Variance bounds for estimators in autoregressive models with constraints

Wolfgang Wefelmeyer (University of Cologne)

based on joint work with
Anton Schick (Binghamton University)
Ursula U. Müller (Texas A & M University)

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

Let $X_{1-p}, \dots, X_n$ be observations of a Markov chain of order p, with a parametric model for the conditional mean,

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

where $\mathbf{X}_{i-1} = (X_{i-p}, \dots, X_{i-1})$ and ϑ is an unknown d-dimensional parameter. An efficient estimator for ϑ in this model is a randomly weighted least squares estimator that solves the estimating equation

$$\sum_{i=1}^{n} \tilde{\sigma}^{-2}(\mathbf{X}_{i-1})\dot{r}_\vartheta(\mathbf{X}_{i-1})\left(X_i - r_\vartheta(\mathbf{X}_{i-1})\right) = 0,$$

where $\dot{r}_\vartheta$ is the vector of partial derivatives of r_ϑ with respect to ϑ, and $\tilde{\sigma}^2(\mathbf{X}_{i-1})$ estimates the conditional variance

$$\sigma^2(\mathbf{X}_{i-1}) = E((X_i - r_\vartheta(\mathbf{X}_{i-1}))^2|\mathbf{X}_{i-1}).$$

Aside: The optimal weights are never *parametric* functions; we always need *nonparametric* estimators. Wef 1996 AS.

The autoregressive model $E(X_i|\mathbf{X}_{i-1}) = r_\vartheta(\mathbf{X}_{i-1})$ can be described through its *transition distribution*

$$A(\mathbf{x}, dy) = T(\mathbf{x}, dy - r_\vartheta(\mathbf{x}))$$

with $\int T(\mathbf{x}, dy)y = 0$ for $\mathbf{x} = (x_1, \ldots, x_p)$.
It can also be written as a nonlinear autoregressive model

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

with ε_i a *martingale increment*: depends on the past through $\mathbf{X}_{i-1}$ only and has conditional distribution $T(\mathbf{x}, dy)$ with $\int T(\mathbf{x}, dy)y = 0$.

We now assume that we know something about the form of T. Then it is useful to describe the model through its *transition distribution*. Optimal estimators not as weighted least squares but as *one-step (Newton–Raphson)* estimators.

(Possible other approaches: constrained M-estimator, Rao/Wu 2009; empirical likelihood, Owen 2001.)

Our model has transition distribution $A(\mathbf{x}, dy) = T(\mathbf{x}, dy - r_\vartheta(\mathbf{x}))$ with $\int T(\mathbf{x}, dy)y = 0$ and *additional constraint*

(1) T is *partially independent* of the past, i.e., $T(\mathbf{x}, dy) = T_0(B\mathbf{x}, dy)$ for a known function $B : \mathbb{R}^p \to \mathbb{R}^q$ with $0 \leq q \leq p$.

(2) T is *invariant* under transformation group $B_j : \mathbb{R}^{p+1} \to \mathbb{R}^{p+1}$, $j = 1, \ldots, m$. (T has density t with $t(\mathbf{z}) = t(B_j\mathbf{z})$ for $\mathbf{z} = (\mathbf{x}, y)$.)

Optimal estimators for ϑ_j (and then jointly) are now constructed *differently*: *First* determine Cramér–Rao bound and influence function in the least favorable one-dimensional submodel; *then* construct one-step estimator with this influence function.

Perturb parameter as $\vartheta_{nu} = \vartheta + n^{-1/2}u$, and transition density as $t_{nv}(\mathbf{x}, y) = t(\mathbf{x}, y)(1 + n^{-1/2}v(\mathbf{x}, y))$. The log-likelihood ratio of the observations $X_{d-1}, \ldots, X_n$ is *locally asymptotically normal*, i.e. approximated by

$$n^{-1/2} \sum_{i=1}^{n} s_{uv}(\mathbf{X}_{i-1}, \varepsilon_i) - \frac{1}{2} E[s_{uv}^2(\mathbf{X}, \varepsilon)],$$

where

$$s_{uv}(\mathbf{x}, y) = u^\top \dot{r}(\mathbf{x}) \ell(\mathbf{x}, y) + v(\mathbf{x}, y)$$

with $\ell = -t'/t$ and $t'(\mathbf{x}, y) = \partial_y t(\mathbf{x}, y)$, and $\dot{r} = \partial_\vartheta r_\vartheta$.

The influence function for ϑ_j in the *least favorable* submodel is the *gradient* of ϑ_j, determined as $s_{u^*v^*}$ such that

$$n^{1/2}(\vartheta_{nu,j} - \vartheta_j) = u = E[s_{u^*v^*}(\mathbf{X}, \varepsilon) s_{uv}(\mathbf{X}, \varepsilon)], \quad \text{all } u, v.$$

The variance bound is Var $s_{u^*v^*}(\mathbf{X}, \varepsilon)$. Constraint on t also constrains the possible perturbations v, which leads to different u^* and v^*.

An efficient estimator $\widehat{\vartheta}$ of ϑ is *asymptotically linear* with *influence function* equal to the gradient $s_{u^*v^*}$,

$$\widehat{\vartheta} = \vartheta + \frac{1}{n}\sum_{i=1}^{n} s_{u^*v^*}(\mathbf{X}_{i-1}, \varepsilon_i) + o_{P_n}(n^{-1/2}).$$

A one-step (Newton–Raphson) improvement $\widehat{\vartheta}$ of an *initial* estimator $\tilde{\vartheta}$ is of the form

$$\widehat{\vartheta} = \tilde{\vartheta} + \frac{1}{n}\sum_{i=1}^{n} \tilde{s}_{u^*v^*}(\mathbf{X}_{i-1}, \tilde{\varepsilon}_i) + o_{P_n}(n^{-1/2})$$

with $\tilde{\varepsilon}_i = X_i - r_{\tilde{\vartheta}}(\mathbf{X}_{i-1})$.

(1) Our model has transition density $a(\mathbf{x}, y) = t(\mathbf{x}, y - r_\vartheta(\mathbf{x}))$ with $\int y t(\mathbf{x}, dy)\, dy = 0$ that is partially independent of the past: $t(\mathbf{x}, y) = t_0(B\mathbf{x}, y)$ for a (known) function $B : \mathbb{R}^p \to \mathbb{R}^q$ with $0 \le q \le p$.

The efficient influence function for ϑ is $\Lambda^{-1}\tau(\mathbf{x}, y)$ with *score vector*

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

and *information matrix* $\Lambda = E[\tau(\mathbf{X}, \varepsilon)\tau^\top(\mathbf{X}, \varepsilon)]$.
Here $\dot{r} = \partial_\vartheta r_\vartheta$, $\ell_0 = -t_0'/t_0$ with $t_0'(\mathbf{x}, y) = \partial_y t_0(\mathbf{x}, y)$,

$$\varrho(\mathbf{b}) = E(\dot{r}(\mathbf{X})|B\mathbf{X} = \mathbf{b}) = \frac{\int_{B\mathbf{x}=\mathbf{b}} \dot{r}_\vartheta(\mathbf{x})g(\mathbf{x})\, d\mathbf{x}}{\int_{B\mathbf{x}=\mathbf{b}} g(\mathbf{x})\, d\mathbf{x}},$$

$$\sigma_0^2(\mathbf{b}) = E(\varepsilon^2|B\mathbf{X} = b) = \frac{\int y^2 h_0(\mathbf{b}, y)\, dy}{\int h_0(\mathbf{b}, y)\, dy},$$

with g and h_0 densities of $\mathbf{X}$ and $(B\mathbf{X}, \varepsilon)$. To estimate ϑ efficiently, we therefore need estimators for the efficient score function, i.e. $(p + 1)$-dimensional density estimators and (generalized) Nadaraya–Watson estimators. No gain if $t_0(\mathbf{b}, \cdot)$ are normal densities.

(2) Our model has transition density $a(\mathbf{x}, y) = t(\mathbf{x}, y - r_\vartheta(\mathbf{x}))$ with $\int yt(\mathbf{x}, dy)\, dy = 0$ that is invariant under a group of transformations:

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

The efficient influence function for ϑ is $\Lambda^{-1}\tau(\mathbf{x}, y)$ with score vector $\tau = \lambda - \lambda_0 + \mu_0$ and information matrix $\Lambda = E[\tau(\mathbf{X}, \varepsilon)\tau^\top(\mathbf{X}, \varepsilon)]$, with symmetrizations

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

of

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

where $\dot r = \partial_\vartheta r_\vartheta$ and $\ell = -t'/t$ with $t'(\mathbf{x}, y) = \partial_y t(\mathbf{x}, y)$, and $\sigma^2(\mathbf{x}) = E(\varepsilon^2|\mathbf{X} = \mathbf{x})$. To estimate ϑ efficiently, we need estimators for these expressions. No gain if $t(\mathbf{b}, \cdot)$ are normal densities.