# Schätzer in heteroskedastischen nichtlinearen autoregressiven Modellen mit zusätzlichen strukturellen Annahmen

Wolfgang Wefelmeyer (Universität zu Köln)

gemeinsam mit
Anton Schick (Binghamton University)
Uschi Müller (Texas A & M University)

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

Seien $X_{p-1}, \ldots, X_n$ Beobachtungen einer Markov-Kette. Ordnung $p$.
Gegeben: parametrisches Modell für den bedingten Erwartungswert:

$$E(X_i | \mathbf{X}_{i-1}) = r_\vartheta(\mathbf{X}_{i-1})$$

mit $\mathbf{X}_{i-1} = (X_{i-p}, \ldots, X_{i-1})$. Äquivalente Beschreibung des Modells:

$$X_i - r_\vartheta(\mathbf{X}_{i-1}) \quad \text{ist Martingal-Inkrement.}$$

Übliche Schätzer: Martingal-Schätzgleichungen

$$\sum_{i=1}^{n} w_\vartheta(\mathbf{X}_{i-1})\Big(X_i - r_\vartheta(\mathbf{X}_{i-1})\Big) = 0.$$

Optimale Gewichte

$$w_\vartheta(\mathbf{X}_{i-1}) = \sigma^{-2}(\mathbf{X}_{i-1})\dot{r}_\vartheta(\mathbf{X}_{i-1})$$

mit $\sigma^2(\mathbf{X}_{i-1}) = E((X_i - r_\vartheta(\mathbf{X}_{i-1}))^2 | \mathbf{X}_{i-1})$ und $\dot{r}_\vartheta$ Gradient von $r_\vartheta$.
Die optimalen Gewichte müssen geschätzt werden: $\sigma^2(\mathbf{X}_{i-1})$ durch Nadaraya–Watson-Schätzer ersetzen. Effizient: Wef 1996 AS.

Jetzt zusätzliche strukturelle Annahmen and die Übergangsverteilung des Martingals. Setze

$$\varepsilon_i = X_i - r_\vartheta(\mathbf{X}_{i-1}).$$

Bezeichne $t(\mathbf{X}_{i-1}, y)$ bedingte Dichte von $\varepsilon_i$. Es gilt $\int yt(\mathbf{x}, y)\, dy = 0$ mit $\mathbf{x} = (x_1, \ldots, x_p)$.

1. $t$ *ist partiell unabhängig von der Vergangenheit*

$$t(\mathbf{x}, y) = t_0(B\mathbf{x}, y), \quad B : \mathbb{R}^p \to \mathbb{R}^q, \quad q \leq p.$$

Beispiel: kürzeres Gedächtnis: $B(x_1, \ldots, x_p) = (x_{p-q}, \ldots, x_p)$.
Beispiel: unabhängige Innovationen: $B\mathbf{x} = 0$.

2. $t$ *ist invariant unter Gruppe von Transformationen*

$$t(\mathbf{z}) = t(B_j\mathbf{z}), \quad B_j : \mathbb{R}^{p+1} \to \mathbb{R}^{p+1}, \quad j = 1, \ldots, m, \quad \mathbf{z} = (\mathbf{x}, y).$$

Beispiel: symmetrisch um 0: $t(\mathbf{x}, y) = t(-\mathbf{x}, -y)$.
Beispiel: bedingt symmetrisch: $t(\mathbf{x}, y) = t(\mathbf{x}, 2x_p - y)$.

Effiziente Schätzer nicht mehr durch Schätzgleichung.
Stattdessen ähnlich wie bei unabhängigen Innovationen,
Koul/Schick 1997 Bernoulli:
Erst effiziente Scorefunktion $s(\mathbf{x}, y)$ bestimmen.
Ein effizienter Schätzer $\hat{\vartheta}$ ist *charakterisiert* durch

$$\hat{\vartheta} = \vartheta + \Lambda^{-1} \frac{1}{n} \sum_{i=1}^{n} s(\mathbf{X}_{i-1}, \varepsilon_i) + o_p(n^{-1/2})$$

mit $\Lambda = E[s(\mathbf{X}, \varepsilon) s^\top(\mathbf{X}, \varepsilon)]$ “Informationsmatrix” für $\vartheta$
(und $\Lambda^{-1} s(\mathbf{x}, y)$ “Gradient” für $\vartheta$).

Er wird *konstruiert* mit one-step (Newton–Raphson)

$$\hat{\vartheta} = \tilde{\vartheta} + \tilde{\Lambda}^{-1} \frac{1}{n} \sum_{i=1}^{n} \tilde{s}(\mathbf{X}_{i-1}, \tilde{\varepsilon}_i).$$

mit Schätzer $\tilde{s}$, geschätzten Innovationen $\tilde{\varepsilon}_i = X_i - r_{\tilde{\vartheta}}(\mathbf{X}_{i-1})$, und

$$\tilde{\Lambda} = \frac{1}{n} \sum_{i=1}^{n} \tilde{s}(\mathbf{X}_{i-1}, \tilde{\varepsilon}_i) \tilde{s}^\top(\mathbf{X}_{i-1}, \tilde{\varepsilon}_i).$$

1. $t$ *ist partiell unabhängig von der Vergangenheit*

$$t(\mathbf{x}, y) = t_0(B\mathbf{x}, y), \quad B : \mathbb{R}^p \to \mathbb{R}^q, \quad q \le p.$$

Setze $t_0'(B\mathbf{x}, y) = \partial_y t_0(B\mathbf{x}, y)$ und

$$\ell_0 = -t_0'/t_0, \quad \sigma_0^2(B\mathbf{X}) = E(\varepsilon^2 | B\mathbf{X}), \quad \varrho(B\mathbf{X}) = E(\dot{r}_\vartheta(\mathbf{X}) | B\mathbf{X}).$$

Effiziente Scorefunktion (das Argument lasse ich weg):

$$s(\mathbf{X}, \varepsilon) = (\dot{r}_\vartheta(\mathbf{X}) - \varrho(B\mathbf{X}))\ell_0(B\mathbf{X}, \varepsilon) + \varrho(B\mathbf{X})\sigma_0^{-2}(B\mathbf{X})\varepsilon.$$

Schätze $t_0$ mit Kernschätzer und Beobachtungen $B\mathbf{X}_i$.
Sei $g$ die Dichte von $\mathbf{X}_{i-1}$. Schreibe

$$\varrho(\mathbf{b}) = \frac{\int_{B\mathbf{x}=\mathbf{b}} \dot{r}(\mathbf{x}) g(\mathbf{x}) \, d\mathbf{x}}{\int_{B\mathbf{x}=\mathbf{b}} g(\mathbf{x}) \, d\mathbf{x}}.$$

Zu schätzen mit verallgemeinertem Nadaraya–Watson-Schätzer.
Ebenso für $\sigma_0^2$.
Beispiel: nichtparametrisch: $B\mathbf{x} = \mathbf{x}$, $t_0 = t$, $\sigma_0 = \sigma$, $\varrho = E[\dot{r}_\vartheta(\mathbf{X})]$.

2. *$t$ ist invariant unter Gruppe von Transformationen*

$$t(\mathbf{z}) = t(B_j\mathbf{z}), \quad B_j : \mathbb{R}^{p+1} \to \mathbb{R}^{p+1}, \quad j = 1, \dots, m, \quad \mathbf{z} = (\mathbf{x}, y).$$

Setze $t'(\mathbf{x}, y) = \partial_y t(\mathbf{x}, y)$, $\ell = -t'/t$, $\sigma^2(\mathbf{X}) = E(\varepsilon^2|\mathbf{X})$ und

$$\lambda(\mathbf{x}, y) = \dot{r}_\vartheta(\mathbf{x})\ell(\mathbf{x}, y), \qquad \mu(\mathbf{x}, y) = \dot{r}_\vartheta(\mathbf{x})\sigma^{-2}(\mathbf{x})y.$$

Symmetrisiere $\lambda$ und $\mu$,

$$\lambda_0(\mathbf{z}) = \frac{1}{n}\sum_{j=1}^{n} \lambda(B_j\mathbf{z}), \qquad \mu_0(\mathbf{z}) = \frac{1}{n}\sum_{j=1}^{n} \mu(B_j\mathbf{z}).$$

Effiziente Scorefunktion:

$$s = \lambda - \lambda_0 + \mu_0.$$

Schätze $t$ mit Kernschätzer, $\sigma^2$ mit Nadaraya–Watson-Schätzer.

Keine Verbesserung, wenn $\varepsilon$ bedingt normalverteilt ist.
Dann $\ell(\mathbf{x}, y) = \sigma^{-2}(\mathbf{x})y$, also $\lambda - \mu = 0$, also $\lambda_0 - \mu_0 = 0$.