# Efficient estimation in a semiparametric heteroscedastic autoregressive model

BY ANTON SCHICK

*Binghamton University*

This paper characterizes and constructs efficient estimators of the autoregression parameter in the heteroscedastic autoregression model of order 1 with unknown innovation density and unknown volatility function.

**1. Introduction.** In this paper I consider a stationary and ergodic semiparametric heteroscedastic autoregressive model of order 1. More precisely, I assume that the observations $X_0, X_1, X_2, \ldots, X_n$ of this model satisfy the structural relation

$$X_t = \rho X_{t-1} + \sigma(X_{t-1})\varepsilon_t, \quad t = 1, 2, \ldots, n$$

for some real parameter $\rho$, some Lipschitz-continuous positive function $\sigma$ that is bounded away from zero, and innovations $\varepsilon_1, \ldots, \varepsilon_n$ which are independent of the initial observation $X_0$ and are independent and identically distributed (iid) with common positive density $\gamma$ that has zero mean, variance 1, finite fourth moment and finite Fisher information for location and scale. The latter means that $\gamma$ is absolutely continuous and

$$\int (1+x^2)\left(\frac{\gamma'(x)}{\gamma(x)}\right)^2 \gamma(x)\,dx < \infty.$$

I also assume that

$$\rho^2 + \limsup_{|x|\to\infty} \frac{\sigma^2(x)}{1+x^2} < 1. \tag{1.1}$$

This condition yields the (geometric) ergodicity of the model as shown by Maercker [10]. Her results establish $V$-uniform ergodicity with $V(x) = 1 + x^2$, $x \in \mathbb{R}$.

In what follows $\rho$, $\sigma$ and $\gamma$ are assumed to be unknown. The goal is to estimate $\rho$ efficiently in the presence of the infinite-dimensional nuisance parameter $(\sigma, \gamma)$. One possible estimator of $\rho$ is the least squares estimator

$$\hat\rho_n^{LS} = \frac{\sum_{j=1}^n X_{j-1}X_j}{\sum_{j=1}^n X_{j-1}^2}.$$

This estimator is $n^{1/2}$-consistent as $n^{1/2}(\hat\rho_n^{LS} - \rho)$ has a limiting normal distribution with mean zero and variance $E[X_1^2\sigma^2(X_1)]/(E[X_1^2])^2$. If $\sigma$ were known one could use the weighted least squares estimator

$$\hat\rho_n^{WLS} = \frac{\sum_{j=1}^n X_{j-1}X_j/\sigma^2(X_{j-1})}{\sum_{j=1}^n X_{j-1}^2/\sigma^2(X_{j-1})}.$$

Supported in part by NSF Grant DMS 0072174

This estimator is also $n^{1/2}$-consistent with $n^{1/2}(\hat{\rho}_n^{WLS} - \rho)$ having a limiting normal distribution with zero mean and variance $1/E[X_1^2/\sigma^2(X_1)]$. An application of the Cauchy-Schwarz Inequality shows that the limiting variance of the weighted least squares estimator is smaller than that of the least squares estimator unless $\sigma$ is constant in which case the two limiting variances are the same. Estimators $\hat{\rho}_n$, which are asymptotically equivalent to the weighted least squares estimator in the sense that $n^{1/2}(\hat{\rho}_n - \hat{\rho}_n^{WLS}) \to 0$ in probability but which do not require the knowledge of $\sigma$, were recently constructed by Wefelmeyer [18] and Schick [16] for slightly more general models in which the innovations are martingale differences. Wefelmeyer's construction relies on martingale arguments and replaces $\sigma^2(X_{j-1})$ by an estimator $v_{j-1}$ based on only the observations $X_0, \ldots, X_{j-1}$ to exploit martingale arguments. Schick's construction uses a generalization of the sample splitting techniques in [12] from the iid case to ergodic Markov chains and replaces $\sigma^2(X_{j-1})$ by $v_2(X_{j-1})$ if $j \le n/2$ and by $v_1(X_{j-1})$ if $j > n/2$ where $v_1(x)$ and $v_2(x)$ are estimators of $\sigma^2(x)$ based on roughly the first and second half of the sample. Wefelmeyer [18] also showed that estimators equivalent to the weighted least squares estimator are efficient in this more general setting. More precisely, he showed that such estimators are regular and least dispersed within the class of regular estimators. However, more efficient estimators can be constructed for models with iid innovations.

Maercker [10] considered this under the assumption that $\gamma$ is also symmetric. She demonstrated that in this case the parameter $\rho$ can be estimated adaptively. More precisely, she constructed an estimator of $\rho$ without the knowledge of $\gamma$ and $\sigma$ that is asymptotically equivalent to the efficient estimator for the parametric model with known $\gamma$ and $\sigma$. If $\gamma$ is the standard normal density this estimator will be equivalent to the weighted least squares estimator, but will improve upon it for other densities. Her construction generalized ideas of Bickel [1]. She used a small initial part of the sample to estimate $\sigma$ and $\gamma$ and hence the influence function and then used the remaining part of the sample to form the average with this estimated influence function. Relying on the above mentioned sample splitting techniques for ergodic Markov chains, Schick [16] was able to make better use of the data by using estimators of the pair $(\sigma, \gamma)$ based on roughly half the data rather than just a small initial part of the sample. He also relaxed some of the assumptions used in [10].

In this paper I shall consider efficient estimation of $\rho$ without the symmetry assumption on $\gamma$. I shall characterize efficient estimators of $\rho$ in Section 4 and then use a modification of the sample splitting techniques of [16] to construct an efficient estimator in Section 5. This efficient estimator will no longer be adaptive in general. However, if $\gamma$ is the standard normal density, the efficient estimator will be equivalent to the weighted least squares estimator and hence be adaptive.

Section 2 introduces notation and important properties of my model such as equi-V-uniform ergodicity and the continuity of the stationary distribution with respect to the parameters. In Section 3 these properties are used to obtain an appropriate LAN Condition of the model that will be needed in the efficiency considerations. This LAN Condition is formulated with respect to the autoregression parameter and parameters which index the volatility function $\sigma$ and the innovation density $\gamma$. This result extends those of Maercker [10], who only considered LAN with respect to the autoregression parameter, and the more general results of [2] and [6], which give sufficient conditions for LAN with respect to the autoregression parameter and parameters indexing the volatility function only.

Throughout this paper $\ell_1$ and $\ell_2$ denote the score functions for location and scale of the innovation density $\gamma$:

$$\ell_1(x) = -\frac{\gamma'(x)}{\gamma(x)} \quad \text{and} \quad \ell_2(x) = -1 - x\frac{\gamma'(x)}{\gamma(x)}, \quad x \in \mathbb{R},$$

and $J_{ij}$ denotes the inner product of $\ell_i$ with $\ell_j$:

$$J_{ij} = \int \ell_i(x)\ell_j(x)\gamma(x)\,dx, \quad i,j = 1,2.$$

Then $J_{11}$ is the Fisher information for location and $J_{22}$ the Fisher information for scale. Under the assumptions on $\gamma$ these quantities are finite and positive. Moreover, the $2 \times 2$ matrix with entries $J_{ij}$ is positive definite.

Let $\mathfrak{L}(\xi \mid P)$ denote the distribution of a random variable $\xi$ under the probability measure $P$ and use $\Rightarrow$ to indicate convergence in distribution. Finally, given random variables $\xi_1, \xi_2, \ldots$ and probability measures $P_1, P_2, \ldots$, I write $\xi_n \xrightarrow{P_n} 0$ to mean $P_n(|\xi_n| > \varepsilon) \to 0$ for all $\varepsilon > 0$.

**2. Preliminaries.** For my model the parameter is $\theta = (\rho, \sigma, \gamma)$, and the parameter set is

$$\Theta = \{(r,s,g) \in \mathbb{R} \times \mathcal{S} \times \mathcal{G} : \; r^2 + \limsup_{x\to\infty} \frac{s^2(x)}{1+x^2} < 1\}.$$

Here $\mathcal{S}$ denotes the set of all Lipschitz continuous functions from $\mathbb{R}$ to $(0,\infty)$ that are bounded away from zero, while $\mathcal{G}$ denotes the set of all positive densities with zero means, unit variances, finite fourth moments and finite Fisher informations for location and scale. For $\vartheta = (r,s,g) \in \Theta$, let $f_\vartheta$ denote the stationary density of the model and $p_\vartheta$ denote the transition density, so that

$$p_\vartheta(x,y) = \frac{1}{s(x)}\, g\left(\frac{y - rx}{s(x)}\right), \quad x, y \in \mathbb{R}.$$

Of course, I can and do take $f_\vartheta$ such that

$$f_\vartheta(y) = \int p_\vartheta(x,y) f_\vartheta(x)\,dx, \quad y \in \mathbb{R}.$$

Also, I let $p_\vartheta^{(j)}$ denote the $j$-step transition density defined iteratively by

$$p_\vartheta^{(j)}(x,y) = \int p_\vartheta^{(j-1)}(z,y) p_\vartheta(x,z)\,dz, \quad x, y \in \mathbb{R}, \; j = 2, 3, \ldots,$$

starting with $p_\vartheta^{(1)} = p_\vartheta$. Let $P_\vartheta$ and $E_\vartheta$ denote the probability measure and expectation associated with the parameter $\vartheta$.

Let $\mathcal{H}$ denote the set of all measurable functions $h$ from $\mathbb{R}$ to $\mathbb{R}$ such that $|h(x)| \leq 1 + x^2$ for all $x \in \mathbb{R}$. The first result gives local equi-$V$-uniform ergodicity of the model for $V(x) = 1 + x^2$, $x \in \mathbb{R}$. This generalizes the $V$-uniform ergodicity result of [10], which is (2.1) with $\vartheta = \theta$ only.

LEMMA 1. *There exist a small positive $\delta$ and positive constants $\zeta$ and $D$, $\zeta < 1$, such that*

$$\sup_{x\in\mathbb{R}}\sup_{h\in\mathcal{H}} \frac{|\int h(y)p_\vartheta^{(j)}(x,y)\,dy - \int h(y)f_\vartheta(y)\,dy|}{1+x^2} \le D\zeta^j, \quad j=1,2,\dots, \tag{2.1}$$

*for all $\vartheta = (r,s,g) \in \Theta$ satisfying*

$$|r-\rho|<\delta, \quad \sup_{x\in\mathbb{R}}\frac{|s(x)-\sigma(x)|}{\sqrt{1+x^2}} < \delta \quad \textit{and} \quad \sup_{x\in\mathbb{R}}\frac{|g(x)-\gamma(x)|}{\gamma(x)} < \delta. \tag{2.2}$$

PROOF. Let $\mathcal{U}_\delta$ denote the set of all $\vartheta = (r,s,g)$ in $\Theta$ satisfying (2.2). I shall use the computable bounds for geometric ergodicity of [11] to derive (2.1). See Theorem 2.2 in [16] for a convenient formulation. In view of Lemma 2.1.1 of [10], one only needs to show that for small enough $\delta$, there are constants $a<1$ and $b<\infty$ and a compact set $C$ such that

$$\sup_{\vartheta\in\mathcal{U}_\delta}\int(1+y^2)p_\vartheta(x,y)\,dy < a(1+x^2)+b\mathbf{1}_C(x), \quad x\in\mathbb{R},$$

and

$$\inf_{x,y\in C}\inf_{\vartheta\in\mathcal{U}_\delta} p_\vartheta(x,y) > 0.$$

But this is easily done generalizing the arguments in the proofs of Lemmas 2.1.1 and 2.1.2 of [10].

The next result establishes the continuity of the stationary density with respect to the full parameter. It generalizes (8.2) in [16] which established continuity with respect to the autoregression parameter only.

LEMMA 2. *Let $\langle\theta_n\rangle = \langle(\rho_n,\sigma_n,\gamma_n)\rangle$ be a sequence in $\Theta$ such that*

$$\rho_n\to\rho, \quad \sup_{x\in\mathbb{R}}\frac{|\sigma_n(x)-\sigma(x)|}{\sqrt{1+x^2}}\to 0 \quad \textit{and} \quad \sup_{x\in\mathbb{R}}\frac{|\gamma_n(x)-\gamma(x)|}{\gamma(x)}\to 0.$$

*Then*

$$\int(1+x^2)|f_{\theta_n}(x)-f_\theta(x)|\,dx\to 0.$$

PROOF. Following the argument in [16] one only needs to show that

$$\int(1+y^2)|p_{\theta_n}(x,y)-p_\theta(x,y)|\,dy\to 0$$

for each $x\in\mathbb{R}$. But this follows as $p_{\theta_n}(x,y)\to p_\theta(x,y)$ for all $x,y$ and

$$\int(1+y^2)p_{\theta_n}(x,y)\,dy = 1+\rho_n^2x^2+\sigma_n^2(x)\to\int(1+y^2)p_\theta(x,y)\,dy$$

for all $x$.

The above continuity will be helpful in establishing LAN. It will be used to show that the information contained in the initial distribution is negligible, i.e., Assumption (A.5) of [6] or Condition (C.1) in [2]. See also (2.8) in [8].

**3. Local asymptotic normality.** I shall now derive the LAN Condition for regular parametric submodels. In view of Lemma 2 I impose (3.1) and (3.2) below.

Let $\{\sigma_\eta : \eta \in (-1,1)\}$ be a subset of $\mathcal{S}$ such that $\sigma_0 = \sigma$ and

$$\sup_{x\in\mathbb{R}} \frac{|\sigma_\eta(x) - \sigma(x)|}{\sqrt{1+x^2}} \to 0 \quad \text{as } \eta \to 0. \tag{3.1}$$

Let $\{\gamma_\tau : \tau \in (-1,1)\}$ be a subset of $\mathcal{G}$ such that $\gamma_0 = \gamma$ and

$$\sup_{x\in\mathbb{R}} \left| \frac{\gamma_\tau(x) - \gamma(x)}{\gamma(x)} \right| \to 0 \quad \text{as } \tau \to 0. \tag{3.2}$$

I call the map $(r, \eta, \tau) \mapsto (r, \sigma_\eta, \gamma_\tau)$ a *path* and denote it by $q$. It follows from (3.1) and (1.1) that there is a $c > 0$ such that

$$\sup_{|r-\rho|<c, |\eta|<c} \left( r^2 + \limsup_{|x|\to\infty} \frac{\sigma_\eta^2(x)}{1+x^2} \right) < 1. \tag{3.3}$$

This shows that $q(\delta) \in \Theta$ for all $\delta = (r, \eta, \tau) \in \Delta = (\rho - c, \rho - c) \times (-c, c) \times (-1, 1)$. Consequently, $\mathcal{P}_q = \{P_{q(\delta)} : \delta \in \Delta\}$ is a parametric submodel of $\mathcal{P} = \{P_\vartheta : \vartheta \in \Theta\}$. The log-likelihood ratio for the observations under this submodel is given by

$$\Lambda_n(\delta, \delta_0) = \log \frac{f_{q(\delta)}(X_0)}{f_{q(\delta_0)}(X_0)} + \sum_{j=1}^n \log \frac{p_{q(\delta)}(X_{j-1}, X_j)}{p_{q(\delta_0)}(X_{j-1}, X_j)}, \quad \delta, \delta_0 \in \Delta.$$

To obtain the LAN condition under this submodel I have to impose additional smoothness requirements on $q$.

I call the path $q$ *regular* if

$$\int \left| \frac{\sigma_\eta(x) - \sigma(x)}{\sigma(x)} - \eta a_q(x) \right|^2 f_\theta(x)\, dx = o(\eta^2) \tag{3.4}$$

for a measurable function $a_q$ such that $0 < \int a_q^2(x) f_\theta(x)\, dx < \infty$, and if

$$\int \left( \sqrt{\gamma_\tau(x)} - \sqrt{\gamma(x)} - \frac{1}{2} \tau b_q(x) \sqrt{\gamma(x)} \right)^2 dx = o(\tau^2) \tag{3.5}$$

for a measurable function $b_q$ such that $0 < \int b_q^2(x)\gamma(x)\, dx < \infty$. In this case, I call the pair $(a_q, b_q)$ the *characteristics* of $q$.

In connection with the Hellinger differentiability condition (3.5) I should mention the following known result; see Lemma 7.2 in [5] for the relevant argument.

LEMMA 3. *Let $\{g_t : |t| \le c\}$ be a family of densities such that*

$$\int \left( \sqrt{g_t(x)} - \sqrt{g_0(x)} - \frac{1}{2} t h(x) \sqrt{g_0(x)} \right)^2 dx = o(t^2), \tag{3.6}$$

*for some measurable function $h$ such that $\int h^2(x) g_0(x)\, dx < \infty$. Let $k$ be a measurable function such that $\limsup_{t\to 0} \int k^2(x) g_t(x)\, dx < \infty$. Then the map $t \mapsto \int k(x) g_t(x)\, dx$ has derivative $\int k(x) h(x) g_0(x)\, dx$ at $0$.*

REMARK 1. In view of this lemma, the function $b_q$ appearing in (3.5) must also satisfy

$$\int x^i b_q(x)\gamma(x)\,dx = 0, \quad i = 0,1,2.$$

This follows as $\int x^i b_q(x)\gamma(x)\,dx$ is the derivative at 0 of the constant function $\tau \mapsto \int x^i \gamma_\tau(x)\,dx$, $i = 0,1,2$.

REMARK 2. Since $\gamma$ has finite Fisher information for location and scale, one obtains that

1. the location model $\{g_t = \gamma(\cdot - t) : t \in \mathbb{R}\}$ satisfies (3.6) with $h = \ell_1$ and
2. the scale model $\{g_t = \gamma(\cdot/(1+t))/(1+t) : |t| < 1\}$ satisfies (3.6) with $h = \ell_2$.

See [4, pages 210-214] for the relevant arguments. Since also $\int \gamma(x-t)\,dx = 1$, $\int x\gamma(x-t)\,dx = t$ and $\int x^2\gamma(x-t)\,dx = 1 + t^2$, an application of Lemma 3 yields

$$\int \ell_1(x)\gamma(x)\,dx = 0, \quad \int x\ell_1(x)\gamma(x)\,dx = 1 \quad \text{and} \quad \int x^2\ell_1(x)\gamma(x)\,dx = 0. \tag{3.7}$$

Similarly, since $\int x^i \gamma(x/(1+t))/(1+t)\,dx = (1+t)^i \int x^i\gamma(x)\,dx$ for $i = 0,1,2$, Lemma 3 yields

$$\int \ell_2(x)\gamma(x)\,dx = 0, \quad \int x\ell_2(x)\gamma(x)\,dx = 0 \quad \text{and} \quad \int x^2\ell_2(x)\gamma(x)\,dx = 2. \tag{3.8}$$

Finally, one can show that

$$\int k'(y)\gamma(y)\,dy = \int k(y)\ell_1(y)\gamma(y)\,dy$$

and

$$\int yk'(y)\gamma(y)\,dy = \int k(y)\ell_2(y)\gamma(y)\,dy$$

for every continuously differentiable function $k$ such that $|k(x)| \le A(1+x^2)$ and $|k'(x)| \le A(1+|x|^3)$ for some $A < \infty$ and all $x \in \mathbb{R}$.

If $q$ is regular, I set

$$S_j(r,q) = \begin{pmatrix} \xi_j \ell_1(\varepsilon_j(r)) \\ a_q(X_{j-1})\ell_2(\varepsilon_j(r)) \\ b_q(\varepsilon_j(r)) \end{pmatrix}, \quad r \in (-1,1),$$

with

$$\xi_j = \frac{X_{j-1}}{\sigma(X_{j-1})} \quad \text{and} \quad \varepsilon_j(r) = \frac{X_j - rX_{j-1}}{\sigma(X_{j-1})}, \quad j = 1,\dots,n,$$

and let

$$W(q) = E_\theta[S_1(\rho,q)S_1^\top(\rho,q)]$$

denote the dispersion matrix of $S_1(\rho,q)$ under $P_\theta$.

In what follows I call a sequence $\langle\rho_n\rangle$ a *local sequence* if $\langle n^{1/2}(\rho_n - \rho)\rangle$ is bounded.

THEOREM 1. *Suppose the path $q$ is regular. Let $\langle\rho_n\rangle$ be a local sequence. Set $\delta_n = (\rho_n, 0, 0)^\top$. Then*

$$\Lambda_n(\delta_n + n^{-1/2}v_n, \delta_n) - n^{-1/2}\sum_{j=1}^{n} v_n^\top S_j(\rho_n, q) + \frac{1}{2}v_n^\top W(q)v_n \xrightarrow{P_\theta} 0$$

*for every bounded sequence $\langle v_n\rangle$ in $\mathbb{R}^3$, and*

$$\mathfrak{L}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n} S_j(\rho_n, q)\,\middle|\, P_{(\rho_n,\sigma,\gamma)}\right) \Rightarrow \mathcal{N}(0, W(q)).$$

*Moreover,*

$$\frac{1}{\sqrt{n}}\sum_{j=1}^{n} S_j(\rho_n, q) - \frac{1}{\sqrt{n}}\sum_{j=1}^{n} S_j(\rho, q) + W(q)\begin{pmatrix} n^{1/2}(\rho_n - \rho) \\ 0 \\ 0 \end{pmatrix} \xrightarrow{P_\theta} 0.$$

PROOF. Let $v_n = (r_n, s_n, t_n)^\top$ be a bounded sequence in $\mathbb{R}^3$. Set $\theta_n = (\rho_n, \sigma, \gamma)^\top$ and $\vartheta_n = q(\delta_n + n^{-1/2}v_n)$. Clearly,

$$\sum_{j=1}^{n}((\rho_n + n^{-1/2}r_n)X_{j-1} - \rho_n X_{j-1} - n^{-1/2}r_n X_{j-1})^2 \xrightarrow{P_\theta} 0.$$

Actually the left hand side is zero. It follows from the regularity of $q$ that

$$\sum_{j=1}^{n}\left(\frac{\sigma_{n^{-1/2}s_n}(X_{j-1}) - \sigma(X_{j-1})}{\sigma(X_{j-1})} - n^{-1/2}s_n a_q(X_{j-1})\right)^2 \xrightarrow{P_\theta} 0.$$

Moreover, by stationarity and square-integrability,

$$\max_{j=1,\dots,n} n^{-1/2}|X_{j-1}| \xrightarrow{P_\theta} 0 \quad \text{and} \quad \max_{j=1,\dots,n} n^{-1/2}|a_q(X_{j-1})| \xrightarrow{P_\theta} 0.$$

Since the distributions of $(X_1, \dots, X_n)$ under $P_{\theta_n}$ and $P_\theta$ are mutually contiguous as established by [10], the four convergence results above remain true if $P_\theta$ is replaced by $P_{\theta_n}$. The proof proceeds now along the lines of [6, 2, 8]. I omit the details, but should mention that the conditions on the initial density required by these papers follow in the present case from Lemma 2. Actually, the first two papers give the present result immediately if $t_n = 0$. The third paper deals with $t_n \neq 0$ but restricts attention to the homoscedastic case.

**4. Efficiency considerations.** Let $\mathcal{Q}$ be the family of all regular paths. The next lemma which is stated without a proof shows that there are many regular paths.

LEMMA 4. *Let $a$ be a measurable function such that $0 < \int a^2 f_\theta(x)\,dx < \infty$ and $b$ be a measurable function such that $0 < \int b^2(x)\gamma(x)\,dx < \infty$ and*

$$\int x^i b(x)\gamma(x)\,dx = 0, \quad i = 0, 1, 2.$$

*Then there exists a regular path $q$ with characteristics $(a, b)$.*

Let $F$ denote the distribution with density $f_\theta$ and $\Gamma$ the distribution with density $\gamma$. For $q \in \mathcal{Q}$, partition the matrix $W(q)$ as

$$W = \begin{bmatrix} W_{11}(q) & W_{12}(q) \\ W_{12}(q) & W_{22}(q) \end{bmatrix}$$

with $W_{11}(q) = E_\theta[(\xi_1 \ell_1(\varepsilon_1))^2]$. Then

$$I(q) = W_{11}(q) - W_{12}(q) W_{22}^{-1}(q) W_{21}(q)$$

is the information for estimating $\rho$ in the subproblem generated by $q$. I am now looking for a *least favorable* path, i.e., a regular path that minimizes the map $q \mapsto I(q)$.

Since $W_{12}(q)W_{22}^{-1}(q)W_{21}(q)$ is the second moment of the projection of $U = \xi_1 \ell_1(\varepsilon_1)$ onto the linear space $\{u\alpha_q(X_0)\ell_2(\varepsilon_1) + vb_q(\varepsilon_1) : u, v \in \mathbb{R}\}$, one sees that

$$I(q) \geq E_\theta[U^2] - E_\theta[V^2] = E_\theta[(U - V)^2], \quad q \in \mathcal{Q},$$

where $V$ is the projection of $U$ onto the closed linear subspace

$$\mathcal{T} = \{a(X_0)\ell_2(\varepsilon_1) + b(\varepsilon_1) : a \in L_2(F), b \in L_2(\Gamma), \int x^i b(x)\, d\Gamma(x) = 0,\ i = 0, 1, 2\}$$

of $L_2(P_\theta)$. In view of (3.8), $\mathcal{T}$ can be written as the sum of the two orthogonal subspaces

$$\mathcal{T}_1 = \{a(X_0)\ell_2(\varepsilon_1) : a \in L_2(F), \int a\, dF = 0\}$$

and

$$\mathcal{T}_2 = \{b(\varepsilon_1) : b \in L_2(\Gamma), \int x^i b(x)\, d\Gamma(x) = 0,\ i = 0, 1\}.$$

Let now

$$\xi(x) = \frac{x}{\sigma(x)}, \quad x \in \mathbb{R}, \quad \text{and} \quad \bar{\xi} = \int \xi\, dF = E_\theta[\xi_1].$$

Then the projection of $U$ onto $\mathcal{T}_1$ is

$$V_1 = \beta(\xi(X_0) - \bar{\xi})\, \ell_2(\varepsilon_1) = \beta(\xi_1 - \bar{\xi})\, \ell_2(\varepsilon_1), \quad \text{with } \beta = \frac{J_{12}}{J_{22}},$$

while the projection of $U$ onto $\mathcal{T}_2$ is

$$V_2 = \bar{\xi}\, (\ell_1(\varepsilon_1) - \varepsilon_1).$$

Since $\mathcal{T}_1$ and $\mathcal{T}_2$ are orthogonal, $V = V_1 + V_2$ and

$$U - V = (\xi_1 - \bar{\xi})\ell_0(\varepsilon_1) + \bar{\xi}\varepsilon_1$$

with

$$\ell_0(x) = \ell_1(x) - \beta\ell_2(x) = \ell_1(x) - \frac{J_{12}}{J_{22}}\, \ell_2(x), \quad x \in \mathbb{R}.$$

This shows that $I(q) \geq I_*$, where

$$I_* = \int (\xi - \bar{\xi})^2\, dF \int \ell_0^2\, d\Gamma + \bar{\xi}^2.$$

Since $\int \ell_0^2\,d\Gamma$ and $\int \xi^2\,dF$ are positive, so is $I_*$. Let $\mu_3$ denote the third moment of $\Gamma$ and define functions $a$ and $b$ by

$$a(x) = \beta(\xi(x) - \bar{\xi}) - \frac{1}{2}\bar{\xi}\mu_3 \quad \text{and} \quad b(x) = \ell_1(x) - x + \frac{1}{2}\mu_3\ell_2(x), \quad x \in \mathbb{R}.$$

Using (3.7) and (3.8) one verifies that $\int x^i b(x)\,d\Gamma(x) = 0$ for $i = 0, 1, 2$. Next, easy calculations show that $a(X_0)\ell_2(\varepsilon_1) + b(\varepsilon_1) = V$. The above shows that each regular path with characteristics $(a, b)$ is least favorable.

Call a function $L$ from $\mathbb{R}$ into $[0,\infty)$ a *loss function* if $L(0) = 0$, $L(-x) = L(x)$, $x \in \mathbb{R}$, and $L$ is nondecreasing on $[0,\infty)$. Let $\mathcal{N}(m, v)$ denote the normal distribution with mean $m$ and variance $v$. We now have the following result.

THEOREM 2. *Let $\langle\hat{\rho}_n\rangle$ be an estimator of $\rho$. Then*

$$\sup_{q\in\mathcal{Q}} \lim_{C\to\infty} \liminf_{n\to\infty} \sup_{(r-\rho)^2+\eta^2+\tau^2\le C/n} E_{q(r,\eta,\tau)}[L(n^{1/2}(\hat{\rho}_n - r))] \ge \int L\,d\mathcal{N}(0, 1/I_*)$$

*for every loss function $L$. Moreover, if*

$$n^{1/2}\left(\hat{\rho}_n - \rho - \frac{1}{n}\sum_{j=1}^{n} I_*^{-1}[(\xi_j - \bar{\xi})\ell_0(\varepsilon_j) + \bar{\xi}\varepsilon_j]\right) \xrightarrow{P_\theta} 0, \tag{4.1}$$

*then*

$$\mathfrak{L}(n^{1/2}(\hat{\rho}_n - \rho_n) \mid P_{q(\delta_n)}) \Rightarrow \mathcal{N}(0, 1/I_*)$$

*for every $q \in \mathcal{Q}$ and every sequence $\langle\delta_n\rangle = \langle(\rho_n, \eta_n, \tau_n)\rangle$ such that $n((\rho_n - \rho)^2 + \eta_n^2 + \tau_n^2)$ is bounded, and this implies*

$$\sup_{q\in\mathcal{Q}} \lim_{C\to\infty} \limsup_{n\to\infty} \sup_{(r-\rho)^2+\eta^2+\tau^2\le C/n} E_{q(r,\eta,\tau)}[L(n^{1/2}(\hat{\rho}_n - r))] = \int L\,d\mathcal{N}(0, 1/I_*)$$

*for every bounded loss function $L$.*

In view of this I call an estimator satisfying (4.1) $\mathcal{Q}$-efficient. The proof of this theorem is the same as that of Theorem 3.2 in [8] and therefore will be omitted.

REMARK 3. The requirement (4.1) is similar to the requirement (2.1) in [15] for efficient estimators in heteroscedastic regression models. Indeed, the regression model most closely related to the present autoregressive model consists of bivariate observations $(Y_1, Z_1), \ldots, (Y_1, Z_n)$ which satisfy the relation

$$Y_j = \alpha Z_j + \sigma(Z)\,\eta_j, \quad j = 1, \ldots, n,$$

where the errors $\eta_1, \ldots, \eta_n$ are independent and identically distributed with common density $\gamma$ and are independent of the covariates $Z_1, \ldots, Z_n$ which are assumed to be independent and identically distributed with finite positive second moment. For this model, requirement (2.1) in Schick [15] becomes

$$\hat{\alpha}_n = \alpha + \frac{1}{n}\sum_{j=1}^{n} \frac{(\xi(Z_j) - \mu)\ell_0(\eta_j) + \mu\eta_j}{J_0 E[(U_1 - \mu)^2] + \mu^2} + o_p(n^{-1/2}),$$

with $\mu = E[\xi(Z_1)]$ and $J_0 = \int \ell_0^2\,d\Gamma$. This is my (4.1) with $Z_j$ replacing $X_{j-1}$.

REMARK 4. If $\gamma$ is the standard normal density, then $\ell_1(x) = x$, $\ell_2(x) = x^2 - 1$ and $\beta = 0$. In this case $I_* = E_\theta[\xi_1^2] = \int \xi^2 \, dF$ and (4.1) simplifies to

$$n^{1/2}\left(\hat{\rho}_n - \rho - \frac{1}{n}\sum_{j=1}^{n} \frac{\xi_j \varepsilon_j}{E_\theta(\xi_1^2)}\right) \xrightarrow{P_\theta} 0.$$

Thus an estimator is efficient at the standard normal innovation density if it is equivalent to the weighted least squares estimate.

REMARK 5. The information for estimating $\rho$ if $(\sigma, \gamma)$ is known is $I = \int \xi^2 \, dF \int \ell_1^2 \, d\Gamma$. The loss in information due to not knowing the nuisance parameter $(\sigma, \gamma)$ is

$$I - I_* = \bar{\xi}^2 (J_{11} - 1) + \beta^2 J_{22} \int (\xi - \bar{\xi})^2 \, d\Gamma.$$

It follows from the previous remark that there is no loss of information if $\gamma$ is the standard normal density. For other densities this can only occur if $\bar{\xi} = 0$ and $J_{12} = 0$. Indeed, if $\gamma$ is not the standard normal density, then $J_{11} - 1 = \int (\ell(x) - x)^2 \gamma(x) \, dx > 0$. Since the stationary density is positive, so is $\int \xi^2 \, d\Gamma$. The claim is now immediate. Note that $J_{12} = 0$ if $\gamma$ happens to be symmetric about 0. The meaning of $\bar{\xi} = 0$ is not so clear.

**5. On the existence of efficient estimators.** Let now $\mathcal{R}$ denote the set of all $r$ such that $(r, \sigma, \gamma) \in \Theta$. Set

$$\zeta_n(r) = r + \frac{1}{nI_*}\sum_{j=1}^{n}\Big[(\xi_j - \bar{\xi})\ell_0(\varepsilon_j(r)) + \bar{\xi}\varepsilon_j(r)\Big], \quad r \in \mathcal{R}.$$

Constructing an efficient estimator amounts to constructing an estimator $\hat{\rho}_n$ such that (4.1) holds:

$$n^{1/2}(\hat{\rho}_n - \zeta_n(\rho)) \xrightarrow{P_\theta} 0.$$

A key to the construction is the fact that

$$n^{1/2}(\zeta_n(\rho_n) - \zeta_n(\rho)) \xrightarrow{P_\theta} 0$$

for every local sequence $\langle \rho_n \rangle$. This follows from Theorem 1 and allows for the following approach. Construct $\hat{\zeta}_n(r)$ such that

$$n^{1/2}(\hat{\zeta}_n(\rho_n) - \zeta_n(\rho_n)) \xrightarrow{P_{(\rho_n, \sigma, \gamma)}} 0 \tag{5.1}$$

for every local sequence $\langle \rho_n \rangle$. Then $\hat{\rho}_n = \hat{\zeta}_n(\tilde{\rho}_n)$ will satisfy (4.1) for every discrete root-n consistent estimator $\tilde{\rho}_n$ of $\rho$. This is a standard argument in the construction of efficient estimators in semiparametric models; see [1, 12, 13, 9, 2, 8]. A possible discrete root-n consistent estimator is obtained by discretizing the least squares estimator. Simply round the value of this estimator to the nearest integer multiple of $cn^{-1/2}$ for some positive $c$. More sophisticated methods of discretization are discussed in [3].

Now fix a local sequence $\langle\rho_n\rangle$ and set $\theta_n = (\rho_n, \sigma, \gamma)$. To construct the desired $\hat{\zeta}_n(\rho_n)$ I shall rely on the sample splitting techniques of [16]. More precisely, I shall use a slightly modified version of the $q$-split. This version improves upon the suggestion made in the Remark prior to Theorem 3.2 in [16] on "wasting of the middle observations" for the case $q = 2$. The key is Lemma 2.1 of [16]. Combined with the contiguity implied by the LAN Condition it yields the mutual contiguity of the sequences $\langle\mathfrak{L}(X_0, \dots, X_{l_n}, X_{l_n+d_n}, \dots, X_n \mid P_\theta)\rangle$ and $\langle\mathfrak{L}(X_0, \dots, X_{l_n} \mid P_{\theta_n}) \times \mathfrak{L}(X_{l_n+d_n}, \dots, X_n \mid P_{\theta_n})\rangle$ whenever $d_n \to \infty$, and $\liminf_n l_n/n > 0$ and $\limsup(l_n + d_n)/n < 1$. Thus in the proofs I may assume that the future $X_{l_n+d_n}, \dots, X_n$ is independent of the past $X_0, \dots, X_{l_n}$ and then apply standard martingale results to sums of the form $\sum_{j=k_n}^{l_n} \Phi_{n,j}(X_j)$ with

$$\Phi_{n,j}(x) = \phi_{n,j}(X_0, \dots, X_{j-1}, x, X_{l_n+d_n}, \dots, X_n), \quad x \in \mathbb{R},$$

and $k_n < l_n$. For example I get

$$\sum_{j=k_n}^{l_n} \Phi_{n,j}(X_j) = \sum_{j=k_n}^{l_n} \int \Phi_{n,j}(y) p_{\theta_n}(X_{j-1}, y)\, dy + O_{P_{\theta_n}}(V_n^{1/2}) \tag{5.2}$$

with

$$V_n = \sum_{j=k_n}^{l_n} \int \Phi_{n,j}^2(y) p_{\theta_n}(X_{j-1}, y)\, dy.$$

To describe the estimator, fix an integer $q$ greater than 2, say $q = 10$. Let $m$ denote the integer part of $n/q$. Let $A_i = \{(i-1)m+1, \dots, im\}$ for $i = 1, \dots, q-1$ and $A_q = \{(q-1)m+1, \dots, n\}$. Set $B_i = \{j = 1, \dots, n : j \le (i-1)m \text{ or } j > (i+1)m\}$, $i = 1, \dots, q$. Let $m_i$ denote the cardinality of $A_i$ and $n_i$ the cardinality of $B_i$. I also need positive tuning parameters $a_n$, $b_n$, $c_n$, $d_n$ tending to zero and $K_n$ tending to infinity. I shall use $a_n$ and $c_n$ as bandwidths in kernel estimators.

I begin by describing kernel type estimators of $\sigma$, $\ell_1$ and $\ell_2$ which are based on the pairs $\{(X_{j-1}, X_j) : j \in B_i\}$ only. Let the kernel $w$ be a Lipschitz-continuous symmetric density with compact support $[-1, 1]$. Put $w_n(x) = w(x/c_n)/c_n$ and

$$\hat{v}_i(x) = \frac{\sum_{j\in B_i} w_n(x - X_{j-1})(X_j - \rho_n X_{j-1})^2}{\sum_{j\in B_i} w_n(x - X_{j-1})}, \quad x \in \mathbb{R}.$$

Since $\hat{v}_i(x)$ is an estimator of $\sigma^2(x)$, I can estimate $\sigma(x)$ by $\hat{\sigma}_i(x) = \sqrt{v_i(x)}$. This estimator will not be reliable if $|x|$ is large or if its denominator is small. For this reason I shall introduce the weight function $\chi_i$ defined by

$$\chi_i(x) = \mathbf{1}_{\{|x|\le K_n, \sum_{j\in B_i} w_n(x - X_{j-1}) \ge n_i d_n\}}, \quad x \in \mathbb{R},$$

which rules out such values. Set

$$\chi_{i,j} = \chi_i(X_{j-1}), \quad M_i = \sum_{j\in A_i} \chi_{i,j} \quad \text{and} \quad N_i = \sum_{j\in B_i} \chi_{i,j}.$$

Put

$$\hat{\sigma}_{i,j} = \hat{\sigma}_i(X_{j-1}) \quad \text{and} \quad \hat{\varepsilon}_{i,j} = \frac{X_j - \rho_n X_{j-1}}{\hat{\sigma}_{i,j}}, \quad j = 1, \dots, n.$$

Let $k$ be the logistic density or any other density with satisfies Condition K of [14] and set $k_n(x) = k(x/a_n)/a_n$. Using the mimicked innovations $\{\hat{\varepsilon}_{i,j} : j \in B_i\}$, I can now estimate $\gamma$ by the kernel estimator

$$\hat{\gamma}_i(x) = \frac{1}{N_i} \sum_{j \in B_i} \chi_{i,j}\, k_n(x - \hat{\varepsilon}_{i,j}) \quad x \in \mathbb{R}.$$

Next, I estimate $\ell_1(x)$ and $\ell_2(x)$ by

$$\hat{\ell}_{1,i}(x) = -\frac{\hat{\gamma}_i(x)}{b_n + \hat{\gamma}_i(x)}, \qquad \hat{\ell}_{2,i}(x) = x\hat{\ell}_{1,i}(x) - 1, \quad x \in \mathbb{R},$$

and $J_{ab}$ by

$$\hat{J}_{ab,i} = \frac{1}{M_i} \sum_{j \in A_i} \chi_{i,j}\, \hat{\varepsilon}_{i,j}^{b-1} \hat{\ell}'_{a,i}(\hat{\varepsilon}_{i,j}), \quad a, b = 1, 2.$$

Finally, take

$$\hat{\zeta}_n(\rho_n) = \rho_n + \frac{1}{n\hat{I}_*} \sum_{i=1}^{q} \sum_{j \in A_i} \chi_{i,j} \Big\{ (\hat{\xi}_{i,j} - \hat{\xi}_{i,*}) \hat{\ell}_{0,i}(\hat{\varepsilon}_{i,j}) + \hat{\xi}_{i,*} \hat{\varepsilon}_{i,j} \Big\} \tag{5.3}$$

with

$$\hat{\xi}_{i,j} = \frac{X_{j-1}}{\hat{\sigma}_i(X_{j-1})}, \quad \hat{\xi}_{i,*} = \frac{1}{M_i} \sum_{j \in A_i} \chi_{i,j}\, \hat{\xi}_{i,j},$$

$$\hat{\ell}_{0,i}(x) = \hat{\ell}_{1,i}(x) - \frac{\hat{J}_{12,i}}{\hat{J}_{22,i}} \hat{\ell}_{2,i}(x), \quad x \in \mathbb{R},$$

$$\hat{I}_* = \frac{1}{q} \sum_{i=1}^{q} \left[ \frac{1}{M_i} \sum_{j \in A_i} \chi_{i,j}\, (\hat{\xi}_{i,j} - \hat{\xi}_{i,*})^2 \left( \hat{J}_{11,i} - \frac{\hat{J}_{12,i}^2}{\hat{J}_{22,i}} \right) + \hat{\xi}_{i,*}^2 \right].$$

I shall now prove (5.1) for this choice under the following assumptions on the tuning sequences. I write $\alpha_n \sim \beta_n$ if the sequences $\langle \alpha_n/\beta_n \rangle$ and $\langle \beta_n/\alpha_n \rangle$ are bounded.

THEOREM 3. *Suppose the tuning sequences satisfy*

$$a_n \sim n^{-1/13}, \quad b_n \sim n^{-3/13}, \quad c_n \sim n^{-1/3}, \quad d_n^{-1} \sim \log n \quad \textit{and} \quad K_n \sim \log n.$$

*Then (5.1) holds for the choice of $\hat{\zeta}_n(\rho_n)$ given in (5.3).*

PROOF. It suffices to prove

$$\hat{I}_* = I_* + o_{P_{\theta_n}}(1), \tag{5.4}$$

and the following statements for $i = 1, \dots, q$:

$$\frac{1}{\sqrt{n}} \sum_{j \in A_i} \Big[ \chi_{i,j} \hat{\xi}_{i,*} \hat{\varepsilon}_{i,j} - \bar{\xi} \varepsilon_j(\rho_n) \Big] = o_{P_{\theta_n}}(1) \tag{5.5}$$

$$\frac{1}{\sqrt{n}} \sum_{j \in A_i} \Big\{ \chi_{i,j} (\hat{\xi}_{i,j} - \hat{\xi}_{i,*}) \hat{\ell}_{0,i}(\hat{\varepsilon}_{i,j}) - (\xi_j - \bar{\xi}) \ell_0(\varepsilon_j(\rho_n)) \Big\} = o_{P_{\theta_n}}(1). \tag{5.6}$$

I shall show later that for $a = 1, 2$, all $\alpha < 2/3$, some $\alpha_* > 1/6$, and some constants $c_0$ and $c$:

$$\frac{1}{m_i} \sum_{j \in A_i} (1 - \chi_{i,j})(1 + X_{j-1}^2) = o_{P_{\theta_n}}(1); \tag{5.7}$$

$$\sup_{x \in \mathbb{R}} \chi_i(x) |\hat{\sigma}_i(x) - \sigma(x)| = o_{P_{\theta_n}}(1); \tag{5.8}$$

$$\frac{1}{m_i} \sum_{j \in A_i} \chi_{i,j}(1 + X_{j-1}^2)(\hat{\sigma}_i(X_{j-1}) - \sigma(X_{j-1}))^2 = o_{P_{\theta_n}}(n^{-\alpha}); \tag{5.9}$$

$$\hat{J}_{ab,i} = \int \hat{\ell}_{a,i}(y) \ell_b(y)\, d\Gamma(y) + o_{P_{\theta_n}}(n^{-\alpha_*}); \tag{5.10}$$

$$\left| \int \Big( \hat{\ell}_{a,i}((1+t)y) - \hat{\ell}_{a,i}(y) - t \hat{\ell}_{a,i}(y) \ell_2(y) \Big)\, d\Gamma(y) \right| \le c_0 t^2 a_n^{-2}; \tag{5.11}$$

$$\int (\hat{\ell}_{a,i}(y) - \ell_a(y))^2\, d\Gamma(y) = o_{P_{\theta_n}}(1); \tag{5.12}$$

$$\frac{|\hat{\ell}_{a,i}(x)|}{1 + |x|} \le c a_n^{-1}, \quad \frac{|\hat{\ell}'_{a,i}(x)|}{1 + |x|} \le c a_n^{-2}, \quad \frac{|\hat{\ell}''_{a,i}(x)|}{1 + |x|} \le c a_n^{-3}, \quad x \in \mathbb{R}. \tag{5.13}$$

Let us now show how to derive (5.4)–(5.6) from these statements. It follows from (5.10) and (5.12) that

$$\hat{J}_{ab,i} = J_{ab} + o_{P_{\theta_n}}(1), \quad a, b = 1, 2. \tag{5.14}$$

Since $\sigma$ is bounded away from zero, (5.7) and (5.8) imply

$$\hat{\xi}_{i,*} = \frac{1}{m_i} \sum_{j \in A_i} \xi_j + o_{P_{\theta_n}}(1) \tag{5.15}$$

and

$$\frac{1}{M_i} \sum_{j \in A_i} \chi_{i,j}\, (\hat{\xi}_{i,j} - \hat{\xi}_{i,*})^2 = \frac{1}{m_i} \sum_{j \in A_i} (\xi_j - \bar{\xi})^2 + o_{P_{\theta_n}}(1). \tag{5.16}$$

In view of the Ergodic Theorem, the above give (5.4).

It is straightforward to verify that

$$\hat{\varepsilon}_{i,j} = (1 + \Delta_{i,j})\, \varepsilon_j(\rho_n) \quad \text{with} \quad \Delta_{i,j} = \frac{\sigma(X_{j-1}) - \hat{\sigma}_{i,j}}{\hat{\sigma}_{i,j}}.$$

Since $\sigma$ is bounded away from 0, one obtains from (5.7) and (5.8) that

$$\frac{1}{n} \sum_{j \in A_i} [\chi_{i,j}(1 + \Delta_{i,j}) - 1]^2 = o_{P_{\theta_n}}(1).$$

Thus it follows from (5.15) and (5.2) applied with $\Phi_{n,j}(X_j) = [\chi_{i,j}(1 + \Delta_{i,j}) - 1]\varepsilon_j(\rho_n)$ that (5.5) holds.

It follows from (5.14) and (5.15) that the left hand side of (5.6) equals

$$Y_{n,1} - \hat{\xi}_{i,*} Z_{n,1} - \frac{\hat{J}_{12,i}}{\hat{J}_{22,i}} \Big( Y_{n,2} - \hat{\xi}_{i,*} Z_{n,2} \Big) + o_{P_{\theta_n}}(1),$$

where, for $a = 1, 2$,

$$Y_{n,a} = \frac{1}{\sqrt{n}} \sum_{i \in A_i} \Big[ \chi_{i,j} \hat{\xi}_{i,j} \hat{\ell}_{a,i}(\hat{\varepsilon}_{i,j}) - \xi_j \ell_a(\varepsilon_j(\rho_n)) \Big],$$

$$Z_{n,a} = \frac{1}{\sqrt{n}} \sum_{i \in A_i} \Big[ \chi_{i,j} \hat{\ell}_{a,i}(\hat{\varepsilon}_{i,j}) - \ell_a(\varepsilon_j(\rho_n)) \Big].$$

In view of (5.14) and (5.15) the desired result (5.6) will follow if one shows that

$$(5.17) \qquad Y_{n,a} = \frac{M_i \hat{\xi}_{i,*}}{\sqrt{n}} \int \hat{\ell}_{a,i} \, d\Gamma + \frac{\hat{J}_{a2,i}}{\sqrt{n}} \sum_{j \in A_i} \chi_{i,j} \hat{\xi}_{i,j} \Delta_{i,j} + o_{P_{\theta_n}}(1), \quad a = 1, 2,$$

and

$$(5.18) \qquad Z_{n,a} = \frac{M_i}{\sqrt{n}} \int \hat{\ell}_{a,i} \, d\Gamma + \frac{\hat{J}_{a2,i}}{\sqrt{n}} \sum_{j \in A_i} \chi_{i,j} \Delta_{i,j} + o_{P_{\theta_n}}(1), \quad a = 1, 2.$$

It follows from (5.7), (5.8), (5.9), (5.12) and (5.13) that

$$\frac{1}{n} \sum_{j \in A_i} \int \Big[ \chi_{i,j} \hat{\xi}_{i,j} \hat{\ell}_{a,i}((1 + \Delta_{i,j})y) - \xi_j \ell_a(y) \Big]^2 \, d\Gamma(y) = o_{P_{\theta_n}}(1).$$

In view of (5.2), this shows that

$$Y_{n,a} = \frac{1}{\sqrt{n}} \sum_{j \in A_i} \chi_{i,j} \hat{\xi}_{i,j} \int \hat{\ell}_{a,i}((1 + \Delta_{i,j})y) \, d\Gamma(y) + o_{P_{\theta_n}}(1).$$

Using (5.11), (5.8) and (5.9) one obtains

$$Y_{n,a} = \frac{1}{\sqrt{n}} \sum_{j \in A_i} \chi_{i,j} \hat{\xi}_{i,j} \int \hat{\ell}_{a,i} \, d\Gamma + \frac{1}{\sqrt{n}} \sum_{j \in A_i} \chi_{i,j} \hat{\xi}_{i,j} \Delta_{i,j} \int \hat{\ell}_{a,i} \ell_2 \, d\Gamma + o_{P_{\theta_n}}(1).$$

The desired (5.17) follows from this, (5.9) and (5.10). Similarly, one verifies (5.18).

I am left to verify (5.7)-(5.12). Since the stationary density $f_\theta$ satisfies $f_\theta(y) = \int p_\theta(x, y) f_\theta(x) \, dx$, $y \in \mathbb{R}$, it has finite Fisher information for location and is thus Hölder continuous of order $1/2$; see [7, page 52] for the latter. It follows from standard arguments that

$$\sup_{|x| \le K_n} \left| \frac{1}{n_i} \sum_{j \in B_i} w_n(x - X_{j-1}) - f_\theta(x) \right| = o_{P_\theta}(d_n^{-1}).$$

This and the fact that $f_\theta$ has finite second moment give (5.7) with $\theta_n = \theta$. The general case follows by contiguity.

Since $\sigma$ is Lipschitz-continuous, there is a positive constant $K$ such that

$$(5.19) \qquad |\sigma^\nu(y) - \sigma^\nu(x)| \le K(1 + |x|^{\nu-1})|y - x|, \quad |y - x| \le 1, \nu = 1, \ldots, 4.$$

Thus one obtains as in [16] that

$$\sup_{x\in\mathbb{R}} \chi_i(x)|v_i(x) - \sigma^2(x)| = o_{P_{\theta_n}}(1)$$

if $K_n n^{-1} c_n^{-5/2} d_n^{-3} \to 0$ and

$$\frac{1}{m_i}\sum_{j\in A_i}^{n} \chi_{i,j}(\hat{v}_i(X_{j-1}) - \sigma^2(X_{j-1}))^2 = O_{P_{\theta_n}}(n^{-1}c_n^{-1}d_n^{-2} + c_n^2).$$

Since $\sigma$ is bounded away from zero, these immediately give (5.8) and (5.9).

The statements in (5.13) follow from the work in [14]. Since

$$\int y\hat{\ell}'_{a,i}(y)\,d\Gamma(y) = \int \hat{\ell}_{a,i}(y)\ell_2(y)\,d\Gamma(y)$$

by Remark 2, a Taylor expansion shows that the left hand side of (5.11) equals the absolute value of

$$t^2\int y^2 \int_0^1\int_0^v \hat{\ell}''_{a,i}((1+ut)y)\,du\,dv\,d\Gamma(y)$$

which by Remark 2 equals

$$t^2\int_0^1\int_0^v (1+ut)^{-2}\int[(1+ut)y\hat{\ell}'_{a,i}((1+ut)y) - \hat{\ell}_{a,i}((1+ut)y)]\ell_2(y)\,d\Gamma(y)\,du\,dv.$$

Thus (5.11) follows from this and (5.13). A similar argument yields that there is a constant $c_*$ such that, for $a,b=1,2$ and $t\in\mathbb{R}$,

$$\left|\int((1+t)y)^{b-1}\hat{\ell}'_{a,i}((1+t)y)\,d\Gamma(y) - \int \hat{\ell}_{a,i}(y)\ell_b(y)\,d\Gamma(y)\right| \le c_*|t|a_n^{-2}.$$

It follows from (5.2), (5.8) and (5.13) that

$$\hat{J}_{ab,i} = \frac{1}{M_i}\sum_{j\in A_i}\chi_{i,j}\int((1+\Delta_{i,j})y)^{b-1}\hat{\ell}'_{a,i}((1+\Delta_{i,j})y)\,d\Gamma(y) + o_{P_{\theta_n}}(n^{-1/2}a_n^{-2}).$$

Combining the above gives (5.10).

Let $\tilde{\ell}_{a,i}$ be defined as $\hat{\ell}_{a,i}$ but with $\hat{\varepsilon}_{i,j}$ replaced by $\varepsilon_j(\rho_n)$ for $j\in B_i$. Then

$$\int(\tilde{\ell}_{a,i}(y) - \ell_a(y))^2\,d\Gamma(y) = o_{P_{\theta_n}}(1), \quad a=1,2.$$

This was shown for the case $a=1$ in [14]. The case $a=2$ is similar; see [17] for the important argument. It follows from (L3) in [14] that

$$\sup_{x\in\mathbb{R}}\frac{|\hat{\ell}_{a,i}(x) - \tilde{\ell}_{a,i}(x)|^2}{1+x^{2a-2}} \le \frac{C}{N_i a_n^5 b_n}\sum_{j\in B_i}\chi_{i,j}(\hat{\varepsilon}_{i,j} - \varepsilon_j(\rho_n))^2, \quad a=1,2,$$

for some constant $C$. Thus (5.12) follows if one shows that

$$\frac{1}{n}\sum_{j\in B_i}\chi_{i,j}(\hat{\varepsilon}_{i,j} - \varepsilon_j(\rho_n))^2 = O_{P_{\theta_n}}(K_n^6 n^{-1}c_n^{-1}d_n^{-2} + c_n^2).$$

In view of (5.8), this is implied by

$$\frac{1}{n}\sum_{j\in B_i}\chi_{i,j}\varepsilon_j^2(\rho_n)(\hat{v}_i(X_{j-1})-\sigma^2(X_{j-1}))^2 = O_{P_{\theta_n}}(K_n^6 n^{-1}c_n^{-1}d_n^{-2}+c_n^2). \tag{5.20}$$

Using (5.19) one finds that the left hand side of (5.20) is at most of the order of

$$\frac{1}{n}\sum_{j\in B_i}\varepsilon_j^2(\rho_n)c_n^2(1+X_{j-1}^2)+\frac{1}{n}\sum_{j\in B_i}\varepsilon_j^2(\rho_n)\Big(\frac{1}{n_i d_n}\sum_{k\in B_i}D_{j,k}U_{n,k}\Big)^2$$

with

$$D_{j,k}=\mathbf{1}_{\{|X_{j-1}|\le K_n\}}w_n(X_{j-1}-X_{k-1}) \quad\text{and}\quad U_{n,k}=\sigma^2(X_{k-1})(\varepsilon_k^2(\rho_n)-1).$$

Clearly, the first term is of order $O_{P_{\theta_n}}(c_n^2)$. The second term can be be bounded by $3T_n^- + 3T_n^\sim + 3T_n^+$, where the terms $T_n^-, T_n^\sim, T_n^+$ are as the original second term but with the summation $k\in B_i$ replaced by $k\in B_{i,j}^- = \{\nu\in B_i : \nu < j-1\}$, $k\in B_{i,j}^\sim = \{\nu\in B_i : |\nu - j|\le 1\}$ and $k\in B_{i,j}^+ = \{\nu\in B_i : \nu > j+1\}$, respectively. By the Lipschitz-continuity of $\sigma$ and the boundedness of $w$, there are finite constants $C_1$ and $C_2$ such that

$$|D_{j,k}U_{n,k}|\le C_1K_n^2 w_n(X_{j-1}-X_{k-1})|\varepsilon_k^2(\rho_n)-1|\le C_2K_n^2c_n^{-1}|\varepsilon_k^2(\rho_n)-1|. \tag{5.21}$$

This, $K_n^4c_n^{-1}d_n^{-2}n^{-1/2}\to 0$ and $n^{-1/2}\max_k|\varepsilon_k^2(\rho_n)-1| = o_{P_{\theta_n}}(1)$ imply that

$$T_n^\sim \le \frac{3}{nn_i^2d_n^2}\sum_{j\in B_i}\varepsilon_j^2(\rho_n)\sum_{k\in B_{i,j}^\sim}D_{j,k}^2U_{n,k}^2 = o_{P_{\theta_n}}(n^{-1}c_n^{-1}).$$

Since $p_{\theta_n}$ is bounded by $\|\gamma\|_\infty\|1/\sigma\|_\infty$ and $w_n(x)\le c_n^{-1}\|w\|_\infty\mathbf{1}_{\{|x|\le c_n\}}$, one sees that

$$\sup_{x,y}\int w_n^2(x-z)p_{\theta_n}(y,z))\,dz = O(c_n^{-1}). \tag{5.22}$$

Using this and the first inequality in (5.21), one obtains that

$$\max_{|j-k|>1}E_{\theta_n}[(1+\varepsilon_j^2(\rho_n))D_{j,k}^2U_{n,k}^2] = O(K_n^4c_n^{-1}). \tag{5.23}$$

Since $E_{\theta_n}[\varepsilon_j^2(\rho_n)D_{j,k}D_{j,l}U_{n,k}U_{n,l}]=0$ for $j<k<l$, one obtains that

$$E_{\theta_n}[T_n^+] = \frac{1}{nn_i^2d_n^2}\sum_{j\in B_i}\sum_{k\in B_{i,j}^+}E_{\theta_n}[\varepsilon_j^2(\rho_n)D_{j,k}^2U_{n,k}^2] = O\Big(\frac{K_n^4}{nd_n^2c_n}\Big).$$

Finally, conditioning on $X_0,\dots,X_{j-1}$ shows that

$$E_{\theta_n}[T_n^-] = \frac{1}{nn_i^2d_n^2}\sum_{j\in B_i}\sum_{l,k\in B_{i,j}^-}E_{\theta_n}[D_{j,l}U_{n,l}D_{j,k}U_{n,k}].$$

For $0<l<k-1<j-2$, one finds via conditioning on $X_0,\dots,X_k$ that

$$\begin{aligned}E_{\theta_n}[D_{j,l}U_{n,l}D_{j,k}U_{n,k}] &= E_{\theta_n}[U_{n,l}U_{n,k}E_{\theta_n}[D_{j,l}D_{j,k}\mid X_0,\dots,X_k]]\\ &= E_{\theta_n}[U_{n,l}U_{n,k}\Delta_{l,k,j}],\end{aligned}$$

where

$$\Delta_{l,k,j} = \int \mathbf{1}_{\{|y|\le K_n\}} w_n(y - X_{l-1}) w_n(y - X_{k-1}) (p_{\theta_n}^{(j-1-k)}(X_k, y) - f_{\theta_n}(y))\, dy.$$

It follows from the equi-V-uniform ergodicity and the properties of $w$ that there is a finite constant $D$ and some $\alpha < 1$ such that $|\Delta_{l,k,j}|$ is bounded by

$$D\alpha^{j-k-1} c_n^{-2}(1 + X_k^2) \mathbf{1}_{\{|X_{k-1}|\le K_n + c_n\}} \mathbf{1}_{\{|X_{l-1}|\le K_n + c_n\}} \mathbf{1}_{\{|X_{k-1} - X_{l-1}|\le 2c_n\}}.$$

Since also

$$X_k^2 \le (\rho_n^2 X_{k-1}^2 + \sigma^2(X_{k-1}))(1 + \varepsilon_k^2(\rho_n)),$$

one finds that

$$\begin{aligned} |E_{\theta_n}[D_{j,l}U_{n,l}D_{j,k}U_{n,k}]| &\le C_3 \alpha^{j-1-k} K_n^6 c_n^{-2} E_{\theta_n}[|\varepsilon_l^2(\rho_n) - 1| \mathbf{1}_{\{|X_{l-1} - X_{k-1}|\le 2c_n\}}] \\ &\le C_4 \alpha^{j-1-k} K_n^6 c_n^{-1} \end{aligned}$$

for finite constants $C_3$ and $C_4$. Combining this with (5.22) one finds that

$$E[T_n^-] = O(K_n^6 d_n^{-2} n^{-1} c_n^{-1}).$$

This completes the proof of (5.20).

## REFERENCES


[1] P. Bickel. On adaptive estimation. *Annals of Statistics*, 10:647–671, 1982.

[2] F.C. Drost, C.A.J. Klaassen, and B.J.M. Werker. Adaptive estimation in time-series models. *Annals of Statistics*, 25:786–817, 1997.

[3] V. Fabian and J. Hannan. On estimation and adaptive estimation for locally asymptotically normal families. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 59:459–478, 1982.

[4] J. Hájek and Z. Šidák. *Theory of Rank Tests.* Academia, Prague, 1967.

[5] I. A. Ibragimov and R. Z. Has'minskij. *Statistical Estimation - Asymptotic Theory.* Springer Verlag, New York, 1981.

[6] P. Jeganathan. Some aspects of asymptotic theory with applications to time series models. *Econometric Theory*, 11:818–887, 1995.

[7] H.L. Koul. *Weighted empiricals and linear models.* Number 21 in Lecture Notes-Monograph Series. Institute of Mathematical Statistics, Hayward, California, 1992.

[8] H.L. Koul and A. Schick. Efficient estimation in nonlinear autoregressive time series models. *Bernoulli*, 3:293–314, 1997.

[9] J.P. Kreiss. On adaptive estimation in stationary arma processes. *Annals of Statistics*, 15:112–133, 1987.

[10] G. Maercker. *Statistical Inference in Conditional Heteroskedastic Autoregressive Models.* Shaker Verlag, Aachen, 1997.

[11] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability.* Springer, London, 1993.

[12] A. Schick. On asymptotically efficient estimation in semiparametric models. *Annals of Statistics*, 14:1139–1151, 1986.

[13] A. Schick. A note on the construction of asymptotically linear estimators. *Journal of Statistical Planning and Inference*, 16:89–105, 1987.

[14] A. Schick. On efficient estimation in regression models. *Annals of Statistics*, 21:1486–1521, 1993.

[15] A. Schick. Efficient estimates in linear and nonlinear regression with heteroscedastic errors. *Journal of Statistical Planning and Inference*, 58:371–387, 1997.

[16] A. Schick. Sample splitting with Markov chains. *Bernoulli*, 7:33–61, 2001.

[17] A. Schick and V. Susarla. Efficient estimation in some missing data problems. *Journal of Statistical Planning and Inference*, 19:217–228, 1988.

[18] W. Wefelmeyer. Adaptive estimators for parameters of the autoregression function of a Markov chain. *Journal of Statistical Planning and Inference*, 58:389–398, 1997.

Anton Schick
Department of Mathematical Sciences
Binghamton University
Binghamton, NY 13902-6000 USA
anton@math.binghamton.edu