Appendix

Kapitel 2

Ökonometrie I
Michael Hauser

Inhalt

- Geometrie der Korrelation
- Freiheitsgrade
- Der OLS Schätzer: Details
- OLS Schätzer für Okuns's law nachgerechnet
- Anforderungen an Theorien und Modelle

Geometrie der Korrelation

Korrelation und Cosinus eines Winkels

Der **Korrelationskoeffizient** zwischen x und y (x_i, y_i), $i = 1, \ldots, n$ in einer Stichprobe ist

$$r_{xy} = \frac{\sum(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum(x_i - \overline{x})^2}\sqrt{\sum(y_i - \overline{y})^2}}$$

Der **Cosinus vom Winkel α zwischen zwei (Spalten-)Vektoren $\boldsymbol{x}$** und $\boldsymbol{y}$, $\alpha(\boldsymbol{x}, \boldsymbol{y})$, ist

$$\cos(\alpha) = \frac{\boldsymbol{x}'\boldsymbol{y}}{|\boldsymbol{x}|\,|\boldsymbol{y}|}$$

wobei $|\boldsymbol{x}|$ die Länge des Vektors bezeichnet, $|\boldsymbol{x}| = \sqrt{\sum x_i^2}$ (nach Pythagoras).

Der Korrelationskoeffizient ist gleich dem Cosinus des Winkels für die mittelwertkorrigierten Datenvektoren.

(Nach dem Verschiebungssatz genügt es, dass eines der Mittel null ist.)

Korrelation und Cosinus eines Winkels

r_{xu}, mit u Fehler der Regression, ist nach Anwendung des Verschiebungssatzes

$$r_{xu} = \frac{\sum(x_i - \overline{x})(u_i)}{\sqrt{\sum(x_i - \overline{x})^2}\sqrt{\sum(u_i)^2}} = \frac{\sum x_i u_i}{\sqrt{\sum(x_i - \overline{x})^2}\sqrt{\sum(u_i)^2}} = 0$$

$$\cos(90^0) = 0$$

Ein Winkel von 90^0 (die Vektoren stehen orthogonal zueinander) entspricht einer Korrelation von null.

Intuition zu Freiheitsgraden

Intuition zu Freiheitsgraden

Freiheitsgrade (FGe) sind die Anzahl der Werte, die in einer Statistik frei variieren können.

Ang es liegen $n = 2$ Beobachtungen vor, und wir nehmen $n\overline{x} = \sum x_i$ als bekannt an. Welche Werte können x_1, und x_2 annehmen, wenn $\overline{x} = 5$ ist?

Da $n\overline{x} = \sum x_i = 10$ ist, ist die Summe der Beobachtungen fixiert. Wählen wir einen beliebigen Wert für x_1, zB $x_1 = -1$, so haben wir keine Freiheit für die Wahl von x_2. $x_2 = 11$, ist fixiert.

Intuition zu Freiheitsgraden

Ang es liegen $n = 3$ Beobachtungen vor, und wir nehmen $n\overline{x} = \sum x_i$ als bekannt an. Welche Werte können x_1, x_2 und x_3 annehmen, wenn $\overline{x} = 5$ ist?

Da $n\overline{x} = \sum x_i = 15$ ist, ist die Summe der Beobachtungen fixiert. Wählen wir einen beliebigen Wert für x_1, zB $x_1 = -1$. So können wir noch x_2 oder x_3 wählen, so dass ihre Summe 16 ergibt.
Wir haben also 2 FGe.

Alternativ ist $\overline{x}$ eine Statistik, die die Qualität eines einzelnen Datenpunkts hat. Damit stehen für die Abweichungen $(x_i - \overline{x})$ nur $(n - 1)$ Werte zur Berechnung der Varianz zur Verfügung.

Herleitung des OLS Schätzers: Details

Minimierung der Fehlerquadratsumme, OLS

Wir minimieren die Summe der quadrierten Residuen.

$$\min_{b_0, b_1} \sum_{t=1}^{n} \widehat{u}_t^2 = \min_{b_0, b_1} S(b_0, b_1)$$

Mathematisch handelt es sich um die Minimierung einer quadratischen Funktion in 2 Variablen (b_0, b_1). Wir berechnen die beiden partiellen Ableitungen 1.Ordnung und setzen sie null (Bedingung 1.Ordnung).

$$S(b_0, b_1) = \sum_{t=1}^{n} \widehat{u}_t^2 = \sum_{t=1}^{n} [y_t - b_0 - b_1\, x_t]^2$$

Die partiellen Ableitungen sind (mit Hilfe der Kettenregel)

$$\partial S / \partial b_0 = -2 \sum (y_t - b_0 - b_1 x_t) = -2 \sum \widehat{u}_t = 0$$

$$\partial S / \partial b_1 = -2 \sum x_t (y_t - b_0 - b_1 x_t) = -2 \sum x_t \widehat{u}_t = 0$$

Herleitung des OLS Schätzers

Vereinfachen ergibt die **Normalgleichungen**

$$\sum y_t = n\, b_0 + b_1 \sum x_t$$

$$\sum x_t y_t = b_0 \sum x_t + b_1 \sum x_t^2$$

Dividieren der ersten Gleichung durch n ergibt

$$b_0 = \overline{y} - b_1 \overline{x}$$

Einsetzen in die zweite und Erweitern mit $(1/n)$

$$b_1 = \frac{\sum (x_t - \overline{x})(y_t - \overline{y})}{\sum (x_t - \overline{x})^2} = r_{xy} \frac{s_y}{s_x}$$

Zwischenrechnungen

$$\frac{1}{n}\sum x = \overline{x}$$

$$\begin{aligned}
\sum x\,y &= (\overline{y} - b_1\,\overline{x})\sum x + b_1\,\sum x^2 \\
&= \overline{y}\,\sum x - b_1\,\overline{x}\,\sum x + b_1\,\sum x^2 \qquad |:n
\end{aligned}$$

$$\frac{1}{n}\sum x\,y - \overline{x}\,\overline{y} = b_1\,[\tfrac{1}{n}\sum x^2 - \overline{x}^2]$$

Verschiebungssatz:

$$\mathrm{cov}(x,y) = \frac{1}{n}\sum(x-\overline{x})(y-\overline{y}) = \frac{1}{n}\sum x\,y - \overline{x}\,\overline{y}$$

$$s_x^2 = \frac{1}{n}\sum(x-\overline{x})^2 = \frac{1}{n}\sum x^2 - \overline{x}^2$$

$$r_{xy} = \mathrm{cov}(x,y)/\sqrt{s_x^2\,s_y^2} \qquad \text{oder}$$

$$\mathrm{cov}(x,y) = r_{xy}\,s_x\,s_y$$

OLS im bivariaten Regressionsmodell: $R^2 = r_{xy}^2$

Es gilt:

$$R^2 = \frac{ESS}{TSS} = \frac{S_{\widehat{y}\widehat{y}}}{S_{yy}}$$

Wir verwenden $b_0 = \overline{y} - b_1\overline{x}$ und $b_1 = r_{xy} s_y / s_x$ in

$$\widehat{y} = b_0 + b_1 x$$

$$\widehat{y} = (\overline{y} - b_1\overline{x}) + b_1 x$$

$$(\widehat{y} - \overline{y}) = b_1(x - \overline{x})$$

$$S_{\widehat{y}\widehat{y}} = b_1^2 S_{xx} \quad \text{und} \quad b_1^2 = r_{xy}^2 S_{yy}/S_{xx}$$

$$R^2 = \frac{S_{\widehat{y}\widehat{y}}}{S_{yy}} = b_1^2 \frac{S_{xx}}{S_{yy}} = r_{xy}^2 \frac{S_{yy}}{S_{xx}} \frac{S_{xx}}{S_{yy}} = r_{xy}^2$$

OLS Schätzer für Okuns's law nachgerechnet

Okun's law: b_0, b_1

Wir rechnen die OLS Schätzung nach in dem wir in die Formeln

$$b_0 = \overline{y} - b_1 \overline{x} \qquad \text{und} \qquad b_1 = S_{xy} / S_{xx}$$

einsetzen. $b_1 = \sum(x_t - \overline{x})(y_t - \overline{y}) \,/\, \sum(x_t - \overline{x})^2$

In EViews:

`DUR` $\rightarrow$ `View` $\rightarrow$ `Descriptive ...` $\rightarrow$ `Histogram ...`

$\overline{y} = \overline{DUR} = 0.140 \qquad \text{und} \qquad \overline{x} = \overline{rho} = 2.158$

$S_{xx} = \sum(x_t - \overline{x})^2 = (n-1)s_x^2 = (n-1)s_{rho}^2 = (34-1)1.714^2$
$\quad = 96.982$

Group aus *DUR* und ρ erstellen $\rightarrow$ `Open` $\rightarrow$ `as Group` $\rightarrow$ `Descriptive ...` $\rightarrow$ `Common Sample`, bzw `Covariance Analysis`

Okun's law: b_0, b_1

$$S_{xy} = \sum (x_t - \overline{x})(y_t - \overline{y}) = n \operatorname{cov}(x, y)$$
$$= n \operatorname{cov}(DUR, \rho) = (34)(-0.471) = -16.027$$

$$b_1 = S_{xy}/S_{xx} = -16.027/96.982 = -0.165$$

$$b_0 = \overline{y} - b_1 \overline{x} = 0.140 - (-0.165)2.158 = 0.500$$

Okun's law: TSS, ESS, RSS, R^2

RSS, sum of squared residuals, ablesen: $RSS = 5.575$.
TSS, ist aus der Standardabweichung der abhängigen Variablen zu berechnen:

$$s_y^2 = \frac{1}{n-1} \sum (y_t - \overline{y})^2 = \frac{1}{n-1} TSS = 0.499^2$$

Mit $n = 34$ ist $TSS = 8.224$ und $ESS = TSS - RSS = 2.649$.
Also

$$TSS = ESS + RSS$$

$$8.224 = 2.649 + 5.575$$

$$R^2 = \frac{2.649}{8.224} = 0.322$$

Okun's law: R^2 und corr(DUR,rho)

Der Zusammenhang zwischen R^2 und $\mathrm{corr}(x, y)$ ist

$$R^2 = \mathrm{corr}(x, y)^2$$

Im `Covariance` Menü von oben `Correlation` wählen:

$$0.322 = R^2 = \mathrm{corr}(DUR, \rho)^2 = (-0.5675)^2 = 0.322$$

Der Zusammenhang zwischen Korrelation von DUR und ρ, und b_1 ist:

$$b_1 = \mathrm{corr}(x, y)\frac{s_y}{s_x} = (-0.567)\frac{0.499}{1.714} = -0.165$$

Anforderungen Theorie und Modelle

Anforderungen an eine Theorie

Eine Theorie hat 2 Kriterien zu genügen:

- *intrinsic closure:* Die inneren Beziehungen der Variablen müssen stabil sein.
- *extrinsic closure:* Dritte Variable haben keinen (relevanten) Einfluss, oder können durch ceteris paribus Überlegungen integriert werden.

Anforderungen an eine Theorie

Eine Theorie kann aus 2 Gründen von Interesse sein:

- *Realism: Kausale Zusammenhänge* zwischen den Variablen stehen im Vordergrund.
- *Instrumentalism:* Die Theorie soll ein Verhalten/einen Verlauf *prognostizieren* können. Dabei wird kein *Allein*'erklärung'sanspruch erhoben. Es sind mehrere verschiedene Theorien für dasselbe Phänomen möglich. Es wird auch kein Anspruch auf Realitätsnähe erhoben.
  Beobachtete Daten werden angesehen, *als ob (as if)* sie aus dieser Theorie stammen.

Ref: Lawson(1989) Oxford Economic Papers

Anforderungen an ein stochastisches Modell

Ein stochastisches Modell ist ein Daten-generierender Prozess. Dh die Grundgesamtheit kann durch eine n-dimensionale Zufallsvariable mit gemeinsamer Wahrscheinlichkeitsverteilung beschrieben werden. Die Daten sind Ziehungen daraus.

ZB $\{y, x\}$ mit der Beziehung: $y = \beta_0 + \beta_1 x + u$, $u \sim \mathrm{N}(0, \sigma^2)$

Jedes Modell ist eine Vereinfachung, und wird nur sehr selten alle kausal relevanten Faktoren enthalten.

Die vereinfachten Zusammenhänge im Modell basieren auf Konventionen, Plausibilität.

Das Kriterium, das *wir* an ein Modell stellen, ist *Nützlichkeit*.

Anforderungen an ein stochastisches Modell

Analog zur Theorie (intrinsic und extrinsic closure) verlangen wir, dass

- die Modellstruktur konstant ist. Dh $\beta_0, \beta_1, \ldots, \sigma_u^2$ sind konstant und die funktionale Form korrekt gewählt.
- alle relevanten dritte Variable enthalten sind.

Dh, dass das Modell korrekt spezifiziert ist.

Solow: Loose fitting positivism

Die überwiegende Mehrheit der empirischen Ökonomie stellt schwächere Anforderungen an ihre Aussagen:

'The study of economy and economic policy through empirically testable models.'

Es tritt die Anforderung der Prognosefähigkeit/-qualität in den Hintergrund. Theorien müssen nicht notwendigerweise nach der Prognosegüte gewählt werden. Sie sollen bestimmte Phänomene reproduzieren können.

Ref: Solow, Robert(1997) How Did Economics Get that Way and What Way Did it Get?, Daedalus, Winter.