ulm university universität
uulm

Universität Ulm | 89069 Ulm | Germany

Fakultät für Mathematik und Wirtschaftswissenschaften

Dipl.-WiWi Christian Peukert
Institut für Wirtschaftspolitik

Ludwig-Erhard-Stiftungsprofessur

Sommersemester 2010

Übung zur Empirischen Wirtschaftsforschung

V. Das Lineare Regressionsmodell

5.1 Das lineare Regressionsmodell
ökonomisches, empirisches, ökonometrisches Modell

5.2 Die Methode der Kleinsten Quadrate
Einfaches lineares Modell, Multiples lineares Modell, Interpretation der Koeffizienten

5.3 Güteeigenschaften des KQ-Schätzers
Erwartungstreue, Konsistenz, Effizienz

5.4 Bestimmtheitsmaß
R^2, $\overline{R^2}$

5.5 Standardfehler
Standardfehler der Schätzung, Standardfehler der Koeffizienten

5.6 Testverfahren

t-Test, Konfidenzintervalle, F-Test

Literatur

Winker, Empirische Wirtschaftsforschung und Ökonometrie, 2007, Kapitel 7

von Auer, Ökonometrie, 2007

Helmholtzstr. 20, Raum E 02
Tel. 0731 50-24265, Fax -24262

http://www.uni-ulm.de/wipo
Christian.Peukert@uni-ulm.de

5.1 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:

einfaches lineares Regressionsmodell: $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$

multiples lineares Regressionsmodell: $y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t} + \varepsilon_t$

dabei sind

- β_0 : Absolutglied (Konstante, Intercept)
- β_1 : Koeffizient der exogenen Variable
- t : Zeitindex mit $t = 1,2, \ldots, T$
- ε : Störterm (Residuum, Fehler).

Wir nehmen an, dass es einen Störterm ε gibt, so dass der funktionale Zusammenhang zwischen y und x nicht zu einer 100%igen Erklärung führt.

Bei der Modellierung unterscheidet man:

ökonomisches Modell: $y_t = \beta_0 + \beta_1 x_t$

empirisches Modell: $y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$

ökonometrisches Modell: $y_t = \hat{\beta}_0 + \hat{\beta}_1 x_t + \hat{\varepsilon}_t$

bzw. $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_t$

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 ($max_{t \in \{1,\ldots,T\}} |\hat{\varepsilon}_t|$)
- ...

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.

5.2 Die Methode der Kleinsten Quadrate (KQ)

5.2.1 Einfaches lineares Regressionsmodell

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

$$\sum_{t=1}^{T} \hat{\varepsilon}_t^2 = \sum_{t=1}^{T} (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)^2$$

Die Minimierung der Summe der Residuenquadrate kann somit über die Minimierung des Ausdrucks auf der rechten Seite der vorherigen Gleichung erfolgen.

Die Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_1$ werden über die partiellen Ableitungen berechnet.

Schritt 1 Partielle Ableitung nach $\hat{\beta}_0$ Null setzen ($\frac{\partial \sum_{t=1}^{T} \hat{\varepsilon}_t^2}{\partial \hat{\beta}_0} \doteq 0$)

$$\begin{aligned}
\frac{\partial \sum_{t=1}^{T} \hat{\varepsilon}_t^2}{\partial \hat{\beta}_0} &= \sum_{t=1}^{T} 2(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)(-1) \doteq 0 \\
2(-1) \sum_{t=1}^{T} (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t) &= 0 \\
\sum_{t=1}^{T} (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t) &= 0 \\
\sum_{t=1}^{T} y_t - \sum_{t=1}^{T} \hat{\beta}_0 - \sum_{t=1}^{T} \hat{\beta}_1 x_t &= 0 \\
\sum_{t=1}^{T} y_t - T\hat{\beta}_0 - \sum_{t=1}^{T} \hat{\beta}_1 x_t &= 0
\end{aligned}$$

mit $\overline{y} = \frac{1}{T} \sum_{t=1}^{T} y_t, \overline{x} = \frac{1}{T} \sum_{t=1}^{T} x_t \Rightarrow \overline{y} - \hat{\beta}_0 - \hat{\beta}_1 \overline{x} = 0$

$$\Rightarrow \hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}$$

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

$$
\begin{aligned}
\frac{\partial \sum_{t=1}^{T} \hat{\varepsilon}_t^2}{\partial \hat{\beta}_1} &= \sum_{t=1}^{T} 2(y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t)(-x_t) \doteq 0 \\
2(-1) \sum_{t=1}^{T} (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t) x_t &= 0 \\
\sum_{t=1}^{T} (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t) x_t &= 0 \\
\sum_{t=1}^{T} y_t x_t - \sum_{t=1}^{T} \hat{\beta}_0 x_t - \sum_{t=1}^{T} \hat{\beta}_1 x_t x_t &= 0 \\
\sum_{t=1}^{T} x_t y_t - \sum_{t=1}^{T} \hat{\beta}_0 x_t - \sum_{t=1}^{T} \hat{\beta}_1 x_t^2 &= 0 \\
\overline{x \cdot y} - \hat{\beta}_0 \overline{x} - \hat{\beta}_1 \overline{x^2} &= 0 \\
\overline{x \cdot y} - (\overline{y} - \hat{\beta}_1 \overline{x}) \overline{x} - \hat{\beta}_1 \overline{x^2} &= 0 \\
\overline{x \cdot y} - \overline{x} \cdot \overline{y} + \hat{\beta}_1 \overline{x} \cdot \overline{x} - \hat{\beta}_1 \overline{x^2} &= 0 \\
\overline{x \cdot y} - \overline{x} \cdot \overline{y} &= \hat{\beta}_1 \overline{x^2} - \hat{\beta}_1 \overline{x}^2 \\
\hat{\beta}_1 (\overline{x^2} - \overline{x}^2) &= \overline{x \cdot y} - \overline{x} \cdot \overline{y} \\
\Rightarrow \hat{\beta}_1 &= \frac{\overline{x \cdot y} - \overline{x} \cdot \overline{y}}{\overline{x^2} - \overline{x}^2}
\end{aligned}
$$

Bekanntlich können Varianz und Kovarianz definiert werden als

$$
\begin{aligned}
Var(x) &= E(x^2) - (E(x))^2 \\
Cov(x, y) &= E(xy) - E(x)E(y) \\
\Rightarrow \hat{\beta}_1 &= \frac{Cov(x, y)}{Var(x)}
\end{aligned}
$$

Schritt 3 Berechnung der Koeffizienten $\hat{\beta}_0$ und $\hat{\beta}_1$ anhand der vorliegenden Daten.

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 $\hat{\beta}_0$:

$$\frac{\partial \sum_{t=1}^{T} \hat{\varepsilon}_t^2}{\partial \hat{\beta}_0} = -2 \sum_{t=1}^{T} (y_t - \hat{\beta}_0 - \hat{\beta}_1 x_t) = -2 \sum_{t=1}^{T} \hat{\varepsilon}_t \doteq 0$$

Interpretation des berechneten Koeffizienten $\hat{\beta}_1$

Additives Modell: $y_t = \beta_0 + \beta_1 x_t$:
Eine Veränderung von x um eine Einheit führt zu einer Veränderung von y von $\hat{\beta}_1$ Einheiten.

Multiplikatives Modell: $log(y_t) = \beta_0 + \beta_1 log(x_t)$:
Ein Veränderung von x um ein Prozent führt zu einem Anstieg von y von $\hat{\beta}_1$ Prozent.

Semi-log Modell: $log(y_t) = \beta_0 + \beta_1 x_t$:
Ein Veränderung von x um eine Einheit führt zu einem Anstieg von y von $\hat{\beta}_1$ Prozent.

5.2.2 Multiples lineares Regressionsmodell

Nehmen wir nun an, dass mehrere Faktoren Einfluss auf die endogene Variable nehmen.

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} \ldots + \beta_k x_{k,t} + \varepsilon_t$$

In Matrixschreibweise ergibt sich:

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix} \boldsymbol{X} = \begin{bmatrix} 1 & x_{1,1} & \ldots & x_{k,1} \\ 1 & x_{1,2} & \ldots & x_{k,2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1,T} & \ldots & x_{k,T} \end{bmatrix} \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}$$

Wie im Modell mit einem Regressor soll auch im multiplen Modell die Summe der Residuenquadrate minimiert werden.

Der Koeffizientenvektor $\hat{\boldsymbol{\beta}}$, der die Summe der Residuenquadrate minimiert, kann wieder über die partiellen Ableitungen berechnet werden.

Unter Verwendung der Ableitungsregeln von Matrizen ergibt sich für den KQ-Schätzer[1]:

$$\frac{\partial \sum_{t=1}^{T} \hat{\varepsilon}_t^2}{\partial \hat{\boldsymbol{\beta}}} = -2\boldsymbol{X}^\top \boldsymbol{y} + 2\boldsymbol{X}^\top \boldsymbol{X}\hat{\boldsymbol{\beta}} = 0$$
$$\Rightarrow \hat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{y}$$

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

[1] Details z.B. in von Auer (2007), S. 182 f.

5.3 Güteeigenschaften des KQ-Schätzers

Der KQ-Schätzer ist der beste lineare erwartungstreue Schätzer, falls gilt:

Die Elemente der Matrix $\boldsymbol{X} \in \mathbb{R}^{T \times k}$ sind exogen vorgeben, es sind alle relevanten Variablen enthalten, kein Element von $\boldsymbol{X}$ ist eine Linearkombination eines anderen Elements von $\boldsymbol{X}$, d.h. $rang(\boldsymbol{X}) = k$.

Die Störterme sind zufällig verteilt mit Erwartungswert $E(\boldsymbol{\varepsilon}) = \mathbf{0}$ und Varianz-Kovarianzmatrix $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^\top) = \sigma^2 \boldsymbol{I_T}$, wobei $\boldsymbol{I_T}$ eine $T \times T$ Einheitsmatrix ist. Das heißt also, dass das Residuum einer Periode ist unkorreliert mit den Regressoren und den Residuen anderer Perioden ist, es existiert folglich keine Systematik in den Fehlern.

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 konvergiert:

$$\lim_{T \to \infty} P(|\hat{\beta} - \beta| < \epsilon) = 1.$$

Linearität:

Linearität bedeutet, dass sich ein Schätzer $\hat{\beta}_i$ als Linearkombination der Beobachtungen der endogenen Variable y_t schreiben lässt:

$$\hat{\beta}_i = c_1 y_1 + \ldots + c_T y_T.$$

Effizienz:

Im Allgemeinen existieren mehrere lineare erwartungstreue Schätzer.

Der Schätzer mit der kleinsten Varianz in dieser Menge von Schätzern wird "effizient" genannt.

5.4 Bestimmtheitsmaß

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

$$\underbrace{\sum_{t=1}^{T}(y_t - \overline{y})^2}_{TSS} = \underbrace{\sum_{t=1}^{T}(\hat{\beta}_0 + \hat{\beta}_1 x_t - \overline{y})^2}_{ESS} + \underbrace{\sum_{t=1}^{T}\hat{\varepsilon}_t^2}_{RSS}$$

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, möglicherweise ohne den inhaltlichen Erklärungsgehalt zu steigern.

Deshalb wird das korrigierte Bestimmtheitsmaß $\overline{R^2}$ um die Anzahl der Freiheitsgrade bereinigt:

$$\overline{R^2} = 1 - \frac{RSS}{TSS} \cdot \frac{T-1}{T-k}$$

Wegen der bestrafenden Berücksichtigung der Anzahl der Regressoren (k) kann das korrigierte Bestimmtheitsmaß auch negative Werte annehmen.

5.5 Standardfehler

Ein wichtiges statistisches Gütekriterium ist der Standardfehler, der 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:

$$\sigma = \sqrt{\frac{1}{T-1} \cdot \sum_{t=1}^{T} (y_t - \overline{y})^2}$$

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

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

Der Standardfehler der Schätzung ist in der Regel kleiner als der Standardfehler der Originalwerte.

Berechnung des Standardfehlers des Absolutglieds $\hat{\beta}_0$:

$$se(\hat{\beta}_0) = \hat{\sigma} \cdot \sqrt{\frac{\sum_{t=1}^{T} x_t^2}{T \cdot \sum_{t=1}^{T} (x_t - \overline{x})^2}}$$

Berechnung des Standardfehlers des Koeffizienten $\hat{\beta}_1$:

$$se(\hat{\beta_1}) = \hat{\sigma} / \sqrt{\sum_{t=1}^{T} (x_t - \overline{x})^2}$$

5.6 Testverfahren

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

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

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):

$$H_0 : \beta_i = 0$$

- Formulierung einer Alternativhypothese. Es können drei Fälle unterschieden werden:

$$H_{A1} : \beta_i \neq \tilde{\beta}_i$$
$$H_{A2} : \beta_i > \tilde{\beta}_i$$
$$H_{A3} : \beta_i < \tilde{\beta}_i$$

- 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 gegeben durch:

$$P(|t\text{-Wert}| > t_c) = \alpha,$$
$$P(t\text{-Wert} > t_c) = \alpha, \text{ bzw.}$$
$$P(t\text{-Wert} < t_c) = \alpha.$$

- Bei gegebenem Signifikanzniveau α und ermittelten kritischen Werten t_c wird H_0 genau dann abgelehnt, wenn

$$|t\text{-Wert}| > t_c,$$
$$t\text{-Wert} > t_c, \text{ bzw.}$$
$$t\text{-Wert} < t_c.$$

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

$\Rightarrow$ Faustregel: Vergleiche, ob der t-Wert betragsmäßig größer als 2 ist, d.h. $|t\text{-Wert}| > 2$.

In EViews wird $se(\hat{\beta}_i)$ 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.

Konfidenzintervalle

Die kritischen Werte t_c können auch dazu genutzt werden ein Konfidenzintervall für β_i zu berechnen.

Allgemein stellt das Intervall

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

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

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

gegeben.

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

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

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

- Durchführung eines F-Tests:
  - Die Nullhypothese des F-Tests lautet $H_0 : \beta_1 = \ldots = \beta_k = 0$,
    kann aber auch als $H_0 : R^2 = 0$ interpretiert werden.
  - 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.