Prediction in moving average processes

Anton Schick and Wolfgang Wefelmeyer

Abstract. For the stationary invertible moving average process of order one with unknown innovation distribution F, we construct root-n consistent plug-in estimators of conditional expectations $E(h(X_{n+1})|X_1,\dots,X_n)$. More specifically, we give weak conditions under which such estimators admit Bahadur type representations, assuming some smoothness of h or of F. For fixed h it suffices that h is locally of bounded variation and locally Lipschitz in $L_2(F)$, and that the convolution of h and F is continuously differentiable. A uniform representation for the plug-in estimator of the conditional distribution function $P(X_{n+1} \le \cdot\,|X_1,\dots,X_n)$ holds if F has a uniformly continuous density. For a smoothed version of our estimator, the Bahadur representation holds uniformly over each class of functions h that have an appropriate envelope and whose shifts are F-Donsker, assuming some smoothness of F. The proofs use empirical process arguments.

AMS 2000 subject classification. Primary: 62M09, 62M10.

Key words and Phrases. Smoothed empirical process, stochastic expansion, asymptotically linear estimator, residual-based density estimator, conditional quantile, conditional absolute moment.

1. Introduction

Let $X_1,\dots,X_n$ be observations from a real-valued stationary time series. Let h be a measurable function such that $E[h^2(X_1)]$ is finite. The best predictor for $h(X_{n+1})$ is the conditional expectation $E(h(X_{n+1})|X_1,\dots,X_n)$. Suppose first that the time series is Markov of known order r. Then the conditional expectation equals $E(h(X_{n+1})|X_{n-r+1},\dots,X_n)$. Convergence rates for kernel estimators of the function $(x_1,\dots,x_r) \mapsto E(h(X_{n+1})|X_{n-r+1} = x_1,\dots,X_n = x_r)$ are in Roussas (1969, 1991), Yakowitz (1985), Masry (1989) and Delecroix and Rosa (1995). Analogous results for estimators of conditional quantiles are in Gannoun, Saracco and Yu (2003). If the observations come from a nonlinear r-order autoregressive process, $X_{t+1} = \varrho_\vartheta(X_{t-r+1},\dots,X_t) + \varepsilon_t$ with independent innovations ε_t with distribution function F, then

$$E(h(X_{n+1})|X_{n-r+1} = x_1,\dots,X_n = x_r) = \int h(y + \varrho_\vartheta(x_1,\dots,x_r))\,dF(y)$$

The research of A. Schick was partially supported by NSF Grant DMS0405791.

can be estimated at the "parametric" root-n rate by plugging in a root-n consistent estimator for ϑ and a residual-based empirical estimator for F. Smoothed and weighted versions of such plug-in estimators are studied in Müller, Schick and Wefelmeyer (2006).

Now let the time series be non-Markovian. Then $E(h(X_{n+1})|X_{n-r+1},\dots,X_n)$ is still an approximation for $E(h(X_{n+1})|X_1,\dots,X_n)$ if r is large enough. Asymptotic results for kernel estimators of $E(h(X_{n+1})|X_{n-r+1}=x_1,\dots,X_n=x_r)$ are obtained by Robinson (1983, 1986), Collomb (1984), Yakowitz (1987), Truong and Stone (1992), Roussas and Tran (1992) and Tran (1993). Estimators of conditional medians are studied in Zhou and Liang (2000, 2003). *Uniform* consistency of set-indexed conditional empirical processes and Bahadur–Kiefer representations for generalized conditional quantile processes are in Polonik and Yao (2000, 2002). If the time series is driven by independent observations, we expect again to obtain root-n consistent plug-in estimators. We show this for a simple non-Markovian invertible linear time series, a stationary moving average process of order one,

$$X_t = \varepsilon_t - \vartheta\varepsilon_{t-1}, \quad t \in \mathbb{Z},$$

with $\vartheta \in (-1,1)$ and independent and identically distributed innovations $\{\varepsilon_t, t \in \mathbb{Z}\}$ with finite mean μ, finite variance σ^2 and distribution function F. We write X and ε for random variables distributed as X_t and ε_t, respectively. Aside from the better convergence rate, our result differs from the above nonparametric results in two respects. We condition on the *full past* $X_1,\dots,X_n$, not just on a string $X_{n-r+1},\dots,X_n$ of fixed length r. For this reason, we estimate the *random variable*

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

not a deterministic function $(x_1,\dots,x_r) \mapsto E(h(X_{n+1})|X_{n-r+1}=x_1,\dots,X_n=x_r)$.

In order to prove that an estimator $\hat{q}(h)$ of $q(h)$ is root-n consistent, we approximate the standardized errors $n^{1/2}(\hat{q}(h)-q(h))$ stochastically by a sequence of random variables that we can show to be tight. Since these sequences involve sums of independent random variables, we call the approximations Bahadur type representations.

Our estimator is constructed as follows. Invertibility of the moving average process allows us to write the innovations as

$$\varepsilon_t = \sum_{s=0}^{\infty} \vartheta^s X_{t-s}, \quad t \in \mathbb{Z}.$$

For non-negative integers r we can write $\varepsilon_t = \varepsilon_{t,r} + \vartheta^{r+1}\varepsilon_{t-r-1}$, where $\varepsilon_{t,r}$ is a truncated version of ε_t,

$$\varepsilon_{t,r} = \sum_{s=0}^{r} \vartheta^s X_{t-s}.$$

In particular, $X_{n+1} = \varepsilon_{n+1} - \vartheta\varepsilon_{n,r} - \vartheta^{r+2}\varepsilon_{n-r-1}$. Since ε_{n+1} is independent of $X_1,\dots,X_n$, we obtain the representation

$$q(h) = E(q_h(\vartheta\varepsilon_{n,r} + \vartheta^{r+2}\varepsilon_{n-r-1})|X_1,\dots,X_n),$$

where

$$q_h(x) = E[h(\varepsilon - x)] = \int h(y-x)\,dF(y), \quad x \in \mathbb{R}.$$

Thus, if q_h is Lipschitz, then $q(h)$ is well approximated by $q_h(\vartheta\varepsilon_{n,r})$ for large integers r. Indeed, with L denoting the Lipschitz constant, we have

$$E[(q(h) - q_h(\vartheta\varepsilon_{n,r}))^2] \le L^2 E[(\vartheta^{r+2}\varepsilon_{n-r-1})^2] = L^2\vartheta^{2r+4}E[\varepsilon^2].$$

Throughout the paper let $\hat\vartheta$ be a root-n consistent estimator of ϑ. We can mimic the innovation ε_j by the truncated residual

$$\hat\varepsilon_j = \sum_{s=0}^{r_n} \hat\vartheta^s X_{j-s}, \quad j = r_n+1, \dots, n.$$

Here r_n is an integer that tends to infinity slowly with the sample size, $r_n \sim \log n \log\log n$. Then $q_h(\vartheta\varepsilon_{n,r_n})$ approximates $q(h)$ up to $o_p(n^{-1/2})$, and we can estimate the conditional expectation $q(h)$ by

$$\hat q(h) = \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} h(\hat\varepsilon_j - \hat\vartheta\hat\varepsilon_n).$$

A stochastic expansion of $\hat q(h)$ is easy to derive by Taylor expansion for a fixed and smooth function h. We do this first, for illustration, and without striving for minimal conditions. Similarly as relation (13) in Schick and Wefelmeyer (2004) one can show that

$$\frac{1}{n-r_n} \sum_{j=r_n+1}^{n} \big(\hat\varepsilon_j - \varepsilon_j - (\hat\vartheta - \vartheta)Y_{j-1}\big)^2 = O_p(n^{-2}) \tag{1.1}$$

with

$$Y_{j-1} = \sum_{s=1}^{\infty} s\vartheta^{s-1}X_{j-s} = \sum_{s=0}^{\infty} \vartheta^s \varepsilon_{j-1-s}.$$

Note that

$$\nu = E[Y_0] = \frac{\mu}{1-\vartheta}.$$

Suppose that h has a bounded second derivative. Then a Taylor expansion yields that

$$\hat q(h) = \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} \Big(h(\varepsilon_j - \vartheta\varepsilon_n) + (\hat\vartheta - \vartheta)(Y_{j-1} - \vartheta Y_{n-1} - \varepsilon_n)h'(\varepsilon_j - \vartheta\varepsilon_n)\Big) + o_p(n^{-1/2}).$$

Since $\varepsilon_n + \vartheta Y_{n-1} = Y_n$ and since Y_{j-1} and ε_j are independent, we derive that

$$\frac{1}{n-r_n} \sum_{j=r_n+1}^{n} (Y_{j-1} - \vartheta Y_{n-1} - \varepsilon_n)h'(\varepsilon_j - \vartheta\varepsilon_n) = (Y_n - \nu)q_h'(\vartheta\varepsilon_n) + o_p(1),$$

where q_h' is the derivative of q_h,

$$q_h'(x) = -E[h'(\varepsilon - x)] = -\int h'(y-x)\,dF(y), \quad x \in \mathbb{R}.$$

We arrive at the Bahadur type representation

$$\hat{q}(h) = \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} h(\varepsilon_j - \vartheta\varepsilon_n) + (\hat{\vartheta} - \vartheta)(Y_n - \nu) q_h'(\vartheta\varepsilon_n) + o_p(n^{-1/2}). \tag{1.2}$$

It implies that $\hat{q}(h)$ is a root-n consistent estimator of $q(h)$,

$$\begin{aligned} \hat{q}(h) = q(h) + \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} \Big(h(\varepsilon_j - \vartheta\varepsilon_n) - \int h(y - \vartheta\varepsilon_n)\, dF(y) \Big) \\ + (\hat{\vartheta} - \vartheta)(Y_n - \nu) q_h'(\vartheta\varepsilon_n) + o_p(n^{-1/2}). \end{aligned} \tag{1.3}$$

The above result applies to $h(x) = x$ and $h(x) = x^2$. For the first choice, expansion (1.3) becomes

$$\begin{aligned} \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} (\hat{\varepsilon}_j - \hat{\vartheta}\hat{\varepsilon}_n) = E(X_{n+1}|X_1, \dots, X_n) + \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} (\varepsilon_j - \mu) \\ + (\hat{\vartheta} - \vartheta)(Y_n - \nu) + o_p(n^{-1/2}), \end{aligned}$$

and for the second choice it becomes

$$\begin{aligned} \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} (\hat{\varepsilon}_j - \hat{\vartheta}\hat{\varepsilon}_n)^2 = E(X_{n+1}^2|X_1, \dots, X_n) + \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} ((\varepsilon_j - \mu)^2 - \sigma^2) \\ + 2(\mu - \vartheta\varepsilon_n) \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} (\varepsilon_j - \mu) - 2(\mu - \vartheta\varepsilon_n)(\hat{\vartheta} - \vartheta)(Y_n - \nu) + o_p(n^{-1/2}). \end{aligned}$$

It is the purpose of this paper to explore minimal conditions under which $\hat{q}(h)$ admits the Bahadur representation (1.2) and is therefore root-n consistent. We give results both for fixed h and uniformly over classes of functions h. We need smoothness of the function q_h, and this can be achieved by assuming some smoothness either of h or of F. In Section 2 we consider a fixed h that is locally of bounded variation and locally $L_2(F)$-Lipschitz and show that (1.2) holds if q_h is continuously differentiable. Examples are conditional absolute moments. Section 3 treats functions $h_z(x) = \mathbf{1}[x \le z]$, estimates the conditional distribution function $P(X_{n+1} \le z|X_1, \dots, X_n)$ by $\hat{q}(h_z) = \hat{\mathbb{F}}(z + \hat{\vartheta}\hat{\varepsilon}_n)$ with $\hat{\mathbb{F}}$ a residual-based empirical distribution function, and gives a Bahadur representation uniformly in z for $\hat{q}(h_z)$, under the assumption that F has a uniformly continuous density f. The result applies to conditional quantiles. Section 4 gives stochastic expansions for residual-based kernel estimators of the density f of F and for estimators of the conditional density of X_{n+1} given $X_1, \dots, X_n$. Section 5 considers general classes of functions h. We use a smoothed version of $\hat{q}(h)$, namely $\hat{q}_s(h) = \int h(y - \hat{\vartheta}\hat{\varepsilon}_n) \hat{f}(y)\, dy$ with $\hat{f}$ a residual-based kernel density estimator of f. We show in particular that (1.2) holds for $\hat{q}_s(h)$ uniformly over $h \in \mathcal{H}$ if $\mathcal{H}$ has an appropriate envelope and $\{h(\cdot - t) : h \in \mathcal{H}, |t| \le C\}$ is F-Donsker for each $C < \infty$, assuming some smoothness of F. The proof uses results of Section 4.

2. Conditional expectations

In this section we prove the Bahadur representation (1.2) of $\hat{q}(h)$ for a fixed function h that need not be smooth. To this end we write

$$\hat{q}(h) - q_h(\vartheta\varepsilon_n) = \hat{\Psi}(\hat{\vartheta}\hat{\varepsilon}_n) + (q_h(\hat{\vartheta}\hat{\varepsilon}_n) - q_h(\vartheta\varepsilon_n)),$$

where

$$\hat{\Psi}(t) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} \big(h(\hat{\varepsilon}_j - t) - E[h(\varepsilon - t)]\big), \qquad t \in \mathbb{R}.$$

We assume that q_h is continuously differentiable. Then for every $C < \infty$ we have

$$\sup_{|x|\le C} |q_h(x+t) - q_h(x) - t q_h'(x)| = o(t). \tag{2.1}$$

We also have $\hat{\vartheta}\hat{\varepsilon}_n - \vartheta\varepsilon_n = (\hat{\vartheta} - \vartheta)\hat{\varepsilon}_n + \vartheta(\hat{\varepsilon}_n - \varepsilon_n) = (\hat{\vartheta} - \vartheta)(\varepsilon_n + \vartheta Y_{n-1}) + o_p(n^{-1/2})$ and thus

$$\hat{\vartheta}\hat{\varepsilon}_n - \vartheta\varepsilon_n = (\hat{\vartheta} - \vartheta)Y_n + o_p(n^{-1/2}). \tag{2.2}$$

Together with $\vartheta\varepsilon_n = O_p(1)$ we obtain

$$q_h(\hat{\vartheta}\hat{\varepsilon}_n) - q_h(\vartheta\varepsilon_n) - (\hat{\vartheta} - \vartheta)Y_n q_h'(\vartheta\varepsilon_n) = o_p(n^{-1/2}), \tag{2.3}$$

$$q_h'(\hat{\vartheta}\hat{\varepsilon}_n) - q_h'(\vartheta\varepsilon_n) = o_p(1). \tag{2.4}$$

Hence the desired expansion (1.2) is valid if we show that

$$\Psi(\hat{\vartheta}\hat{\varepsilon}_n) = \Psi(\vartheta\varepsilon_n) + o_p(n^{-1/2}) \tag{2.5}$$

and

$$\sup_{|t|\le C} |\hat{\Psi}(t) - \Psi(t) + (\hat{\vartheta} - \vartheta)\nu q_h'(t)| = o_p(n^{-1/2}) \tag{2.6}$$

for all $C < \infty$, where

$$\Psi(t) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} \big(h(\varepsilon_j - t) - E[h(\varepsilon - t)]\big), \qquad t \in \mathbb{R}.$$

If the function h can be written as a linear combination of monotone right-continuous functions, it suffices to study the behavior of $\hat{q}(h)$ for the latter functions. We have the following result.

THEOREM 1. *Let h be a non-decreasing right-continuous function such that $\int h^2\,dF$ is finite, q_h is continuously differentiable, and there is a non-decreasing function L on $(0, \infty)$ so that*

$$\int (h(y-t) - h(y-s))^2\,dF(y) \le L(C)|t-s|^2, \qquad |t|, |s| < C, \tag{2.7}$$

for all $C < \infty$. Then

$$\hat{q}(h) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} h(\varepsilon_j - \vartheta\varepsilon_n) + (\hat{\vartheta} - \vartheta)(Y_n - \nu)q_h'(\vartheta\varepsilon_n) + o_p(n^{-1/2}).$$

PROOF. It remains to show (2.5) and (2.6). It follows from (2.7) that $nE[(\Psi(t)-\Psi(s))^2] \le L(C)|t-s|^2$ for all positive finite C and all $|s|,|t| \le C$. Thus it follows from Theorems 12.4 and 15.6 in Billingsley (1968) that, for each finite positive C, the sequence $\{n\Psi(t) : |t| \le C\}$ of processes converges in distribution in $D[-C,C]$ to a Gaussian process with continuous sample paths. Thus we have

$$\sup_{|t|\le C} |\Psi(t+\xi_n) - \Psi(t)| = o_p(n^{-1/2}) \tag{2.8}$$

for all $C < \infty$ and every sequence $\xi_n = o_p(1)$. The desired (2.5) is now immediate.

Let us now verify (2.6). Set

$$\Psi(\Delta, t) = H(\Delta, t) - \bar{H}(\Delta, t), \qquad \Delta, t \in \mathbb{R},$$

where

$$H(\Delta, t) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} h(\varepsilon_j + n^{-1/2}\Delta Y_{j-1} - t),$$

$$\bar{H}(\Delta, t) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} \int h(y + n^{-1/2}\Delta Y_{j-1} - t)\, dF(y).$$

Note that $\Psi(0,t) = \Psi(t)$. As a first step we shall show that for $C < \infty$,

$$\sup_{|\Delta|,|t|\le C} |\Psi(\Delta, t) - \Psi(0,t)| = o_p(n^{-1/2}). \tag{2.9}$$

To prove (2.9), fix a constant $C < \infty$ and a positive integer M. Set $\eta = C/M$ and let

$$\Delta_i = -C + (2i-1)\eta, \quad U_{ij} = \Delta_i Y_{j-1} - \eta|Y_{j-1}|, \quad V_{ij} = \Delta_i Y_{j-1} + \eta|Y_{j-1}|$$

for $i = 1,\dots,M$ and $j = r_n+1,\dots,n$. For Δ in the sub-interval $[\Delta_i - \eta, \Delta_i + \eta]$ of $[-C,C]$ we obtain

$$|\Psi(\Delta,t) - \Psi(0,t)| \le |\Psi(\Delta_i,t) - \Psi(0,t)| + |\Psi(\Delta,t) - \Psi(\Delta_i,t)|$$

and $U_{ij} \le \Delta Y_{j-1} \le V_{ij}$ for all j. Thus, exploiting the monotonicity of h,

$$|\Psi(\Delta,t) - \Psi(\Delta_i,t)| \le K_i(t) + \bar{K}_i(t) \le K_i(t) - \bar{K}_i(t) + 2\bar{K}_i(t),$$

where

$$K_i(t) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} \big(h(\varepsilon_j + n^{-1/2}V_{ij} - t) - h(\varepsilon_j + n^{-1/2}U_{ij} - t)\big),$$

$$\bar{K}_i(t) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} \int \big(h(y + n^{-1/2}V_{ij} - t) - h(y + n^{-1/2}U_{ij} - t)\big)\, dF(y).$$

This shows that the left-hand side of (2.9) is bounded by

$$\sup_{|t|\le C} \max_{1\le i\le M} (|\Psi(\Delta_i,t) - \Psi(0,t)| + |K_i(t) - \bar{K}_i(t)| + 2\bar{K}_i(t)).$$

Since the variables Y_j are stationary with finite second moment, we have

$$P\Big(\max_{r_n<j\le n}|Y_{j-1}|>\varepsilon n^{1/2}\Big)\le \varepsilon^{-2}E[Y_0^2\mathbf{1}[|Y_0|>\varepsilon n^{1/2}]]=o_p(1). \tag{2.10}$$

Now let Y_{nj} denote a truncated version of Y_{j-1}, defined by

$$Y_{nj}=Y_{j-1}\mathbf{1}[|Y_{j-1}|\le n^{1/2}]+\operatorname{sign}(Y_{j-1})n^{1/2}\mathbf{1}[|Y_{j-1}|>n^{1/2}].$$

Since we have $Y_{j-1}=Y_{nj}$ for all j except on the event $\{\max_{r_n<j\le n}|Y_{j-1}|>n^{1/2}\}$ whose probability tends to zero, we see that we may work with versions of the above processes in which the variables Y_{j-1} are replaced by Y_{nj}. Let $\Psi^*(\Delta,t)$, $K_i^*(t)$, $\bar K_i^*(t)$, U_{ij}^* and V_{ij}^* denote these versions. Since $|n^{-1/2}Y_{nj}|\le 1$, we obtain with the help of a martingale argument that

$$\begin{aligned}(n-r_n)&E[(\Psi^*(\Delta,t)-\Psi^*(\Delta,s))^2]\\ &\le E\Big[\int\big(h(y+n^{-1/2}\Delta Y_{n1}-t)-h(y+n^{-1/2}\Delta Y_{n1}-s)\big)^2\,dF(y)\Big]\\ &\le L(C+|\Delta|)|t-s|^2,\quad |s|,|t|\le C.\end{aligned}$$

Thus, by Theorem 15.6 of Billingsley (1968), the sequence $\{n^{1/2}(\Psi^*(\Delta,t)-\Psi^*(0,t)):|t|\le C\}$ of processes is tight in $D[-C,C]$ for each Δ. Since, for each $t\in\mathbb{R}$,

$$\begin{aligned}(n-r_n)E[(\Psi^*(\Delta,t)-\Psi^*(0,t))^2]&\le E\Big[\int\big(h(y+n^{-1/2}\Delta Y_{n1}-t)-h(y-t)\big)^2\,dF(y)\Big]\\ &\le L(|t|+|\Delta|)n^{-1}\Delta^2E[Y_0^2]\to 0,\end{aligned}$$

we obtain that

$$\max_{1\le i\le M}\sup_{|t|\le C}|\Psi^*(\Delta_i,t)-\Psi^*(0,t)|=o_p(n^{-1/2}).$$

Similarly, one verifies

$$(n-r_n)E[(K_i^*(t)-\bar K_i^*(t)-K_i^*(s)+\bar K_i^*(s))^2]\le 4L(2C)|t-s|^2,\qquad |s|,|t|\le C,$$

and

$$\begin{aligned}(n-r_n)E[(K_i^*(t)-\bar K_i^*(t))^2]&\le E\Big[\int\big(h(y+n^{-1/2}V_{i1}^*-t)-h(y+n^{-1/2}U_{i1}^*-t)\big)^2\,dF(y)\Big]\\ &\le L(|t|+C)4n^{-1}\eta^2E[Y_0^2]\to 0.\end{aligned}$$

Hence we obtain as above that

$$\max_{1\le i\le M}\sup_{|t|\le C}|K_i^*(t)-\bar K_i^*(t)|=o_p(n^{-1/2}).$$

Finally, for $|t|\le C$ we find that

$$\begin{aligned}n(\bar K_i^*(t))^2&\le\frac{n}{n-r_n}\sum_{j=r_n+1}^n\int\big(h(y+n^{-1/2}V_{ij}^*-t)-h(y+n^{-1/2}U_{ij}^*-t)\big)^2\,dF(y)\\ &\le L(2C)4\eta^2\frac{1}{n-r_n}\sum_{j=r_n+1}^n Y_{j-1}^2\end{aligned}$$

and obtain

$$\max_{1\le i\le n}\sup_{|t|\le C} n^{1/2}\bar{K}_i^*(t) \le \eta K_C + o_p(1),$$

where $K_C = 2(L(2C)E[Y_0^2])^{1/2}$. Combining the above we see that

$$n^{1/2}\sup_{|\Delta|,|t|\le C}|\Psi(\Delta,t)-\Psi(0,t)| \le \frac{CK_C}{M} + o_p(1).$$

This holds for all positive integers M and thus yields the desired result (2.9).

In view of (2.8) and (2.9) we obtain for each $C<\infty$ that

$$\sup_{|\Delta|,|t|\le C}|\Psi(\Delta,t+\xi_n)-\Psi(0,t)| = o_p(n^{-1/2})$$

for any random variable $\xi_n = o_p(1)$. Because of $\hat\Delta = n^{1/2}(\hat\vartheta-\vartheta) = O_p(1)$ we then obtain

$$A(\xi_n) = \sup_{|t|\le C}|\Psi(\hat\Delta,t+\xi_n)-\Psi(0,t)| = o_p(n^{-1/2})$$

for any random variable $\xi_n = o_p(1)$. Assume now that $\xi_n = o_p(n^{-1/2})$. Then one also has

$$B(\xi_n) = \sup_{|t|\le C}|\bar{H}(\hat\Delta,t+\xi_n)-\bar{H}(0,t)+(\hat\vartheta-\vartheta)\nu q_h'(t)| = o_p(n^{-1/2}).$$

This follows if we show that

$$\sup_{|\Delta|,|t|\le C}|\bar{H}(\Delta,t+\xi_n)-\bar{H}(0,t)+n^{-1/2}\Delta\nu q_h'(t)| = o_p(n^{-1/2})$$

for finite C. The left-hand side can be bounded by T_1+T_2, where

$$T_1 = \sup_{|\Delta|,|t|\le C}\Big|\frac{1}{n-r_n}\sum_{j=r_n+1}^{n}\big(q_h(t+\xi_n-n^{-1/2}\Delta Y_{j-1})-q_h(t)-(\xi_n-n^{-1/2}\Delta Y_{j-1})q_h'(t)\big)\Big|,$$

$$T_2 = \sup_{|t|\le C}|q_h'(t)|\Big(|\xi_n| + Cn^{-1/2}\Big|\frac{1}{n-r_n}\sum_{j=r_n+1}^{n}Y_{j-1}-\nu\Big|\Big).$$

It is clear that $T_2 = o_p(n^{-1/2})$. It follows from (2.1) that $T_1 = o_p(n^{-1/2})$.

Set

$$R_j = \hat\varepsilon_j - \varepsilon_j - (\hat\vartheta-\vartheta)Y_{j-1}, \quad j = r_n+1,\dots,n.$$

Then we have

$$R_n^* = \max_{r_n<j\le n}|R_j| = o_p(n^{-1/2}). \tag{2.11}$$

The monotonicity of h yields the bounds

$$H(\hat\Delta,t+R_n^*)-\bar{H}(0,t) \le \hat\Psi(t) \le H(\hat\Delta,t-R_n^*)-\bar{H}(0,t)$$

for all real t. Using this and $\Psi(t)=\Psi(0,t)$, we find that

$$\sup_{|t|\le C}|\hat\Psi(t)-\Psi(t)+(\hat\vartheta-\vartheta)\nu q_h'(t)| \le \max\big(A(-R_n^*),A(R_n^*)\big)+\max\big(B(-R_n^*),B(R_n^*)\big).$$

The desired (2.6) is now immediate. □

REMARK 1. Theorem 1 applies to estimating conditional absolute moments. Let $\beta \geq 1$ and $h(y) = |y|^\beta$. Then $q(h) = E(|X_{n+1}|^\beta | X_1, \dots, X_n)$. Assume that F has a moment of order 2β. We can write h as the difference $h_1 - h_2$ of the continuous non-decreasing functions h_1 and h_2 defined by $h_1(y) = h(y)1[y > 0]$ and $h_2 = -h(y)1[y < 0]$, $y \in \mathbb{R}$. Then the assumptions of Theorem 1 hold with $h = h_1$ and $h = h_2$ if $\beta > 1$ and require continuity of the distribution function F in the case $\beta = 1$. Our estimator is

$$\hat{q}(h) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} |\hat{\varepsilon}_j - \hat{\vartheta}\hat{\varepsilon}_n|^\beta.$$

We have $q_h(x) = E[|\varepsilon - x|^\beta]$ and $q_h'(x) = \beta E[\operatorname{sign}(\varepsilon - x)|\varepsilon - x|^{\beta-1}]$.

3. Conditional distribution function

For indicator functions $h_z(y) = \mathbf{1}[y \leq z]$ and $s < t$, the integral

$$\int (h_z(y-t) - h_z(y-s))^2 \, dF(y) = P(a + s \leq y < a + t)$$

is of order $t - s$, and assumption (2.7) on h does not hold. Hence Theorem 1 does not apply to estimating the conditional distribution function $q(h_z) = P(X_{n+1} \leq z | X_1, \dots, X_n)$ of X_{n+1} given $X_1, \dots, X_n$ at z. In this section we show that the Bahadur representation still holds for $\hat{q}(h_z)$ if F is smooth.

Let $\mathcal{H}$ be a class of functions h that is closed under shifts. The residual-based empirical estimator for the (unconditional) expectation $E[h(\varepsilon)]$ is

$$\hat{m}(h) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} h(\hat{\varepsilon}_j).$$

Suppose that we have a Bahadur representation for $\hat{m}(h)$ uniformly over $h \in \mathcal{H}$,

$$\sup_{h \in \mathcal{H}} \Big| \hat{m}(h) - \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} h(\varepsilon_j) + (\hat{\vartheta} - \vartheta)\nu q_h'(0) \Big| = o_p(n^{-1/2}).$$

Let

$$\mathbb{B}(h) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} (h(\varepsilon_j) - E[h(\varepsilon)]), \quad h \in \mathcal{H}.$$

If the process $\{n^{1/2}\mathbb{B}(h) : h \in \mathcal{H}\}$ is tight and q_h is smooth uniformly in $h \in \mathcal{H}$ in an appropriate sense, then the Bahadur representation (1.2) for the estimator $\hat{q}(h)$ of the conditional expectation $q(h)$ follows from the above representation for $\hat{m}(h)$, and it is uniform in $h \in \mathcal{H}$.

We illustrate this with the problem of estimating the conditional distribution function $q(h_z)$. Then $q_{h_z}(x) = F(z+x)$. The plug-in estimator for $q(h_z)$ is $\hat{q}(h_z) = \hat{\mathbb{F}}(z + \hat{\vartheta}\hat{\varepsilon}_n)$, where

$$\hat{\mathbb{F}}(z) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} \mathbf{1}[\hat{\varepsilon}_j \le z], \quad t \in \mathbb{R},$$

denotes the empirical distribution function based on the residuals $\hat{\varepsilon}_{r_n+1}, \dots, \hat{\varepsilon}_n$. We are interested in a version of the Bahadur representation (1.2) that is uniform in z.

Assume that F has a uniformly continuous density f. Then we have

$$\sup_{t\in\mathbb{R}} \big| F(t+s) - F(t) - sf(t) \big| = o(s).$$

From this we get the following uniform version of (2.3),

$$\sup_{z\in\mathbb{R}} \big| F(z + \hat{\vartheta}\hat{\varepsilon}_n) - F(z + \vartheta\varepsilon_n) - (\hat{\vartheta} - \vartheta) Y_n f(z + \vartheta\varepsilon_n) \big| = o_p(n^{-1/2}).$$

By the stochastic equi-continuity of the empirical process we have

$$\sup_{z\in\mathbb{R}} | \, \mathbb{F}(z + \hat{\vartheta}\hat{\varepsilon}_n) - \mathbb{F}(z + \vartheta\varepsilon_n) | = o_p(n^{-1/2}).$$

The desired uniform version of (1.2) thus follows if we show that

$$\sup_{z\in\mathbb{R}} | \, \hat{\mathbb{F}}(z) - \mathbb{F}(z) + (\hat{\vartheta} - \vartheta)\nu f(z) | = o_p(n^{-1/2}), \tag{3.1}$$

where

$$\mathbb{F}(z) = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} \mathbf{1}[\varepsilon_j \le z], \quad z \in \mathbb{R},$$

is the empirical distribution function based on the true innovations. The stochastic expansion was obtained by Boldin (1989) under the assumption that f has a bounded derivative. He also assumed that $E[\varepsilon] = 0$ and therefore $\nu = 0$. Kreiss (1991) generalizes Boldin's result to linear processes with parametric coefficients, including ARMA(p,q) models. Koul (1992), Corollary 7.2.3, shows for ARMA(1,1) that it suffices to assume that f is uniformly continuous. (His assumption that f is almost everywhere positive can be omitted). See also Koul and Ossiander (1994), Koul (2002) and Koul and Ling (2006). We therefore have the following result.

THEOREM 2. *Suppose f is uniformly continuous. Then*

$$\sup_{z\in\mathbb{R}} \big| \hat{\mathbb{F}}(z + \hat{\vartheta}\hat{\varepsilon}_n) - \mathbb{F}(z + \vartheta\varepsilon_n) - (\hat{\vartheta} - \vartheta)(Y_n - \nu) f(z + \vartheta\varepsilon_n) \big| = o_p(n^{-1/2}).$$

For $u \in (0,1)$ let $\psi(u)$ denote the conditional u-quantile of X_{n+1} given $X_1, \dots, X_n$. Write G^{-1} for the right-continuous inverse of a distribution function G. An estimator for $\psi(u)$ is the u-quantile of $\hat{\mathbb{F}}(\cdot + \hat{\vartheta}\hat{\varepsilon}_n)$, which can be written $\hat{\mathbb{F}}^{-1}(u) - \hat{\vartheta}\hat{\varepsilon}_n$. Assume that f is positive. By Proposition 1 of Gill (1989) on compact differentiability of quantile functions we obtain from (3.1) and (2.2) the following Bahadur representation.

THEOREM 3. *Suppose f is uniformly continuous and positive. Let $0 < a \le b < 1$. Then*

$$\sup_{a \le u \le b} \Big| \hat{\mathbb{F}}^{-1}(u) - \hat{\vartheta}\hat{\varepsilon}_n - F^{-1}(u) + \vartheta\varepsilon_n + \frac{1}{f(F^{-1}(u))} \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} (\mathbf{1}[\varepsilon_j \le F^{-1}(u)] - u) - (\hat{\vartheta} - \vartheta)(Y_n - \nu) \Big| = o_p(n^{-1/2}).$$

4. Conditional density

In this section we derive properties of residual-based kernel estimators of the innovation density f that are needed in Section 5. We also apply these properties to estimators of the conditional density of X_{n+1} given $X_1, \dots, X_n$. The required conditions on f are expressed in terms of a norm defined as follows. Let V be a continuous function on $\mathbb{R}$ with $V(0) = 1$ and such that

$$V(x + y) \le V(x)V(y), \quad x, y \in \mathbb{R}, \tag{4.1}$$

$$\sup_{|s| \le 1} V(sx) \le V(x), \quad x \in \mathbb{R}. \tag{4.2}$$

These conditions imply that

$$|V(x + y) - V(x)| \le V(x)(V(y) - 1), \quad x, y \in \mathbb{R}. \tag{4.3}$$

With the function V we associate the *V-norm*

$$\|g\|_V = \int V(x)|g(x)|\, dx.$$

If g has finite V-norm, so does the shifted function $S_t g = g(\cdot - t)$,

$$\|S_t g\|_V \le V(t)\|g\|_V. \tag{4.4}$$

By Lemma 4 in Schick and Wefelmeyer (2006a), the shift is continuous in the V-norm,

$$\lim_{t \to 0} \|S_t g - g\|_V = 0. \tag{4.5}$$

Finally, the convolution $g_1 * g_2$ of two functions g_1 and g_2 with finite V-norms has finite V-norm and we have

$$\|g_1 * g_2\|_V \le \|g_1\|_V \|g_2\|_V.$$

We say g is *V-Lipschitz* (with *constant* L) if

$$\|S_t g - g\|_V \le L|t|V(t), \quad t \in \mathbb{R}.$$

By Lemma 6 in Schick and Wefelmeyer (2006a), if g is absolutely continuous and its a.e. derivative g' has finite V-norm, then g is V-Lipschitz with constant $\|g'\|_V$. Weaker sufficient conditions for the V-Lipschitz property are given in Lemma 4.4 in Schick and Wefelmeyer (2006c). For example, functions of bounded variation are V-Lipschitz for bounded V.

Now let $\hat f$ denote the kernel estimator of f based on the residuals $\hat\varepsilon_{r_n+1}, \dots, \hat\varepsilon_n$,

$$\hat f(y) = \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} k_{b_n}(y - \hat\varepsilon_j), \quad y \in \mathbb{R},$$

where $k_{b_n}(y) = k(y/b_n)/b_n$ for some kernel k and some bandwidth b_n. Let $\tilde f$ denote the kernel estimator based on the true innovations $\varepsilon_{r_n+1}, \dots, \varepsilon_n$,

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

We impose the following conditions on k and f.

(K) *The kernel k is a symmetric density with support $[-1,1]$ and is three times continuously differentiable.*

(F1) *The density f satisfies $\int \big((1+|x|)^\alpha V^2(x) + |x|^\xi\big) f(x)\,dx < \infty$ for some $\alpha > 1$ and some $\xi > 16/7$.*

(F2) *The density f is absolutely continuous, and its a.e. derivative f' has finite V-norm.*

Remark 2. If $V(x) = (1+|x|)^r$ for some non-negative r, then (F1) simplifies to a moment condition. Indeed, (F1) is then equivalent to f having a finite moment of order $\beta = \max\{\xi, 2r+\alpha\}$ for some $\alpha > 1$ and some $\xi > 16/7$. If $r < 9/14$, then $\beta > 16/7$ suffices; if $r \ge 9/14$, then $\beta > 2r+1$ suffices.

Lemma 1. *Suppose* (F1), (F2) *and* (K) *hold and* $b_n \sim (n \log n)^{-1/4}$. *Then*

$$\|\hat f - \tilde f + (\hat\vartheta - \vartheta)\nu f'\|_V = o_p(n^{-1/2}), \tag{4.6}$$

$$\|\hat f' - f'\|_V = o_p(1). \tag{4.7}$$

Proof. Set

$$\hat f_*(y) = \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} k_{b_n}(y - \varepsilon_j - (\hat\vartheta - \vartheta)Y_{j-1}), \quad y \in \mathbb{R}.$$

Let $R_j = \hat\varepsilon_j - \varepsilon_j - (\hat\vartheta - \vartheta)Y_{j-1}$. By (2.11) we have $R_n^* = \max_{r_n < j \le n} |R_j| = o_p(1)$. Relation (2.10) and the root-n consistency of $\hat\vartheta$ imply $S_n^* = \max_{r_n<j\le n} |\hat\vartheta - \vartheta||Y_{j-1}| = o_p(1)$. From (F1) we obtain

$$\frac{1}{n-r_n} \sum_{j=r_n+1}^{n} V^2(\varepsilon_j) = O_p(1).$$

Note that k_{b_n} is V-Lipschitz with constant $\|k'_{b_n}\|_V = O(b_n^{-1})$. This, (4.1) and (4.4) yield

$$\|\hat f - \hat f_*\|_V \le \|k'_{b_n}\|_V \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} |R_j| V(R_j) V(\varepsilon_j + (\hat\vartheta - \vartheta) Y_{j-1})$$
$$\le \|k'_{b_n}\|_V V(R_n^*) V(S_n^*) \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} |R_j| V(\varepsilon_j).$$

The above and (1.1) imply

$$\|\hat f - \hat f_*\|_V = O_p(n^{-1} b_n^{-1}) = o_p(n^{-1/2}). \tag{4.8}$$

Similarly, one verifies

$$\|\hat f' - \hat f'_*\|_V = O_p(n^{-1} b_n^{-2}) = o_p(1). \tag{4.9}$$

It follows with the arguments of Lemmas 4 and 3 in Schick and Wefelmeyer (2006b) that

$$\|\hat f'_* - \tilde f'\|_V = O_p(n^{-1} b_n^{-3}) = o_p(1); \tag{4.10}$$
$$\|\tilde f' - f * k'_{b_n}\|_V = O_p(n^{-1/2} b_n^{-3/2}) = o_p(1). \tag{4.11}$$

Since $f * k'_{b_n} = f' * k_{b_n}$ and f' has finite V-norm, we have

$$\|f * k'_{b_n} - f'\|_V \to 0. \tag{4.12}$$

It follows from (4.9), (4.10), (4.11) and (4.12) that (4.7) holds. As in the proof of Lemma 5 in Schick and Wefelmeyer (2006b) we obtain

$$\|\hat f_* - \tilde f + (\hat\vartheta - \vartheta)\bar\Gamma\|_V = o_p(n^{-1/2}) \quad \text{with } \bar\Gamma = f * k'_{b_n} \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} Y_{j-1}. \tag{4.13}$$

Relation (4.12) implies that $\|\bar\Gamma - \nu f'\|_V = o_p(1)$. This, (4.8) and (4.13) imply (4.6). □

An estimator for the conditional density of X_{n+1} given $X_1, \dots, X_n$ at y is $\hat f(y + \hat\vartheta \hat\varepsilon_n)$. We show that it admits a stochastic expansion similar to expansion (4.6) for $\hat f$.

THEOREM 4. *Suppose* (F1), (F2) *and* (K) *hold and* $b_n \sim (n \log n)^{-1/4}$. *Then*

$$\|\hat f(\cdot + \hat\vartheta \hat\varepsilon_n) - \tilde f(\cdot + \vartheta \varepsilon_n) - (\hat\vartheta - \vartheta)(Y_n - \nu) f'(\cdot + \vartheta \varepsilon_n)\|_V = o_p(n^{-1/2}).$$

PROOF. Let $\Delta = \hat\vartheta \hat\varepsilon_n - \vartheta \varepsilon_n$. Then $\hat f(y + \hat\vartheta \hat\varepsilon_n) = \hat f(y + \vartheta \varepsilon_n + \Delta)$. In view of (2.2), (4.4), (4.6) and $V(\vartheta \varepsilon_n) = O_p(1)$ it suffices to show $\|\hat\psi\|_V = o_p(n^{-1/2})$ with

$$\hat\psi(y) = \hat f(y + \Delta) - \hat f(y) - \Delta f'(y) = \Delta \int_0^1 \big(\hat f'(y + s\Delta) - f'(y)\big)\, ds.$$

We have the bound

$$\|\hat\psi\|_V \le |\Delta| \int_0^1 \big(\|\hat f'(\cdot + s\Delta) - f'(\cdot + s\Delta)\|_V + \|f'(\cdot + s\Delta) - f'\|_V\big)\, ds.$$

Hence the desired $\|\hat\psi\|_V = o_p(n^{-1/2})$ follows from (4.4), (4.5), (4.7) and $\Delta = O_p(n^{-1/2})$. □

5. Smoothed predictors

In this section we obtain a uniform version of the stochastic expansion (1.2) over large classes of functions that are not necessarily smooth. For this we require some smoothness of the innovation density and work with a smoothed version of $\hat{q}(h)$, namely

$$\hat{q}_s(h) = \int h(y - \hat{\vartheta}\hat{\varepsilon}_n)\hat{f}(y)\, dy,$$

where $\hat{f}$ is the residual-based kernel estimator of Section 4. It is easy to verify that

$$\hat{q}_s(h) = \hat{q}(h_n),$$

where $h_n = h * k_{b_n}$ is the convolution of h and k_{b_n}, which we can write as

$$h_n(y) = \int h(y - b_n u)k(u)\, du, \quad y \in \mathbb{R}.$$

We show that $\hat{q}_s(h)$ has the same stochastic expansion as $\hat{q}(h)$, uniformly over certain classes of functions h.

As shown in Lemma 6 of Schick and Wefelmeyer (2006a), it follows from (F2) that

$$\|S_t f - f + tf'\|_V = o(t).$$

From this, (4.4) and (4.5) we derive that, for each measurable h bounded by a multiple of V, the function q_h is continuously differentiable with derivative

$$q'_h(x) = \int h(y - x)f'(y)\, dy = \int h(y)f'(y + x)\, dy, \quad x \in \mathbb{R}.$$

Theorem 5. *Suppose* (F1), (F2) *and* (K) *hold and* $b_n \sim (n \log n)^{-1/4}$. *Let* $\mathcal{H}$ *be a class of measurable functions that has envelope* cV *for some positive* c *and such that for all* $C < \infty$ *the class of shifts* $\mathcal{H}_C = \{h(\cdot - t) : h \in \mathcal{H}, |t| \le C\}$ *is F-Donsker, and*

$$\varrho_n(C) := \sup_{h \in \mathcal{H}_C} \Big| \int h(y)(f * k_{b_n}(y) - f(y))\, dy \Big| = o(n^{-1/2}). \tag{5.1}$$

Then

$$\sup_{h \in \mathcal{H}} \Big| \hat{q}_s(h) - \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} h(\varepsilon_j - \vartheta\varepsilon_n) - (\hat{\vartheta} - \vartheta)(Y_n - \nu)q'_h(\vartheta\varepsilon_n) \Big| = o_p(n^{-1/2}). \tag{5.2}$$

Proof. Let $h_n = h * k_{b_n}$. Then we can show that

$$\int h(y)\tilde{f}(y + \vartheta\varepsilon_n)\, dy = \frac{1}{n - r_n} \sum_{j=r_n+1}^{n} h_n(\varepsilon_j - \vartheta\varepsilon_n).$$

Using this and the above representation for q'_h we can express the term inside the absolute values of (5.2) as the sum of the following three terms:

$$\hat{T}_1(h) = \int h(y)\Big(\hat{f}(y+\hat{\vartheta}\hat{\varepsilon}_n) - \tilde{f}(y+\vartheta\varepsilon_n) - (\hat{\vartheta}-\vartheta)(Y_n-\nu)f'(y+\vartheta\varepsilon_n)\Big)\,dy,$$

$$\hat{T}_2(h) = \frac{1}{n-r_n}\sum_{j=r_n+1}^{n}\big(h_n(\varepsilon_j-\vartheta\varepsilon_n) - h(\varepsilon_j-\vartheta\varepsilon_n)\big) - \hat{T}_3(h)$$

$$\hat{T}_3(h) = \int \big(h_n(y-\vartheta\varepsilon_n) - h(y-\vartheta\varepsilon_n)\big)f(y)\,dy = \int h(y-\vartheta\varepsilon_n)(f * k_{b_n}(y) - f(y))\,dy.$$

Thus it suffices to show that

$$\hat{D}_i = \sup_{h\in\mathcal{H}}|\hat{T}_i(h)| = o_p(n^{-1/2}), \quad i = 1,\dots,3.$$

Since $\mathcal{H}$ has envelope cV, we obtain $\hat{D}_1 = o_p(n^{-1/2})$ from Theorem 4. For $\eta > 0$ and $C < \infty$,

$$\begin{aligned} P(\hat{D}_3 \ge \eta n^{-1/2}) &\le P(|\vartheta\varepsilon_n| > C) + \mathbf{1}[\varrho_n(C) > \eta n^{-1/2}] \\ &\le \frac{\vartheta^2 E[\varepsilon^2]}{C^2} + \mathbf{1}[\varrho_n(C) > \eta n^{-1/2}]. \end{aligned}$$

This shows that $\hat{D}_3 = o_p(n^{-1/2})$.

Consider the stochastic process

$$\mathbb{A}(x,h) = \frac{1}{n-r_n}\sum_{j=r_n+1}^{n}(h(\varepsilon_j - x) - E[h(\varepsilon - x)]), \quad x\in\mathbb{R},\ h\in\mathcal{H}.$$

We can write

$$\hat{T}_2(h) = \mathbb{A}(\vartheta\varepsilon_n, h_n) - \mathbb{A}(\vartheta\varepsilon_n, h) = \int (\mathbb{A}(\vartheta\varepsilon_n + b_n u, h) - \mathbb{A}(\vartheta\varepsilon_n, h))k(u)\,du$$

and obtain

$$P(\hat{D}_2 > \eta n^{-1/2}) \le P(|\vartheta\varepsilon_n| > C) + P\Big(\sup_{h\in\mathcal{H},|x|\le C,|t|\le b_n}|\mathbb{A}(x+t,h) - \mathbb{A}(x,h)| > \eta n^{-1/2}\Big)$$

for all $\eta > 0$ and $C < \infty$. Define the empirical process

$$\mathbb{B}(h) = \frac{1}{n-r_n}\sum_{j=r_n+1}^{n}(h(\varepsilon_j) - E[h(\varepsilon)]), \quad h\in\mathcal{H}_C.$$

For $|x| \le C$ we can write $\mathbb{A}(x,h) = \mathbb{B}(S_x h)$. Thus we have for all $\eta > 0$ and $C < \infty$,

$$P(\hat{D}_2 > \eta n^{-1/2}) \le \frac{\vartheta^2 E[\varepsilon_0^2]}{C^2} + P\Big(\sup_{h\in\mathcal{H}_C,|t|\le b_n}|\mathbb{B}(S_t h) - \mathbb{B}(h)| > \eta n^{-1/2}\Big).$$

Write d_f for the metric induced by the $L_2(F)$-norm, i.e.

$$d_f(g_1,g_2) = \Big(\int (g_1(y) - g_2(y))^2 f(y)\,dy\Big)^{1/2}, \qquad g_1, g_2 \in L_2(F).$$

Since $\mathcal{H}_C$ is an F-Donsker class, we have *stochastic equi-continuity*: For every $\eta > 0$ there is a $\delta > 0$ (which depends on η and C) such that

$$\sup_n P\Big(\sup_{h,g\in\mathcal{H}_C, d_f(h,g)<\delta} |\mathbb{B}(h) - \mathbb{B}(g)| > \eta n^{-1/2}\Big) < \eta.$$

By (4.3) we have

$$\int (V(y+t) - V(y))^2 f(y)\,dy \le (V(t)-1)^2 \int V^2(y) f(y)\,dy \to 0, \qquad t \to 0.$$

Note that $\mathcal{H}_C$ has envelope $cV(C)V$. Thus we get from Lemma 7.1 in Müller, Schick and Wefelmeyer (2006) that

$$\sup_{h\in\mathcal{H}_C} d_f(S_t h, h) \to 0, \qquad t \to 0.$$

In view of this we have

$$P\Big(\sup_{h\in\mathcal{H}_C, |t|\le b_n} |\mathbb{B}(S_t h) - \mathbb{B}(h)| > \eta n^{-1/2}\Big) \to 0.$$

Since this is true for every $C < \infty$, we obtain $\hat{D}_2 = o_p(n^{-1/2})$. □

Remark 3. Under the other conditions of Theorem 5, a sufficient condition for (5.1) is that f' is V-Lipschitz. In the terminology of Schick and Wefelmeyer (2006a), f is then *V-smooth of order 2*. Hence their Lemma 7 yields that $\|f * k_{b_n} - f\|_V = O(b_n^2) = o_p(n^{-1/2})$. Since $\mathcal{H}_C$ has envelope $cV(C)V$, the desired (5.1) follows as $\varrho_n(C) \le cV(C)\|f * k_{b_n} - f\|_V$.

Remark 4. Suppose we have

$$\sup_{h\in\mathcal{H}_C} |h(y+t) - h(y)| \le M_C V(y)|t|, \quad y \in \mathbb{R}, |t| \le C. \tag{5.3}$$

Then (5.1) follows from (F2). Indeed, using the absolute continuity of f and the fact that k has mean zero, we can write

$$f * k_{b_n}(y) - f(y) = \int \int_0^1 b_n u (f'(y - s b_n u) - f'(u))\,ds\,k(u)\,du$$

and thus obtain

$$\begin{aligned}\int h(y)(f * k_{b_n}(y) - f(y))\,dy &= \int \int b_n u \int_0^1 h(y)(f'(y - s b_n u) - f'(y))\,ds\,k(u)\,du\,dy \\ &= \int_0^1 \int b_n u \int (h(y + s b_n u) - h(y)) f'(y)\,dy\,k(u)\,du\,ds.\end{aligned}$$

Using (5.3) we obtain for large n that

$$\rho_n(C) \le M_C \|f'\|_V \int u^2 k(u)\,du\; b_n^2$$

and thus (5.1) by the choice of bandwidth in Theorem 5.

REMARK 5. Theorem 5 applies to estimating the conditional distribution function. Let $h(y) = h_z(y) = \mathbf{1}[y \le z]$ with z. Then $q(h_z) = P(X_{n+1} \le z | X_1, \dots, X_n)$. We have

$$q_{h_z}(x) = F(z+x), \qquad q'_{h_z}(x) = f(z+x).$$

Our estimator is

$$\hat{q}_s(h_z) = \int_{-\infty}^{z+\hat{\vartheta}\hat{\varepsilon}_n} \hat{f}(y)\,dy = \frac{1}{n-r_n} \sum_{j=r_n+1}^{n} K_{b_n}(z + \hat{\vartheta}\hat{\varepsilon}_n - \hat{\varepsilon}_j),$$

where K is the distribution function of k and $K_b(y) = K(y/b)$. The class $\mathcal{H} = \{\mathbf{1}[\cdot \le z] : z \in \mathbb{R}\}$ is closed under shifts and F-Donsker. Here we can take $V(x) = 1$. Then (F1) holds if f has a moment of order $\xi > 16/7$, and (F2) holds if f is absolutely continuous with integrable a.e. derivative. If f' can be chosen to be bounded, then

$$\rho_n(C) = \sup_{z \in \mathbb{R}} |F * k_{b_n}(z) - F(z)| \le b_n^2 \|f'\|_\infty \int y^2 k(y)\,dy$$

and (5.1) holds for the choice of b_n in Theorem 5. For the *unsmoothed* estimator $\hat{\mathbb{F}}(z + \hat{\vartheta}\hat{\varepsilon}_n)$ of $q(h_z)$ we obtain from Theorem 2 the same Bahadur representation as in Theorem 5 under the assumption that f is uniformly continuous.

REMARK 6. Theorem 5 applies to estimating conditional absolute moments. Let $\beta \ge 1$ and $h(y) = |y|^\beta$. Then $q(h) = E(|X_{n+1}|^\beta | X_1, \dots, X_n)$. We have

$$q_h(x) = \int |y-x|^\beta f(y)\,dy = E[|\varepsilon - x|^\beta], \qquad q'_h(x) = \int |y-x|^\beta f'(y)\,dy.$$

Our estimator is

$$\hat{q}_s(h) = \int |y - \hat{\vartheta}\hat{\varepsilon}_n|^\beta \hat{f}(y)\,dy.$$

We can take $V(x) = (1+|x|)^\beta$. As $\beta \ge 1$, (F1) holds if f has a moment of order greater than $2\beta + 1$. Here $\mathcal{H}$ consists of a single function, and $\mathcal{H}_C$ is clearly F-Donsker for all $C < \infty$. It is also easy to check that (5.3) holds for each $C < \infty$, so that (F2) implies (5.1). For the *unsmoothed* estimator $\hat{q}(h)$, we obtain from Theorem 1 the same Bahadur representation as in Theorem 5, assuming only that the innovation distribution has a moment of order 2β.

References

[1] Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, Chichester.

[2] Boldin, M. V. (1989). On testing hypotheses in the sliding average scheme by the Kolmogorov–Smirnov and ω^2 tests. *Theory Probab. Appl.* **34**, 699–704.

[3] Collomb, G. (1984). Propriétés de convergence presque complète de prédicteur à noyau. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* **66**, 441–460.

[4] Delecroix, M. and Rosa, A. C. (1995). Ergodic processes prediction via estimation of the conditional distribution function. *Publ. Inst. Statist. Univ. Paris* **39**, 35–56.

[5] Gannoun, A., Saracco, J. and Yu, K. (2003). Nonparametric prediction by conditional median and quantiles. *J. Statist. Plann. Inference* **117**, 207–223.

[6] 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.

[7] Koul, H. L. (1992). *Weighted Empiricals and Linear Models.* Institute of Mathematical Statistics Lecture Notes—Monograph Series 21, Institute of Mathematical Statistics, Hayward, CA.

[8] Koul, H. L. (2002). *Weighted Empirical Processes in Dynamic Nonlinear Models.* Lecture Notes in Statistics 166. Springer-Verlag, New York.

[9] Koul, H. L. and Ling, S. (2006). Fitting an error distribution in some heteroscedastic time series models. *Ann. Statist.* **34**, 994–1012.

[10] Koul, H. L. and Ossiander, M. (1994). Weak convergence of randomly weighted dependent residual empiricals with applications to autoregression. *Ann. Statist.* **22**, 540–562.

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

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

[13] Müller, U. U., Schick, A. and Wefelmeyer, W. (2006). Efficient prediction for linear and nonlinear autoregressive models. To appear in: *Ann. Statist.* **34**.

[14] Polonik, W. and Yao, Q. (2000). Conditional minimum volume predictive regions for stochastic processes. *J. Amer. Statist. Assoc.* **95**, 509–519.

[15] Polonik, W. and Yao, Q. (2002). Set-indexed conditional empirical and quantile processes based on dependent data. *J. Multivariate Anal.* **80**, 234–255.

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

[17] 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.

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

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

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

[21] Schick, A. and Wefelmeyer, W. (2004). Root n consistent and optimal density estimators for moving average processes. *Scand. J. Statist.* **31**, 63–78.

[22] Schick, A. and Wefelmeyer, W. (2006a). Root-n consistent density estimators of convolutions in weighted L_1-norms. To appear in: *J. Statist. Plann. Inference.*

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

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

[25] Tran, L. T. (1993). Nonparametric function estimation for time series by local average estimators. *Ann. Statist.* **21**, 1040–1057.

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

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

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

[29] Zhou, Y. and Liang, H. (2000). Asymptotic normality for L_1 norm kernel estimator of conditional median under α-mixing dependence. *J. Multivariate Anal.* **73**, 136–154.

[30] Zhou, Y. and Liang, H. (2003). Asymptotic properties for L_1 norm kernel estimator of conditional median under dependence. *J. Nonparametr. Stat.* **15**, 205–219.

Anton Schick, Binghamton University, Department of Mathematical Sciences, Binghamton, NY 13902-6000, USA

Wolfgang Wefelmeyer, Mathematisches Institut, Universität zu Köln, Weyertal 86-90, 50931 Köln, Germany