2.3 Schätzeigenschaften der OLS-Methode

Jede Schätzmethode weist bestimmte **Güteeigenschaften** auf, die von der Erfüllung bestimmter Voraussetzungen abhängen. Was die **gewöhnliche Methode der kleinsten Quadrate (OLS-Methode)** betrifft, sind bestimmte **Schätzeigenschaften** bekannt, sofern die Standardannahmen des multiplen Regressionsmodells erfüllt sind. Ist die Störvariable im Mittel gleich 0 und ist ihre Varianz konstant, so ist der **OLS-Schätzer** der Regressionskoeffizienten nach dem Gauss-Markov-Theorem bei fehlender Autokorrelation die **beste lineare unverzerrte Schätzfunktion** (best linear unbiased (blu) estimator)**.**

Unverzerrt bedeutet dabei erwartungstreu. Im Mittel stimmen die OLS-geschätzten Regressionskoeffizienten mit den unbekannten Regressionskoeffizienten der Grundgesamtheit überein. Bester Schätzer bezieht sich auf die Genauigkeit der Schätzung und bedeutet Effizienz. Unter den Standardannahmen ist der OLS-Schätzer im linearen Regressionsmodell also **erwartungstreu** und **effizient**. Es gibt dann keinen anderen linearen und erwartungstreuen Schätzer der Regressionskoeffizienten, der zu einer geringeren Varianz führt.

Darüber hinaus ist die OLS-Methode **konsistent**. Konsistenz heißt, dass sich die geschätzten Regressionskoeffizienten mit wachsendem Stichprobenumfang immer besser den unbekannten Regressionskoeffizienten der Grundgesamtheit annähern.

● Erwartungstreue

Ein Schätzer ist erwartungstreu, wenn sein Erwartungswert mit dem Parameter der Grundgesamtheit übereinstimmt.

Um die Erwartungstreue des OLS-Schätzers $\hat{\boldsymbol{\beta}}$ zu zeigen, setzen wir zunächst das Regressionsmodell (2.5) in (2.15) ein:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'y} = (\mathbf{X'X})^{-1}\mathbf{X'}(\mathbf{X}\boldsymbol{\beta} + \mathbf{u})$$

Nach Ausmultiplizieren erhält man

$$\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\mathbf{X'X}\boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'u}$$

so dass

(2.33) $$\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'u}$$

folgt.

Nun kann der Erwartungswert von $\hat{\boldsymbol{\beta}}$ bestimmt werden:

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \mathrm{E}\left[\boldsymbol{\beta} + (\mathbf{X'X})^{-1}\mathbf{X'u})\right] = \mathrm{E}(\boldsymbol{\beta}) + \mathrm{E}\left[(\mathbf{X'X})^{-1}\mathbf{X'u}\right]$$

$$= \boldsymbol{\beta} + E\left[(\mathbf{X'X})^{-1}\mathbf{X'u}\right] = \boldsymbol{\beta} + (\boldsymbol{X'X})^{-1}\,\mathbf{X'E}(\mathbf{u})$$

Mit E(**u**) = **0** erhält man schließlich

(2.34) $$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

Damit ist die Erwartungstreue des OLS-Schätzers $\hat{\boldsymbol{\beta}}$ gezeigt. Aus der ausführlichen Schreibweise von (2.34)

$$\begin{pmatrix} \mathrm{E}(\hat{\beta}_1) \\ \mathrm{E}(\hat{\beta}_2) \\ \vdots \\ \mathrm{E}(\hat{\beta}_k) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix},$$

ist ersichtlich, dass die Erwartungswerte der OLS-Schätzer $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ mit den unbekannten Regressionskoeffizienten des ökonometrischen Modells übereinstimmen.

● **Effizienz**

Ein erwartungstreuer Schätzer, der im Vergleich zu allen alternativen unverzerrten Schätzern die kleinste Varianz hat, heißt effizient. Man bezeichnet eine solche Schätzfunktion als besten unverzerrten Schätzer. Der OLS-Schätzer $\hat{\boldsymbol{\beta}}$ ist in der Klasse der linearen Schätzfunktionen effizient.

Es sei $\hat{\beta}_j$ der OLS-Schätzer für den unbekannten Regressionskoeffizienten βj und $\tilde{\beta}_j$ ein beliebiger anderer linearer Schätzer. Effizienz bedeutet dann, dass Die Varianz des OLS-Schätzers $\hat{\beta}_j$ am geringsten ist:

(2.35) $\quad \mathrm{Var}(\hat{\beta}_j) \leq \mathrm{Var}(\tilde{\beta}_j)$, j=1,2,…,k.

● **Konsistenz**

Wenn ein Schätzer mit zunehmendem Stichprobenumfang einen Parametervektor immer genauer treffen soll, dann ist insbesondere sein Grenzverhalten für n -> ∞ von Interesse, also eine Eigenschaft bei großen Stichproben. Bei der Konsistenz geht es um die Frage, ob ein Schätzer bei einem über alle Grenzen wachsenden Stichprobenumfang schließlich mit dem Parametervektor **β** zusammenfällt:

$$\lim_{n\to\infty} \mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \quad \text{und} \quad \lim_{n\to\infty} \mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$$

2.4 Bestimmtheitsmaß und multipler Korrelationskoeffizient

● **Bestimmtheitsmaß (Determinationskoeffizient)**

- Beurteilung der globalen Güte der Anpassung des Regressionsmodells

- Quantifizierung des Erklärungsgehalt der exogenen Variablen in einem ökonometrischen Modell

Varíanz der abhängigen Variablen Y:

(2.36) $$s_y^2 = \frac{1}{n}\sum(y_t - \bar{y})^2 = \frac{1}{n}\sum y_t^2 - \bar{y}^2 = \frac{1}{n}\mathbf{y}'\mathbf{y} - \bar{y}^2$$

Varianzzerlegung:

(2.37) $$s_y^2 = s_{\hat{y}}^2 + s_{\hat{u}}^2$$

mit

(2.38) $$s_{\hat{y}}^2 = \frac{1}{n}\sum(\hat{y}_t - \bar{y})^2 = \frac{1}{n}\sum \hat{y}_t^2 - \bar{y}^2 = \frac{1}{n}\hat{\mathbf{y}}'\hat{\mathbf{y}} - \bar{y}^2 \qquad \text{wegen } \bar{\hat{y}} = \bar{y}$$

(2.39) $$s_{\hat{u}}^2 = \frac{1}{n}\sum \hat{u}_t^2 = \frac{1}{n}\hat{\mathbf{u}}'\hat{\mathbf{u}} \qquad \text{wegen } \bar{\hat{u}} = 0$$

Bestimmtheitsmaß (Determinationskoeffizient):

(2.40a) $$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$ oder (2.40b) $$R^2 = 1 - \frac{s_{\hat{u}}^2}{s_y^2}$$

Wertebereich: $0 \leq R^2 \leq 1$

Rechenformeln:

(2.41a) $$R^2 = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}} - n \cdot \bar{y}^2}{\mathbf{y}'\mathbf{y} - n \cdot \bar{y}^2} = \frac{n\hat{\mathbf{y}}'\hat{\mathbf{y}} - \left(\sum y_t\right)^2}{n\mathbf{y}'\mathbf{y} - \left(\sum y_t\right)^2}$$

oder

(2.41b) $$R^2 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n \cdot \bar{y}^2}{\mathbf{y}'\mathbf{y} - n \cdot \bar{y}^2} = \frac{n\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \left(\sum y_t\right)^2}{n\mathbf{y}'\mathbf{y} - \left(\sum y_t\right)^2}$$

Beispiel:

Gesucht ist der Anteil der Varianz des Privaten Verbrauchs, der auf die Streuung des verfügbaren Einkommens zurückgeführt werden kann. Um hierüber eine Aussage machen zu können, wird das Bestimmtheitsmaß für die makroökonomische Konsumfunktion für den Stützbereich von 1994 bis 2012 unter Verwendung der Berechnungsformel (2.41b) ermittelt. Unter Verwendung von

$$\hat{\beta}' = \begin{bmatrix} -38,5222 & 0,93354 \end{bmatrix},$$

$$X'y = \begin{bmatrix} \sum y_t \\ \sum x_t y_t \end{bmatrix} = \begin{bmatrix} 24034,68 \\ 33960226,83 \end{bmatrix} \quad y'y = \sum y_t^2 = 30778325$$

berechnen wir

$$\hat{y}'\hat{y} = \hat{\beta}' X'y = 30777206$$

Mit n=19 und $\bar{y} = 1264,98$ ergibt sich damit ein Bestimmtheitsmaß von

$$R^2 = \frac{19 \cdot 30777206 - 24034,68^2}{19 \cdot 30778325 - 24034,68^2} = \frac{583166740}{583187986} = 0,9968$$

● Einfacher und multipler Korrelationskoeffizient

- Einfache Regression:

(2.42) $$R^2 = r_{yx}^2$$

Aufgrund der Lineartransformation $\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_t$ gilt

$$s_{\hat{y}}^2 = \hat{\beta}_2^2 \cdot s_x^2$$

womit (2.42) einfach gezeigt werden kann:

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{\hat{\beta}_2^2 \cdot s_x^2}{s_y^2} = \frac{\left(\frac{s_{xy}}{s_x^2}\right)^2 \cdot s_x^2}{s_y^2} = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = r_{xy}^2$$

Die Wurzel des Bestimmtheitsmaßes gibt bei der einfachen Regression daher den absoluten Wert des Korrelationskoeffizienten zwischen dem Regressanden y und dem Regressor x wieder:

(2.43) $$\left| r_{yx} \right| = \sqrt{R^2}$$

- Multiple Regression:

Das Bestimmtheitsmaß ist dann das Quadrat des multiplen Korrelationskoeffizienten $r_{y\hat{y}}$, der die Korrelation zwischen der abhängigen Variablen y und allen unabhängigen Variablen $x_1, x_2, \ldots, x_k$ (=Korrelation zwischen der abhängigen Variablen y_t und den Regressionswerte $\hat{y}_t$) wiedergibt:

(2.44) $$R^2 = r_{y\hat{y}}^2 = r_{yx_{1,2,\ldots k}}^2$$

Das Vorzeichen des multiplen Korrelationskoeffizienten $r_{yx_{1,2,\ldots k}}$ bleibt unbestimmt, da positive und negative Einflüsse der Regressoren gleichzeitig auftreten können:

(2.45) $$\left|r_{y\hat{y}}\right| = \left|r_{yx_{1,2,\ldots k}}\right| = \sqrt{R^2}$$

● Korrigiertes Bestimmtheitsmaß

Obwohl das Bestimmtheitsmaß gut interpretierbar ist, eignet sich nur eingeschränkt für einen Vergleich von Regressionsmodellen mit einer unterschiedlichen Anzahl von unabhängigen Variablen. Wenn zusätzliche Regressoren in ein Regressionsmodell aufgenommen werden, erhöht sich R^2 auch wenn diese keine ökonomische Relevanz besitzen.

Man kann dies vermeiden, indem man die Zahl der Freiheitsgrade bei der Berechnung der Determination berücksichtigt. Das **korrigierte Bestimmtheitsmaß**

(2.46) $$\bar{R}^2 = 1 - \frac{\frac{1}{n-k}\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\frac{1}{n-1}\left(\mathbf{y}'\mathbf{y} - n\bar{y}^2\right)}$$

kann im Gegensatz zu R^2 im Falle einer Einbeziehung irrelevanter Regressoren sinken.

Eine Berechung des korrigierten Bestimmheitsmaßes kann über R^2 erfolgen:

(2.47) $$\bar{R}^2 = 1 - \frac{n-1}{n-k}\left(1 - R^2\right)$$