Das Verhalten von Schätzern in falsch spezifizierten Regressionsmodellen

Wolfgang Wefelmeyer (Universität zu Köln)

gemeinsam mit
Priscilla E. Greenwood (University of British Columbia)
Karl Küsters (Universität zu Köln)

mailto:wefelm@math.uni-koeln.de
http://www.mi.uni-koeln.de/~wefelm/

Seien $(X_1, Y_1), \ldots, (X_n, Y_n)$ Beobachtungen eines nichtlinearen Regressionsmodells

$$Y = r_\vartheta(X) + \varepsilon$$

mit $E\varepsilon = 0$ und X, ε unabhängig. Ein Schätzer für ϑ ist der *Kleinste-Quadrate-Schätzer*, eine Lösung der Schätzgleichung

$$\sum_{i=1}^{n} \dot{r}_\vartheta(X_i)(Y_i - r_\vartheta(X_i)) = 0.$$

Ein Schätzer für $Eh(Y) = Eh(r_\vartheta(X) + \varepsilon)$ ist die *von-Mises-Statistik*

$$\hat{M} = \frac{1}{n^2} \sum_{i,j=1}^{n} h(r_{\hat{\vartheta}}(X_i) - r_{\hat{\vartheta}}(X_j) + Y_j).$$

Wie verhalten sich die Schätzer, wenn entweder

(1) die wahre Regressionsfunktion nicht von der Form r_ϑ ist oder

(2) X, ε nicht unabhängig sind oder

(3) nicht $E\varepsilon = 0$ gilt?

Der Kleinste-Quadrate-Schätzer löst

$$\sum_{i=1}^{n} \dot{r}_\vartheta(X_i)(Y_i - r_\vartheta(X_i)) = 0,$$

konvergiert also *immer* zur Lösung von

$$E[\dot{r}_\vartheta(X)(Y - r_\vartheta(X))] = 0.$$

Das ist ϑ im richtigen Modell.

a) Wenn nicht $E\varepsilon = 0$ gilt, läßt sich das Modell schreiben als

$$Y = r_\tau(X) + \mu + \varepsilon \quad \text{mit } E\varepsilon = 0.$$

Also ist (3) ein Spezialfall von (1): Die wahre Regressionsfunktion ist nicht von der Form r_ϑ. Wir lassen (3) im folgenden weg.

b) Wenn die wahre Regressionsfunktion nicht von der Form r_ϑ ist, werden wir auch einen "Fehler" erhalten, der von X abhängt. Also lassen wir auch weg: (2) X, ε nicht unabhängig.

$\widehat{\vartheta}$ im richtigen Modell.

Durch stochastische Entwicklung der Schätzgleichung ergibt sich

$$\begin{aligned} 0 &= \frac{1}{n}\sum_{i=1}^{n} \dot{r}_{\widehat{\vartheta}}(X_i)(Y_i - r_{\widehat{\vartheta}}(X_i)) \\ &= \frac{1}{n}\sum_{i=1}^{n} \dot{r}_{\vartheta}(X_i)\varepsilon_i - (\widehat{\vartheta} - \vartheta)\frac{1}{n}\sum_{i=1}^{n} \dot{r}_{\vartheta}^2(X_i) + o_p(n^{-1/2}). \end{aligned}$$

Also ist $\widehat{\vartheta}$ asymptotisch linear,

$$\widehat{\vartheta} = \vartheta + (E\dot{r}_{\vartheta}^2(X))^{-1}\frac{1}{n}\sum_{i=1}^{n} \dot{r}_{\vartheta}(X_i)\varepsilon_i + o_p(n^{-1/2}).$$

$\widehat{\vartheta}$**, wenn** r_ϑ **falsch ist.** Das wahre Modell sei

$$Y = r(X) + \varepsilon, \quad X, \varepsilon \text{ unabhängig}, \quad E\varepsilon = 0.$$

Sei P die Verteilung von (X, Y). Dann schätzt $\widehat{\vartheta}$ die Lösung $\vartheta(P)$ von $E[\dot{r}_\vartheta(X)(Y - r_\vartheta(X))] = 0$. Der "Fehler" $\varepsilon_P = Y - r_{\vartheta(P)}(X)$ hat die Zerlegung $\varepsilon_P = \varepsilon + r(X) - r_{\vartheta(P)}(X)$, ist also weder unabhängig von X noch bedingt zentriert. Deshalb hat die stochastische Entwicklung der Schätzgleichung einen zusätzlichen Term, mit $\ddot{r}_{\vartheta(P)}$,

$$\begin{aligned} 0 \;=\; & \frac{1}{n}\sum_{i=1}^n \dot{r}_{\widehat{\vartheta}}(X_i)(Y_i - r_{\widehat{\vartheta}}(X_i)) = \frac{1}{n}\sum_{i=1}^n \dot{r}_{\vartheta(P)}(X_i)\varepsilon_{Pi} \\ & -(\widehat{\vartheta} - \vartheta(P))\Big(\frac{1}{n}\sum_{i=1}^n \dot{r}^2_{\vartheta(P)}(X_i) - \frac{1}{n}\sum_{i=1}^n \ddot{r}_{\vartheta(P)}(X_i)\varepsilon_{Pi}\Big) \\ & + o_p(n^{-1/2}), \end{aligned}$$

Resultat für $\widehat{\vartheta}$, wenn r_ϑ falsch ist:

Wenn die Regressionsfunktion r statt r_ϑ ist, schätzt $\widehat{\vartheta}$ die Lösung $\vartheta(P)$ von $E[\dot{r}_\vartheta(X)(Y - r_\vartheta(X))] = 0$ und ist asymptotisch linear,

$$\widehat{\vartheta} \;=\; \vartheta(P) + c^{-1}(P)\frac{1}{n}\sum_{i=1}^{n} \dot{r}_{\vartheta(P)}(X_i)\varepsilon_{Pi} + o_p(n^{-1/2}),$$

und

$$c(P) = E\dot{r}^2_{\vartheta(P)}(X) - E[\ddot{r}_{\vartheta(P)}(X)\varepsilon_P].$$

hängt auch von der *zweiten* Ableitung $\ddot{r}_{\vartheta(P)}$ ab.

Schätzen von $Eh(Y)$.

Der übliche Schätzer für einen Erwartungswert $Eh(Y)$ ist der empirische Schätzer $\frac{1}{n}\sum_{i=1}^n h(Y_i)$. Er verwendet das Modell nicht und ist robust gegen jede Fehlspezifikation. Die Faltungsdarstellung $Y = r_\vartheta(X) + \varepsilon$ wird verwendet von der *Von-Mises-Statistik*

$$\hat{M} = \frac{1}{n^2}\sum_{i,j=1}^n h(r_{\hat\vartheta}(X_i) + \hat\varepsilon_j) = \frac{1}{n^2}\sum_{i,j=1}^n h(r_{\hat\vartheta}(X_i) - r_{\hat\vartheta}(X_j) + Y_j)$$

mit *Residuen*

$$\hat\varepsilon_j = Y_j - r_{\hat\vartheta}(X_j).$$

($\hat{M}$ wird effizient, wenn $\hat\vartheta$ effizient für ϑ ist und die Information $E\varepsilon = 0$ durch Owen weights oder additive Korrektur verwendet wird.)

Von-Mises-Statistik für $Eh(Y)$ im richtigen Modell.

Erst $\hat{M}$ als Funktion von $\hat{\vartheta}$ um ϑ entwickeln,

$$\begin{aligned}\hat{M} &= \frac{1}{n^2}\sum_{i,j=1}^{n} h(r_{\hat{\vartheta}}(X_i)+\hat{\varepsilon}_j) = \frac{1}{n^2}\sum_{i,j=1}^{n} h(r_{\vartheta}(X_i)+\varepsilon_j) \\ &\quad +(\hat{\vartheta}-\vartheta)\frac{1}{n^2}\sum_{i,j=1}^{n}(\dot{r}_{\vartheta}(X_i)-\dot{r}_{\vartheta}(X_j))h'(r_{\vartheta}(X_i)+\varepsilon_j)+o_p(n^{-1/2}) \\ &= \frac{1}{n^2}\sum_{i,j=1}^{n} h(r_{\vartheta}(X_i)+\varepsilon_j)+(\hat{\vartheta}-\vartheta)H+o_p(n^{-1/2})\end{aligned}$$

mit $H = E[\dot{r}_{\vartheta}(X)h'(Y)] - E\dot{r}_{\vartheta}(X)Eh'(Y)$.
Dann Hoeffding-Zerlegung,

$$\frac{1}{n^2}\sum_{i,j=1}^{n} h(r_{\vartheta}(X_i)+\varepsilon_j) = M + \frac{1}{n}\sum_{i=1}^{n}\Big(h_{X_i}-M+h_{\varepsilon_i}-M\Big)+o_p(n^{-1/2})$$

mit

$$h_X = E(h(Y)|X), \quad h_{\varepsilon} = E(h(Y)|\varepsilon).$$

Von-Mises-Statistik für $Eh(Y)$, wenn r_ϑ falsch ist. Das wahre Modell sei

$$Y = r(X) + \varepsilon, \quad X, \varepsilon \text{ unabhängig}, \quad E\varepsilon = 0.$$

Sei P die Verteilung von (X, Y). Dann schätzt $\hat{\vartheta}$ die Lösung $\vartheta(P)$ von $E[\dot{r}_\vartheta(X)(Y - r_\vartheta(X))] = 0$, und

$$\varepsilon_P = Y - r_{\vartheta(P)}(X) = \varepsilon + r(X) - r_{\vartheta(P)}(X) \quad \text{und} \quad r_{\vartheta(P)}(X)$$

sind *abhängig*. Die Von-Mises-Statistik

$$\widehat{M} = \frac{1}{n^2} \sum_{i,j=1}^{n} h(r_{\hat{\vartheta}}(X_i) + \hat{\varepsilon}_j)$$

wird aber von den Termen mit $i \neq j$ dominiert, schätzt also

$$M(P) = Eh(r_{\vartheta(P)}(X_1) + \varepsilon_{P2}) = \int h(r_{\vartheta(P)}(x) + y) P^X \otimes P^{\varepsilon_P}(dx, dy).$$

Stochastische Entwicklung von $\widehat{M}$:

$$\begin{aligned}
\widehat{M} &= \frac{1}{n^2}\sum_{i,j=1}^{n} h(r_{\widehat{\vartheta}}(X_i)+\widehat{\varepsilon}_j) = \frac{1}{n^2}\sum_{i,j=1}^{n} h(r_{\vartheta(P)}(X_i)+\varepsilon_{Pj}) \\
&\quad +(\widehat{\vartheta}-\vartheta(P))\frac{1}{n^2}\sum_{i,j=1}^{n}(\dot{r}_{\vartheta(P)}(X_i)-\dot{r}_{\vartheta(P)}(X_j))h'(r_{\vartheta(P)}(X_i)+\varepsilon_{Pj}) \\
&\quad +o_p(n^{-1/2}) \\
&= \frac{1}{n^2}\sum_{i,j=1}^{n} h(r_{\vartheta(P)}(X_i)+\varepsilon_{Pj}) + (\widehat{\vartheta}-\vartheta(P))H(P)+o_p(n^{-1/2})
\end{aligned}$$

mit

$$\begin{aligned}
H(P) &= E[r_{\vartheta(P)}(X_1)h'(r_{\vartheta(P)}(X_1)+\varepsilon_{P2})] \\
&\quad -E[r_{\vartheta(P)}(X_2)h'(r_{\vartheta(P)}(X_1)+\varepsilon_{P2})].
\end{aligned}$$

Hoeffding-Zerlegung:

$$\frac{1}{n^2}\sum_{i,j=1}^{n} h(r_{\vartheta(P)}(X_i)+\varepsilon_{Pj})$$
$$= M(P)+\frac{1}{n}\sum_{i=1}^{n}\Big(h_{PX_i}-M(P)+h_{P\varepsilon_{Pi}}-M(P)\Big)+o_p(n^{-1/2})$$

mit

$$\begin{aligned} h_{PX_1} &= E(h(r_{\vartheta(P)}(X_1)+\varepsilon_{P2})|X_1), \\ h_{P\varepsilon_{P2}} &= E(h(r_{\vartheta(P)}(X_1)+\varepsilon_{P2})|\varepsilon_{P2}). \end{aligned}$$

Beispiel: Lineare Regression, Fehler nicht zentriert. $Y=\vartheta X+\varepsilon$ mit $E\varepsilon\neq 0$ ist äquivalent zu $Y=\vartheta X+\mu+\varepsilon$ mit $E\varepsilon=0$.
Die Von-Mises-Statistik $\widehat{M}$ schätzt dann

$$M(P)=Eh\Big(Y_2+\Big(\vartheta+\mu\frac{EX}{EX^2}\Big)(X_1-X_2)\Big).$$