vorherigen Gleichung erfolgen.
- Die Koeffizienten $\hat{\beta}_1$ und $\hat{\beta}_2$ werden über die partiellen Ableitungen berechnet.
- Vorgehensweise:

Schritt 1: Partielle Ableitung nach $\hat{\beta}_1$ Null setzen ($\frac{\partial \sum \hat{\varepsilon}_t^2}{\partial \hat{\beta}_1} \doteq 0$)

Schritt 2: Partielle Ableitung nach $\hat{\beta}_2$ Null setzen ($\frac{\partial \sum \hat{\varepsilon}_t^2}{\partial \hat{\beta}_2} \doteq 0$)

Schritt 3: Berechnung der Koeffizienten $\hat{\beta}_1$ und $\hat{\beta}_2$ anhand der vorliegenden Daten

Schritt 1

$$\frac{\partial \sum \hat{\varepsilon}_t^2}{\partial \hat{\beta}_1} = \sum 2 \cdot (y_t - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_t) \cdot (-1) \doteq 0$$

$$\Rightarrow 2 \cdot (-1) \cdot \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_t) = 0$$

$$\Rightarrow \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_t) = 0$$

$$\Rightarrow \sum y_t - \sum \hat{\beta}_1 - \sum \hat{\beta}_2 \cdot x_t = 0$$

$$\Rightarrow \sum y_t - \hat{\beta}_1 \cdot T - \sum \hat{\beta}_2 \cdot x_t = 0$$

$$\Rightarrow \bar{y} - \hat{\beta}_1 - \hat{\beta}_2 \cdot \bar{x} = 0$$

$$\Rightarrow \hat{\beta}_1 = \bar{y} - \hat{\beta}_2 \cdot \bar{x}$$

Schritt 2

$$\begin{aligned}\frac{\partial \sum \hat{\varepsilon}_t^2}{\partial \hat{\beta}_2} &= \sum 2 \cdot (y_t - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_t) \cdot (-x_t) \stackrel{!}{=} 0 \\ \Rightarrow 2 \cdot (-1) \cdot \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_t) \cdot x_t &= 0 \\ \Rightarrow \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_t) \cdot x_t &= 0 \\ \Rightarrow \sum y_t \cdot x_t - \sum \hat{\beta}_1 \cdot x_t - \sum \hat{\beta}_2 \cdot x_t \cdot x_t &= 0 \\ \Rightarrow \sum x_t \cdot y_t - \sum \hat{\beta}_1 \cdot x_t - \sum \hat{\beta}_2 \cdot x_t^2 &= 0 \\ \Rightarrow \overline{x \cdot y} - \hat{\beta}_1 \cdot \bar{x} - \hat{\beta}_2 \cdot \overline{x^2} &= 0 \\ \Rightarrow \overline{x \cdot y} - (\bar{y} - \hat{\beta}_2 \cdot \bar{x}) \cdot \bar{x} - \hat{\beta}_2 \cdot \overline{x^2} &= 0 \\ \Rightarrow \overline{x \cdot y} - \bar{x} \cdot \bar{y} + \hat{\beta}_2 \cdot \bar{x} \cdot \bar{x} - \hat{\beta}_2 \cdot \overline{x^2} &= 0 \\ \Rightarrow \overline{x \cdot y} - \bar{x} \cdot \bar{y} = \hat{\beta}_2 \cdot \overline{x^2} - \hat{\beta}_2 \cdot \bar{x}^2 & \\ \Rightarrow \hat{\beta}_2 \cdot (\overline{x^2} - \bar{x}^2) = \overline{x \cdot y} - \bar{x} \cdot \bar{y} & \\ \Rightarrow \hat{\beta}_2 = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} &\end{aligned}$$

Schritt 3

Berechnung von $\hat{\beta}_2$ und $\hat{\beta}_1$ anhand der vorliegenden Werte von x_t und y_t .

- Die **Schätzwerte der abhängigen Variable** \hat{y}_t können anhand der Koeffizienten $\hat{\beta}_1$ und $\hat{\beta}_2$ für jeden Zeitpunkt t berechnet werden.
- Die Schätzwerte \hat{y}_t und die Originalwerte y_t stimmen in der Regel nicht überein. Die Residuen $\hat{\varepsilon}_t$ entsprechen der Differenz von Schätz- und Originalwerten.
- Die **Summe der Residuen** ist immer Null. Dies folgt aus der Null gesetzten partiellen Ableitung nach $\hat{\beta}_1$

$$\frac{\partial \sum \hat{\varepsilon}_t^2}{\partial \hat{\beta}_1} = (-2) \sum (y_t - \hat{\beta}_1 - \hat{\beta}_2 \cdot x_t) = (-2) \sum \hat{\varepsilon}_t \doteq 0.$$

- Wenn $\hat{\beta}_1, \hat{\beta}_2$ anders gewählt werden, ist die Summe der Residuenquadrate größer.
- **Interpretation** des berechneten Koeffizienten $\hat{\beta}_2$:
 - **Additives Modell** ($y_t = \beta_1 + \beta_2 \cdot x_t$):
Wenn die Variable x eine Einheit höher liegt, liegt die endogene Variable y um β_2 Einheiten höher.
 - **Multiplikatives Modell** ($\log(y_t) = \beta_1 + \beta_2 \cdot \log(x_t)$):
Wenn die Variable x ein Prozent höher liegt, liegt die endogene Variable y um β_2 Prozent höher.
 - **Semi-log Modell** ($\log(y_t) = \beta_1 + \beta_2 \cdot x_t$):
Wenn die Variable x eine Einheit höher liegt, liegt die endogene Variable y um β_2 Prozent höher.

- Im Normalfall haben mehrere Faktoren Einfluss auf die endogene Variable.

$$\Rightarrow y_t = \beta_1 + \beta_2 x_{2,t} + \beta_3 x_{3,t} \dots + \beta_k x_{k,t} + \varepsilon_t$$

- In Matrixschreibweise ergibt sich:

$$Y = X\beta + \varepsilon$$

- Abkürzungen:

Y : Vektor der Beobachtungen der endogenen Variable

X : Matrix der Beobachtungen der erklärenden Variablen

β : Vektor der zu schätzenden Koeffizienten

ε : Vektor der Störterme

- Wie im Modell mit einem Regressor soll auch im multiplen Modell die Summe der Residuenquadrate minimiert werden.
- Der Koeffizientenvektor $\hat{\beta}$, der die Summe der Residuenquadrate minimiert, kann wieder über die partiellen Ableitungen berechnet werden.

- Unter Verwendung der Ableitungsregeln von Matrizen und Vektoren ergeben sich folgende Normalgleichungen:

$$\frac{\partial \sum \hat{\varepsilon}_i^2}{\partial \hat{\beta}} = -2X^T Y + 2X^T X \hat{\beta} = 0$$

- Somit ergibt sich für $\hat{\beta}$ der KQ-Schätzer

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

- Ob eine Interpretation der Koeffizienten ökonomisch sinnvoll ist, hängt vom Modell, statistischen Gütekriterien und ökonometrischen Testverfahren ab!

3 Güteeigenschaften des KQ-Schätzers

- Der KQ-Schätzer ist der **beste lineare erwartungstreue Schätzer**, falls gilt:
 - Die Matrix der Regressoren (X) ist exogen vorgegeben, enthält alle relevanten Variablen und besitzt vollen Spaltenrang (k).
 - Das Residuum einer Periode ist unkorreliert mit den Regressoren und den Residuen anderer Perioden, d.h. es existiert keine Systematik in den Fehlern.
 - Die Residuen sind normalverteilt mit dem Erwartungswert $E(\varepsilon) = 0$ und der Varianz-Kovarianzmatrix $E(\varepsilon^2) = \sigma^2 \cdot I_T$, wobei I_T die Einheitsmatrix der Größe $T \times T$ ist.

3.1 Was bedeutet „bester linearer erwartungstreuer Schätzer“?

- **Erwartungstreue**:
 - Ist ein Schätzer erwartungstreu, so entspricht der Erwartungswert des Schätzers dem wahren Parameterwert ($E(\hat{\beta}) = \beta$).
 - Liegen mehrere Schätzungen vor ergibt sich im Durchschnitt der wahre Parameterwert.
 - Ist es nicht möglich einen erwartungstreuen Schätzer zu erhalten, sollte er zumindest die Eigenschaft der **Konsistenz** aufweisen.
 - Ein Schätzer $\hat{\beta}$ ist konsistent, falls der Schätzer mit zunehmendem Stichprobenumfang in Wahrscheinlichkeit gegen den wahren Parameterwert β geht
 $\Rightarrow \lim_{T \rightarrow \infty} P(|\hat{\beta} - \beta| < \epsilon) = 1.$
- **Linearität**:
 - Linearität bedeutet, dass sich ein Schätzer $\hat{\beta}_i$ als Linearkombination der Beobachtungen der endogenen Variable y_t schreiben lässt ($\hat{\beta}_i = c_1 y_1 + \dots + c_T y_T$).
- **Effizienz**:
 - Im Allgemeinen existieren mehrere lineare erwartungstreue Schätzer.
 - Der Schätzer mit der kleinsten Varianz in dieser Menge von Schätzern wird „effizient“ genannt.

4 Bestimmtheitsmaß

- Im einfachen linearen Regressionsmodell kann die Variation der zu erklärenden Variable (TSS) in zwei Komponenten zerlegt werden

$$\underbrace{\sum_{t=1}^T (y_t - \bar{y})^2}_{TSS} = \underbrace{\sum_{t=1}^T (\hat{\beta}_1 + \hat{\beta}_2 \cdot x_t - \bar{y})^2}_{ESS} + \underbrace{\sum_{t=1}^T \hat{\varepsilon}_t^2}_{RSS}$$

- Das Bestimmtheitsmaß R^2 gibt den Anteil der erklärten Varianz (ESS) an, d.h. wieviel Prozent der Varianz durch das Modell erklärt werden

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}.$$

- Das R^2 bewegt sich im Wertebereich $[0; 1]$. Je größer der ermittelte Wert, desto größer ist die Erklärungskraft des Modells.
- Normalerweise ist das R^2 bei Zeitreihendaten höher als bei Querschnittsdaten.
- Nimmt man weitere exogene Variablen in das Modell auf, erhöht sich das R^2 der Schätzung. Da ein Modell mit zu vielen erklärenden Variablen nicht wünschenswert ist, wird das korrigierte Bestimmtheitsmaß \bar{R}^2 um die Anzahl der Freiheitsgrade bereinigt

$$\bar{R}^2 = 1 - \left(\frac{RSS}{TSS}\right) \left(\frac{T-1}{T-k}\right).$$

- Wegen der bestrafenden Berücksichtigung der Anzahl der Regressoren kann das korrigierte Bestimmtheitsmaß auch negative Werte annehmen.

5 Standardfehler

- Ein wichtiges statistisches Gütekriterium ist der Standardfehler, der als Wurzel aus der Varianz wichtige Hinweise über die Streuung der Daten liefert.
- Je geringer der Standardfehler, desto höher die Zuverlässigkeit der Schätzergebnisse und deren Interpretation.
- Zu unterscheiden sind der Standardfehler der Originalwerte der endogenen Variable, der Standardfehler der Schätzung sowie der Standardfehler der Koeffizienten.
- Berechnung des Standardfehlers σ der Originalwerte:

$$\sigma^2 = \frac{1}{T-1} \cdot \sum (y_t - \bar{y})^2$$

- Berechnung des Standardfehlers $\hat{\sigma}$ der Schätzung:

$$\hat{\sigma}^2 = \frac{1}{T-k} \cdot \sum \hat{\varepsilon}_t^2$$

- Der Standardfehler der Schätzung ist in der Regel kleiner als der Standardfehler der Originalwerte.
- Berechnung des Standardfehlers des Absolutglieds $\hat{\beta}_1$:

$$\widehat{Var}(\hat{\beta}_1) = \hat{\sigma}^2 \cdot \frac{\bar{x}^2}{T \cdot (\bar{x}^2 - \bar{x}^2)}$$

- Berechnung des Standardfehlers des Koeffizienten $\hat{\beta}_2$:

$$\widehat{Var}(\hat{\beta}_2) = \hat{\sigma}^2 \cdot \frac{1}{T \cdot (\bar{x}^2 - \bar{x}^2)}$$

6 Testverfahren

6.1 Der t-Test

- Mit dem t -Test kann überprüft werden, ob ein geschätzter Koeffizient einem theoretisch angenommenen Wert entspricht. Hierzu wird der sogenannte **t-Wert** genutzt.
- Es kann gezeigt werden, dass die Größe

$$t\text{-Wert} = \frac{\hat{\beta}_i - \beta_i}{\sqrt{\widehat{Var}(\hat{\beta}_i)}}$$

einer t -Verteilung mit $T - k$ Freiheitsgraden folgt.

- Die t -Verteilung gleicht sich mit steigender Anzahl an Freiheitsgraden immer mehr der $\mathcal{N}(0, 1)$ -Verteilung an, weshalb die t -Werte ab mindestens 30 Beobachtungen mit dem kritischen Wert der $\mathcal{N}(0, 1)$ -Verteilung verglichen werden können.
- Durchführung eines t -Tests:
 - Einsetzen eines ökonomisch interessierenden Wertes $\tilde{\beta}_i$ für β_i in t -Wert.
 - Die Nullhypothese (H_0) lautet also $\beta_i = \tilde{\beta}_i$. In den meisten Fällen soll getestet werden, ob ein Regressor überhaupt einen signifikanten Beitrag zur Erklärung der endogenen Variable liefert (**Signifikanztest**).

$$\Rightarrow \beta_i = 0$$

- Formulierung einer Alternativhypothese. Es werden folgende drei Fälle unterschieden:
 - * $\beta_i \neq \tilde{\beta}_i$
 - * $\beta_i > \tilde{\beta}_i$
 - * $\beta_i < \tilde{\beta}_i$
- Wählen eines Signifikanzniveaus α . Ein Signifikanzniveau von α bedeutet, dass die Wahrscheinlichkeit, die Nullhypothese zu verwerfen, obwohl sie richtig ist, kleiner als α ist. Übliche Werte für α sind 10%, 5% und 1%.
- Ob die Nullhypothese verworfen wird, wird anhand kritischer Werte bestimmt. Diese sind bei gewähltem α durch
 - * $P(|t\text{-Wert}| > t_c) = \alpha$,
 - * $P(t\text{-Wert} > t_c) = \alpha$ bzw.
 - * $P(t\text{-Wert} < t_c) = \alpha$ gegeben.
- Bei gegebenem Signifikanzniveau α und ermittelten kritischen Werten t_c wird H_0 genau dann abgelehnt, wenn
 - * $|t\text{-Wert}| > t_c$,
 - * $t\text{-Wert} > t_c$ bzw.
 - * $t\text{-Wert} < t_c$.

- Wird ein beidseitiger Signifikanztest zum 5% Signifikanzniveau durchgeführt, so ist der kritische Wert $t_c = 1,96$.

⇒ **Faustregel**: Vergleiche, ob der t -Wert betragsmäßig größer als 2 ist, d.h. t -Wert $> |2|$.

- In EViews wird $\sqrt{\widehat{Var}(\hat{\beta}_i)}$ in der Spalte „Std. Error“ und der t -Wert für $\tilde{\beta}_i = 0$ in der Spalte „t-Statistic“ angegeben.
- In der letzten Spalte „Prob.“ wird angegeben zu welchem Signifikanzniveau H_0 verworfen werden kann. Ist der Wert beispielsweise $< 5\%$, so kann H_0 zu einem Signifikanzniveau von 5% verworfen werden.
- Die kritischen Werte t_c können auch dazu genutzt werden ein **Konfidenzintervall** für β_i zu berechnen.
- Allgemein stellt das Intervall

$$[\hat{\beta}_i - t_c \cdot \sqrt{\widehat{Var}(\hat{\beta}_i)}, \hat{\beta}_i + t_c \cdot \sqrt{\widehat{Var}(\hat{\beta}_i)}]$$

zu einem gewählten Signifikanzniveau von α und dem entsprechenden kritischen Wert t_c das $(1 - \alpha)$ -Konfidenzintervall dar, in dem der wahre Parameterwert von β_i mit einer Wahrscheinlichkeit von $1 - \alpha$ enthalten ist.

- Demnach ist das 95%-Konfidenzintervall für β_i annäherungsweise durch

$$[\hat{\beta}_i - 2 \cdot \sqrt{\widehat{Var}(\hat{\beta}_i)}, \hat{\beta}_i + 2 \cdot \sqrt{\widehat{Var}(\hat{\beta}_i)}]$$

gegeben.

6.2 Der F -Test

- Im Gegensatz zum t -Test, der den Einfluss jeder einzelnen erklärenden Variable überprüft, kann mit dem F -Test der gemeinsame Einfluss aller erklärender Variablen mit Ausnahme des Absolutgliedes getestet werden.
- Es kann gezeigt werden, dass die F -Statistik

$$F_{k-1, T-k} = \frac{R^2}{1-R^2} \frac{T-k}{k-1}$$

einer F -Verteilung mit $(k-1, T-k)$ Freiheitsgraden folgt.

- Durchführung eines F -Tests:
 - Die Nullhypothese des F -Tests lautet $\beta_2 = \dots = \beta_k = 0$.
 - Es wird wie beim t -Test ein Signifikanzniveau α gewählt und die kritischen Werte ermittelt.
 - Der Wert der F -Statistik wird mit den kritischen Werten der F -Verteilung mit $(k-1, T-k)$ Freiheitsgraden verglichen.
 - In EViews wird sowohl der Wert der F -Statistik („F-Statistic“) als auch das Signifikanzniveau, ab dem die Nullhypothese verworfen werden kann („Prob(F-Statistic)“) angegeben. Ist der Eintrag bei „Prob(F-Statistic)“ beispielsweise 4,5%, so kann die Nullhypothese zu einem Signifikanzniveau von 5% verworfen werden.