Prof. Dr. Peter von der Lippe

Übersicht zur Schätz- und Testtheorie in der Regressionsanalyse

Für Hörer der Ökonometrie-Vorlesung am Standort Duisburg (revidierte Fassung Febr. 2007)

Wenn Formeln nummeriert sind, bezieht sich die Zahl auf das Lehrbuch von Ludwig von Auer

---

A. Modell, Parameter

Modell (Grundgesamtheit) $y = \alpha + \beta_1 x_1 + \ldots + \beta_K x_K + u, \quad \sigma^2 = \text{var}(u)$

bei einfacher Regression ist K = 1

Notation $\sum \hat{u}_t^2 = S_{\hat{u}\hat{u}}$, entsprechend $\sum (y_t - \bar{y})^2 = S_{yy}$ usw.

Begriffe: S_{yy} = "**Variation**", s_y^2 = **Varianz** (es gilt: $S_{yy} = T \cdot s_y^2$);

geschätzte Gleichung (Stichprobe) $y = \hat{\alpha} + \hat{\beta}_1 x_1 + \ldots + \hat{\beta}_K x_K + \hat{u}$

geschätzte Varianz der Störgröße $\hat{\sigma}^2 = \dfrac{\sum \hat{u}_t^2}{T-K-1} = \dfrac{S_{\hat{u}\hat{u}}}{T-(K+1)}$

B. Eigenschaften der Schätzer

KQ (OLS) Schätzer sind erwartungstreu und effizient. Bei Geltung der B-Annahmen ist die (geschätzte) Varianz - Kovarianzmatrix $\hat{\mathbf{V}} = \hat{\sigma}^2 (\mathbf{X'X})^{-1}$. Im Fall der einfachen Regression bedeutet das $\hat{\sigma}^2 (\mathbf{X'X})^{-1} = \dfrac{\hat{\sigma}^2}{|\mathbf{X'X}|} \begin{bmatrix} \sum x_t^2 & -\sum x_t \\ -\sum x_t & T \end{bmatrix}$ mit $|\mathbf{X'X}| = T \cdot S_{xx} = T^2 \cdot s_x^2$. Somit gilt

$$\boxed{\hat{\sigma}^2_{\hat{\beta}} = \frac{T\hat{\sigma}^2}{|X'X|} = \frac{\hat{\sigma}^2}{S_{xx}} \quad \text{und} \quad \hat{\sigma}^2_{\hat{\alpha}} = \overline{X}^2 \hat{\sigma}^2_{\hat{\beta}} = \hat{\sigma}^2 \left(\frac{1}{T} + \frac{\bar{x}^2}{S_{xx}} \right).}$$

C. Konfidenzintervall und Test für die Varianz der Störgröße

Bei Annahme B4 ist $\sum \dfrac{u^2}{\sigma} \sim \chi^2_T$ und $\sum \dfrac{\hat{u}^2}{\sigma^2} \sim \chi^2_{T-(K+1)}$

G_u und G_o ist die untere und obere Signifikanzschranke bei χ^2_{T-K-1} Verteilung (bei einf. Regression T-2, da K = 1). Die Grenzen des Konfidenzintervalls sind dann: $\dfrac{\sum \hat{u}}{G_o} \le \sigma^2 \le \dfrac{\sum \hat{u}}{G_u}$.

Punktschätzung $\hat{\sigma}_{\hat{u}}^2 = \sum \hat{u}_t^2 / (T-K-1)$ bei einfacher Regression $\sum \hat{u}_t^2$/(T-2).

Die Prüfgröße für H_0: $\sigma^2 = \sigma_0^2$ ist $\dfrac{\sum \hat{u}_t^2}{\sigma_0^2} \sim \chi^2_{T-K-1}$

D. Konfidenzintervall und Test für *einen* Regressionskoeffizienten oder für *eine* (L = 1) Linearkombination von Regressionskoeffizienten

Test **eines** Regressionskoeffizienten $Ho: \beta_k = q$ (etwa $q = 0$): Prüfgröße $\dfrac{\hat{\beta}_k - q}{\hat{\sigma}_{\hat{\beta}_k}} \sim t_{T-K-1}$

mit $\hat{\sigma}^2_{\beta_k}$ als Diagonalelement aus der Matrix $\hat{\sigma}^2 (\mathbf{X'X})^{-1}$. Für die KQ Schätzer und deren Fehler gilt im Falle der "Zweifachregression" ($\hat{y} = \alpha + \beta_1 x_1 + \beta_2 x_2, \quad K = 2$):

$\hat{\alpha}=\bar{y}-\hat{\beta}_1\bar{x}_1-\bar{\beta}_2\bar{x}_2$	$\text{var}(\hat{\alpha})=\frac{\sigma^2}{T}+(\bar{x}_1)^2\text{var}(\hat{\beta}_1)+(\bar{x}_2)^2\text{var}(\hat{\beta}_2)+2\bar{x}_1\bar{x}_2\text{cov}(\hat{\beta}_1\hat{\beta}_2)$	
$\hat{\beta}_1=(S_{22}S_{1y}-S_{12}S_{2y})/N$	$\text{var}(\hat{\beta}_1)=\frac{\sigma^2}{S_{11}(1-R^2_{1.2})}$	Bei der geschätzten Varianz ist dann jeweils σ^2 durch $\hat{\sigma}^2=\frac{S_{\hat{u}\hat{u}}}{T-3}$ zu ersetzen. Bei K = 2 ist natürlich $R^2_{1.2}=r^2_{12}$
$\hat{\beta}_2=(S_{11}S_{2y}-S_{12}S_{1y})/N$	$\text{var}(\hat{\beta}_2)=\frac{\sigma^2}{S_{22}(1-R^2_{1.2})}$	
$N=S_{11}S_{22}-S^2_{12}$	$\text{cov}(\hat{\beta}_1,\hat{\beta}_2)=\frac{-\sigma^2R^2_{1.2}}{S_{12}(1-R^2_{1.2})}$	

Test **einer** (L = 1) Linearkombination, etwa H_0: $r_1\beta_1+r_2\beta_2=q$ oder $\begin{bmatrix}r_1 & r_2\end{bmatrix}\begin{bmatrix}\beta_1\\ \beta_2\end{bmatrix}=r'\beta=q$ in Matrixschreibweise: Die geschätzte Standardabweichung $\hat{\sigma}_{lk}$ (lk = Linearkombination) ist dann $\hat{\sigma}^2_{lk}=r_1^2\frac{\hat{\sigma}^2}{S_{11}(1-R^2_{1.2})}+r_2^2\frac{\hat{\sigma}^2}{S_{22}(1-R^2_{1.2})}+2r_1r_2\frac{-\hat{\sigma}^2R^2_{1\cdot 2}}{S_{12}(1-R^2_{1.2})}$ und die t- verteilte Prüfgröße lautet $t=\frac{r_1\hat{\beta}_1+r_2\hat{\beta}_2-q}{\hat{\sigma}_{lk}}\sim t_{T-K-1}$ (bei Zweifachregression K = 2, also T - 3).

Hier, wie auch im Folgenden (Abschnitt E usw.) besteht stets ein einfacher Zusammenhang zwischen Tests und Intervallschätzung in der folgenden Art: Für letztere (Schätzung des Konfidenzintervalls) gelten die Grenzen $r_1\hat{\beta}_1+r_2\hat{\beta}_2\pm z_\alpha\hat{\sigma}_{lk}$, wobei z_α die der entsprechenden Verteilung (hier also der t_{T-3} - Verteilung bei K = 2) entnommene Signifikanzschranke bei einem Signifikanzniveau von α ist.

E. Simultane Tests von mehreren (L > 1) Linearkombinationen

Beispiel bei L = 2: $H_0:\beta_1=\beta_2=0$ (Hypothese der Irrelevanz aller Regressoren bei K = 2) oder $H_0:\beta_1+\beta_2=1$ ***und*** $\alpha=1$ (das sind L = 2 Restriktionen[1]); das bedeutet allgemein in Matrixschreibweise $\mathbf{R}\boldsymbol{\beta}=\mathbf{q}$ mit $\mathbf{R}=\begin{bmatrix}1&0&0\\0&1&1\end{bmatrix}$, $\boldsymbol{\beta}'=\begin{bmatrix}\alpha & \beta_1 & \beta_2\end{bmatrix}$ und $\mathbf{q}'=\begin{bmatrix}1 & 1\end{bmatrix}$. Ferner ist

$S_{\hat{u}\hat{u}}$ die Summe der Quadrate der Abweichungen $\left(\sum\hat{u}_t^2\right)$ im unrestringierten Modell,

$S^0_{\hat{u}\hat{u}}$ ist die entsprechende Summe im restringierten Modell ("bei Geltung von H_0").

Die Prüfgröße eines solchen Tests ist dann ganz allgemein (Gl. 10.14 bei v. Auer):

$$\boxed{F=\frac{(S^0_{\hat{u}\hat{u}}-S_{\hat{u}\hat{u}})/L}{S_{\hat{u}\hat{u}}/(T-K-1)}\sim F_{L,T-K-1}}$$

Der Test mit einer Linearkombination, also L = 1 (lineare Restriktion) ist der Spezialfall $t^2=F\sim F_{1,T-K-1}$, wobei diese Verteilung auch zugleich die t_{T-K-1} - Verteilung ist (man kann Fragestellung D auch mit einem F-Test statt einem t-Test bearbeiten; dann aber Probleme mit einer einseitigen Fragestellung).

In der gleichen Art: Test der Signifikanz hinzugenommener Regressoren[2] (z.B. Hinzukommen von und x_3, x_4, zum bisherigen Modell $\bar{y}=\alpha+\beta_1x_1+\beta_2x_2$). Zu prüfen H_0: $\beta_3=\beta_4=0$.

[1] Faustregel: Anzahl der Gleichheitszeichen zählen.
[2] Joint significance test on a subgroup of regression coefficients.

Das bisherige Modell gilt als restringiertes Modell (d.h. bei Geltung von H_0 oder der Irrelevanz der hinzugekommenen Regressoren) und man erhält dafür $S^0_{\hat{u}\hat{u}}$. Dagegen liefert das unrestringierte Modell $\hat{y} = \alpha + \beta_1 x_1 + \beta_3 x_3 + \beta_4 x_4$ die residuale Variation $S_{\hat{u}\hat{u}}$. Auch hier erfolgt die Konstruktion der Prüfgröße F nach dem folgenden Muster:

$F = \frac{\text{(zusätzliche) erklärte Varianz}}{\text{Residualvarianz}}$. Dabei ist eine Varianz stets eine Summe von Abweichungsquadraten (Variation) dividiert durch die Anzahl der Freiheitsgrade.

Im Beispiel $F = \frac{\left(S^0_{\hat{u}\hat{u}} - S_{\hat{u}\hat{u}}\right)/L}{S_{\hat{u}\hat{u}}/(T-K-1)} = \frac{\left(S^0_{\hat{u}\hat{u}} - S_{\hat{u}\hat{u}}\right)/2}{S_{\hat{u}\hat{u}}/(T-5)}$

F. Bestimmtheiten, Korrelationskoeffizienten

einfache Regression bedeutet einen (K = 1) Regressor

Variationsquelle	Freiheitsgrade	Varianz
erklärt Q1 = $S_{\hat{y}\hat{y}} = R^2_{y.123...} S_{yy}$	K	Q1 / K = V1
Residual Q2 = $S_{\hat{u}\hat{u}} = 1 - R^2_{y.123...} S_{yy}$	T – K - 1	Q2 / (T – K – 1) = V2

Die Prüfgröße[3] $\left(H_0 : \beta_1 = \beta_2 = ... = \beta_K = 0\right)$ $F = \frac{V_1}{V_2} = \frac{R^2/K}{\left(1-R^2\right)/\left(T-K-1\right)} = \frac{R^2(T-K-1)}{(1-R^2)K}$ ist F-verteilt mit K und T – K – 1 Freiheitsgraden.

Der Test von $\beta = 0$ bei einfacher Regression (oben D) ist ein Spezialfall hiervon. Er ist auch identisch mit dem folgenden Test des Korrelationskoeffizienten bei einfacher Regression.

$H_0 : \rho = 0$ Prüfgröße $(n \leq 50)$ $\quad t = \frac{r\sqrt{T-2}}{\sqrt{1-r^2}} \sim t_{T-2}$.

Bei H_0: $\rho \neq 0$ ist mit r und ρ die z-Transformation nach Fisher durchzuführen mit $r^* = \frac{1}{2}\ln\left(\frac{1+r}{1-r}\right)$ $\left(\rho \to \rho^* \text{ entsprechend}\right)$ Prüfgröße $\left(r^* - \rho^*\right)\sqrt{T-3} \sim N(0,1)$.

G. Prognoseintervall

1. ***Einfache Regression***

a) Schätzung des *Regresswerts* (durchschnittl. Prognose) $\hat{y}_o = E\left(Y|x_o\right)$: Hiervon (von $\hat{y}_0$) ist die geschätzte Varianz: $\hat{\sigma}^2\left[\frac{1}{T} + \frac{(x_o - \bar{x})^2}{S_{xx}}\right] = \hat{\sigma}^2_{\hat{y}_0}$ mit $\hat{\sigma}^2 = S_{\hat{u}\hat{u}}/(T-2)$ weil gilt K = 1

b) die geschätzte Varianz der Schätzung *eines individuellen Werts* $y_o = \hat{y}_o + u_o$ ist wegen der Varianz von u_0 um $\hat{\sigma}^2$ größer als im Fall a

(7.4) $$\hat{\sigma}^2_{\hat{y}o} + \hat{\sigma}^2 = \hat{\sigma}^2\left[1 + \frac{1}{T} + \frac{(x_o - \bar{x})^2}{S_{xx}}\right] = \widehat{\text{var}}\left(y_o\right)$$

[3] Diese Prüfgröße ist nicht identisch mit dem korrigierten Bestimmtheitsmaß (Gl. 13.11, S. 254 bei v. Auer)

2. ***Multiple Regression*** (K >1) ist die geschätzte Varianz von $\hat{y}_0$ (Fall b, bei a ohne die 1) $\hat{\sigma}^2\left[1+\mathbf{x}_o'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}_o\right]$; bei einfacher Regression liefert das mit $\mathbf{x}_0=\begin{bmatrix}1\\x_0\end{bmatrix}$ Formel 7.4.

H. Konfidenzellipsoide (gemeinsame Konfidenzintervalle für alle Regressionskoeffizienten

Einfache Regression: gemeinsames Konfidenzintervall für α und β

$\frac{\sum\left[(\hat{\alpha}-\alpha)+(\hat{\beta}-\beta)x\right]^2}{\sigma^2}\sim\chi_2^2$, bzw. $\frac{\sum_t\left[(\hat{\alpha}-\alpha)+(\hat{\beta}-\beta)x_t\right]^2}{2\hat{\sigma}^2}\sim F_{2,\,n-2}$, wenn σ^2 nicht bekannt ist und mit $\hat{\sigma}^2=S_{\hat{u}\hat{u}}/(T-2)$ geschätzt werden muß. Die Grenzen des Konfidenzintervalls sind beim Konfidenzellipsoid (im zweidimensionalen Fall eine Konfidenzellipse) weiter als die eindimensionalen Konfidenzintervalle für α und β gem. dem Aufgabentyp unter D. Man kann diese Aufgaben (siehe Abschn. D) als Spezialfälle des hier diskutieren simultanen Konfidenzbereichs betrachten).

Was im konkreten Fall die Berechnung einer Konfidenzellipse (z.B. bei einer Klausur) aufwändig macht ist dass $T(\hat{\alpha}-\alpha)^2+2(\hat{\alpha}-\alpha)(\hat{\beta}-\beta)\sum x_t+(\hat{\beta}-\beta)^2\sum x_t^2=2\hat{\sigma}^2F$ eine *quadratische* Gleichung ist, in der als Unbekannte α^2, $\alpha\beta$ und β^2 vorkommen (die übrigen Größen T, $\sum x_t, \sum x_t^2, \hat{\sigma}^2$ und F sind gegebene Größen [aufgrund der Stichprobendaten bzw. der Tabelle wie im Falle von F]).

Multiple Regression: man erhält ein K+1-dimensionales Konfidenzellipsoid bei K Regressoren $x_1, x_2, \ldots, x_K$ mit der quadratischen Form $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'\mathbf{V}^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\sim\chi_{K+1}^2$ wenn σ^2 bekannt ist, bzw. realistischer (wenn σ^2 unbekannt) mit $\frac{(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})'(\mathbf{X}'\mathbf{X})^{-1}(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})}{(K+1)\hat{\sigma}^2}\sim F_{K+1,\,T-K-1}$ als den geometrischen Ort aller Schätzwerte $\hat{\alpha},\hat{\beta}_1,\ldots,\hat{\beta}_K$ des Vektors $\hat{\boldsymbol{\beta}}$ (die Koeffizienten des Parametervektors $\boldsymbol{\beta}$ sind die Achsen) auffassen. Für ihn gilt, dass alle Kombinationen von $\hat{\alpha},\hat{\beta}_1,\ldots,\hat{\beta}_K$ Werten (also alle Punkte) auf der Kontur (es gilt =) oder innerhalb (es gilt <) eines (K+1)-dimensionalen Ellipsoids liegen, das mit $(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\mathbf{X}'\mathbf{X})(\hat{\boldsymbol{\beta}}-\boldsymbol{\beta})\le F\cdot(K+1)\cdot\hat{\sigma}^2$ gegeben ist. Gleichbedeutend ist: H_o: $\alpha=\beta_1=\ldots=\beta_K=0$ annehmen, wenn das Ellipsoid, das mit $1-\alpha$ (Konfidenzwahrscheinlichkeit) den wahren Parametervektor $\boldsymbol{\beta}$ umfasst, auch den Nullpunkt des Koordinatensystems umfasst.

I. Simultane Konfidenzregion einer Auswahl von m Regressionskoeffizienten (m < K+1): m-dimensionale Konfidenzellipsoide

Das ist die allgemeinste Betrachtungsweise: alle Aufgabentypen kann man als Spezialfälle hiervon betrachten. Durch eine geeignete Transformationsmatrix **T** vom Typ m × (K+1) kann man $\hat{\boldsymbol{\beta}}$ und $\boldsymbol{\beta}$ transformieren in $\hat{\boldsymbol{\beta}}_1$ bzw. $\boldsymbol{\beta}_1$, etwa in der folgenden Art:

$$\mathbf{T}\boldsymbol{\beta}=\begin{bmatrix}0&1&0&0&0&0\\0&0&0&1&0&0\end{bmatrix}\boldsymbol{\beta}=\begin{bmatrix}\beta_1\\\beta_3\end{bmatrix}=\boldsymbol{\beta}_1 \text{ mit } \boldsymbol{\beta}'=\begin{bmatrix}\alpha&\beta_1&\beta_2&\beta_3&\beta_4\end{bmatrix}.$$

Dann gilt: $(\mathbf{T}\hat{\boldsymbol{\beta}}-\mathbf{T}\boldsymbol{\beta})(\mathbf{TVT}')^{-1}(\mathbf{T}\hat{\boldsymbol{\beta}}-\mathbf{T}\boldsymbol{\beta})\sim\chi_m^2$ wenn σ^2 bekannt, bzw. wenn σ^2 unbekannt ist und mit $\hat{\sigma}^2$ geschätzt wird gilt $\frac{(\mathbf{T}\hat{\boldsymbol{\beta}}-\mathbf{T}\boldsymbol{\beta})\left[\mathbf{T}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{T}'\right](\mathbf{T}\hat{\boldsymbol{\beta}}-\mathbf{T}\boldsymbol{\beta})}{m\hat{\sigma}^2}\sim F_{m,\,n-k}$ für das Konfidenzellipsoid.