EFFICIENT ESTIMATION IN INVERTIBLE LINEAR PROCESSES

ANTON SCHICK AND WOLFGANG WEFELMEYER

Abstract. An invertible causal linear process is a process which has infinite order moving average and autoregressive representations. We assume that the coefficients in these representations depend on a Euclidean parameter, while the corresponding innovations have an unknown centered distribution with some moment restrictions. We discuss efficient estimation of differentiable functionals in such a semiparametric model. For this we first obtain a suitable semiparametric version of local asymptotic normality and then use Hájek's convolution theorem to characterize efficient estimators. Then we apply this result to construct efficient estimators of the Euclidean parameter and of linear functionals of the innovation distribution.

Key words: Time series, nonlinear process of increasing order, empirical estimator, local asymptotic normality, asymptotically linear estimator, influence function, adaptive estimator, regular estimator, least dispersed estimator, contiguity.

AMS 2000 Subject Classification: Primary 62M09, 62M10; secondary 62G05, 62G20.

1. Introduction

An invertible causal linear process is a family $\{Y_t : t \in \mathbb{Z}\}$ of random variables which has infinite order moving average and autoregressive representations

$$Y_t = X_t + \sum_{s=1}^{\infty} \alpha_s X_{t-s}, \quad t \in \mathbb{Z}, \tag{1.1}$$

$$Y_t = X_t - \sum_{s=1}^{\infty} \beta_s Y_{t-s}, \quad t \in \mathbb{Z}. \tag{1.2}$$

Here the *innovations* $\{X_t : t \in \mathbb{Z}\}$ are independent and identically distributed random variables with zero mean and finite variance. We assume that the distribution function F of the innovations is unknown otherwise. We also assume that the coefficients $\alpha_1, \alpha_2, \dots$ and $\beta_1, \beta_2, \dots$ are summable, and that they depend on a Euclidean parameter ϑ. Of course, the coefficients $\beta_1, \beta_2, \dots$ are determined by the coefficients $\alpha_1, \alpha_2, \dots$. This is a semiparametric model. Examples are the classical AR(p) models, the MA(q) models and the ARMA(p,q) models. We study efficient estimation of differentiable functionals of (ϑ, F). To this end we first derive, in Section 2, an appropriate semiparametric version of local asymptotic normality in the sense of Le Cam.

Estimation of ϑ is well-studied in a number of special cases. It is known that ϑ can be estimated *adaptively*: There are estimators for ϑ that are asymptotically as good as the best estimator for ϑ when F is *known.* In order to prove efficiency of estimators for ϑ, it is therefore enough to show local

First author partially supported by NSF Grant DMS 0072174.

asymptotic normality for fixed F, i.e. parametrically, perturbing only ϑ. This is done for AR(1) by Akahira [1], for AR(p) by Akritas and Johnson [2], and for ARMA(p,q) by Kreiss [15]. An extension to AR(p) with regression trend is in Swensen [25], to MA(q) with regression trend in Garel [7], and to vector ARMA(p,q) with regression trend in Hallin and Puri [9] and Garel and Hallin [8]. Nonlinear autoregressive models are studied in Hwang and Basawa [11] and Jeganathan [13]; a heteroscedastic generalization in Drost, Klaassen and Werker [5]. Kreiss [17, 18] considers invertible causal linear processes (1.1) with arbitrary coefficients $\alpha_1, \alpha_2, \dots$ and F known or known up to scale.

Here we are interested in estimating functionals which depend not only on ϑ but also on F. This requires a *semiparametric* version of local asymptotic normality, perturbing both ϑ and F. Such a result has already been obtained in special cases: for AR(1) by Huang [10], for AR(p) by Kreiss [16]. These cases are covered by the nonlinear autoregressive model considered in Koul and Schick [14]. Our proof of local asymptotic normality for linear processes relies partly on their arguments.

Section 3 briefly recalls a characterization of least dispersed regular estimators in the context of our model. Section 4 constructs an efficient estimator for the expectation of a square-integrable function under the innovation distribution. So far, such a result has been obtained for the AR(1) model by Wefelmeyer [26], and for nonlinear and heteroscedastic autoregressive models by Schick and Wefelmeyer [24]. Section 5 contains an efficient estimator for ϑ. The construction avoids some unpleasant features of previous estimators: sample splitting, truncation of the score function, assumption of symmetry of the innovation distribution.

Another interesting application of Section 2 is efficient estimation of the *stationary* law, specifically of linear functionals of this law. The construction requires space and uses additional results of independent interest, and we give it elsewhere, Schick and Wefelmeyer [23]. In Section 6, however, we discuss a special case which can be tackled by a different and direct method, namely efficient estimation of moments of the stationary distribution.

2. Local asymptotic normality

In this section we prove local asymptotic normality for invertible causal linear processes. An essential step is the approximation by nonlinear autoregressive processes of increasing order. Let us formally introduce our model. We fix a measurable function ψ from $\mathbb{R}$ to $\mathbb{R}$. As finite-dimensional parameter space we take an open subset Θ of $\mathbb{R}^k$. As infinite-dimensional parameter space we take the collection $\mathfrak{F}$ of all innovation distributions with zero mean, finite variance and finite Fisher information for location, and for which ψ is square-integrable. The condition on ψ is needed if we want to estimate the expectation of ψ, e.g. moments of the innovation distribution. Recall that a distribution function F has *finite Fisher information $I(F)$ for location* if F has an absolutely continuous density f and

$$I(F) = \int \ell_F^2 \, dF < \infty, \quad \text{where} \quad \ell_F = f'/f. \tag{2.1}$$

Let $\alpha = (\alpha_1, \alpha_2, \dots)$ and $\beta = (\beta_1, \beta_2, \dots)$ be functions from Θ into the Banach space of absolutely summable sequences:

$$\sum_{s=1}^{\infty} |\alpha_s(\vartheta)| < \infty \quad \text{and} \quad \sum_{s=1}^{\infty} |\beta_s(\vartheta)| < \infty, \quad \vartheta \in \Theta.$$

Set $\beta_0(\vartheta) = 1$ and $\alpha_0(\vartheta) = 1$ for $\vartheta \in \Theta$. For each (ϑ, F) in $\Theta \times \mathfrak{F}$, we let $P_{\vartheta,F}$ denote a probability measure for which the process $\{Y_t : t \in \mathbb{Z}\}$ has the infinite order moving average and autoregressive representations

$$Y_t = X_t(\vartheta) + \sum_{s=1}^{\infty} \alpha_s(\vartheta) X_{t-s}(\vartheta), \quad t \in \mathbb{Z}, \tag{2.2}$$

$$Y_t = X_t(\vartheta) - \sum_{s=1}^{\infty} \beta_s(\vartheta) Y_{t-s}, \quad t \in \mathbb{Z}, \tag{2.3}$$

with innovations $\{X_t(\vartheta) : t \in \mathbb{Z}\}$ that are independent and identically distributed with distribution function F. These assumptions imply that $\{Y_t : t \in \mathbb{Z}\}$ is a centered process with autocovariance function

$$\gamma_t(\vartheta, F) = E_{\vartheta,F}(Y_0, Y_t) = \int x^2 \, dF(x) \sum_{s=0}^{\infty} \alpha_s(\vartheta)\alpha_{s+t}(\vartheta), \quad t = 0, 1, 2, \ldots.$$

By our assumption on α, the autocovariances are absolutely summable and

$$\sum_{s=0}^{\infty} |\gamma_s(\vartheta, F)| \le \int x^2 \, dF(x) \Big(\sum_{t=0}^{\infty} |\alpha_t(\vartheta)| \Big)^2. \tag{2.4}$$

We assume that we can only observe the string $Y_0, \ldots, Y_n$. We now single out a point τ in Θ and an innovation distribution function G in $\mathfrak{F}$ with density g and derive local asymptotic normality at (τ, G) for a parametric submodel obtained by restricting the innovation distribution functions to a subset $\{F_\eta : \eta \in \Delta\}$ of $\mathfrak{F}$ with $F_0 = G$ and Δ an open neighborhood of the origin in $\mathbb{R}^q$. We call the map $\eta \mapsto F_\eta$ a (*q-dimensional*) *path* (*through* G). For the applications in Sections 4 and 5, it suffices to take $q = 1$. Since F_η belongs to $\mathfrak{F}$, it has finite Fisher information for location, and thus a density, which we denote by f_η. We say that the path is *smooth* if the following four conditions hold:

1. The variances are continuous at the origin:

$$\lim_{\eta \to 0} \int x^2 \, dF_\eta(x) = \int x^2 \, dG(x).$$

2. The second moments of ψ are continuous at the origin:

$$\lim_{\eta \to 0} \int \psi^2(x) \, dF_\eta(x) = \int \psi^2(x) \, dG(x).$$

3. The Fisher informations are continuous at the origin:

$$\lim_{\eta \to 0} I(F_\eta) = I(G).$$

4. The map $\eta \mapsto F_\eta$ is Hellinger differentiable at the origin: There is a measurable function ζ from $\mathbb{R}$ to $\mathbb{R}^q$ such that $\int \|\zeta\|^2 \, dG < \infty$ and

$$\int \Big[f_\eta^{1/2}(x) - g^{1/2}(x) - \frac{1}{2}\eta^T \zeta(x) g^{1/2}(x) \Big]^2 dx = o(\|\eta\|^2). \tag{2.5}$$

The map ζ is called *Hellinger derivative.* It follows from Ibragimov and Has'minskii [12, Chapter 1, Lemma 7.2] that

$$\int \zeta \, dG = 0 \quad \text{and} \quad \int x\zeta(x)\, dG(x) = 0.$$

Let

$$H = \{h \in L_2(G) : \int h \, dG = 0 \quad \text{and} \quad \int xh(x)\, dG(x) = 0\}. \tag{2.6}$$

The following lemma and corollary show that H is the *tangent space* of $\mathfrak{F}$ at G.

Lemma 1. *For each ζ in H^q there exists a smooth path $\eta \mapsto F_\eta$ which has ζ as its Hellinger derivative.*

Proof. Fix $\zeta \in H^q$. If ζ is bounded and Lipschitz, we can take $dF_\eta = (1 + \eta^\top \zeta)\, dG$ for η in $\mathbb{R}^q$ with $\|\eta\|$ sufficiently small so that $1 + \eta^\top \zeta$ is indeed a density. Otherwise we have to replace ζ by a bounded and Lipschitz continuous version in H^q. This is achieved by first truncating ζ, then smoothing this version and finally projecting the smoothed version back into H^q. For this fix a in $(0, 1/2)$ and let φ denote the standard normal density. For $\eta \in \mathbb{R}^q$, $\eta \neq 0$, the smoothed version

$$\bar\zeta_\eta(x) = \int \zeta^*_\eta(x - \|\eta\|^a y)\varphi(y)\, dy, \quad x \in \mathbb{R},$$

of $\zeta^*_\eta = \zeta\, \mathbf{1}[\|\zeta\| \le \|\eta\|^{-a}]$ is Lipschitz (with Lipschitz constant $L_\eta = O(\|\eta\|^{-2a})$) and bounded ($\|\bar\zeta_\eta\| \le \|\eta\|^{-a}$) and converges to ζ in quadratic mean ($\int \|\bar\zeta_\eta - \zeta\|^2 \, dG \to 0$ as $\eta \to 0$). But $\bar\zeta_\eta$ may not belong to H^q. Therefore we need to modify it slightly. Let $\gamma(x) = (1, x)^\top$ and $\gamma_\eta(x) = (1, -\|\eta\|^{-a} \vee x \wedge \|\eta\|^{-a})^\top$, and put

$$\zeta_\eta = \bar\zeta_\eta - \int \bar\zeta_\eta \gamma^\top \, dG \Big(\int \gamma_\eta \gamma^\top \, dG \Big)^{-1} \gamma_\eta$$

for η close to the origin. As $\int \zeta_\eta \gamma^\top \, dG = 0$, we see that $\zeta_\eta \in H^q$. Moreover, we have $\|\eta^\top \zeta_\eta\| \le \|\eta\|^{1-a} + o(\|\eta\|^{1-a})$ and $\int \|\zeta_\eta - \zeta\|^2 \, dG \to 0$ as $\eta \to 0$. Define F_η by $dF_\eta = (1 + \eta\zeta_\eta) dG$ for η close enough to 0 so that $1 + \eta^\top \zeta_\eta \ge 0$. It is now easy to check that F_η is a smooth path. □

In Sections 4 and 5 we consider one-dimensional sequences rather than q-dimensional paths. For an appropriate version of Lemma 1, set $q = 1$, $\zeta = h$, $\eta = n^{-1/2}$ and $a = 1/4$ in the proof of Lemma 1 to obtain the following.

Corollary 1. *For each h in H there exists a sequence $\langle h_n \rangle$ in H such that*

$$|h_n| \le n^{1/8}, \qquad \int |h_n - h| \, dG \to 0,$$

and such that the distribution function $F_{n,h}$ with G-density $1 + n^{-1/2} h_n$ has finite Fisher information $I(F_{n,h})$ which converges to $I(G)$.

To obtain local asymptotic normality we need to impose some smoothness assumptions on the maps α and β.

Assumption 1. The function α is locally Lipschitz at τ: There is a constant L such that for all ϑ in a neighborhood of τ,

$$\sum_{s=1}^{\infty} |\alpha_s(\vartheta) - \alpha_s(\tau)| \leq L\|\vartheta - \tau\|.$$

Assumption 2. The maps $\beta_1, \beta_2, \dots$ are continuously differentiable with gradients $\dot{\beta}_1, \dot{\beta}_2, \dots$, and for some $\delta > 0$,

$$\sum_{s=1}^{\infty} \sup_{\|\vartheta-\tau\|<\delta} \|\dot{\beta}_s(\vartheta)\| < \infty.$$

We also need a condition which guarantees a good approximation of the actual innovation $X_t(\vartheta)$ for large t by the truncated innovation defined by

$$\xi_t(\vartheta) = Y_t + \sum_{s=1}^{t} \beta_s(\vartheta) Y_{t-s}, \quad t = 1, 2, \dots. \tag{2.7}$$

Note that

$$E_{\vartheta,F}|X_t(\vartheta) - \xi_t(\vartheta)| \leq \int |x|\, dF(x) \sum_{s=t+1}^{\infty} |\beta_s(\vartheta)| \sum_{k=0}^{\infty} |\alpha_k(\vartheta)|, \tag{2.8}$$

$$E_{\vartheta,F}[(X_t(\vartheta) - \xi_t(\vartheta))^2] \leq \int x^2\, dF(x) \Big(\sum_{s=t+1}^{\infty} |\beta_s(\vartheta)| \sum_{k=0}^{\infty} |\alpha_k(\vartheta)| \Big)^2. \tag{2.9}$$

Call a sequence $\langle \vartheta_n \rangle$ in Θ *local* (*at* τ) if $\langle n^{1/2}(\vartheta_n - \tau) \rangle$ is bounded.

Assumption 3. There are functions r_n from Θ into the set of positive integers such that for all local sequences $\langle \vartheta_n \rangle$ and $\langle \vartheta_n^* \rangle$,

$$n^{-1/2} r_n(\vartheta_n^*) \to 0 \quad \text{and} \quad n \sum_{s=r_n(\vartheta_n^*)+1}^{\infty} |\beta_s(\vartheta_n)| \to 0.$$

Remark 1. We should point out that the above three assumptions are met by the AR, MA and ARMA models under the usual assumptions. To be more transparent, let us look at the simplest such models and give explicit formulas for α and β.

(i) Consider first the AR(1) process $Y_t = X_t + \vartheta Y_{t-1}$ with $|\vartheta| < 1$. In this case, $\alpha_s(\vartheta) = \vartheta^s$, while $\beta_1(\vartheta) = -\vartheta$ and $\beta_s(\vartheta) = 0$ for $s \geq 2$.

(ii) For the MA(1) process $Y_t = X_t + \vartheta X_{t-1}$ with $|\vartheta| < 1$, we have $\alpha_1(\vartheta) = \vartheta$ and $\alpha_s(\vartheta) = 0$ for $s \geq 2$, while $\beta_s(\vartheta) = (-\vartheta)^s$.

(iii) Finally consider the ARMA(1,1) process $Y_t - \vartheta_1 Y_{t-1} = X_t - \vartheta_2 X_{t-1}$ with $\Theta = \{(\vartheta_1, \vartheta_2) : \vartheta_1, \vartheta_2 \in (-1, 1), \vartheta_1 \neq \vartheta_2\}$. Then $\alpha_s(\vartheta) = (\vartheta_1 - \vartheta_2)\vartheta_1^{s-1}$ and $\beta_s(\vartheta) = (\vartheta_2 - \vartheta_1)\vartheta_2^{s-1}$.

From these formulas for α and β it is easy to see that the above three assumptions hold.

It follows from Assumption 2 that the maps $\vartheta \mapsto X_t(\vartheta)$ and $\vartheta \mapsto \xi_t(\vartheta)$ are continuously differentiable in a neighborhood of τ with gradients

$$\dot X_t(\vartheta) = \sum_{s=1}^{\infty} \dot\beta_s(\vartheta) Y_{t-s} \quad \text{and} \quad \dot\xi_t(\vartheta) = \sum_{s=1}^{t} \dot\beta_s(\vartheta) Y_{t-s}.$$

We will also use the following consequences of Assumption 2.

Lemma 2. *Let $\langle\vartheta_n\rangle$ be a local sequence. Then for s_n tending to infinity,*

$$\sum_{j=s_n+1}^{n} E_{\tau,G}[|\xi_j(\vartheta_n) - \xi_j(\tau) - (\vartheta_n - \tau)^\top \dot\xi_j(\tau)|^2] \quad \to \quad 0, \tag{2.10}$$

$$(n-s_n)^{-1} \sum_{j=s_n+1}^{n} E_{\tau,G}[\|\dot\xi_j(\vartheta_n) - \dot X_j(\tau)\|^2] \quad \to \quad 0. \tag{2.11}$$

Proof. The spectral norm of a non-negative definite matrix is bounded by the largest absolute row sum. This and (2.4) show that the spectral norms of the dispersion matrices $[\gamma_{|r-s|}(\tau, G)]_{r,s=0,\dots,t}$ of $(Y_0, \dots, Y_t)$ under $P_{\tau,G}$ are bounded uniformly in t by a constant K. Using this we find that the left-hand side of (2.10) is bounded by

$$Kn \sum_{j=0}^{n} |\beta_j(\vartheta_n) - \beta_j(\tau) - (\vartheta_n - \tau)^\top \dot\beta_j(\tau)|^2,$$

which tends to zero by Assumption 2. Similarly, one verifies that

$$(n-s_n)^{-1} \sum_{j=s_n+1}^{n} E_{\tau,G}[\|\dot\xi_j(\vartheta_n) - \dot\xi_j(\tau)\|^2] \le K \sum_{j=0}^{\infty} \|\dot\beta_j(\vartheta_n) - \dot\beta_j(\tau)\|^2,$$

which also converges to zero by Assumption 2. Finally, we have

$$(n-s_n)^{-1} \sum_{j=s_n+1}^{n} E_{\tau,G}[\|\dot\xi_j(\tau) - \dot X_j(\tau)\|^2] \le K \sum_{j=s_n+1}^{\infty} \|\dot\beta_j(\tau)\|^2 \to 0.$$

This completes the proof of (2.11). $\square$

Note that under $P_{\vartheta,G}$ the process $\{\dot X_t(\vartheta) : t \in \mathbb{Z}\}$ is centered with finite second moments,

$$E_{\vartheta,G}[\dot X_0(\vartheta)] = 0 \quad \text{and} \quad E_{\vartheta,G}[\|\dot X_0(\vartheta)\|^2] < \infty,$$

and has dispersion matrix

$$V(\vartheta, G) = E_{\vartheta,G}[\dot X_0(\vartheta) \dot X_0^\top(\vartheta)].$$

We can use this and a martingale central limit theorem to show that

$$S_n(\vartheta, G) = n^{-1/2} \sum_{j=1}^{n} \begin{pmatrix} \dot X_j(\vartheta) \ell_G(X_j(\vartheta)) \\ \zeta(X_j(\vartheta)) \end{pmatrix}$$

is asymptotically normal under $P_{\vartheta,G}$ with mean vector zero and dispersion matrix

$$W(\vartheta,G)=\begin{bmatrix} V(\vartheta,G)I(G) & 0 \\ 0 & \int \zeta\zeta^\top\,dG\end{bmatrix}.$$

Let $Q_{n,\vartheta,\eta}$ denote the distribution of $(Y_0,\dots,Y_n)$ under P_{ϑ,F_η}.

Theorem 1. *Let $\langle\vartheta_n\rangle$ be a local sequence, and let $\langle w_n\rangle=\langle(u_n^\top,v_n^\top)^\top\rangle$ be a bounded sequence in $\mathbb{R}^k\times\mathbb{R}^q$. Then*

$$\begin{aligned}\log\frac{dQ_{n,\vartheta_n+n^{-1/2}u_n,n^{-1/2}v_n}}{dQ_{n,\vartheta_n,0}}(Y_0,\dots,Y_n) &= w_n^\top S_n(\vartheta_n,G)-\frac{1}{2}w_n^\top W(\tau,G)w_n+o_{P_{\vartheta_n,G}}(1),\\ \mathfrak{L}(S_n(\vartheta_n,G)|P_{\vartheta_n,G}) &\Longrightarrow N(0,W(\tau,G)).\end{aligned}$$

Theorem 1 establishes local asymptotic normality locally uniformly in ϑ. This is in the spirit of Jeganathan [13], Drost, Klaassen and Werker [5] and Koul and Schick [14].

To prove Theorem 1 we shall show first that the distributions $Q_{n,\vartheta,\eta}$ are asymptotically equivalent to other distributions $\bar{Q}_{n,\vartheta,\eta}$ for which local asymptotic normality is essentially known. Note that $Q_{n,\vartheta,\eta}$ has density

$$q_{n,\vartheta,\eta}(y_0,\dots,y_n)=E_{\vartheta,F_\eta}\Big[\prod_{j=0}^n f_\eta\Big(\sum_{i=0}^j\beta_i(\vartheta)y_{j-i}+\sum_{i=j+1}^\infty\beta_i(\vartheta)Y_{j-i}\Big)\Big],\quad y_0,\dots,y_n\in\mathbb{R}.$$

Since this density is hard to work with, we shall now show that it can be replaced by the more manageable density

$$\bar{q}_{n,\vartheta,\eta}(y_0,\dots,y_n)=q_{r_n(\vartheta_n^*),\vartheta,\eta}(y_0,\dots,y_{r_n(\vartheta_n)})\prod_{j=r_n(\vartheta_n^*)+1}^n f_\eta\Big(\sum_{i=0}^j\beta_i(\vartheta)y_{j-i}\Big),\quad y_0,\dots,y_n\in\mathbb{R},$$

for some fixed local sequence $\langle\vartheta_n^*\rangle$ and r_n as in Assumption 3. Indeed, we have for every constant C,

$$\sup_{\|\vartheta-\tau\|+\|\eta\|\le Cn^{-1/2}}\int|q_{n,\vartheta,\eta}-\bar{q}_{n,\vartheta,\eta}|\,d\lambda^{n+1}\to 0. \tag{2.12}$$

Here λ denotes the Lebesgue measure on the Borel sets of $\mathbb{R}$ and λ^m its m-fold product. The above is a simple consequence of the following lemma and Assumption 3.

Lemma 3. *For every $\vartheta\in\Theta$ and $\eta\in\Delta$,*

$$\int|q_{n,\vartheta,\eta}-\bar{q}_{n,\vartheta,\eta}|\,d\lambda^{n+1}\le(I(F_\eta))^{1/2}\sum_{j=r_n(\vartheta_n^*)+1}^n E_{\vartheta,F_\eta}\Big|\sum_{i=1}^\infty\beta_{j+i}(\vartheta)Y_{-i}\Big|.$$

Proof. Fix $\vartheta\in\Theta$ and $\eta\in\Delta$. Set

$$R_j=\sum_{i=1}^\infty\beta_{j+i}(\vartheta)Y_{-i},\quad j\ge 0.$$

Using the inequality $|\prod_{i=1}^m a_i - \prod_{i=1}^m b_i| \le \sum_{k=1}^m \prod_{i=1}^{k-1} |a_i||a_k - b_k| \prod_{j=k+1}^m |b_j|$ and then the substitutions $x_k = \sum_{i=0}^k \beta_k(\vartheta) y_{k-i}$ we can calculate

$$\int |q_{n,\vartheta,\eta} - \bar{q}_{n,\vartheta,\eta}|\, d\lambda^{n+1} \le \sum_{j=r_n(\vartheta_n^*)+1}^{n} E_{\vartheta,F_\eta} \int_{-\infty}^{\infty} |f_\eta(x_j + R_j) - f_\eta(x_j)|\, dx_j.$$

Since F_η has finite Fisher information, we may choose f_η to be absolutely continuous and have $\int |f_\eta'|\, d\lambda \le (I(F_\eta))^{1/2}$. Using this and Fubini's theorem we can bound

$$\int_{-\infty}^{\infty} |f_\eta(x+R) - f_\eta(x)|\, dx = \int_{-\infty}^{\infty} \Big| \int_0^1 R f_\eta'(x + sR) ds \Big|\, dx \le (I(F_\eta))^{1/2} |R|.$$

The desired result is now immediate. □

We also need the following result.

Lemma 4. *There exists $\delta > 0$ and a finite constant C such that for all positive integers r and whenever $\|\vartheta - \tau\| + \|\eta\| < \delta$,*

$$\int |q_{r,\vartheta,\eta} - q_{r,\tau,0}|\, d\lambda^{r+1} \le (r+1) C (\|\vartheta - \tau\| + \|\eta\|).$$

Proof. Note that

$$R_j(\vartheta) = \sum_{i=1}^{\infty} \beta_{j+i}(\vartheta) Y_{-i} = \sum_{k=1}^{\infty} \Phi_{j,k}(\vartheta) X_{-k}(\vartheta) \qquad \text{with} \quad \Phi_{j,k}(\vartheta) = \sum_{i=0}^{k-1} \beta_{j+k-i}(\vartheta) \alpha_i(\vartheta).$$

Easy calculations show that, for all $j \ge 1$,

$$\sum_{k=1}^{\infty} |\Phi_{j,k}(\vartheta)| \le \sum_{k=1}^{\infty} |\beta_k(\vartheta)| \sum_{i=0}^{\infty} |\alpha_i(\vartheta)|, \tag{2.13}$$

$$\begin{aligned} \sum_{k=1}^{\infty} |\Phi_{j,k}(\vartheta) - \Phi_{j,k}(\tau)| \le & \sum_{i=0}^{\infty} |\alpha_i(\vartheta)| \sum_{k=1}^{\infty} |\beta_k(\vartheta) - \beta_k(\tau)| \\ & + \sum_{k=1}^{\infty} |\beta_k(\tau)| \sum_{i=0}^{\infty} |\alpha_i(\vartheta) - \alpha_i(\tau)|. \end{aligned} \tag{2.14}$$

Let $U_1, U_2, \dots$ be independent random variables each uniformly distributed on the open interval $(0,1)$. Let F^{-1} denote the quantile function of the distribution function F, i.e. $F^{-1}(u) = \inf\{s \in \mathbb{R} : F(s) \ge u\}$ for $0 < u < 1$. Then $R_j(\vartheta)$ has the same distribution under P_{ϑ,F_η} as

$$\tilde{R}_j(\vartheta,\eta) = \sum_{k=1}^{\infty} \Phi_{j,k}(\vartheta) F_\eta^{-1}(U_k).$$

This allows us to write

$$q_{r,\vartheta,\eta}(y_0,\dots,y_r) = E\Big[\prod_{j=0}^{r} f_\eta\Big(\sum_{i=0}^{j} \beta_i(\vartheta) y_{j-i} + \tilde{R}_j(\vartheta,\eta) \Big) \Big], \quad y_0,\dots,y_r \in \mathbb{R}.$$

Using the fact that

$$\int |f_\eta(x+s+t) - g(x+s)|\,dx \le \int |f_\eta(x) - g(x)|\,dx + (I(G))^{1/2}|t|, \quad s,t \in \mathbb{R},$$

we can now infer that

$$\int |q_{r,\vartheta,\eta} - q_{r,\tau,0}|\,d\lambda^{r+1} \le (r+1)\int |f_\eta - g|\,d\lambda + (I(G))^{1/2}(A_r(\vartheta,\eta) + B_r(\vartheta,\eta)),$$

where

$$\begin{aligned} A_r(\vartheta,\eta) &= \sum_{j=1}^{r}\sum_{i=1}^{j} |\beta_i(\vartheta) - \beta_i(\tau)| E_{\vartheta,\eta}[|Y_{j-i}|] \\ &\le r\sum_{k=1}^{\infty} |\beta_k(\vartheta) - \beta_k(\tau)| \sum_{i=0}^{\infty} |\alpha_i(\vartheta)| \int |x|\,dF_\eta(x) \end{aligned}$$

and

$$\begin{aligned} B_r(\vartheta,\eta) &= \sum_{j=0}^{r} E[|\tilde{R}_j(\vartheta,\eta) - \tilde{R}_j(\tau,0)|] \\ &\le \sum_{j=0}^{r}\sum_{k=1}^{\infty} \Big(|\Phi_{j,k}(\vartheta) - \Phi_{j,k}(\tau)| \int |x|\,dF_\eta(x) + |\Phi_{j,k}(\tau)| \int |x||f_\eta(x) - g(x)|\,dx \Big). \end{aligned}$$

Here we have used the fact that

$$\begin{aligned} \int_0^1 |F_\eta^{-1}(u) - G^{-1}(u)|\,du &= \int_{-\infty}^{\infty} |F_\eta(x) - G(x)|\,dx \\ &\le \int_0^{\infty} \int_{|t|>x} |f_\eta(t) - g(t)|\,dt\,dx \\ &= \int |x||f_\eta(x) - g(x)|\,dx. \end{aligned}$$

The first equality is known, see for example Bickel and Freedman [3, relation (8.1)], and the last equality follows from an application of Fubini's theorem. To get the inequality, we have used the fact that $\int_{-\infty}^{x} (f_\eta(t) - g(t))\,dt = -\int_x^{\infty} (f_\eta(t) - g(t))\,dt$.

It follows from the properties of a regular path that

$$\int (1+|x|)|f_\eta(x) - g(x)|\,dx = O(\|\eta\|).$$

The desired result is now immediate in view of this, the above inequalities and Assumptions 1 and 2. □

Let $\bar{Q}_{n,\vartheta,\eta}$ denote the distribution with density $\bar{q}_{n,\vartheta,\eta}$. It follows from (2.12) that for $\langle\vartheta_n\rangle$ local and $n^{1/2}\eta_n$ bounded, $\langle\bar{Q}_{n,\vartheta_n,\eta_n}\rangle$ and $\langle Q_{n,\vartheta_n,\eta_n}\rangle$ are contiguous and

$$\log \frac{d\bar{Q}_{n,\vartheta_n,\eta_n}}{dQ_{n,\vartheta_n,\eta_n}}(Y_0,\dots,Y_n) = o_{P_{\vartheta_n,\eta_n}}(1). \tag{2.15}$$

Note that $\bar{Q}_{n,\vartheta,\eta}$ is the distribution of $(Y_0,\dots,Y_n)$ under a probability measure $P_{r_n(\vartheta_n^*),\vartheta,F_\eta}$ for which $\{Y_t : t \le r_n(\vartheta_n^*)\}$ have the same distribution as under P_{ϑ,F_η}, and $\{\xi_t(\vartheta) : t > r_n(\vartheta_n^*)\}$ are independent and identically distributed with common distribution function F_η and independent of $\{Y_t : t \le r_n(\vartheta_n^*)\}$. In other words, the truncated innovations $\{\xi_t(\vartheta) : r_n(\vartheta_n^*) < t \le n\}$ under the measure P_{ϑ,F_η} become the actual innovations under the measure $P_{r_n(\vartheta_n^*),\vartheta,F_\eta}$. Thus under these alternative measures our observations form a nonlinear autoregressive process of order $r_n(\vartheta_n^*)$. Sufficient conditions for local asymptotic normality of such models have been given in Koul and Schick [14] if $r_n(\vartheta_n^*)$ equals a constant. Their conditions, however, extend easily to the present case with $r_n(\vartheta_n^*)$ tending to infinity. Indeed, we have the following result.

Theorem 2. *Let $\langle\vartheta_n\rangle$ and $\langle\tau_n\rangle$ be local sequences and $\langle w_n\rangle = \langle(u_n^\top, v_n^\top)^\top\rangle$ a bounded sequence in $\mathbb{R}^k \times \mathbb{R}^q$. Then, with $P_n^* = P_{r_n(\vartheta_n^*),\vartheta_n,G}$,*

$$\int |q_{r_n(\vartheta_n^*),\vartheta_n,n^{-1/2}v_n} - q_{r_n(\vartheta_n^*),\tau,0}|\, d\lambda^{r_n(\vartheta_n)+1} \;\to\; 0, \tag{2.16}$$

$$\sum_{j=r_n(\vartheta_n^*)+1}^{n} |\xi_j(\tau_n) - \xi_j(\vartheta_n) - (\tau_n-\vartheta_n)^\top \dot\xi_j(\vartheta_n)|^2 = o_{P_n^*}(1), \tag{2.17}$$

$$\max_{r_n(\vartheta_n^*)<j\le n} n^{-1/2}\|\dot\xi_j(\vartheta_n)\| = o_{P_n^*}(1), \tag{2.18}$$

$$(n-r_n(\vartheta_n^*))^{-1} \sum_{j=r_n(\vartheta_n^*)+1}^{n} \dot\xi_j(\vartheta_n) = o_{P_n^*}(1), \tag{2.19}$$

$$(n-r_n(\vartheta_n^*))^{-1} \sum_{j=r_n(\vartheta_n^*)+1}^{n} \dot\xi_j(\vartheta_n)\dot\xi_j^\top(\vartheta_n) = V(\tau,G) + o_{P_n^*}(1). \tag{2.20}$$

Consequently,

$$\log \frac{d\bar{Q}_{n,\vartheta_n+n^{-1/2}u_n,n^{-1/2}v_n}}{d\bar{Q}_{n,\vartheta_n,0}}(Y_0,\dots,Y_n) = w_n^\top \bar{S}_n(\vartheta_n,G) - \frac{1}{2}w_n^\top W(\tau,G)w_n + o_{P_n^*}(1), \tag{2.21}$$

$$\mathfrak{L}(\bar{S}_n(\vartheta_n,G)|\bar{P}_{n,\vartheta_n}) \implies N(0,W(\tau,G)), \tag{2.22}$$

where

$$\bar{S}_n(\vartheta,G) = (n-r_n(\vartheta_n^*))^{-1/2} \sum_{j=r_n(\vartheta_n^*)+1}^{n} \begin{pmatrix} \dot\xi_j(\vartheta)\ell_G(\xi_j(\vartheta)) \\ \zeta(\xi_j(\vartheta)) \end{pmatrix}.$$

Proof. We need only show that (2.16) to (2.20) hold. The remaining statements (2.21) and (2.22) follow then as in Koul and Schick [14]. Of course, (2.16) follows from Lemma 4 and the fact that $r_n(\vartheta_n^*)(\|\vartheta_n-\tau\| + \|n^{-1/2}v_n\|) \to 0$ by Assumption 3.

It follows from (2.10) and (2.11) and contiguity of $\langle\bar{Q}_{n,\tau,0}\rangle$ and $\langle Q_{n,\tau,0}\rangle$ that (2.17) to (2.20) hold with $\vartheta_n = \tau$. Thus (2.21) and (2.22) hold with $\vartheta_n = \tau$. This yields contiguity of $\langle\bar{Q}_{n,\vartheta_n,0}\rangle$ and $\langle\bar{Q}_{n,\tau,0}\rangle$. (2.10) and (2.11) imply also (2.17) to (2.20) with P_n^* replaced by $P_{\tau,G}$. This gives (2.17) and (2.20) as $\langle\bar{Q}_{n,\vartheta_n,0}\rangle$ and $\langle Q_{n,\tau,0}\rangle$ are contiguous. $\square$

It follows from Theorem 2 and (2.15) that the sequences $\langle Q_{n,\vartheta_n,\eta_n}\rangle$ and $\langle Q_{n,\tau,0}\rangle$ are contiguous whenever $\langle\vartheta_n\rangle$ is local and $n^{1/2}\eta_n$ is bounded. From this and (2.15) we obtain that

$$\log\frac{d\bar{Q}_{n,\vartheta_n+n^{-1/2}u_n,n^{-1/2}v_n}}{d\bar{Q}_{n,\vartheta_n,0}}(Y_0,\dots,Y_n)-\log\frac{dQ_{n,\vartheta_n+n^{-1/2}u_n,n^{-1/2}v_n}}{dQ_{n,\vartheta_n,0}}(Y_0,\dots,Y_n)=o_{P_{\vartheta_n,G}}(1).$$

Thus Theorem 1 follows if we show that

$$\bar{S}_n(\vartheta_n,G)-S_n(\vartheta_n,G)=o_{P_{\vartheta_n,G}}(1). \tag{2.23}$$

Of course, this statement is easily verified with the help of Assumption 3 if ℓ_G and all the components of ζ are Lipschitz continuous. The general result follows from the fact that the Lipschitz continuous functions are dense in $L_2(G)$.

Remark 2. Theorem 2 establishes local asymptotic normality locally uniformly in ϑ. This yields the following consequence. Under the assumptions of the above theorem, for $\langle\vartheta_n\rangle$ local,

$$\bar{S}_n(\vartheta_n,G)-\bar{S}_n(\tau,G)=M(\tau,G)(\vartheta_n-\tau)+o_{P_{\vartheta,G}}(1),$$

where $M(\tau,G)$ is the matrix consisting of the first k columns of $W(\tau,G)$. This and (2.23) yield for such sequences $\langle\vartheta_n\rangle$ and $\vartheta_n^*=\vartheta_n$:

$$(n-r_n(\vartheta_n))^{-1/2}\sum_{j=r_n(\vartheta_n)+1}^{n}\zeta(\xi_j(\vartheta_n))=n^{-1/2}\sum_{j=1}^{n}\zeta(X_j(\tau))+o_{P_{\tau,G}}(1), \tag{2.24}$$

and, if $V(\tau,G)$ is invertible,

$$\bar{U}_n(\vartheta_n,G)=U_n(\tau,G)+o_{P_{\tau,G}}(n^{-1/2}) \tag{2.25}$$

with

$$\bar{U}_n(\vartheta_n,G) = \vartheta_n+(n-r_n(\vartheta_n))^{-1}\sum_{j=r_n(\vartheta_n)+1}^{n}(V(\tau,G)I(G))^{-1}\dot{\xi}_j(\vartheta_n)\ell_G(\xi_j(\vartheta_n)), \tag{2.26}$$

$$U_n(\tau,G) = \tau+\frac{1}{n}\sum_{j=1}^{n}(V(\tau,G)I(G))^{-1}\dot{X}_j(\tau)\ell_G(X_j(\tau)). \tag{2.27}$$

3. Characterization of efficient estimators

Let κ denote a function from $\Theta\times\mathfrak{F}$ into $\mathbb{R}^m$. In this section we recall a characterization of least dispersed regular estimators of κ in the context of our model. Fix τ in Θ and G in $\mathfrak{F}$, let Assumptions 1 to 3 hold and assume that $V(\tau,G)$ is positive definite. Let H denote the tangent space (2.6) of $\mathfrak{F}$ at G. For each $t\in\mathbb{R}^k$ fix a sequence $\langle\vartheta_{n,t}\rangle$ in Θ such that $n^{1/2}(\vartheta_{n,t}-\tau)\to t$. For each $h\in H$ fix a sequence $\langle h_n\rangle$ in H such that $|h_n|\le n^{1/8}$ and $\int(h_n-h)^2\,dG\to 0$, and such that the distribution function $F_{n,h}$ with G-density $1+n^{-1/2}h_n$ has finite Fisher information $I(F_{n,h})$ which converges to $I(G)$. Such a sequence exists by Corollary 1. Let $Q_{t,h}^{(n)}$ denote the distribution of $(Y_0,\dots,Y_n)$ under $P_{\vartheta_{n,t},F_{n,h}}$. By Theorem 1, $Q^{(n)}_{t,h}$ is locally asymptotically normal,

$$\log \frac{dQ^{(n)}_{t,h}}{dQ^{(n)}_{0,0}}(Y_0,\dots,Y_n) = n^{-1/2}\sum_{j=1}^n [t^\top \dot X_j(\tau)\ell_G(X_j(\tau)) + h(X_j(\tau)) - \frac{1}{2}\sigma^2(t,h) + o_{P_{\tau,G}}(1),$$

$$\mathfrak{L}\Big(n^{-1/2}\sum_{j=1}^n [t^\top \dot X_j(\tau)\ell_G(X_j(\tau)) + h(X_j(\tau))]|P_{\tau,G}\Big) \implies N(0,\sigma^2(t,h)),$$

with squared *LAN norm*

$$\sigma^2(t,h) = t^\top V(\tau,G)I(G)t + \int h^2\,dG. \tag{3.1}$$

We assume that κ is *differentiable* at (τ,G) in the sense that there are an $m\times k$ matrix A and a vector $b\in H^m$ such that

$$n^{1/2}(\kappa(\vartheta_{n,t},F_{n,h}) - \kappa(\tau,G)) \to At + \int bh\,dG, \quad (t,h)\in\mathbb{R}^k\times H. \tag{3.2}$$

An estimator $\hat\kappa$ of κ is called *regular* at (τ,G) with *limit* Q if Q is a distribution such that

$$\mathfrak{L}(n^{1/2}(\hat\kappa - \kappa(\vartheta_{n,t},F_{n,h}))|P_{\vartheta_{n,t},F_{n,h}}) \implies Q, \quad (t,h)\in\mathbb{R}^k\times H.$$

We now express the right-hand side of (3.2) in terms of the inner product induced by the LAN norm (3.1),

$$At + \int bh\,dG = A(V(\tau,G)I(G))^{-1}V(\tau,G)I(G)t + \int bh\,dG.$$

It follows from the Hájek–LeCam convolution theorem that a regular estimator $\hat\kappa$ for κ is least dispersed at (τ,G) if and only if

$$\hat\kappa = \kappa(\tau,G) + \frac{1}{n}\sum_{j=1}^n [A(V(\tau,G)I(G))^{-1}\dot X_j(\tau)\ell_G(X_j(\tau)) + b(X_j(\tau))] + o_{P_{\tau,G}}(n^{-1/2}). \tag{3.3}$$

Also, any estimator $\hat\kappa$ fulfilling (3.3) is regular. For the required semiparametric version of the convolution theorem we refer to Bickel et al. [4, Section 3.3]; the result there is stated for the i.i.d. case but easily seen to be valid for any locally asymptotically normal model.

4. Efficient estimators of a linear functional of the innovation distribution

In this section we construct a least dispersed regular estimator of the functional κ_ψ defined by

$$\kappa_\psi(\vartheta,F) = \int \psi\,dF, \quad \vartheta\in\Theta, F\in\mathfrak{F}.$$

Fix τ in Θ and G in $\mathfrak{F}$, let Assumptions 1 to 3 hold and assume that $V(\tau,G)$ is positive definite. The functional is obviously differentiable at (τ,G) with $A=0$ and $b=\psi_*$, where ψ_* is the projection of ψ onto H:

$$\psi_*(x) = \psi(x) - \int \psi\,dG - c_* x, \quad x\in\mathbb{R},$$

with

$$c_* = \frac{\int x\psi(x)\, dG(x)}{\int x^2\, dG(x)}.$$

Hence, by characterization (3.3), an estimator $\hat{\psi}_n$ of the functional κ_ψ is least dispersed and regular at (τ, G) if and only if

$$\hat{\psi}_n = \int \psi\, dG + \frac{1}{n}\sum_{j=1}^{n} \psi_*(X_j(\tau)) + o_{P_{\tau,G}}(n^{-1/2}). \tag{4.1}$$

It suggests itself to construct such an estimator as in the following theorem. By an estimator of the Euclidean parameter we mean a sequence $\langle\tilde{\vartheta}_n\rangle$ with $\tilde{\vartheta}_n = T_n(Y_0, \dots, Y_n)$ for some measurable function T_n from $\mathbb{R}^{n+1}$ into Θ. Call $\tilde{\vartheta}_n$ *discretized* if $cn^{1/2}\tilde{\vartheta}_n$ takes values in $\mathbb{Z}^k$ for some positive c. Call it $n^{1/2}$-consistent at (τ, G) if $n^{1/2}(\tilde{\vartheta}_n - \tau) = O_{P_{\tau,G}}(1)$. Each estimator $n^{1/2}$-consistent at (τ, G) has an obvious discretized $n^{1/2}$-consistent modification.

Theorem 3. *Let $\tilde{\vartheta}_n$ be a discretized estimator of the Euclidean parameter that is $n^{1/2}$-consistent at (τ, G). Define*

$$\begin{aligned}
\tilde{\psi}_n(\vartheta) &= (n - r_n(\vartheta))^{-1} \sum_{j=r_n(\vartheta)+1}^{n} [\psi(\xi_j(\vartheta)) - \hat{c}_n(\vartheta)\xi_j(\vartheta)], \quad \vartheta \in \Theta,\\
\hat{c}_n(\vartheta) &= \frac{\sum_{j=r_n(\vartheta)+1}^{n} \xi_j(\vartheta)\psi(\xi_j(\vartheta))}{\sum_{j=r_n(\vartheta)+1}^{n} \xi_j^2(\vartheta)}, \quad \vartheta \in \Theta.
\end{aligned}$$

Then $\tilde{\psi}_n(\tilde{\vartheta}_n)$ satisfies (4.1) and is therefore a least dispersed and regular estimator of κ_ψ at (τ, G).

Proof. Since the estimator $\tilde{\vartheta}_n$ is discrete and $n^{1/2}$-consistent at (τ, G), it suffices to show that the "estimator" $\tilde{\psi}_n(\vartheta_n)$ satisfies (4.1) for every local sequence $\langle\vartheta_n\rangle$. Fix such a local sequence $\langle\vartheta_n\rangle$. In view of (2.24) with $\zeta = \psi_*$, it suffices to show that

$$\tilde{\psi}_n(\vartheta_n) = \int \psi\, dG + (n - r_n(\vartheta_n))^{-1} \sum_{j=r_n(\vartheta_n)+1}^{n} \psi_*(\xi_j(\vartheta_n)) + o_{P_{\tau,G}}(n^{-1/2}).$$

This is equivalent to

$$(\hat{c}_n(\vartheta_n) - c_*)(n - r_n(\vartheta_n))^{-1} \sum_{j=r_n(\vartheta_n)+1}^{n} \xi_j(\vartheta_n) = o_{P_{\tau,G}}(n^{-1/2}).$$

Using contiguity, we can verify the latter under the sequence $\langle\bar{P}_n\rangle = \langle P_{r_n(\vartheta_n),\vartheta_n,G}\rangle$. Since the variables $\xi_{r_n(\vartheta_n)+1}(\vartheta_n), \dots, \xi_n(\vartheta_n)$ are independent and identically distributed under $\bar{P}_n$, we obtain

$$(n - r_n(\vartheta_n))^{-1} \sum_{j=r_n(\vartheta_n)+1}^{n} \xi_j(\vartheta_n) = O_{\bar{P}_n}(n^{-1/2})$$

and

$$\hat{c}_n(\vartheta_n) = c_* + o_{\bar{P}_n}(1),$$

which yields the desired result. □

Example 1. Suppose we want to estimate the m-th moment of the error distribution for some $m > 1$. Then we need to take $\psi(x) = x^m$. The least dispersed regular estimator in this case has the simple form

$$\hat{\mu}_{n,m} - \frac{\hat{\mu}_{n,m+1}}{\hat{\mu}_{n,2}}\, \hat{\mu}_{n,1}$$

where

$$\hat{\mu}_{n,\nu} = \frac{1}{n - r_n(\tilde{\vartheta}_n)} \sum_{j=r_n(\tilde{\vartheta}_n)+1}^{n} \xi_j^{\nu}(\tilde{\vartheta}_n), \quad \nu = 1, 2, \dots. \tag{4.2}$$

Example 2. An important example is given by the choice $\psi(x) = \mathbf{1}[x \le t]$, for which $\kappa_\psi(\vartheta, F) = F(t)$ is the value of the distribution function F at t. Here t is a fixed real number. The corresponding least dispersed regular estimator is given by $\hat{F}_n(t, \tilde{\vartheta}_n)$ with $\tilde{\vartheta}_n$ as in Theorem 3 and

$$\hat{F}_n(t, \vartheta) = (n - r_n(\vartheta))^{-1} \sum_{j=r_n(\vartheta)+1}^{n} \big(\mathbf{1}[\xi_j(\vartheta) \le t] - \hat{c}_n(t, \vartheta)\xi_j(\vartheta)\big)$$

with

$$\hat{c}_n(t, \vartheta) = \frac{\sum_{j=r_n(\vartheta)+1}^{n} \xi_j(\vartheta)\mathbf{1}[\xi_j(\vartheta) \le t]}{\sum_{j=r_n(\vartheta)+1}^{n} \xi_j^2(\vartheta)}.$$

Of course, $\hat{F}_n(t, \vartheta)$ is just an improved version of the empirical estimator based on the pseudo-innovations $\xi_{r_n(\vartheta)+1}(\vartheta), \dots, \xi_n(\vartheta)$ that incorporates the constraint that the innovations have zero mean. See Schick and Wefelmeyer [24] for a variant of this result for nonlinear autoregression models.

It follows from Theorem 3 that

$$n^{1/2} |\hat{F}_n(t, \tilde{\vartheta}_n) - F_n(t)| = o_{P_{\tau,G}}(1),$$

where

$$F_n(t) = \frac{1}{n} \sum_{j=1}^{n} \big(\mathbf{1}[X_j(\tau) \le t] - c_*(t) X_j(\tau)\big) \quad \text{with} \quad c_*(t) = \frac{\int x \mathbf{1}[x \le t]\, dG(x)}{\int x^2\, dG(x)}.$$

We shall now strengthen this to uniform convergence.

Note that the maps $t \mapsto \hat{F}_n(t, \vartheta)$, F_n and G have obvious extensions to $[-\infty, \infty]$. It is easy to check that the process $\{n^{1/2}(F_n(t) - G(t)) : t \in [-\infty, \infty]\}$, viewed as element of the space $D([-\infty, \infty])$ endowed with the Skorohod topology, converges in distribution under $P_{\tau,G}$ to a centered continuous Gaussian process with covariance function

$$(s, t) \mapsto G(s \wedge t) - G(s)G(t) - c_*(s)c_*(t) \int x^2\, dG(x).$$

Let $F_{n,*}(t, \vartheta)$ denote the version of $\hat{F}_n(t, \vartheta)$ obtained by replacing $\hat{c}_n(t, \vartheta)$ by $c_*(t)$. A similar argument as for F_n shows that the process $\{n^{1/2}(F_{n,*}(t, \vartheta_n) - G(t)) : t \in [-\infty, \infty]\}$ converges in distribution under $\bar{P}_{n,\vartheta_n} = P_{r_n(\vartheta_n),\vartheta_n,G}$ to the same Gaussian process whenever $\langle \vartheta_n \rangle$ is local, and is therefore tight under $P_{\tau,G}$ by contiguity. Finally, a simple argument shows that for every local $\langle\vartheta_n\rangle$,

$$\sup_{t\in\mathbb{R}} |\hat{c}_n(t,\vartheta_n) - c_*(t)| = o_{\bar{P}_{n,\vartheta_n}}(1).$$

From this we can conclude that for every local $\langle\vartheta_n\rangle$,

$$\sup_{t\in\mathbb{R}} n^{1/2}|\hat{F}_n(t,\vartheta_n) - F_{n,*}(t,\vartheta_n)| = o_{\bar{P}_{n,\vartheta_n}}(1).$$

We obtain from the above that

$$\sup_{t\in\mathbb{R}} n^{1/2}|\hat{F}_n(t,\tilde{\vartheta}_n) - F_n(t)| = o_{P_{\tau,G}}(1).$$

Thus $\langle\hat{F}_n(t,\tilde{\vartheta}_n)\rangle$ is also a least dispersed and regular estimator of the infinite dimensional functional $(\vartheta, F) \mapsto F$ at (τ, G). See Schick and Susarla [22] and Bickel et al. [4, Section 5.2] for appropriate infinite-dimensional versions of the convolution theorem.

5. Efficient estimation of the Euclidean parameter

Fix τ in Θ and G in $\mathfrak{F}$, let Assumptions 1 to 3 hold and assume that $V(\tau,G)$ is positive definite. We also assume that a preliminary $n^{1/2}$-consistent estimator of the Euclidean parameter is available. In this section we construct a least dispersed regular estimator of the Euclidean parameter. The functional associated with the Euclidean parameter maps (ϑ, F) to ϑ. This functional is differentiable at (τ, G) with A the $k \times K$ identity matrix and $b = 0$. By characterization (3.3), an estimator $\hat{\vartheta}_n$ of the Euclidean parameter is least dispersed and regular at (τ, G) if and only if

$$\hat{\vartheta}_n = U_n(\tau,G) + o_{P_{\tau,G}}(n^{-1/2}), \tag{5.1}$$

where $U_n(\tau,G)$ is defined in (2.27).

Such estimators have been constructed by Kreiss [15, 16], Drost, Klaassen and Werker [5] and Koul and Schick [14] in special cases under various assumptions. They all require the availability of preliminary $n^{1/2}$-consistent estimators and use Le Cam's discretization technique. We shall follow this approach as well. Our construction does not need additional assumptions and does not require truncation arguments as used in Koul and Schick [14], or sample splitting techniques as used in Koul and Schick [14] and Drost, Klaassen and Werker [5], or symmetry of the innovation density as in Jeganathan [13].

As we will be using a discretized version of the preliminary estimator, it suffices to construct functions u_n from $\Theta \times \mathbb{R}^{n+1}$ into $\mathbb{R}^d$ such that, with $\hat{U}_n(\vartheta) = u_n(\vartheta, Y_0, \ldots, Y_n)$ for $\vartheta \in \Theta$, we have for every local sequence $\langle\vartheta_n\rangle$:

$$\hat{U}_n(\vartheta_n) = U_n(\tau,G) + o_{P_{\tau,G}}(n^{-1/2}). \tag{5.2}$$

From this one immediately obtains that $\hat{\vartheta}_n = \hat{U}_n(\tilde{\vartheta}_n)$ satisfies (5.1) for every discretized $n^{1/2}$-consistent estimator $\tilde{\vartheta}_n$. In view of Remark 2 and contiguity, relation (5.2) is equivalent to

$$\hat{U}_n(\vartheta_n) = \bar{U}_n(\vartheta_n, G) + o_{\bar{P}_{n,\vartheta_n}}(n^{-1/2}), \tag{5.3}$$

with $\bar{U}_n(\vartheta_n, G)$ as in (2.26) and $\bar{P}_{n,\vartheta_n} = P_{r_n(\vartheta_n),\vartheta_n,G}$. This suggests to take

$$\hat{U}_n(\vartheta) = \vartheta + (n - r_n(\vartheta))^{-1} \sum_{j=r_n(\vartheta)+1}^{n} (\hat{V}_n(\vartheta)\hat{I}_n(\vartheta))^{-1}\dot{\xi}_j(\vartheta)\hat{L}_{n,\vartheta}(\xi_j(\vartheta)), \tag{5.4}$$

where

$$\hat{I}_n(\vartheta) = (n - r_n(\vartheta))^{-1} \sum_{j=r_n(\vartheta)+1}^{n} \hat{L}^2_{n,\vartheta}(\xi_j(\vartheta)), \tag{5.5}$$

$$\hat{V}_n(\vartheta) = (n - r_n(\vartheta))^{-1} \sum_{j=r_n(\vartheta)+1}^{n} \dot{\xi}_j(\vartheta)\dot{\xi}_j(\vartheta)^\top, \tag{5.6}$$

and where

$$\hat{L}_{n,\vartheta}(x) = L_{n-r_n(\vartheta)}(x; \xi_{r_n(\vartheta)+1}(\vartheta), \dots, \xi_n(\vartheta)) \tag{5.7}$$

is an estimator of $\ell_G(x)$ constructed from the truncated innovations $\xi_{r_n(\vartheta)+1}(\vartheta), \dots, \xi_n(\vartheta)$. We shall take

$$L_m(x; y_1, \dots, y_m) = \frac{\sum_{j=1}^m a_m^{-2}\rho'(a_m^{-1}(x - y_j))}{mb_m + \sum_{j=1}^m a_m^{-1}\rho(a_m^{-1}(x - y_j))}, \quad x, y_1, \dots, y_m \in \mathbb{R}, \tag{5.8}$$

where a_m and b_m are positive constants and ρ is a symmetric continuously differentiable density with finite second moment and such that $|\rho'(x)| \le Ck(x)$ and $|\rho''(x)| \le C\rho(x)$ for all $x \in \mathbb{R}$ and some constant C. A possible choice is the logistic density.

Theorem 4. *Let $\tilde{\vartheta}_n$ be a preliminary estimator of the Euclidean parameter that is discretized and $n^{1/2}$-consistent at (τ, G). Let $\hat{U}_n(\vartheta)$ be defined as in (5.4) to (5.8) with $a_n \to 0$, $b_n \to 0$ and $na_n^4 b_n^2 \to \infty$. Then $\hat{U}_n(\tilde{\vartheta}_n)$ satisfies (5.1). This estimator is thus a least dispersed and regular estimator of the Euclidean parameter at (τ, G). It is also locally asymptotically minimax adaptive at (τ, G) in the sense of Fabian and Hannan* [6].

To prove this theorem we shall rely on the following lemma.

Lemma 5. *Let $\{\varepsilon_t : t \in \mathbb{Z}\}$ be independent random variables with common distribution function G, and let $c_{m,j}(i)$ with $m, j, i = 1, 2, \dots$ and $j \le m$ be real numbers such that*

$$K = \sup_{m,j} \sum_{i=1}^{\infty} |c_{m,j}(i)| < \infty.$$

Set

$$\zeta_{m,j} = \sum_{i=1}^{\infty} c_{m,j}(i)\varepsilon_{j-i}.$$

Then, as $m \to \infty$,

$$\Delta_m = E\Big[\int L_m(x;\varepsilon_1,\dots,\varepsilon_m) - \ell_G(x))^2\,dG(x)\Big] \to 0, \tag{5.9}$$

$$\frac{1}{m}\sum_{j=1}^m L_m^2(\varepsilon_j;\varepsilon_1,\dots,\varepsilon_m) = I(G) + o_p(1), \tag{5.10}$$

$$m^{-1/2}\sum_{j=1}^m \zeta_{m,j}\Big[L_m(\varepsilon_j;\varepsilon_1,\dots,\varepsilon_m) - \int L_m(x;\varepsilon_1,\dots,\varepsilon_m)\,dG(x) - \ell_G(\varepsilon_j)\Big] = o_p(1). \tag{5.11}$$

Proof. The first two results follow from Schick [20, 3.7. Example 1]. Now let

$$\tilde L_m(x) = L_m(x;\varepsilon_1,\dots,\varepsilon_m) - \int L_m(z;\varepsilon_1,\dots,\varepsilon_m)\,dG(z) - \ell_G(x), \quad x \in \mathbb{R}.$$

For distinct elements $i_1,\dots,i_r$ of $\{1,\dots,m\}$, let $\tilde L_{m,i_1,\dots,i_r}$ be defined as $\tilde L_m$ but with the variables $\varepsilon_{i_1},\dots,\varepsilon_{i_r}$ replaced by $\varepsilon_{m+1},\dots,\varepsilon_{m+r}$. It follows from Schick [20] that

$$|\tilde L_m(x) - \tilde L_{m,i_1,\dots,i_r}(x)| \le r d_m, \qquad \text{where} \quad d_m = 4C\sup_{t\in\mathbb{R}}|\rho(t)| m^{-1} a_m^{-2} b_m^{-1}. \tag{5.12}$$

Thus it is easy to see that (5.11) is equivalent to

$$D_m = m^{-1/2}\sum_{j=1}^m \zeta_{m,j}\tilde L_{m,j}(\varepsilon_j) = o_p(1). \tag{5.13}$$

The second moment of the left-hand side is

$$E[D_m^2] = \frac{1}{m}\sum_{i=1}^m\sum_{j=1}^m\sum_{k=1}^\infty\sum_{l=1}^\infty c_{m,i}(k)c_{m,j}(l)E[\varepsilon_{i-k}\varepsilon_{j-k}\tilde L_{m,i}(\varepsilon_i)\tilde L_{m,j}(\varepsilon_j)].$$

Using (5.12) and the inequality $a^2 \le 2b^2 + 2(a-b)^2$ we get

$$E[\varepsilon_l^2\tilde L_{m,j}^2(\varepsilon_j)] \le 2E[\varepsilon_1^2]\Big(E\Big[\int \tilde L_{m,j,l}^2(x)\,dG(x)\Big] + 4d_m^2\Big) \le 2E[\varepsilon_1^2][\Delta_m + 4d_m^2], \quad l < j \le m,$$

and with the Cauchy–Schwarz inequality,

$$|E[\varepsilon_k\varepsilon_l\tilde L_{m,i}(\varepsilon_i)\tilde L_{m,j}(\varepsilon_j)]| \le 2E[\varepsilon_1^2][\Delta_m + 4d_m^2], \quad k < i \le m,\ l < j \le m.$$

If all the indices are distinct we can do even better. Then, utilizing $\int \tilde L_{m,i_1,\dots,i_r}(x)\,dG(x) = 0$, we get

$$E[\varepsilon_k\varepsilon_l\tilde L_{m,i}(\varepsilon_i)\tilde L_{m,j}(\varepsilon_j)] = E[\varepsilon_k\varepsilon_l(\tilde L_{m,i}(\varepsilon_i) - \tilde L_{m,i,j}(\varepsilon_i))(\tilde L_{m,j}(\varepsilon_j) - \tilde L_{m,j,i}(\varepsilon_j))]$$

and obtain

$$|E[\varepsilon_k\varepsilon_l\tilde L_{m,i}(\varepsilon_i)\tilde L_{m,j}(\varepsilon_j)]| \le 4d_m^2E[\varepsilon_1^2], \quad k < i < j \le m,\ l < j,\ l \ne i.$$

Using these bounds we arrive at

$$E[D_m^2] \le (8K^2md_m^2 + 4K^2\Delta_m)E[\varepsilon_1^2] \to 0.$$

This implies (5.13). □

Proof of Theorem 4. Fix a local sequence $\langle \vartheta_n \rangle$. We need only prove (5.3). Abbreviate $r_n(\vartheta_n)$ by r and $n - r_n(\vartheta_n)$ by m. Set

$$\mathbf{Y}_n = (Y_0, \dots, Y_n), \quad \mathbf{Z}_n = (\xi_{r+1}(\vartheta_n), \dots, \xi_n(\vartheta_n)), \quad \mathbf{X}_n = (X_{r+1}(\vartheta_n), \dots, X_n(\vartheta_n)).$$

Keep in mind that the sequences $\langle \mathfrak{L}(\mathbf{Y}_n \mid P_{r,\vartheta_n,G}) \rangle$, $\langle \mathfrak{L}(\mathbf{Y}_n \mid P_{\vartheta_n,G}) \rangle$ and $\langle \mathfrak{L}(\mathbf{Y}_n \mid P_{\tau,G}) \rangle$ are contiguous.

Since the variables $\xi_{r+1}(\vartheta_n), \dots, \xi_n(\vartheta_n)$ are independent with common distribution function G under $P_{r,\vartheta_n,G}$, it follows from the previous lemma that

$$\hat{I}_n(\vartheta_n) = I(G) + o_{P_{r,\vartheta_n,G}}(1).$$

Using this, (2.20) and contiguity, we see that the desired (5.3) is implied by

$$(n-r)^{-1} \sum_{j=r+1}^{n} \dot{\xi}_j(\vartheta_n)[L_m(\xi_j(\vartheta_n); \mathbf{Z}_n) - \ell_G(\xi_j(\vartheta_n))] = o_{P_{\vartheta_n,G}}(n^{-1/2}). \tag{5.14}$$

Our next goal is to show that (5.14) is equivalent to the following version in which we have replaced truncated residuals by actual residuals:

$$(n-r)^{-1} \sum_{j=r+1}^{n} \dot{\xi}_j(\vartheta_n)[L_m(X_j(\vartheta_n); \mathbf{X}_n) - \ell_G(X_j(\vartheta_n)] = o_{P_{\vartheta_n,G}}(n^{-1/2}). \tag{5.15}$$

For this we need the bound

$$\sum_{j=r+1}^{n} |L_m(x; \mathbf{Z}_n) - L_m(x; \mathbf{X}_n)|^2 \le B(a_m^{-4} + a_m^{-5} b_m^{-1}) \sum_{j=r+1}^{n} |\xi_j(\vartheta_n) - X_j(\vartheta_n)|^2$$

for some constant B, which follows from inequalities (L1) and (L3) in Schick [21]. Now use (2.23), (2.9), Assumption 3 and the Cauchy–Schwarz inequality to conclude the equivalence of (5.14) and (5.15). Using the identity $\dot{\xi}_j(\vartheta_n) = \sum_{i=0}^{j-1} \dot{\beta}_{j-i}(\vartheta_n) Y_i$ and (2.4) we find by straightforward calculations that

$$E_{\vartheta_n,G}\Big(\Big\| \sum_{j=r+1}^{n} \dot{\xi}_j(\vartheta_n) \Big\|^2\Big) \le 2n \Big(\sum_{i=1}^{\infty} \|\dot{\beta}_i(\vartheta_n)\| \Big)^2 \int x^2 \, dG(x) \Big(\sum_{j=0}^{\infty} |\alpha_j(\vartheta_n)| \Big)^2 = O(1). \tag{5.16}$$

Since the innovations $X_1(\vartheta_n), \dots, X_n(\vartheta_n)$ are independent with common distribution function G under $P_{\vartheta_n,G}$, we obtain from (5.9) that $\int L_m(x; \mathbf{X}_n) \, dG(x) = o_{P_{\vartheta_n,G}}(1)$. This and (5.16) show that (5.15) is equivalent to

$$(n-r)^{-1} \sum_{j=r+1}^{n} \dot{\xi}_j(\vartheta_n) \Big[L_m(X_j(\vartheta_n); \mathbf{X}_n) - \int L_m(x; \mathbf{X}_n) \, dG(x) - \ell_G(X_j(\vartheta_n)) \Big] = o_{P_{\vartheta_n,G}}(n^{-1/2}).$$

But this follows by applying (5.11) to the components of the left-hand side. This application is justified by the fact that

$$\dot{\xi}_j(\vartheta_n) = \sum_{s=1}^{\infty} \delta_{j,s}(\vartheta_n) X_{j-s}(\vartheta_n) \quad \text{with} \quad \delta_{j,s}(\vartheta_n) = \sum_{i=0}^{s \wedge j} \dot{\beta}_i(\vartheta_n) \alpha_{s-i}(\vartheta_n) \tag{5.17}$$

under $P_{\vartheta_n,G}$ and that

$$\sum_{s=1}^{\infty} \|\delta_{j,s}(\vartheta_n)\| \le \sum_{i=1}^{\infty} \|\dot\beta_i(\vartheta_n)\| \sum_{t=1}^{\infty} |\alpha_t(\vartheta_n)|. \tag{5.18}$$

6. Estimating moments of the stationary distribution

In this section we show how the results of the previous sections can be used to construct least dispersed regular estimators of moments of the stationary distribution. Note that the first moment of the stationary distribution is zero. Thus we focus on estimating the m-th moment for some integer m greater than one. We choose $\mathfrak{F}$ as in Section 2, with $\psi(x) = x^m$. Then the error distributions have finite $2m$-th moments. We shall write $\mu_\nu(F)$ for the ν-th moment of a distribution F in $\mathfrak{F}$, $\nu = 1, \dots, m$. We shall also require that the functions A_ν from Θ into $\mathbb{R}$ defined by

$$A_\nu(\vartheta) = \sum_{j=0}^{\infty} \alpha_j^\nu(\vartheta), \quad \vartheta \in \Theta,$$

are differentiable for $\nu = 2, \dots, m$.

The functional associated with the m-th moment of the stationary distribution is

$$\kappa(\vartheta, F) = E_{\vartheta,F}(Y_1^m), \quad \vartheta \in \Theta, F \in \mathfrak{F}.$$

It follows from the moving average representation (2.2) that

$$\kappa(\vartheta, F) = \chi(A_2(\vartheta), \dots, A_m(\vartheta), \mu_2(F), \dots, \mu_m(F)),$$

for some continuously differentiable function χ from $\mathbb{R}^{2m-2}$ to $\mathbb{R}$. For example, for $m = 2$ we have

$$\kappa(\vartheta, F) = A_2(\vartheta)\mu_2(F), \quad \vartheta \in \Theta, F \in \mathfrak{F},$$

while for $m = 4$ we have

$$\kappa(\vartheta, F) = A_4(\vartheta)(\mu_4(F) - 3\mu_2(F)^2) + 3(A_2(\vartheta)\mu_2(F))^2, \quad \vartheta \in \Theta, F \in \mathfrak{F}.$$

We shall also assume that we have at our disposal a least dispersed regular estimator $\hat\vartheta_n$ of the Euclidean parameter, and consequently also a discretized $n^{1/2}$-consistent estimator $\tilde\vartheta_n$ of the Euclidean parameter. We have already seen in Example 1 that a least dispersed regular estimators of the ν-th moment of the error distribution is given by

$$\hat\mu^*_{n,\nu} = \hat\mu_{n,\nu} - \frac{\hat\mu_{n,\nu+1}}{\hat\mu_{n,2}} \hat\mu_{n,1}, \quad \nu = 2, \dots, m,$$

where $\hat\mu_{n,\nu}$ is defined in (4.2). In view of the differentiability of the functions A_ν, a least dispersed regular estimators of A_ν is given by $A_\nu(\hat\vartheta_n)$, $\nu = 2, \dots, n$. By the differentiability of χ we obtain now that

$$\chi(A_2(\hat\vartheta_n), \dots, A_m(\hat\vartheta_n), \hat\mu^*_{n,2}, \dots, \hat\mu^*_{n,m})$$

is a least dispersed regular estimator of κ.

References

[1] M. Akahira, *On the asymptotic efficiency of estimators in an autoregressive process*, Ann. Inst. Statist. Math., 28 (1976), 35–48.

[2] M.G. Akritas and R.A. Johnson, *Efficiencies of tests and estimators for p-order autoregressive processes when the error distribution is non-normal*, Ann. Inst. Statist. Math., 34 (1982), 579–589.

[3] P.J. Bickel and D. Freedman, *Some asymptotic theory for the bootstrap*, Ann. Statist., 9 (1981), 1196-1217.

[4] P.J. Bickel, C.A.J. Klaassen, Y. Ritov and J.A. Wellner, *Efficient and Adaptive Estimation for Semiparametric Models*, Springer, New York 1998.

[5] F.C. Drost, C.A.J. Klaassen and B.J.M. Werker, *Adaptive estimation in time-series models*, Ann. Statist., 25 (1997), 786–817.

[6] V. Fabian and J. Hannan, *On estimation and adaptive estimation for locally asymptotically normal families*, Z. Wahrsch. Verw. Gebiete, 59 (1982), 459–478.

[7] B.B. Garel, *The asymptotic distribution of the likelihood ratio for M.A. processes with a regression trend*, Statist. Decisions, 7 (1989, 167–184.

[8] B.B. Garel and M. Hallin, *Local asymptotic normality of multivariate ARMA processes with a linear trend*, Ann. Inst. Statist. Math., 47 (1995), 147–169.

[9] M. Hallin and M.L. Puri, *Aligned rank tests for linear models with autocorrelated error terms*, J. Multivariate Anal., 50 (1994), 175–237.

[10] W.-M. Huang, *A characterization of limiting distributions of estimators in an autoregressive process*, Ann. Inst. Statist. Math., 38 (1986), 137–144.

[11] S.Y. Hwang and I.V. Basawa, *Asymptotic optimal inference for a class of nonlinear time series models*, Stochastic Process. Appl., 46 (1993), 91–113.

[12] I.A. Ibragimov add R.Z. Has'minskii, *Statistical Estimation. Asymptotic Theory*, Springer, New York 1981.

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

[14] H.L. Koul and A. Schick, *Efficient estimation in nonlinear autoregressive time series models*, Bernoulli, 3 (1997), 247–277.

[15] J.-P. Kreiss, *On adaptive estimation in stationary ARMA processes*, Ann. Statist., 15 (1987), 112–133.

[16] J.-P. Kreiss, *On adaptive estimation in autoregressive models when there are nuisance functions*, Statist. Decisions, 5 (1987), 59–76.

[17] J.-P. Kreiss, *Asymptotic Statistical Inference for a Class of Stochastic Processes*, Habilitationsschrift, Fachbereich Mathematik, Universität Hamburg, 1988.

[18] J.-P. Kreiss, *Local asymptotic normality for autoregression with infinite order*, J. Statist. Plann. Inference, 26 (1990), 185–219.

[19] J.-P. Kreiss, *Estimation of the distribution function of noise in stationary processes*, Metrika, 38 (1991), 285–297.

[20] A. Schick, *A note on the construction of asymptotically linear estimators*, J. Statist. Plann. Inference, 16 (1987), 89–105, Correction: 22 (1989), 269–270.

[21] A.Schick, *On efficient estimation in regression models*, Ann. Statist., 21 (1993), 1486–1521, Correction: 23 (1995), 1862-1863.

[22] A. Schick and V. Susarla, *An infinite dimensional convolution theorem with applications to random censoring and missing data models*, J. Statist. Plann. Inference, 24 (1990), 13–23.

[23] A. Schick and W. Wefelmeyer, *Estimating invariant laws of linear processes by U-statistics*, Technical Report, Department of Mathematical Sciences, Binghamton University, 2000.

[24] A. Schick and W. Wefelmeyer, *Estimating the innovation distribution in nonlinear autoregressive models*, To appear in: Ann. Inst. Statist. Math., 53 (2002).

[25] A.R. Swensen, *The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend*, J. Multivariate Anal., 16 (1985), 54–70.

[26] W. Wefelmeyer, *An efficient estimator for the expectation of a bounded function under the residual distribution of an autoregressive process*, Ann. Inst. Statist. Math., 46 (1994), 309–315.

Anton Schick, Binghamton University, Department of Mathematical Sciences, Binghamton, NY 13902-6000, USA
E-mail address: anton@math.binghamton.edu
URL: http://math.binghamton.edu/anton/index.html

Wolfgang Wefelmeyer, Universität Siegen, Fachbereich 6 Mathematik, Walter-Flex-Str. 3, 57068 Siegen, Germany
E-mail address: wefelmeyer@mathematik.uni-siegen.de
URL: http://www.math.uni-siegen.de/statistik/wefelmeyer.html