Prediction in invertible linear processes

Anton Schick *
Binghamton University

Wolfgang Wefelmeyer
Universität zu Köln

Abstract

We construct root-n consistent plug-in estimators for conditional expectations of the form $E(h(X_{n+1}, \ldots, X_{n+m})|X_1, \ldots, X_n)$ in invertible linear processes. More specifically, we prove a Bahadur type representation for such estimators, uniformly over certain classes of not necessarily bounded functions h. We obtain in particular a uniformly root-n consistent estimator for the m-dimensional conditional distribution function. The proof uses empirical process techniques.

Keywords. Von Mises statistic, kernel smoothed empirical process, residual-based kernel density estimator, stochastic expansion, infinite-order moving average process, infinite-order autoregressive process.

1 Introduction

Let $X_1, \ldots, X_n$ be observations from a real-valued stationary time series. Let m be a positive integer and h a measurable function on $\mathbb{R}^m$ such that $E[h^2(X_{n+1}, \ldots, X_{n+m})]$ is finite. The best predictor for $h(X_{n+1}, \ldots, X_{n+m})$ is the conditional expectation

$$q(h) = E(h(X_{n+1}, \ldots, X_{n+m})|X_1, \ldots, X_n).$$

Convergence rates for kernel estimators of $E(h(X_{n+1}, \ldots, X_{n+m})|X_{n-r+1} = x_1, \ldots, X_n = x_r)$ for fixed $x_1, \ldots, x_r$ and fixed r are e.g. in Roussas (1969, 1991), Robinson (1983, 1986), Yakowitz (1985, 1987), Masry (1989), Roussas and Tran (1992), Tran (1992), and Truong and Stone (1992).

If the time series is driven by independent innovations, one can construct estimators for conditional expectations that converge at the "parametric" root-n rate. For nonlinear autoregression see Müller et al. (2006). For the MA(1) model $X_t = \varepsilon_t - \vartheta\varepsilon_{t-1}$ with $|\vartheta| < 1$ and innovations ε_t, $t \in \mathbb{Z}$, that are i.i.d. with finite variance, Schick and Wefelmeyer (2006b) construct root-n consistent estimators for the random variable $q(h)$ when $m = 1$. We generalize their result to arbitrary invertible linear processes and to arbitrary m.

*Supported in part by NSF Grant DMS 0405791.

2 Result

Consider a real-valued stationary linear process with infinite-order moving average representation

$$X_t = \varepsilon_t + \sum_{s=1}^{\infty} \varphi_s \varepsilon_{t-s}, \quad t \in \mathbb{Z}, \tag{2.1}$$

with i.i.d. innovations ε_t, $t \in \mathbb{Z}$, that have mean zero, finite variance, and density f. Let F denote the corresponding distribution function. Assume that the characteristic series $\varphi(z) = 1 + \sum_{s=1}^{\infty} \varphi_s z^s$ is bounded and bounded away from zero on the complex unit disk D. Then $\varrho(z) = 1/\varphi(z) = 1 + \sum_{s=1}^{\infty} \varrho_s z^s$ is also bounded and bounded away from zero on D. Hence the innovations have the infinite-order moving average representation

$$\varepsilon_t = X_t + \sum_{s=1}^{\infty} \varrho_s X_{t-s}, \quad t \in \mathbb{Z}, \tag{2.2}$$

which is an infinite-order autoregressive representation for the process X_t, $t \in \mathbb{Z}$.

First we derive a tractable approximation of the conditional expectation $q(h)$ defined in the Introduction. Set $\varphi_0 = \varrho_0 = 1$. The backshift operator B is defined by $BX_t = X_{t-1}$. For $k = 1, 2, \dots$ we decompose the representation (2.1) as

$$X_{n+k} = \varphi(B)\varepsilon_{n+k} = \varphi_{<k}(B)\varepsilon_{n+k} + \varphi_{\geq k}(B)\varepsilon_{n+k} \tag{2.3}$$

with

$$\varphi_{<k}(z) = \sum_{s=0}^{k-1} \varphi_s z^s, \qquad \varphi_{\geq k}(z) = \sum_{s=k}^{\infty} \varphi_s z^s.$$

Using the representation (2.2) we obtain

$$\varphi_{\geq k}(B)\varepsilon_{n+k} = \varphi_{\geq k}(B)\varrho(B)X_{n+k} = \sum_{s=k}^{\infty}\sum_{t=0}^{\infty} \varphi_s \varrho_t X_{n+k-s-t} = \sum_{t=0}^{\infty} c_{tk} X_{n-t} \tag{2.4}$$

with

$$c_{tk} = \sum_{s=0}^{t} \varphi_{k+s} \varrho_{t-s}.$$

Fix $m \in \mathbb{N}$. Introduce the vectors

$$\mathbf{X}_{n+1} = (X_{n+1}, \dots, X_{n+m})^\top, \qquad \boldsymbol{\varepsilon}_{n+1} = (\varepsilon_{n+1}, \dots, \varepsilon_{n+m})^\top,$$
$$\mathbf{c}_t = (c_{t1}, \dots, c_{tm})^\top, \qquad \boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_{m-1})^\top$$

and the $m \times m$ matrix of moving average coefficients

$$M_{\boldsymbol{\varphi}} = \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ \varphi_1 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \vdots \\ \varphi_{m-1} & \cdots & \cdots & \varphi_1 & 1 \end{pmatrix}.$$

From (2.3) and (2.4) we obtain the decomposition

$$\mathbf{X}_{n+1} = M_{\boldsymbol{\varphi}}\, \boldsymbol{\varepsilon}_{n+1} + \mathbf{Z}_n$$

with

$$\mathbf{Z}_n = \sum_{t=0}^{\infty} \mathbf{c}_t X_{n-t}.$$

Since $\boldsymbol{\varepsilon}_{n+1}$ is independent of $X_1, \dots, X_n$, we can write

$$q(h) = E(q_h(\boldsymbol{\varphi}, \mathbf{Z}_n) | X_1, \dots, X_n)$$

with

$$q_h(\boldsymbol{\varphi}, \mathbf{z}) = \int h(M_{\boldsymbol{\varphi}}\, \mathbf{y} + \mathbf{z})\, dF_m(\mathbf{y})$$

and $F_m(\mathbf{y}) = F(y_1) \cdots F(y_m)$ the distribution function of $\boldsymbol{\varepsilon}_{n+1}$. Decompose $\mathbf{Z}_n = \mathbf{Z}_{nr} + \mathbf{R}_n$ with

$$\mathbf{Z}_{nr} = \sum_{t=0}^{r} \mathbf{c}_t X_{n-t}, \qquad \mathbf{R}_n = \sum_{t=r+1}^{\infty} \mathbf{c}_t X_{n-t}.$$

If q_h is Lipschitz with constant L, we have

$$E[(q(h) - q_h(\boldsymbol{\varphi}, \mathbf{Z}_{nr}))^2] \le L^2 E[\|\mathbf{R}_n\|^2].$$

We will choose $r = r_n$ increasing so that the right-hand side is $o(n^{-1})$. Then we arrive at the desired approximation,

$$q(h) = q_h(\boldsymbol{\varphi}, \mathbf{Z}_{nr}) + o_p(n^{-1/2}). \tag{2.5}$$

We can now construct an estimator for the conditional expectation $q(h)$ via the approximation (2.5) as follows. Let $\hat{\varphi}_1, \hat{\varphi}_2, \dots$ be estimators for the moving average coefficients $\varphi_1, \varphi_2, \dots$. Let $\hat{\varrho}_1, \hat{\varrho}_2, \dots$ be estimators for the autoregression coefficients $\varrho_1, \varrho_2, \dots$. Set

$$\hat{c}_{tk} = \sum_{s=0}^{t} \hat{\varphi}_{k+s} \hat{\varrho}_{t-s}, \qquad \hat{\mathbf{c}}_t = (\hat{c}_{t1}, \dots, \hat{c}_{tm})^\top,$$
$$\hat{\mathbf{Z}}_{nr} = \sum_{t=0}^{r} \hat{\mathbf{c}}_t X_{n-t}, \qquad \hat{\boldsymbol{\varphi}} = (\hat{\varphi}_1, \dots, \hat{\varphi}_{m-1})^\top.$$

We choose $p = p_n$ with $p/n \to 0$ and estimate the innovation ε_j by the residual

$$\hat{\varepsilon}_j = X_j + \sum_{s=1}^{p} \hat{\varrho}_s X_{j-s}, \quad j = p+1, \dots, n.$$

Introduce the residual-based and the innovation-based empirical distribution functions as

$$\hat{\mathbb{F}}(y) = \frac{1}{n-p} \sum_{j=p+1}^{n} \mathbf{1}[\hat{\varepsilon}_j \le y], \qquad \mathbb{F}(y) = \frac{1}{n-p} \sum_{j=p+1}^{n} \mathbf{1}[\varepsilon_j \le y].$$

Set $\hat{\mathbb{F}}_m(\mathbf{y}) = \hat{\mathbb{F}}(y_1)\cdots\hat{\mathbb{F}}(y_m)$. Then an estimator for $q(h)$ is

$$\hat{q}(h) = \int h(M_{\hat{\boldsymbol{\varphi}}}\,\mathbf{y} + \hat{\mathbf{Z}}_{nr})\, d\hat{\mathbb{F}}_m(\mathbf{y}).$$

In order to prove root-n consistency of the estimator $\hat{q}(h)$, we derive a Bahadur type representation for it. We do this first heuristically, for a fixed and smooth function h. An expansion of the product $\hat{\mathbb{F}}_m(\mathbf{y})$ gives

$$\hat{\mathbb{F}}_m(\mathbf{y}) - F_m(\mathbf{y}) = \sum_{k=1}^{m}(\hat{\mathbb{F}}(y_k) - F(y_k))\prod_{i\neq k} F(y_i) + o_p(n^{-1/2}). \tag{2.6}$$

Since $\hat{\varepsilon}_j - \varepsilon_j$ is approximated by $\sum_{s=1}^{p}(\hat{\varrho}_s - \varrho_s)X_{j-s}$ with $E[X_{j-s}] = 0$, the residual-based empirical distribution function $\hat{\mathbb{F}}$ is asymptotically equivalent to the distribution function $\mathbb{F}$ based on the true innovations. Hence we obtain from (2.6) the expansion

$$\hat{q}(h) = \frac{1}{n-p}\sum_{j=p+1}^{n}\overline{h}(\varepsilon_j, \hat{\boldsymbol{\varphi}}, \hat{\mathbf{Z}}_{nr}) + o_p(n^{-1/2})$$

with

$$\overline{h}(y, \boldsymbol{\varphi}, \mathbf{z}) = \sum_{k=1}^{m} T_k h(y, \boldsymbol{\varphi}, \mathbf{z})$$

and

$$T_k h(y, \boldsymbol{\varphi}, \mathbf{z}) = E(h(M_{\boldsymbol{\varphi}}\,\boldsymbol{\varepsilon} + \mathbf{z})|\varepsilon_k = y),$$

where $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_m)^\top$. Let $\overline{h}^{(1)}(y, \boldsymbol{\varphi}, \mathbf{z})$ and $\overline{h}^{(2)}(y, \boldsymbol{\varphi}, \mathbf{z})$ denote the gradients of $\overline{h}(y, \boldsymbol{\varphi}, \mathbf{z})$ as functions of $\boldsymbol{\varphi}$ and $\mathbf{z}$, respectively. By Taylor expansion, we arrive at the Bahadur type representation

$$\begin{aligned}\hat{q}(h) = \frac{1}{n-p}\sum_{j=p+1}^{n}\overline{h}(\varepsilon_j, \boldsymbol{\varphi}, \mathbf{Z}_{nr}) &+ (\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi})^\top \int \overline{h}^{(1)}(y, \boldsymbol{\varphi}, \mathbf{Z}_{nr})\, dF(y)\\ &+ (\hat{\mathbf{Z}}_{nr} - \mathbf{Z}_{nr})^\top \int \overline{h}^{(2)}(y, \boldsymbol{\varphi}, \mathbf{Z}_{nr})\, dF(y) + o_p(n^{-1/2}).\end{aligned}$$

This shows that $\hat{q}(h)$ is root-n consistent for appropriate choices of $\hat{\boldsymbol{\varphi}}$ and $\hat{\mathbf{c}}_t$.

Let $q_h^{(1)}(\boldsymbol{\varphi}, \mathbf{z})$ and $q_h^{(2)}(\boldsymbol{\varphi}, \mathbf{z})$ denote the gradients of $q_h(\boldsymbol{\varphi}, \mathbf{z})$ as a function of $\boldsymbol{\varphi}$ and $\mathbf{z}$, respectively. Taking derivatives under the integral, we have

$$q_h^{(i)}(\boldsymbol{\varphi}, \mathbf{z}) = \int \overline{h}^{(i)}(y, \boldsymbol{\varphi}, \mathbf{z})\, dF(y), \quad i = 1, 2.$$

For non-smooth h, the derivatives $\overline{h}^{(1)}$ and $\overline{h}^{(2)}$ may no longer exist and we may have to replace the integrals on the right-hand side in the stochastic expansion of $\hat{q}(h)$ by $q_h^{(1)}(\boldsymbol{\varphi}, \mathbf{Z}_{nr})$ and $q_h^{(2)}(\boldsymbol{\varphi}, \mathbf{Z}_{nr})$, the existence of which can be guaranteed by smoothness on f.

In order to cover estimation of $\mathbf{t} \mapsto P(\mathbf{X}_n \le \mathbf{t}|X_1, \dots, X_n)$, we prove root-$n$ consistency uniformly over large classes of not necessarily smooth functions h. Then it is convenient to work instead with a smoothed version of $\hat{q}(h)$,

$$\hat{q}_s(h) = \int h(M_{\hat{\boldsymbol{\varphi}}}\mathbf{y} + \hat{\mathbf{Z}}_{nr})\hat{f}_m(\mathbf{y})\,d\mathbf{y}.$$

Here $\hat{f}_m(\mathbf{y}) = \hat{f}(y_1)\cdots\hat{f}(y_m)$, and $\hat{f}$ is the residual-based kernel estimator for the innovation density f given by

$$\hat{f}(y) = \frac{1}{n-p}\sum_{j=p+1}^{n} K_b(y - \hat{\varepsilon}_j),$$

where $K_b(y) = K(y/b)/b$ with K a kernel and b a bandwidth.

In order to cover estimation of conditional moments and absolute moments, we must consider unbounded functions h. We therefore use a weighted version of the L_1 norm as follows. Let V denote the function defined by

$$V(y) = (1 + |y|)^{\gamma}, \quad y \in \mathbb{R},$$

for some non-negative γ and set

$$W(\mathbf{y}) = V(\|\mathbf{y}\|) = (1 + \|\mathbf{y}\|)^{\gamma}, \quad \mathbf{y} \in \mathbb{R}^m.$$

The V-*norm* of a measurable function g on $\mathbb{R}$ is $\|g\|_V = \int V(x)|g(x)|\,dx$.

The stochastic expansion of $\hat{q}(h)$ will be shown to be uniform over h in a class $\mathscr{H}$ of measurable functions on $\mathbb{R}^m$ with the following properties.

(H) *The class $\mathscr{H}$ has envelope cW for some positive c. There is a positive α such that, for all $k = 1, \dots, m$ and all (large) C, the class*

$$\mathscr{H}_{k,C} = \{T_k h(\cdot, \boldsymbol{\psi}, \mathbf{z}) : h \in \mathscr{H}, \|\boldsymbol{\psi} - \boldsymbol{\varphi}\| \le \alpha, \|\mathbf{z}\| \le C\}$$

is F-Donsker. Finally,

$$\lim_{\mathbf{s}\to 0, \mathbf{t}\to 0} \sup_{h\in\mathscr{H}} \sup_{\|\mathbf{z}\|\le C} \int |T_k h(y, \boldsymbol{\varphi} + \mathbf{s}, \mathbf{z} + \mathbf{t}) - T_k h(y, \boldsymbol{\varphi}, \mathbf{z})|^2 f(y)\,dy = 0. \tag{2.7}$$

We impose the following assumptions on the density f. Recall that γ is the exponent in the definition of V.

(F) *The density f has mean zero and a finite moment of order β with $\beta \ge \max\{4, 2+2\gamma)$ and is absolutely continuous with an (almost everywhere) derivative f' that satisfies $\|f'\|_V < \infty$ and is V-Lipschitz, which means that there is a constant L such that*

$$\int V(x)|f'(x+t) - f'(x)|\,dx \le LV(t)|t|, \quad t \in \mathbb{R}. \tag{2.8}$$

These assumptions on f have the following implications. If follows from (2.8) that

$$\|f * K_b - f\|_V = O(b^2) \tag{2.9}$$

for any symmetric density K with $\int u^2 V(u)K(u)\,du$ finite. This is stated as Lemma 1(3) in Schick and Wefelmeyer (2006c) and follows from Lemma 6 in Schick and Wefelmeyer (2007). Since the transformation $\mathbf{y} \mapsto M_{\boldsymbol{\psi}}\mathbf{y} + \mathbf{z}$ has Jacobian 1, the random vector $M_{\boldsymbol{\psi}}\boldsymbol{\varepsilon} + \mathbf{z}$ has density $f_{\boldsymbol{\psi},\mathbf{z}}$ given by

$$f_{\boldsymbol{\psi},\mathbf{z}}(\mathbf{y}) = f_m(M_{\boldsymbol{\psi}}^{-1}(\mathbf{y} - \mathbf{z})), \quad \mathbf{y} \in \mathbb{R}^m.$$

One can now show by a standard argument that for every finite constant C,

$$\sup_{\|\boldsymbol{\psi}\|+\|\mathbf{z}\|\le C} \int W(\mathbf{y})\big|f_{\boldsymbol{\psi}+\mathbf{s},\mathbf{z}+\mathbf{t}}(\mathbf{y}) - f_{\boldsymbol{\psi},\mathbf{z}}(\mathbf{y}) - \mathbf{s}^\top \chi^{(1)}_{\boldsymbol{\psi},\mathbf{z}}(\mathbf{y}) - \mathbf{t}^\top \chi^{(2)}_{\boldsymbol{\psi},\mathbf{z}}(\mathbf{y})\big|\,d\mathbf{y} = o(\|\mathbf{s}\| + \|\mathbf{t}\|),$$

where $\chi^{(1)}_{\boldsymbol{\psi},\mathbf{z}}(\mathbf{y})$ and $\chi^{(2)}_{\boldsymbol{\psi},\mathbf{z}}(\mathbf{y})$ are the gradients of $\chi_{\boldsymbol{\psi},\mathbf{z}}$ as functions of $\boldsymbol{\psi}$ and $\mathbf{z}$, respectively (which exist for almost all $\mathbf{y}$). Since $q_h(\boldsymbol{\psi}, \mathbf{z})$ equals $\int h(\mathbf{y}) f_{\boldsymbol{\psi},\mathbf{z}}(\mathbf{y})\,d\mathbf{y}$ and $\mathscr{H}$ has envelope cW, we immediately see that the map q_h is uniformly differentiable in the following sense:

$$\sup_{h\in\mathscr{H},\|\mathbf{z}\|\le C} \Big|q_h(\boldsymbol{\varphi} + \mathbf{s}, \mathbf{z} + \mathbf{t}) - q_h(\boldsymbol{\varphi}, \mathbf{z}) - \mathbf{s}^\top q_h^{(1)}(\boldsymbol{\varphi}, \mathbf{z}) + \mathbf{t}^\top q_h^{(2)}(\boldsymbol{\varphi}, \mathbf{z})\Big| = o(\|\mathbf{s}\| + \|\mathbf{t}\|) \tag{2.10}$$

for every finite C, with $q_h^{(i)}(\boldsymbol{\varphi}, \mathbf{z}) = \int h(\mathbf{y})\chi^{(i)}_{\boldsymbol{\psi},\mathbf{z}}(\mathbf{y})\,d\mathbf{y}$.

Finally, we use the following conditions which were used in part by Schick and Wefelmeyer (2006a).

(Q) *The autoregression coefficients fulfill* $\sum_{s>p} |\rho_s| = O(n^{-1/2-\zeta})$ *for some* $\zeta > 0$.

(R) *The estimators* $\hat{\varrho}_i$ *of the autoregression coefficients* ϱ_i *fulfill*

$$\sum_{i=1}^{p} (\hat{\varrho}_i - \varrho_i)^2 = O_p(qn^{-1})$$

for some $q = q_n$ *with* $1 \le q \le p$.

(K) *The kernel* K *is a three times differentiable symmetric density with compact support.*

(B) *The bandwidth* $b = b_n$ *satisfies* $b_n \sim (n \log n)^{-1/4}$.

We can now state our result.

Theorem 1. *Suppose (F), (H), (Q), (R), (K) and (B) hold and* $p^6 q^6 n^{-1} \log^3 n \to 0$. *Let* $n^{1/2}(\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi}) = O_p(1)$ *and* $n^{1/2}(\hat{\mathbf{Z}}_{nr} - \mathbf{Z}_{nr}) = O_p(1)$. *Then*

$$\sup_{h\in\mathscr{H}} \Big|\hat{q}_s(h) - \frac{1}{n-p}\sum_{j=p+1}^{n} \overline{h}(\varepsilon_j, \boldsymbol{\varphi}, \mathbf{Z}_{nr}) - (\hat{\boldsymbol{\varphi}} - \boldsymbol{\varphi})^\top q_h^{(1)}(\boldsymbol{\varphi}, \mathbf{Z}_{nr}) - (\hat{\mathbf{Z}}_{nr} - \mathbf{Z}_{nr})^\top q_h^{(2)}(\boldsymbol{\varphi}, \mathbf{Z}_{nr})\Big| = o_p(n^{-1/2}).$$

Let us consider applications and special cases. The simplest case is $m = 1$. Then the conditional expectation to be estimated is $q(h) = E(h(X_{n+1})|X_1, \dots, X_n)$. We have $X_{n+1} = \varepsilon_{n+1} + Z_n$ with

$$Z_n = \sum_{t=0}^{\infty} c_t X_{n-t}, \qquad c_t = \sum_{s=0}^{t} \varphi_{1+s} \varrho_{t-s},$$

and $q_h(z) = \int h(y + z)\, dF(y)$. Our estimator for $q(h)$ is

$$\hat{q}_s(h) = \int h(y + \hat{Z}_{nr}) \hat{f}(y)\, dy$$

with

$$\hat{Z}_{nr} = \sum_{t=0}^{r} \hat{c}_t X_{n-t}, \qquad \hat{c}_t = \sum_{s=0}^{t} \hat{\varphi}_{1+s} \hat{\varrho}_{t-s}.$$

Here $\overline{h}(y, \boldsymbol{\varphi}, z) = h(y + z)$, and we obtain the stochastic expansion

$$\hat{q}_s(h) = \frac{1}{n-p} \sum_{j=p+1}^{n} h(\varepsilon_j + Z_{nr}) + (\hat{Z}_{nr} - Z_{nr}) q_h'(Z_{nr}) + o_p(n^{-1/2}),$$

with q_h' the derivative of q_h.

In particular, for $h_t(y) = \mathbf{1}[y \le t]$, the conditional expectation $q(h_t)$ is the conditional distribution function $q(t) = P(X_{n+1} \le t|X_1, \dots, X_n)$. Let G denote the distribution function of the kernel K. Then

$$\hat{\mathbb{F}}_s(t) = \int_{-\infty}^{t} \hat{f}(y)\, dy = \frac{1}{n-p} \sum_{j=p+1}^{n} G\Big(\frac{t - \hat{\varepsilon}_j}{b}\Big)$$

defines the distribution function of $\hat{f}$. Note that $\hat{\mathbb{F}}_s$ is a smoothed version of $\hat{\mathbb{F}}$. Our estimator for $q(t)$ is $\hat{q}_s(t) = \hat{\mathbb{F}}_s(t - \hat{Z}_{nr})$, and its stochastic expansion is

$$\hat{q}_s(t) = \frac{1}{n-p} \sum_{j=p+1}^{n} \mathbf{1}[\varepsilon_j \le t - Z_{nr}] - (\hat{Z}_{nr} - Z_{nr}) f(t - Z_{nr}) + o_p(n^{-1/2})$$

uniformly for $t \in \mathbb{R}$.

For $u \in (0, 1)$, an estimator for the conditional u-quantile of X_{n+1} given $X_1, \dots, X_n$ is the u-quantile $\hat{\mathbb{F}}_s^{-1}(u) + \hat{Z}_{nr}$ of $t \mapsto \hat{\mathbb{F}}_s(t - \hat{Z}_{nr})$. By Gill (1989) the quantile function is compactly differentiable, and we obtain the stochastic expansion

$$\hat{\mathbb{F}}_s^{-1}(u) = F^{-1}(u) - \frac{1}{f(F^{-1}(u))} \frac{1}{n-p} \sum_{j=p+1}^{n} (\mathbf{1}[\varepsilon_j \le F^{-1}(u)] - u) + o_p(n^{-1/2})$$

uniformly for $0 < a \le u \le b < 1$.

For $m = 2$ we have

$$q_h(\varphi_1, \mathbf{z}) = \iint h(y_1 + z_1, \varphi_1 y_1 + y_2 + z_2)\, dF(y_1) dF(y_2),$$

and our estimator is the smoothed von Mises statistic

$$\hat{q}_s(h) = \iint h(y_1 + \hat{Z}_{nr1}, \hat{\varphi}_1 y_1 + y_2 + \hat{Z}_{nr2}) \hat{f}(y_1) \hat{f}(y_2)\, dy_1\, dy_2.$$

The stochastic expansion of $\hat{q}_s(h)$ holds with

$$T_1 h(y, \varphi_1, \mathbf{z}) = E[h(y + z_1, \varphi_1 y + \varepsilon + z_2)],$$
$$T_2 h(y, \varphi_1, \mathbf{z}) = E[h(\varepsilon + z_1, \varphi_1 \varepsilon + y + z_2)].$$

For arbitrary m and $h_{\mathbf{t}}(\mathbf{y}) = \mathbf{1}[\mathbf{y} \le \mathbf{t}]$ the conditional expectation is the m-dimensional conditional distribution function $q(\mathbf{t}) = P(\mathbf{X}_n \le \mathbf{t} | X_1, \dots, X_n)$, and our estimator is

$$\hat{q}_s(\mathbf{t}) = \hat{\mathbb{F}}_{sm}(M_{\hat{\boldsymbol{\varphi}}}^{-1}(\mathbf{t} - \hat{\mathbf{Z}}_{nr}))$$

with $\hat{\mathbb{F}}_{sm}(\mathbf{y}) = \hat{\mathbb{F}}_s(y_1) \cdots \hat{\mathbb{F}}_s(y_m)$. The stochastic expansion of $\hat{q}_s(\mathbf{t})$ holds with

$$T_k h_{\mathbf{t}}(y, \boldsymbol{\varphi}, \mathbf{z}) = \mathbf{1}[y \le m_k(\mathbf{t}, \boldsymbol{\varphi}, \mathbf{z})] \prod_{i \ne k} F(m_i(\mathbf{t}, \boldsymbol{\varphi}, \mathbf{z})),$$

where $m_i(\mathbf{t}, \boldsymbol{\varphi}, \mathbf{z})$ is the i-th component of $M_{\boldsymbol{\varphi}}^{-1}(\mathbf{t} - \mathbf{z})$.

Suppose $h(X_{n+1}, \dots, X_{n+m})$ depends only on X_{n+m}, so the conditional expectation to be estimated is $q(h) = E(h(X_{n+m}) | X_1, \dots, X_n)$. Examples are conditional moments and absolute moments $E(X_{n+m}^\alpha | X_1, \dots, X_n)$ and $E(|X_{n+m}|^\alpha | X_1, \dots, X_n)$, and the one-dimensional conditional distribution function $P(X_{n+m} \le t | X_1, \dots, X_n)$. Then our estimator is

$$\hat{q}_s(h) = \int h(y_m + \hat{\varphi}_1 y_{m-1} + \dots + \hat{\varphi}_{m-1} y_1 + \hat{Z}_{nmr}) \hat{f}_m(\mathbf{y})\, d\mathbf{y}$$

with

$$\hat{Z}_{nmr} = \sum_{t=0}^{r} \hat{c}_{tm} X_{n-t},$$

and we have

$$T_k h(y, \boldsymbol{\varphi}, z) = E(h(\varepsilon_m + \varphi_1 \varepsilon_{m-1} + \dots + \varphi_{m-1} \varepsilon_1 + z) | \varepsilon_k = y).$$

3 Proof

Note that $\mathbf{Z}_{nr} = O_p(1)$. In view of this, the uniform differentiability (2.10), and the properties of $\hat{\boldsymbol{\varphi}}$ and $\hat{\mathbf{Z}}_{nr}$, it suffices to show the following two statements,

$$\sup_{h \in \mathscr{H}} \Big| \hat{q}_s(h) - \frac{1}{n-p} \sum_{j=p+1}^{n} \overline{h}_{\hat{\boldsymbol{\varphi}}, \hat{\mathbf{Z}}_{nr}}(\varepsilon_j) \Big| = o_p(n^{-1/2}) \tag{3.1}$$

and

$$(3.2)\quad \sup_{h\in\mathscr{H}}\Big|\frac{1}{n-p}\sum_{j=p+1}^{n}(\overline{h}(\varepsilon_j,\hat{\boldsymbol{\varphi}},\hat{\mathbf{Z}}_{nr})-\overline{h}(\varepsilon_j,\boldsymbol{\varphi},\mathbf{Z}_{nr}))-q_h(\hat{\boldsymbol{\varphi}},\hat{\mathbf{Z}}_{nr})+q_h(\boldsymbol{\varphi},\mathbf{Z}_{nr})\Big| = o_p(n^{-1/2}).$$

Let $\tilde{f}$ denote the kernel density estimator based on the true innovations,

$$\tilde{f}(y) = \frac{1}{n-p}\sum_{j=p+1}^{n} K_b(y-\varepsilon_j), \quad y\in\mathbb{R}.$$

We begin by recalling results about $\hat{f}$ and $\tilde{f}$. It follows from (2.9) and the proof of Theorem 10.1 in Schick and Wefelmeyer (2006a) that $\|\tilde{f}-f\|_V = O(b^2)+O_p(n^{-1/2}b^{-1/2}) = o_p(n^{-1/4})$. Since the observations have a finite fourth moment by (F), we can improve on the bound on $\|\hat{f}-\tilde{f}\|_V$ given in their proof. Indeed, proceeding as in their Lemma 9.2 with $a_n = K_b$, but using a second-order Taylor expansion instead of the first-order Taylor expansion used there, and utilizing the result of their Lemma 9.3, one can bound $\|\hat{f}-\tilde{f}\|_V$ by

$$\frac{1}{2}\|\hat{\Delta}^\top\mathbb{B}_{n2}\hat{\Delta}\|_V + O_p(p^{1/2}q^{1/2}n^{-1}b^{-3/2}) + O_p(n^{-1/2-\zeta}b^{-1}) + O_p(b^{-3}\|\hat{\Delta}\|^3E[\|\mathbf{X}_0\|^3]),$$

where $\hat{\Delta} = (\hat{\varrho}_1-\varrho_1,\dots,\hat{\varrho}_p-\varrho_p)^\top$ and $\mathbf{X}_{j-1} = (X_{j-1},\dots,X_{j-p})^\top$, and where

$$\mathbb{B}_{n2}(x) = \frac{1}{n-p}\sum_{j=p+1}^{n}\mathbf{X}_{j-1}\mathbf{X}_{j-1}^\top a_n''(x-\varepsilon_j), \quad x\in\mathbb{R}.$$

Note that $E[\|\mathbf{X}_0\|^3] = O(p^{3/2})$. Because of the identity $\int K'(u)\,du = 0$ we derive that

$$a_n''*f(x) = a_n'*f'(x) = b^{-1}\int(f'(x-bu)-f'(x))K'(u)\,du, \quad x\in\mathbb{R},$$

and obtain $\|a_n''*f\|_V = O(1)$ in view of the V-Lipschitz property of f'. Using this we can show that $\|\hat{\Delta}^\top\overline{\mathbb{B}}_{n2}\hat{\Delta}\|_V = O_p(pqn^{-1})$, where

$$\overline{\mathbb{B}}_{n2}(x) = \frac{1}{n-p}\sum_{j=p+1}^{n}\mathbf{X}_{j-1}\mathbf{X}_{j-1}^\top a_n''*f(x), \quad x\in\mathbb{R}.$$

Since $\mathbf{X}_{j-1}\mathbf{X}_{j-1}^\top(a_n''(x-\varepsilon_j)-a_n''*f(x))$ are uncorrelated for $j=p+1,\dots,n$, we find that

$$(n-p)E[\|\mathbb{B}_{n2}(x)-\overline{\mathbb{B}}_{n2}(x)\|^2] \le p^2E[X_0^4](a_n'')^2*f(x)$$

and thus obtain as in the proof of Lemma 9.3 of Schick and Wefelmeyer (2006a) that

$$\|\hat{\Delta}^\top(\mathbb{B}_{n2}-\overline{\mathbb{B}}_{n2})\hat{\Delta}\|_V = O_p(pqn^{-3/2}b^{-5/2}).$$

Consequently, we have

$$(3.3)\qquad \|\hat{f}-\tilde{f}\|_V = o_p(n^{-1/2})$$

and thus also

$$\|\hat{f} - f\|_V = o_p(n^{-1/4}). \tag{3.4}$$

The following result about smoothed empirical processes based on $\tilde{f}$ is Proposition 2.1 of Müller et al. (2006).

Lemma 1. *Let $\mathscr{G}$ denote a class of measurable functions on $\mathbb{R}$ with envelope G. Suppose that the following conditions are met.*

(G1) The envelope G belongs to $L_2(F)$ and is translation-continuous in $L_2(F)$:

$$\lim_{t\to 0} \int |G(x+t) - G(x)|^2 \, dF(x) = 0.$$

(G2) The enlarged class $\mathscr{G}_\eta = \{g(\cdot - t) : g \in \mathscr{G}, |t| \le \eta\}$ is F-Donsker for some $\eta > 0$.

(G3) The bias is uniformly negligible:

$$\sup_{g\in\mathscr{G}} \Big| \int g(y)(f * K_b(y) - f(y))\, dy \Big| = o_p(n^{-1/2}). \tag{3.5}$$

Then

$$\sup_{g\in\mathscr{G}} \Big| \int g(y)\tilde{f}(y)\, dy - \frac{1}{n-p} \sum_{j=p+1}^{n} g(\varepsilon_j) \Big| = o_p(n^{-1/2}). \tag{3.6}$$

For $\mathbf{\Delta} = (\mathbf{s}^\top, \mathbf{t}^\top)^\top \in \mathbb{R}^{m-1} \times \mathbb{R}^m$, we write $g_{\mathbf{\Delta}}$ for the affine transformation

$$g_{\mathbf{\Delta}}(\mathbf{y}) = M_{\boldsymbol{\varphi}+\mathbf{s}}\, \mathbf{y} + \mathbf{t}, \quad \mathbf{y} \in \mathbb{R}^m.$$

Fix a positive constant C and a δ in $(0, \alpha/2)$ with α as in (H). Set

$$\mathscr{U}_\delta = \{\mathbf{\Delta} \in \mathbb{R}^{2m-1} : \|\mathbf{\Delta}\| \le \delta\}, \qquad \mathscr{H}_C = \{h(\cdot + \mathbf{z}) : \|\mathbf{z}\| \le C\}$$

and

$$\mathscr{W} = \{w_{h,\mathbf{\Delta}} = h \circ g_{\mathbf{\Delta}} : h \in \mathscr{H}_C, \mathbf{\Delta} \in \mathscr{U}_\delta\}.$$

Let $\bar{\mathscr{W}} = \{\overline{w} : w \in \mathscr{W}\}$ and $\bar{\mathscr{W}}_k = \{\overline{w}_k : w \in \mathscr{W}\}$, where $\overline{w} = \overline{w}_1 + \cdots + \overline{w}_m$ and $\overline{w}_k(y) = T_k w(y) = E(w(\boldsymbol{\varepsilon})|\varepsilon_k = y)$.

It is easy to check that there is a constant B such that

$$\|g_{\mathbf{\Delta}}(\mathbf{y})\| \le B(1 + \|\mathbf{y}\|), \quad \mathbf{y} \in \mathbb{R}^m, \mathbf{\Delta} \in \mathscr{U}_\delta.$$

Thus $W(g_{\mathbf{\Delta}}(\mathbf{y}) + \mathbf{z}) \le V(B+C)W(\mathbf{y})$ for all $\mathbf{y}$ if $\|\mathbf{\Delta}\| \le \delta$ and $\|\mathbf{z}\| \le C$. This shows that $\mathscr{W}$ has envelope $a\mathscr{W}$ for some positive constant a. It is easy to check that

$$W(\mathbf{y}) \le V(y_1) \cdots V(y_m), \quad \mathbf{y} = (y_1, \dots, y_m) \in \mathbb{R}^m. \tag{3.7}$$

This shows that the sets $\bar{\mathscr{W}}_1, \dots, \bar{\mathscr{W}}_m$ have envelope aV and $\bar{\mathscr{W}}$ has envelope maV. Using the latter and (3.4) we obtain as in the proof of Theorem 2.1 of Müller et al. (2006) that

$$\sup_{w \in \mathscr{W}} \left| \int w(\mathbf{y}) \hat{f}_m(\mathbf{y})\, d\mathbf{y} - \int \overline{w}(y) \hat{f}(y)\, dy \right| = o_p(n^{-1/2}). \tag{3.8}$$

Next, we derive from (3.3) and the fact that $\bar{\mathscr{W}}$ has envelope maV that

$$\sup_{\overline{w} \in \bar{\mathscr{W}}} \left| \int \overline{w}(y)(\hat{f}(y) - \tilde{f}(y))\, dy \right| = o_p(n^{-1/2}). \tag{3.9}$$

By the moment assumption on f, the function V is translation-continuous in $L_2(F)$. Fix $k \in \{1, \dots, m\}$. Note that the enlarged class $\bar{\mathscr{W}}_{k,\eta} = \{\overline{w}_k(\cdot - t) : \overline{w}_k \in \bar{\mathscr{W}}_k, |t| \le \eta\}$, is a subset of $\mathscr{H}_{k,2C}$ for small enough $\eta > 0$ and hence F-Donsker by (H). It follows from (2.9) that

$$\sup_{\overline{w}_k \in \bar{\mathscr{W}}_k} \left| \int \overline{w}_k(y)(f * K_b(y) - f(y))\, dy \right| = o_p(n^{-1/2}).$$

Thus Lemma 1, applied with $\mathscr{G} = \bar{\mathscr{W}}_k$, yields the expansion

$$\sup_{\overline{w}_k \in \bar{\mathscr{W}}_k} \left| \int \overline{w}_k(y) \tilde{f}(y)\, dy - \frac{1}{n-p} \sum_{j=p+1}^{n} \overline{w}_k(\varepsilon_j) \right| = o_p(n^{-1/2}). \tag{3.10}$$

Since k was arbitrary, we now have from (3.8), (3.9) and (3.10) that

$$\sup_{w \in \mathscr{W}} \int w(\mathbf{y}) \hat{f}_m(\mathbf{y})\, d\mathbf{y} - \frac{1}{n-p} \sum_{j=p+1}^{n} \overline{w}(\varepsilon_j) \Big| = o_p(n^{-1/2}).$$

This holds for all finite C, and thus implies (3.1).

Now fix again $k \in \{1, \dots, m\}$ and set

$$S_{nk}(h, \mathbf{\Delta}) = \frac{1}{n-p} \sum_{j=p+1}^{n} (T_k w_{h,\mathbf{\Delta}}(\varepsilon_j) - E[T_k w_{h,\mathbf{\Delta}}(\varepsilon)]), \quad h \in \mathscr{H}_C, \mathbf{\Delta} \in \mathscr{U}_\delta.$$

It follows from (2.7) that

$$\lim_{\delta \to 0} \lim_{n \to \infty} P\Big(\sup_{h \in \mathscr{H}_C} \sup_{\mathbf{\Delta} \in \mathscr{U}_\delta} n^{1/2} |S_{nk}(h, \mathbf{\Delta}) - S_{nk}(h, 0)| > \epsilon \Big) = 0$$

for every $\epsilon > 0$. Since this is valid for all k and finite C, we derive by the properties of $\mathbf{Z}_{nr}$, $\hat{\mathbf{Z}}_{nr}$ and $\hat{\boldsymbol{\varphi}}$ that (3.2) holds. This completes the proof of Theorem 1.

References

Gill, R. D. (1989). Non- and semi-parametric maximum likelihood estimators and the von Mises method. I. With a discussion by J. A. Wellner and J. Præstgaard and a reply by the author. *Scand. J. Statist.* **16**, 97–128.

Masry, E. (1989). Nonparametric estimation of conditional probability densities and expectations of stationary processes: Strong consistency and rates. *Stochastic Process. Appl.* **32**, 109–127.

Müller, U. U., Schick, A. and Wefelmeyer, W. (2006). Efficient prediction for linear and nonlinear autoregressive models. *Ann. Statist.* **34**, 2496-2533.

Robinson, P. M. (1983). Nonparametric estimators for time series. *J. Time Ser. Anal.* **4**, 185–207.

Robinson, P. M. (1986). On the consistency and finite-sample properties of nonparametric kernel time series regression, autoregression and density estimators. *Ann. Inst. Statist. Math.* **38**, 539–549.

Roussas, G. G. (1969). Nonparametric estimation of the transition distribution function of a Markov process. *Ann. Math. Statist.* **40**, 1386-1400.

Roussas, G. G. (1991). Recursive estimation of the transition distribution function of a Markov process: Asymptotic normality. *Statist. Probab. Lett.* **11**, 435–447.

Roussas, G. G. and Tran, L. T. (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions. *Ann. Statist.* **20**, 98–120.

Schick, A. and Wefelmeyer, W. (2006a). Root-n consistency in weighted L_1-spaces for density estimators of invertible linear processes. Technical Report, Department of Mathematical Sciences, Binghamton University.

Schick, A. and Wefelmeyer, W. (2006b). Prediction in moving average processes. To appear in: *J. Statist. Plann. Inference.*

Schick, A. and Wefelmeyer, W. (2006c). Plug-in estimators for higher-order transition densities in autoregression. Technical Report, Department of Mathematical Sciences, Binghamton University.

Schick, A. and Wefelmeyer, W. (2007). Root-n consistent density estimators of convolutions in weighted L_1-norms. *J. Statist. Plann. Inference* **137**, 1765–1774.

Tran, L. T. (1992). Kernel density estimation for linear processes. *Stochastic Process. Appl.* **41**, 281–296.

Truong, Y. K. and Stone, C. J. (1992). Nonparametric function estimation involving time series. *Ann. Statist.* **20**, 77–97.

Yakowitz, S. (1985). Nonparametric density estimation, prediction, and regression for Markov sequences. *J. Amer. Statist. Assoc.* **80**, 215–221.

Yakowitz, S. (1987). Nearest-neighbour methods for time series analysis. *J. Time Ser. Anal.* **8**, 235–247.

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

Wolfgang Wefelmeyer
Mathematisches Institut
Universität zu Köln
50931 Köln
Germany
wefelm@math.uni-koeln.de
http://www.mi.uni-koeln.de/∼wefelm/