Efficient estimators in nonlinear and heteroscedastic autoregressive models with constraints

Wolfgang Wefelmeyer
University of Cologne

jointly with
Ursula U. Müller (Texas A&M University)
and Anton Schick (Binghamton University)

A *nonlinear and heteroscedastic autoregressive model* (of order 1, for simplicity) is a first-order Markov chain with parametric models for the conditional mean and variance,

$$\begin{aligned} E(X_i|X_{i-1}) &= r_\vartheta(X_{i-1}), \\ E((X_i - r_\vartheta(X_{i-1}))^2|X_{i-1}) &= s_\vartheta^2(X_{i-1}). \end{aligned}$$

The model is also called *quasi-likelihood model*. We want to estimate ϑ efficiently. (For simplicity, ϑ is one-dimensional.) The *least squares estimator* minimizes

$$\sum_{i=1}^n (X_i - r_\vartheta(X_{i-1}))^2,$$

i.e. it solves the *martingale estimating equation*

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

(The dot means derivative w.r.t. ϑ.)

The least squares estimator is improved by weighing the terms in the estimating equation with inverse conditional variances,

$$\sum_{i=1}^{n} s_\vartheta^{-2}(X_{i-1})\dot r_\vartheta(X_{i-1})(X_i - r_\vartheta(X_{i-1})) = 0.$$

This *quasi-likelihood estimator* is still inefficient; it ignores the information in the model for the conditional variance. Better estimators are obtained from estimating equations of the form

$$\sum_{i=1}^{n} v(X_{i-1})(X_i - r_\vartheta(X_{i-1})) + w(X_{i-1})\Big((X_i - r_\vartheta(X_{i-1}))^2 - s_\vartheta^2(X_{i-1})\Big) = 0.$$

The best weights (not given explicitly here) minimize the asymptotic variance; they involve third and fourth conditional moments

$$E((X_i - r_\vartheta(X_{i-1}))^k | X_{i-1}), \quad k = 3, 4,$$

which must be estimated nonparametrically (by Nadaraya–Watson). The resulting estimator for ϑ is efficient, W. 1996, Müller/W. 2002. The improvement over the quasi-likelihood estimator can be large.

In this talk we are interested in models with *additional* information on the transition density. Then the above approach breaks down. We also use a different *description* of the model. Let $t(x, y)$ denote a standardized *conditional innovation density* (mean 0, variance 1). Introduce *conditional location and scale parameters*,

$$q(x, y) = \frac{1}{s_\vartheta(x)} t\Big(\frac{y - r_\vartheta(x)}{s_\vartheta(x)}\Big).$$

This describes the quasi-likelihood model.

We can now put constraints on t:

1. $t(x, y) = f(y)$: heteroscedastic and nonlinear regression with independent innovations.
2. no constraint.
3. $t(x, y) = t(x, 2x - y)$: symmetric innovations.
4. $t(x, y) = t(Ax, y)$ for a known A: partial invariance.

(Models 1. and 2. are known but treated differently here.)

For simplicity, in the following we treat only the *homoscedastic model*: We have a Markov chain with transition density

$$q(x,y) = t(x, y - r_\vartheta(x))$$

where t has conditional mean zero, $\int y t(x,y)\,dy = 0$.

Equivalently, we have a Markov chain with conditional mean

$$E(X_i|X_{i-1}) = r_\vartheta(X_{i-1}).$$

With no further information on t, an efficient estimator of ϑ is the weighted least squares estimator

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

where $\tilde{\sigma}^2(x)$ is a Nadaraya–Watson estimator for $\sigma^2(x)$.

We characterize efficient estimators using the Hájek–Le Cam approach via *local asymptotic normality*. Perturb the parameters ϑ and t as $\vartheta + n^{-1/2}u$ and $t(x,y)(1+n^{-1/2}v(x,y))$ with $u \in \mathbb{R}$ and v in a space V that depends on what we know about t. Write $\varepsilon_i = X_i - r_\vartheta(X_{i-1})$. We get for the log-likelihood

$$\log \frac{dP_{nuv}}{dP_n} = n^{-1/2} \sum_{i=1}^{n} s_{uv}(X_{i-1}, \varepsilon_i) - \frac{1}{2} E s_{uv}^2(X, \varepsilon) + o_{P_n}(1)$$

with $s_{uv}(X,\varepsilon) = u\dot{r}(X)\ell(X,\varepsilon) + v(X,\varepsilon)$ and $\ell = -t'/t$.
An efficient estimator $\hat{\vartheta}$ for ϑ is characterized by

$$n^{1/2}(\hat{\vartheta} - \vartheta) = n^{-1/2} \sum_{i=1}^{n} g(X_{i-1}, \varepsilon_i) + o_{P_n}(1)$$

with $g = s_{u_* v_*}(X, \varepsilon)$ determined by

$$n^{1/2}((\vartheta + n^{-1/2}u) - \vartheta) = u = E s_{u_* v_*}(X,\varepsilon) s_{uv}(X,\varepsilon), \quad u \in \mathbb{R},\, v \in V.$$

I.e. we express the perturbation of ϑ in terms of the inner product induced by the LAN variance.

1. $t(x,y)=f(y)$: **heteroscedastic and nonlinear regression with independent innovations**.

We obtain the efficient influence function

$$g(X,\varepsilon)=\Lambda^{-1}\Big((\dot r(X)-\mu)\ell(\varepsilon)+\sigma^{-2}\mu\varepsilon\Big)$$

with $\ell=-f'/f$, $\mu=E\dot r(X)$, $\sigma^2=E\varepsilon^2$ and $\Lambda=J(R-\mu^2)+\sigma^{-2}\mu^2$ with $J=E\ell^2(\varepsilon)$ and $R=E\dot r^2(X)$. Different route in Koul and Schick 1997. An efficient estimator $\hat\vartheta$ of ϑ can be obtained as one-step improvement of an initial estimator $\tilde\vartheta$ (e.g. least squares),

$$\hat\vartheta=\tilde\vartheta+\frac{1}{n}\sum_{i=1}^{n}\tilde g(X_{i-1},\tilde\varepsilon_i)$$

with $\tilde g(X,\varepsilon)=\tilde\Lambda^{-1}\Big((\dot r_{\tilde\vartheta}(X)-\tilde\mu)\tilde\ell(\varepsilon)+\tilde\sigma^{-2}\tilde\mu\varepsilon\Big)$, residual estimators $\tilde\varepsilon_i=X_i-r_{\tilde\vartheta}(X_{i-1})$, empirical estimators $\tilde\mu=\frac{1}{n}\sum_{i=1}^{n}\dot r_{\tilde\vartheta}(X_{i-1})$ and $\tilde\sigma^2=\frac{1}{n}\sum_{i=1}^{n}\tilde\varepsilon_i^2$, and $\tilde\ell=-\tilde f'/\tilde f$ for a kernel estimator $\tilde f$, and with $\tilde\Lambda=\tilde J(\tilde R-\tilde\mu^2)+\tilde\sigma^{-2}\tilde\mu^2$ and $\tilde J=\frac{1}{n}\sum_{i=1}^{n}\tilde\ell^2(\tilde\varepsilon_i)$, $\tilde R=\frac{1}{n}\sum_{i=1}^{n}\dot r_{\tilde\vartheta}^2(X_{i-1})$.

2. **no constraint on** $t(x,y)$.

We obtain the efficient influence function

$$g(X,\varepsilon) = M^{-1}\dot r(X)\sigma^{-2}(X)\varepsilon$$

with $\sigma^2(x) = \int y^2 t(x,y)\,dy$ and $M = E\sigma^{-2}(X)\dot r^2(X)$. We have already obtained an efficient estimator as an appropriately weighted least squares estimator $\sum_{i=1}^n \tilde\sigma^{-2}(X_{i-1})\dot r_\vartheta(X_{i-1})(X_i - r_\vartheta(X_{i-1})) = 0$. Here we obtain another efficient estimator as one-step improvement of an initial estimator $\tilde\vartheta$ (e.g. least squares),

$$\hat\vartheta = \tilde\vartheta + \frac{1}{n}\sum_{i=1}^n \tilde g(X_{i-1},\tilde\varepsilon_i)$$

with

$$\tilde g(X,\varepsilon) = \tilde M^{-1}\dot r_{\tilde\vartheta}(X)\tilde\sigma^{-2}(X)\varepsilon, \quad \tilde M = \frac{1}{n}\sum_{i=1}^n \tilde\sigma^{-2}(X_{i-1})\dot r^2_{\tilde\vartheta}(X_{i-1})$$

and $\tilde\sigma^2(x)$ the Nadaraya–Watson estimator for $\sigma^2(x)$.

3. $t(x,y) = t(x, 2x-y)$: **symmetric innovations.**

We obtain the efficient influence function

$$g(X,\varepsilon) = T^{-1}\dot{r}(X)\ell(X,\varepsilon)$$

with $T = E\dot{r}^2(X)\ell^2(X,\varepsilon)$.

We obtain an efficient estimator $\hat{\vartheta}$ of ϑ as one-step improvement of an initial estimator $\tilde{\vartheta}$ (e.g. least squares),

$$\hat{\vartheta} = \tilde{\vartheta} + \frac{1}{n}\sum_{i=1}^{n} \tilde{g}(X_{i-1}, \tilde{\varepsilon}_i)$$

with

$$\tilde{g}(X,\varepsilon) = \tilde{T}^{-1}\dot{r}_{\tilde{\vartheta}}(X)\tilde{\ell}(X,\varepsilon),$$
$$\tilde{T} = \frac{1}{n}\sum_{i=1}^{n} \dot{r}_{\tilde{\vartheta}}^2(X_{i-1})\tilde{\ell}^2(X_{i-1},\tilde{\varepsilon}_i)$$

and $\tilde{\ell} = -\tilde{t}'/\tilde{t}$ with $\tilde{t}$ a Nadaraya–Watson estimator for t.