WEIGHTED RESIDUAL-BASED DENSITY ESTIMATORS FOR NONLINEAR AUTOREGRESSIVE MODELS

URSULA U. MÜLLER, ANTON SCHICK, AND WOLFGANG WEFELMEYER

ABSTRACT. This paper considers residual-based and randomly weighted kernel estimators for innovation densities of nonlinear autoregressive models. The weights are chosen to make use of the information that the innovations have mean zero. Rates of convergence are obtained in weighted L_1-norms. These estimators give rise to smoothed and weighted empirical distribution functions and moments. It is shown that the latter are efficient if an efficient estimator for the autoregression parameter is used to construct the residuals.

AMS 2000 subject classifications. Primary 62G07, 62G20, 62M05, 62M10.

Key words and Phrases. Weighted residual-based density estimator, empirical likelihood estimator, Owen estimator, plug-in estimator, efficient estimator.

1. INTRODUCTION

Consider a nonlinear autoregressive model $X_i = r_\vartheta(\mathbf{X}_{i-1}) + \varepsilon_i$ of order p, where $\mathbf{X}_{i-1} = (X_{i-p}, \dots, X_{i-1})$ and ϑ is a q-dimensional parameter. Assume that the innovations ε_i are independent and identically distributed (i.i.d.) and have mean zero, finite variance and positive density f. We are interested in estimating f and study weighted kernel estimators based on estimated innovations $\hat\varepsilon_i = X_i - r_{\hat\vartheta}(\mathbf{X}_{i-1})$,

$$\hat{\hat{f}}_w(y) = \frac{1}{n} \sum_{i=1}^n \hat{w}_i k_{b_n}(y - \hat\varepsilon_i),$$

where $k_{b_n}(y) = k(y/b_n)/b_n$ for an appropriate kernel k and bandwidth b_n, and $\hat{w}_i$ are nonnegative random weights that average to one. The ordinary kernel estimator uses weights $\hat{w}_i = 1$. Residual-based density estimators in time series are studied in Robinson (1986, 1987) and Liebscher (1999). Since the innovations have mean zero, we take weights for which our kernel estimator also has mean zero, i.e., $\int y \hat{\hat{f}}_w(y)\, dy = 0$. Motivated by Owen (1988, 2001), we take $\hat{w}_i$ of the form

$$(1.1) \qquad \hat{w}_i = \frac{1}{1 + \hat\lambda \hat\varepsilon_i},$$

where $\hat\lambda$ is chosen such that $\sum_{i=1}^n \hat{w}_i \hat\varepsilon_i = 0$. Our density estimator can be written as $\hat{\hat{f}}_w(y) = \int k_{b_n}(y - u)\, d\hat{\hat{F}}_w(u)$, where $\hat{\hat{F}}_w(t) = \frac{1}{n} \sum_{i=1}^n \hat{w}_i \mathbf{1}[\hat\varepsilon_i \leq t]$ is Owen's empirical likelihood estimator of the distribution function F, here based on the residuals $\hat\varepsilon_i$. In the literature, weighted kernel

density estimators have been studied for i.i.d. observations, e.g., for the actual innovations; see Chen (1997), Zhang (1998), and Hall and Presnell (1999).

We study the behavior of our density estimators in the V-norm

$$\|g\|_V = \int V(y)|g(y)|\,dy$$

for a measurable function $V \geq 1$. The choice $V = 1$ yields the usual L_1-norm. Convergence in this norm suffices when estimating the innovation distribution function. The choice $V(y) = (1 + |y|)^\gamma$ for some $\gamma > 0$ is useful when estimating moments of the innovation distribution.

We first derive, in Section 2, convergence rates for weighted kernel estimators based on i.i.d. observations. These results will be used in Section 3 to establish our main result, convergence rates of residual-based weighted kernel estimators. In the V-norm, the difference between the weighted residual-based kernel estimator and the (unweighted) kernel estimator based on the actual innovations is of order $n^{-1/2}$, which is faster than the convergence rate of kernel estimators. Nevertheless, if kernel estimators are plugged into smooth functionals of the density, the resulting plug-in estimators have rate $n^{-1/2}$. Moreover, using the above weights leads to an asymptotic variance reduction, and even to efficient estimators if efficient estimators of the autoregression parameter are used. This will be illustrated in the last two sections. In Sections 4 and 5 we use $\hat{\hat{f}}_w$ to construct estimators $\int_{-\infty}^t \hat{\hat{f}}_w(y)\,dy$ and $\int y^m \hat{\hat{f}}_w(y)\,dy$ for the distribution function $F(t)$ and the m-th moment of the innovations. These estimators will be $n^{1/2}$-consistent if $\hat{\vartheta}$ is, and will be efficient if $\hat{\vartheta}$ is. In these sections we also demonstrate in some special cases that these weighted linear functionals have significantly smaller asymptotic variance than their unweighted counterparts. Müller, Schick and Wefelmeyer (2003) use the results of Section 3 to obtain $n^{-1/2}$-consistent and efficient estimators for conditional expectations.

2. Weighted kernel estimators based on i.i.d. observations

Let $\varepsilon_1, \ldots, \varepsilon_n$ be i.i.d. observations with mean zero, finite variance σ^2 and density f. We write $\hat{f}$ for the usual kernel density estimator and $\hat{f}_w$ for the weighted kernel estimator:

$$\hat{f}(y) = \frac{1}{n}\sum_{i=1}^n k_{b_n}(y - \varepsilon_i) \quad \text{and} \quad \hat{f}_w(y) = \frac{1}{n}\sum_{i=1}^n w_i k_{b_n}(y - \varepsilon_i).$$

We restrict attention to positive weights w_i of the form

$$w_i = \frac{1}{1 + \lambda\varepsilon_i},$$

where λ is chosen such that $\sum_{i=1}^n w_i\varepsilon_i = 0$. As shown by Owen (2001), such a λ exists with probability tending to one. Our density estimator can be written as $\hat{f}_w(y) = \int k_{b_n}(y - u)\,d\hat{F}_w(u)$, where $\hat{F}_w(t) = \frac{1}{n}\sum_{i=1}^n w_i\mathbf{1}[\varepsilon_i \leq t]$ is Owen's empirical likelihood estimator of $F(t)$.

Let $V \geq 1$ be a measurable function. We assume throughout that V is f-square integrable. We study the behavior of kernel density estimators in the V-norm. For this we require additional

assumptions on V. Let D be the function defined by

$$D(s) = \sup_{|t| \le |s|} \sup_{y \in \mathbb{R}} \frac{|V(y+t) - V(y)|}{V(y)}, \quad s \in \mathbb{R},$$

Condition V1. The function D is continuous at 0.

Condition V2. For some $\alpha > 1$ and $W(y) = (1 + |y|)^\alpha V^2(y)$, $\|f\|_W = \int W(y) f(y) \, dy < \infty$.

These two conditions are met with $\alpha = 2$ by $V(x) = 1$ as we have assumed f to have finite variance. We are also interested in the choice $V(x) = (1 + |x|)^\gamma$ for $\gamma > 0$. One verifies that this function satisfies Condition V1 with $D(s) \le |s|\gamma$ for $0 < \gamma < 1$, and with $D(s) \le |s|\gamma(1 + |s|)^{\gamma-1}$ for $\gamma \ge 1$. Condition V2 holds provided $\int |y|^{2\gamma+\alpha} f(y) \, dy$ is finite for some $\alpha > 1$. For $\gamma < 1/2$, this integral is finite for $\alpha < 2(1 - \gamma)$ as f has finite variance.

Throughout we impose the following conditions on the bandwidth and the kernel.

Condition K. The bandwidth b_n satisfies $b_n \to 0$ and $nb_n \to \infty$. The kernel k is a bounded measurable function,

$$(2.1) \qquad \int k(u) \, du = 1, \quad \int u k(u) \, du = 0,$$

and, for some $\beta > 3$,

$$(2.2) \qquad \int (1 + |u|)^\beta (1 + D(u))^2 |k(u)| \, du < \infty.$$

The requirements on k are met by a symmetric bounded density with compact support contained in $\{D < \infty\}$, but Condition K allows for kernels of higher order and does not require a compact support if $D(s)$ is finite for all $s \in \mathbb{R}$. For the choice $V(x) = (1 + |x|)^\gamma$ for some $\gamma > 0$, (2.2) is implied by $\int (1 + |u|)^{2\gamma_1+\beta} |k(u)| \, du < \infty$ with $\gamma_1 = \max\{1, \gamma\}$.

It follows from Condition V1 that

$$(2.3) \qquad V(y + s) \le (1 + D(s)) V(y), \quad s, y \in \mathbb{R}.$$

Since $\|\hat{f}\|_V \le \frac{1}{n} \sum_{i=1}^n \int V(y) |k_{b_n}(y - \varepsilon_i)| \, dy = \frac{1}{n} \sum_{i=1}^n \int V(\varepsilon_i + b_n u) |k(u)| \, du$, we obtain from (2.3) that

$$(2.4) \qquad \|\hat{f}\|_V \le \|k\|_{1+D} \frac{1}{n} \sum_{i=1}^n V(\varepsilon_i), \quad b_n \le 1.$$

Note that $\|k\|_{1+D}$ is finite under (2.2). The expected value of $\hat{f}(y)$ is $f * k_{b_n}(y) = \int f(y - b_n u) k(u) \, du$. Since $\int V(y) |f * k_{b_n}(y)| \, dy \le \iint V(y + b_n u) f(y) |k(u)| \, du \, dy$, we derive from (2.3) that

$$(2.5) \qquad \|f * k_{b_n}\|_V \le \|k\|_{1+D} \|f\|_V, \quad b_n \le 1.$$

The following lemma shows that $\|f * k_{b_n} - f\|_V \to 0$.

Lemma 2.1. *Suppose Condition V1 holds. Then, for every (measurable) function g with finite V-norm and every (measurable) function h with finite $(1+D)$-norm, we have*

$$(2.6) \qquad \int V(y) \int |g(y-bu) - g(y)||h(u)| \, du \, dy \to 0 \quad as \ b \to 0.$$

Proof. Since g has finite V-norm, the product Vg is integrable. Thus the map $s \to \int |(Vg)(y-s) - (Vg)(y)| \, dy$ is bounded by $2\|g\|_V$, and continuous in view of the L_1-continuity of translation, see Rudin ((1974), Theorem 9.5). It now follows from the Lebesgue Dominated Convergence Theorem that $I_1(b) = \iint |(Vg)(y-bu) - (Vg)(y)| \, dy \, |h(u)| \, du \to 0$ as $b \to 0$. By Condition V1, for each $u \in \mathbb{R}$ we have $D(bu) \le D(u)$ for $|b| \le 1$ and $D(bu) \to 0$ as $b \to 0$. Hence the substitution $v = y - bu$, inequality (2.3), and again the Lebesgue Dominated Convergence Theorem give that

$$I_2(b) = \iint |V(y) - V(y-bu)||g(y-bu)| \, dy \, |h(u)| \, du \le \|g\|_V \int D(bu)|h(u)| \, du \to 0 \quad as \ b \to 0.$$

Since the left-hand side of (2.6) is bounded by $I_1(b) + I_2(b)$, the desired result follows. $\qquad \square$

We say a function g is *V-Lipschitz* if there is a positive constant L such that

$$\|g(\cdot - s) - g\|_V = \int V(y)|g(y-s) - g(y)| \, dy \le L(1+D(s))|s|, \quad s \in \mathbb{R}.$$

If f is V-Lipschitz and $\int (1+D(s))|sk(s)| \, ds$ is finite, then $\|f * k_{b_n} - f\|_V = O(b_n)$. A slightly stronger result is possible if f is V-smooth. We say a function g is *V-smooth* if g is absolutely continuous and its almost everywhere derivative g' has finite V-norm. It is easy to check that a V-smooth function is V-Lipschitz.

Lemma 2.2. *Suppose Conditions K and V1 hold. If f is V-smooth, then $\|f * k_{b_n} - f\|_V = o(b_n)$. If, in addition, f' is V-Lipschitz, then $\|f * k_{b_n} - f\|_V = O(b_n^2)$.*

Proof. Since f is absolutely continuous, we have $f(y-s) - f(y) = -sf'(y) - s\int_0^1 (f'(y-ts) - f'(y)) \, dt$. This and (2.1) give

$$f * k_{b_n}(y) - f(y) = -b_n \int_0^1 \int (f'(y - tb_n u) - f'(y))uk(u) \, du \, dt.$$

It follows from (2.2) that $\int (1+D(s))|sk(s)| \, ds$ is finite. Now use the previous lemma and the Lebesgue Dominated Convergence Theorem to conclude that $\|f * k_{b_n} - f\|_V = o(b_n)$. If f' is V-Lipschitz, we get the faster rate $O(b_n^2)$ using again (2.2). $\qquad \square$

Even faster rates are possible if we require f to be V-smooth of higher order. We say a measurable function g is *V-smooth of order zero* if $\|g\|_V < \infty$ and define V-smoothness of order m for positive integers m recursively: g is *V-smooth of order* $m+1$ if g is absolutely continuous with an almost sure derivative g' that is V-smooth of order m. In particular, a V-smooth function is V-smooth of order one. We say the kernel is of *type m* if m is a positive integer and

$$\int u^i k(u) \, du = 0, \quad i = 1, \dots, m, \quad \text{and} \quad \int (1+D(u))(1+|u|)^{m+1}|k(u)| \, du < \infty.$$

The following result is now immediate.

Lemma 2.3. *Suppose Conditions K and V1 hold. Let f be V-smooth of order m for some $m > 1$, and let k be of type m. Then $\|f * k_{b_n} - f\|_V = o(b_n^m)$. If, in addition, $f^{(m)}$ is V-Lipschitz, then $\|f * k_{b_n} - f\|_V = O(b_n^{m+1})$.*

It follows from (2.3) and simple calculations that

$$(2.7) \qquad W(y + s) \leq D_\alpha(s) W(y), \quad s, y \in \mathbb{R},$$

with $D_\alpha(s) = (1 + |s|)^\alpha (1 + D(s))^2$, $s \in \mathbb{R}$. From this we now derive inequalities analogous to (2.4) and (2.5),

$$(2.8) \qquad \|\hat{f}\|_W \leq \|k\|_{D_\alpha} \frac{1}{n} \sum_{i=1}^n W(\varepsilon_i), \quad b_n \leq 1,$$

$$(2.9) \qquad \|f * k_{b_n}\|_W \leq \|k\|_{D_\alpha} \|f\|_W, \quad b_n \leq 1.$$

Note that $\|k\|_{D_\alpha}$ is finite under (2.2) for $\alpha \leq \beta$.

The following result is known for the case $V = 1$; see Devroye (1992).

Lemma 2.4. *Suppose Conditions K, V1 and V2 hold. Then $\|\hat{f} - f * k_{b_n}\|_V = O_p(n^{-1/2} b_n^{-1/2})$.*

Proof. The Cauchy–Schwarz inequality yields that, for measurable g,

$$(2.10) \qquad \|g\|_V^2 \leq C_\alpha \|g^2\|_W$$

with $C_\alpha = \int (1 + |y|)^{-\alpha} \, dy$. We may assume that $b_n \leq 1$ and that $\alpha \leq \beta$. Since k is bounded, the latter and (2.2) imply that $\|k^2\|_{D_\alpha}$ is finite. We calculate

$$nE[\|(\hat{f} - f * k_{b_n})^2\|_W] = \|nE[(\hat{f} - f * k_{b_n})^2]\|_W \leq \|k_{b_n}^2 * f\|_W \leq b_n^{-1} \|k^2\|_{D_\alpha} \|f\|_W.$$

Thus, the above inequalities yield the bound $E[\|\hat{f} - f * k_{b_n}\|_V^2] = O(n^{-1} b_n^{-1})$ which implies the desired result. $\qquad \square$

If follows from Lemmas 2.1 and 2.4 that $\|\hat{f} - f\|_V = o_p(1)$ for every V that satisfies Conditions V1 and V2. Actually, a stronger result is possible. Let $V_*(y) = (1 + |y|) V(y)$, $y \in \mathbb{R}$.

Corollary 2.1. *Suppose Conditions K, V1 and V2 hold. Then $\|\hat{f} - f\|_{V_*} = o_p(1)$.*

Proof. An application of the Cauchy–Schwarz inequality yields

$$\|\hat{f} - f\|_{V_*}^2 \leq \int (1 + |y|) V^2(y) |\hat{f}(y) - f(y)| \, dy \int (1 + |y|) |\hat{f}(y) - f(y)| \, dy.$$

The first factor on the right-hand side is bounded by $\|\hat{f}\|_W + \|f\|_W$, which is $O_p(1)$ in view of (2.8) and Condition V2. Using the Cauchy–Schwarz inequality again we find that the square of the second factor is bounded by $\int (1 + |y|)^2 (|\hat{f}(y)| + f(y)) \, dy \int |\hat{f}(y) - f(y)| \, dy$. Since f and $|k|$ have finite second moments, $\int (1 + |y|)^2 (|\hat{f}(y)| + f(y)) \, dy = O_p(1)$. The desired result now follows as $\|\hat{f} - f\|_1 \leq \|\hat{f} - f\|_V = o_p(1)$. $\qquad \square$

Of course rates in the V-norm are also possible. Combining Lemmas 2.3 and 2.4, we obtain the following result which gives rates analogous to those for pointwise estimation of densities.

Corollary 2.2. *Suppose Conditions K, V1 and V2 hold. Let f be V-smooth of order m for some non-negative integer m, and let $f^{(m)}$ be V-Lipschitz. Assume that k is of type m if $m > 1$. Let $b_n \sim n^{-1/(2m+3)}$. Then $\|\hat{f} - f\|_V = O_p(n^{-(m+1)/(2m+3)})$.*

Let us now look at the weighted density estimator $\hat{f}_w$. Owen ((2001), pp. 219–221) has shown that the λ appearing in the definition of the weights w_i satisfies

$$(2.11) \qquad \lambda = \sigma^{-2} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i + o_p(n^{-1/2}),$$

and that the weights w_i are uniformly close to one:

$$(2.12) \qquad w_* = \max_{1 \le i \le n} |w_i - 1| = o_p(1).$$

We use this to compare $\hat{f}_w$ and $\hat{f}$.

Lemma 2.5. *Suppose Conditions K, V1 and V2 hold. Then $\|\hat{f}_w - \hat{f} + \lambda\psi\|_V = o_p(n^{-1/2})$, where $\psi(y) = yf(y)$.*

Proof. Since $w_i - 1 = -\lambda\varepsilon_i w_i$, we can write $\hat{f}_w - \hat{f} = -\lambda\hat{\psi}_w$, where $\hat{\psi}_w(y) = \frac{1}{n}\sum_{i=1}^{n} w_i\varepsilon_i k_{b_n}(y - \varepsilon_i)$. As $n^{1/2}\lambda = O_p(1)$ by (2.11), it suffices to show that $\|\hat{\psi}_w - \psi\|_V = o_p(1)$. Let $\hat{\psi}$ be the version of $\hat{\psi}_w$ with $w_i = 1$ for all i. We have

$$\|\hat{\psi}_w - \hat{\psi}\|_V \le w_* \frac{1}{n}\sum_{i=1}^{n} |\varepsilon_i| \int V(y)|k_{b_n}(y - \varepsilon_i)|\,dy = w_* \frac{1}{n}\sum_{i=1}^{n} |\varepsilon_i| \int V(\varepsilon_i + b_n u)|k(u)|\,du.$$

Thus, in view of (2.3), (2.12), and $E[|\varepsilon|V(\varepsilon)] < \infty$, we obtain $\|\hat{\psi}_w - \hat{\psi}\|_V = o_p(1)$. Let $\tilde{\psi}(y) = y\hat{f}(y)$. It follows from Corollary 2.1 that $\|\tilde{\psi} - \psi\|_V \le \|\hat{f} - f\|_{V_*} = o_p(1)$. Thus we are left to show that $\|\hat{\psi} - \tilde{\psi}\|_V = o_p(1)$.

It is easy to check that $\tilde{h}(y) = (\hat{\psi}(y) - \tilde{\psi}(y))/b_n = \frac{1}{n}\sum_{i=1}^{n} \tilde{k}_{b_n}(y - \varepsilon_i)$ with $\tilde{k}(y) = -yk(y)$ and $\tilde{k}_{b_n} = \tilde{k}(y/b_n)/b_n$. Repeating the arguments of Lemma 2.4 with k replaced by $\tilde{k}$ and with $\alpha \le \beta - 2$ to guarantee the finiteness of $\|\tilde{k}^2\|_{D_\alpha}$, we obtain that $\|\tilde{h} - f * \tilde{k}_{b_n}\|_V = O_p(n^{-1/2}b_n^{-1/2})$. It follows from (2.3) that $\|f * \tilde{k}_{b_n}\|_V \le \|f\|_V\|\tilde{k}\|_{1+D} < \infty$. Combining the above yields the desired $\|\hat{\psi} - \tilde{\psi}\|_V = o_p(1)$. $\qquad\square$

3. Residual-based weighted kernel estimators

Now consider observations $X_{1-p}, \ldots, X_n$ from a stationary and ergodic nonlinear autoregressive process of order p,

$$X_i = r_\vartheta(\mathbf{X}_{i-1}) + \varepsilon_i,$$

with $\mathbf{X}_{i-1} = (X_{i-p}, \ldots, X_{i-1})$ and ϑ a q-dimensional parameter. Assume that the innovations $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. with mean zero, finite variance σ^2, and density f, and are independent of

$\mathbf{X}_0$. Write $(\varepsilon, \mathbf{X})$ for random variables distributed as $(\varepsilon_i, \mathbf{X}_{i-1})$. Then ε and $\mathbf{X}$ are independent. Denote the distribution functions of ε and $\mathbf{X}$ by F and G. We make the following assumptions on the autoregression function.

Condition R. The function $\tau \mapsto r_\tau(\mathbf{x})$ is differentiable for all $\mathbf{x}$ with gradient $\tau \mapsto \dot{r}_\tau(\mathbf{x})$. For each constant C,

$$(3.1) \qquad \sup_{|\tau - \vartheta| \leq Cn^{-1/2}} \sum_{i=1}^{n} \left(r_\tau(\mathbf{X}_{i-1}) - r_\vartheta(\mathbf{X}_{i-1}) - \dot{r}_\vartheta(\mathbf{X}_{i-1})^\top (\tau - \vartheta) \right)^2 = o_p(1).$$

Moreover, $E[|\dot{r}_\vartheta(\mathbf{X})|^{5/2}] < \infty$ and the matrix $E[\dot{r}_\vartheta(\mathbf{X})\dot{r}_\vartheta(\mathbf{X})^\top] = \int \dot{r}_\vartheta \dot{r}_\vartheta^\top \, dG$ is positive definite.

Example 3.1. Consider the classical autoregressive model $X_t = \vartheta X_{t-1} + \varepsilon_t$ of order one with $|\vartheta| < 1$. Here $p = q = 1$ and $r_\vartheta(x) = \vartheta x$. Thus $\dot{r}_\vartheta(x) = x$ and the left-hand side of (3.1) equals zero. The moment condition $E[|\dot{r}_\vartheta(\mathbf{X})|^{5/2}] = E[|X_0|^{5/2}] < \infty$ follows from

$$(3.2) \qquad \int |y|^{5/2} f(y) \, dy < \infty.$$

Of course, $E[X_0^2] > 0$. This shows that the autoregressive process of order one satisfies Condition R if (3.2) holds. The same can be shown for higher order autoregressive models.

Example 3.2. Condition R holds for self-exciting threshold autoregressive models with known thresholds. Let us look at the simplest such model, namely the SETAR(2,1,1) model

$$X_t = \vartheta_1 X_{t-1} I[X_{t-1} \leq 0] + \vartheta_2 X_{t-1} I[X_{t-1} > 0] + \varepsilon_t,$$

with $\vartheta_1 < 1$, $\vartheta_2 < 1$ and $\vartheta_1 \vartheta_2 < 1$. The conditions on the parameter yield ergodicity of the model, see Petrucelli and Woolford (1984). Here $p = 1$, $q = 2$, $r_\vartheta(x) = \vartheta_1 x I[x \leq 0] + \vartheta_2 x I[x > 0]$, $\dot{r}_\vartheta(x)^\top = (x I[x \leq 0], x I[x > 0])$, and the left-hand side of (3.1) equals zero. The moment condition is equivalent to $E[|X_0|^{5/2}] < \infty$ and is implied by (3.2). It is easy to check that the matrix $\int \dot{r}_\vartheta \dot{r}_\vartheta^\top \, dG$ is diagonal with positive diagonal entries. This shows that Condition R is satisfied if (3.2) holds.

Example 3.3. Now look at the EXPAR(1) model $X_t = [\vartheta_1 + \vartheta_2 \exp(-\vartheta_3 X_{t-1}^2)] X_{t-1} + \varepsilon_t$ with $|\vartheta_1| < 1$ and $\vartheta_3 > 0$. Chan and Tong (1985) have shown that this model is geometrically ergodic. Here $p = 1$, $q = 3$, $r_\vartheta(x) = [\vartheta_1 + \vartheta_2 \exp(-\vartheta_3 x^2)]x$ and $\dot{r}_\vartheta(x)^\top = (x, x \exp(-\vartheta_3 x^2), -\vartheta_2 x^3 \exp(-\vartheta_3 x^2))$. It is easy to see that this gradient satisfies a Lipschitz condition: there are positive H and δ such that $|\dot{r}_\tau(x) - \dot{r}_\vartheta(x)| \leq H|\tau - \vartheta|$ for all $x \in \mathbb{R}$ and all τ with $|\tau - \vartheta| < \delta$. From this we immediately derive that the left-hand side of (3.1) is $O_p(n^{-1})$. The moment condition $E[\dot{r}_\vartheta(\mathbf{X})|^{5/2}] < \infty$ follows from $E[|X_0|^{5/2}]$, which in turn is implied by (3.2). The matrix $\int \dot{r}_\vartheta \dot{r}_\vartheta^\top \, dG$ is positive definite unless $\vartheta_2 = 0$. Thus the EXPAR(1) model satisfies Condition R if $\vartheta_2 \neq 0$ and (3.2) holds.

Remark 3.1. In the previous examples we have actually verified that (3.1) holds with $o_p(1)$ replaced by 0 or by $O_p(n^{-1})$. These faster rates are consequences of the smoothness of the functions $\tau \mapsto \dot{r}_\tau(\mathbf{x})$. Indeed, suppose these functions satisfy a *Hölder condition* at ϑ with exponent $\zeta > 0$ in

the following sense: there is a $\delta > 0$ and an $A \in L_2(G)$ such that $|\dot{r}_\tau(\mathbf{x}) - \dot{r}_\vartheta(\mathbf{x})| \leq |\tau - \vartheta|^\zeta A(\mathbf{x})$, for $\mathbf{x} \in \mathbb{R}^p$, $|\tau - \vartheta| < \delta$. Then (3.1) holds with $o_p(1)$ replaced by $O_p(n^{-\zeta})$.

We do not know the innovations and estimate them by the residuals $\hat{\varepsilon}_i = X_i - r_{\hat{\vartheta}}(\mathbf{X}_{i-1})$, $i = 1, \ldots, n$, where $\hat{\vartheta}$ is a $n^{1/2}$-consistent estimator of ϑ. These residuals are uniformly close to the actual innovations:

$$(3.3) \qquad \max_{1 \leq i \leq n} |\hat{\varepsilon}_i - \varepsilon_i| = o_p(1).$$

To see this, introduce $\varepsilon_i^* = \varepsilon_i - \dot{r}_\vartheta(\mathbf{X}_{i-1})^\top(\hat{\vartheta} - \vartheta)$. It follows from Condition R and the $n^{1/2}$-consistency of $\hat{\vartheta}$ that

$$(3.4) \qquad \sum_{i=1}^n (\hat{\varepsilon}_i - \varepsilon_i^*)^2 = o_p(1),$$

and hence

$$(3.5) \qquad \zeta_1 = \max_{1 \leq i \leq n} |\hat{\varepsilon}_i - \varepsilon_i^*| = o_p(1).$$

It follows from stationarity and finiteness of $E[|\dot{r}_\vartheta(\mathbf{X})|^{5/2}]$, that

$$(3.6) \qquad \zeta_2 = \max_{1 \leq i \leq n} |\varepsilon_i^* - \varepsilon_i| \leq |\hat{\vartheta} - \vartheta| \max_{1 \leq i \leq n} |\dot{r}_\vartheta(\mathbf{X}_{i-1})| = o_p(n^{-1/10}).$$

Relations (3.5) and (3.6) imply (3.3).

The kernel estimators based on $\hat{\varepsilon}_i$ and ε_i^* are defined by $\hat{\hat{f}}(y) = \frac{1}{n} \sum_{i=1}^n k_{b_n}(y - \hat{\varepsilon}_i)$ and $\hat{f}^*(y) = \frac{1}{n} \sum_{i=1}^n k_{b_n}(y - \varepsilon_i^*)$. The next lemma compares these two in the V-norm.

Lemma 3.1. *Suppose Conditions K, R and V1 hold, and k is continuously differentiable with $\|k'\|_{1+D}$ finite. Then $\|\hat{\hat{f}} - \hat{f}^*\|_V^2 = O_p\left(n^{-1}b_n^{-2} \sum_{i=1}^n (\hat{\varepsilon}_i - \varepsilon_i^*)^2\right) = o_p(n^{-1}b_n^{-2})$.*

Proof. We may assume that $b_n \leq 1$. First, for a non-negative (measurable) function g and random variables $\xi_1, \ldots, \xi_n$, one has the inequalities

$$(3.7) \qquad \int V(y) \frac{1}{b_n} g\left(\frac{y - z - \xi_i}{b_n}\right) dy \leq \left(1 + D\left(\max_{1 \leq i \leq n} |\xi_i|\right)\right) \|g\|_{1+D} V(z), \quad i = 1, \ldots, n.$$

To see this, make the substitution $y = z + \xi_i + b_n u$ and then use (2.3) twice. In view of (3.7) we have

$$
\begin{aligned}
\|\hat{\hat{f}} - \hat{f}^*\|_V &\leq \frac{1}{n} \sum_{i=1}^n \int V(y) \big|k_{b_n}(y - \hat{\varepsilon}_i) - k_{b_n}(y - \varepsilon_i^*)\big| \, dy \\
&\leq \frac{1}{n} \sum_{i=1}^n |\hat{\varepsilon}_i - \varepsilon_i^*| \int V(y) \int_0^1 \big|k'_{b_n}(y - \varepsilon_i^* - s(\hat{\varepsilon}_i - \varepsilon_i^*))\big| \, ds \, dy \\
&\leq (1 + D(\zeta_1 + \zeta_2)) \|k'\|_{1+D} \frac{1}{n b_n} \sum_{i=1}^n |\hat{\varepsilon}_i - \varepsilon_i^*| V(\varepsilon_i).
\end{aligned}
$$

Since $D(\zeta_1 + \zeta_2) = o_p(1)$ by Condition V1, (3.5) and (3.6), and since V is assumed to be in $L_2(F)$, an application of the Cauchy–Schwarz inequality gives the desired result. $\qquad\square$

Next we compare $\hat{f}^*$ and $\hat{f}$. We do so under an additional assumption on the kernel k.

Condition S. The kernel k is twice continuously differentiable with k' and k'' having finite $(1+D)$-norms and $(k')^2$ and $(k'')^2$ having finite D_α-norms.

Recall that q denotes the dimension of the parameter ϑ. Set $\hat{\gamma}(y) = \frac{1}{n} \sum_{i=1}^n k'_{b_n}(y - \varepsilon_i) \dot{r}_\vartheta(\mathbf{X}_{i-1})$, $y \in \mathbb{R}$.

Lemma 3.2. *Suppose Conditions K, R, S, V1 and V2 hold and f is V-smooth. Let $0 \le \xi \le 1/2$ and $\phi = (20q - 10\xi q + 50 - 20\xi)/(14 + 5q)$. Suppose that $nb_n^\phi \to \infty$. Then $\|\hat{f}^* - \hat{f} - (\hat{\vartheta} - \vartheta)^\top \hat{\gamma}\|_V = o_p(n^{-1/2} b_n^{-\xi})$.*

Proof. We may assume that $b_n \le 1$. It is easy to check that ϕ increases with q. Thus $\phi \ge (70 - 30\xi)/19 \ge 55/19 > 2$. Consequently, $nb_n^\phi \to \infty$ implies $nb_n^2 \to \infty$. Set $r_{ni} = n^{-1/2} \dot{r}_\vartheta(\mathbf{X}_{i-1})$. To stress the dependence on $\hat{\Delta} = n^{1/2}(\hat{\vartheta} - \vartheta)$, we express $\hat{f}^*(y) - \hat{f}(y) - (\hat{\vartheta} - \vartheta)^\top \hat{\gamma}(y)$ as $R_{\hat{\Delta}}(y)$, where

$$R_\Delta(y) = \frac{1}{n} \sum_{i=1}^n \left[k_{b_n}(y - \varepsilon_i + \Delta^\top r_{ni}) - k_{b_n}(y - \varepsilon_i) - \Delta^\top r_{ni} k'_{b_n}(y - \varepsilon_i) \right].$$

In view of the $n^{1/2}$-consistency of $\hat{\vartheta}$, it suffices to show that, for each (large) constant C,

$$(3.8) \qquad \sup_{|\Delta| \le C} \|R_\Delta\|_V = o_p(n^{-1/2} b_n^{-\xi}).$$

Fix now such a C. A Taylor expansion shows that

$$R_\Delta(y) = \frac{1}{n} \sum_{i=1}^n \Delta^\top r_{ni} \int_0^1 \left(k'_{b_n}\big(y - \varepsilon_i + v\Delta^\top r_{ni}\big) - k'_{b_n}(y - \varepsilon_i) \right) dv$$

$$= \frac{1}{n} \sum_{i=1}^n (\Delta^\top r_{ni})^2 \int_0^1 \int_0^v k''_{b_n}\big(y - \varepsilon_i + u\Delta^\top r_{ni}\big) \, du \, dv$$

and that

$$R_{\Delta+\tilde{\Delta}}(y) - R_\Delta(y) = \frac{1}{n} \sum_{i=1}^n \tilde{\Delta}^\top r_{ni} \int_0^1 \left(k'_{b_n}\big(y - \varepsilon_i + (\Delta + v\tilde{\Delta})^\top r_{ni}\big) - k'_{b_n}(y - \varepsilon_i) \right) dv$$

$$= \frac{1}{n} \sum_{i=1}^n \tilde{\Delta}^\top r_{ni} \int_0^1 (\Delta + v\tilde{\Delta})^\top r_{ni} \int_0^v \left(k''_{b_n}\big(y - \varepsilon_i + u(\Delta + v\tilde{\Delta})^\top r_{ni}\big) \right) du \, dv.$$

Let a_n be a sequence of positive numbers such that $C \ge a_n \to 0$. It follows from (3.7) that

$$\sup_{|\Delta| \le C} \sup_{|\tilde{\Delta}| \le a_n} \|R_{\Delta+\tilde{\Delta}} - R_\Delta\|_V \le 2a_n C \left(1 + D\Big(2C \max_{1 \le i \le n} |r_{ni}|\Big) \right) b_n^{-2} \|k''\|_{1+D} \frac{1}{n} \sum_{i=1}^n |r_{ni}|^2 V(\varepsilon_i).$$

Since $\max_{1\le i\le n}|r_{ni}| = o_p(n^{-1/10})$ and $E[|r_{ni}|^2 V(\varepsilon_i)] = n^{-1}E[\dot r_\vartheta(\mathbf{X})|^2]\|f\|_V$, we see that

$$(3.9) \qquad \sup_{|\Delta|\le C}\ \sup_{|\tilde\Delta|\le a_n}\|R_{\Delta+\tilde\Delta}-R_\Delta\|_V = O_p(a_n n^{-1} b_n^{-2}).$$

Let now $R_\Delta^*(y)$ be defined as $R_\Delta(y)$, but with r_{ni} replaced by $r_{ni}^* = r_{ni}\mathbf{1}[|r_{ni}|\le n^{-1/10}]$. Then $R_\Delta^*(y)$ and $R_\Delta(y)$ can differ only on the event $\{\max_{1\le i\le n}|r_{ni}| > n^{-1/10}\}$ which has probability tending to zero. This shows that

$$(3.10) \qquad \sup_{|\Delta|\le C}\|R_\Delta^*-R_\Delta\|_V = o_p(n^{-1/2}b_n^{-\xi}).$$

Now set

$$\bar R_\Delta^*(y) = \frac{1}{n}\sum_{i=1}^n (\Delta^\top r_{ni}^*)^2\int_0^1\int_0^v\int k_{b_n}''(y-z+u\Delta^\top r_{ni}^*)f(z)\,dz\,du\,dv.$$

It is easy to check that $\int k_{b_n}''(y-z)f(z)\,dz = \int k_{b_n}'(y-z)f'(z)\,dz$. Using this and (3.7), we obtain $\|\bar R_\Delta^*\|_V \le \frac{1}{n}\sum_{i=1}^n (\Delta^\top r_{ni}^*)^2 b_n^{-1}(1+D(|\Delta|n^{-1/10}))\|k'\|_{1+D}\|f'\|_V$. Since $\|f'\|_V < \infty$ and $nb_n^2\to\infty$, we obtain that

$$(3.11) \qquad \sup_{|\Delta|\le C}\|\bar R_\Delta^*\|_V = O_p(n^{-1}b_n^{-1}) = o_p(n^{-1/2}b_n^{-\xi}).$$

Next,

$$E[\|(R_\Delta^*-\bar R_\Delta^*)^2\|_W] \le \int_0^1\int_0^v\int W(y)E[\Gamma_\Delta^2(y,u)]\,dy\,du\,dv,$$

with $\Gamma_\Delta(y,u) = \frac{1}{n}\sum_{i=1}^n(\Delta^\top r_{ni}^*)^2\Big(k_{b_n}''\big(y-\varepsilon_i+u\Delta^\top r_{ni}^*\big)-\int k_{b_n}''\big(y-z+u\Delta^\top r_{ni}^*\big)f(z)\,dz\Big)$ a martingale. Since $E[\Gamma_\Delta^2(y,u)] \le n^{-3}|\Delta|^4 E[|\dot r_\vartheta(\mathbf{X})|^4\mathbf{1}[|\dot r_\vartheta(\mathbf{X})|\le n^{2/5}](k_{b_n}''(y-\varepsilon+u\Delta^\top n^{-1/2}\dot r_\vartheta(\mathbf{X})))^2]$, we obtain from (3.7) with W in place of V that

$$\sup_{|\Delta|\le C}E[\|(R_\Delta^*-\bar R_\Delta^*)^2\|_W] \le D_\alpha(Cn^{-1/10})C^4 n^{-3+3/5}b_n^{-5}E[|\dot r_\vartheta(\mathbf{X})|^{5/2}]\|f\|_W\|(k'')^2\|_{D_\alpha}.$$

In view of (2.10) we then have, for every finite subset D_n of $\{\Delta\in\mathbb{R}^q : |\Delta|\le C\}$ with M_n elements,

$$P\Big(\max_{\Delta\in D_n} n^{1/2}b_n^\xi\|R_\Delta^*-\bar R_\Delta^*\|_V > \eta\Big) \le \sum_{\Delta\in D_n} P(n^{1/2}b_n^\xi\|R_\Delta^*-\bar R_\Delta^*\|_V > \eta)$$

$$\le \sum_{\Delta\in D_n} P(nb_n^{2\xi}C_\alpha\|(R_\Delta^*-\bar R_\Delta^*)^2\|_W > \eta^2)$$

$$\le \sum_{\Delta\in D_n} \frac{nb_n^{2\xi}C_\alpha}{\eta^2}E[\|(R_\Delta^*-\bar R_\Delta^*)^2\|_W], \quad \eta > 0.$$

This shows that, for every $\eta > 0$ and every finite subset D_n as above,

$$(3.12) \qquad P\Big(\max_{\Delta\in D_n} n^{1/2}b_n^\xi\|R_\Delta^*-\bar R_\Delta^*\|_V > \eta\Big) = O(M_n n^{-7/5}b_n^{-5+2\xi}).$$

Now take D_n to be such that the balls of radius a_n centered at elements of D_n cover the ball $\{\Delta : |\Delta|\le C\}$. We can choose D_n such that $M_n = O(a_n^{-q})$. Thus, if $a_n^q n^{7/5}b_n^{5-2\xi}\to\infty$ and

$a_n^{-1}n^{1/2}b_n^{2-\xi} \to \infty$, then we obtain from (3.9)–(3.12) the desired (3.8). But if we take $a_n = b_n^{(3-4\xi)/(14+5q)}$, then $a_n^q n^{7/5}b_n^{5-2\xi} = (nb_n^\phi)^{7/5} \to \infty$ and $a_n^{-1}n^{1/2}b_n^{2-\xi} = (nb_n^\phi)^{1/2} \to \infty$. $\square$

Lemma 3.3. *Suppose Conditions K, R, S, V1 and V2 hold and f is V-smooth. Then*

$$\|\hat{\gamma} - f'E[\dot{r}_\vartheta(\mathbf{X})]\|_V = O_p(n^{-1/2}b_n^{-3/2}) + o_p(1).$$

Proof. Let $f'_{b_n} = k'_{b_n} * f = f' * k_{b_n}$ and set $\bar{\gamma}(y) = \frac{1}{n}\sum_{i=1}^n f'_{b_n}(y)\dot{r}_\vartheta(\mathbf{X}_{i-1})$. It suffices to show

(3.13) $$\|\hat{\gamma} - \bar{\gamma}\|_V = O_p(n^{-1/2}b_n^{-3/2}),$$

(3.14) $$\|\bar{\gamma} - f'E[\dot{r}_\vartheta(\mathbf{X})]\|_V = o_p(1).$$

By Lemma 2.1, we have $\|f'_{b_n} - f'\|_V = \|f' * k_{b_n} - f'\|_V \to 0$. Relation (3.14) follows from this and the ergodic theorem.

To prove (3.13) we may assume that $b_n \le 1$. Since $\hat{\gamma}(y) - \bar{\gamma}(y)$ is a martingale, we have

$$nb_n^3 E[|\hat{\gamma}(y) - \bar{\gamma}(y)|^2] \le b_n^3 E[|\dot{r}_\vartheta(\mathbf{X})|^2]\int (k'_{b_n}(y-z))^2 f(z)\,dz$$

$$\le E[|\dot{r}_\vartheta(\mathbf{X})|^2]\int f(y-b_n z)(k'(z))^2\,dz$$

and thus, in view of (2.7), $nb_n^3 \int W(y)E[|\hat{\gamma}(y) - \bar{\gamma}(y)|^2]\,dy \le E[|\dot{r}_\vartheta(\mathbf{X})|^2]\|(k')^2\|_{D_\alpha}\|f\|_W$. Relation (3.13) follows from this and (2.10). $\square$

Corollary 3.1. *Suppose Conditions K, R, S, V1 and V2 hold and f is V-smooth. Let $nb_n^{\phi_1} \to \infty$ with $\phi_1 = (40+15q)/(14+5q)$. Then $\|\hat{f}^* - \hat{f}\|_V = o_p(n^{-1/2}b_n^{-1/2})$.*

Proof. Note that ϕ_1 equals ϕ of Lemma 3.2 with $\xi = 1/2$, and $55/19 \le \phi_1 < 3$. Thus $\phi_1 > 2$, and $nb_n^{\phi_1} \to \infty$ implies that $nb_n^2 \to \infty$. Consequently we obtain from Lemma 3.3 that $\|\hat{\gamma}^\top(\hat{\vartheta} - \vartheta)\|_V = O_p(n^{-1}b_n^{-3/2}) + O_p(n^{-1/2}) = o_p(n^{-1/2}b_n^{-1/2})$. This and Lemma 3.2 give the desired result. $\square$

Corollary 3.2. *Suppose Conditions K, R, S, V1 and V2 hold and f is V-smooth. Let $nb_n^{\phi_0} \to \infty$ with $\phi_0 = (50+20q)/(14+5q)$. Then $\|\hat{f}^* - \hat{f} - f'E[\dot{r}_\vartheta(\mathbf{X})]^\top(\hat{\vartheta} - \vartheta)\|_V = o_p(n^{-1/2})$.*

Proof. Note that ϕ_0 equals ϕ of Lemma 3.2 with $\xi = 0$ and $70/19 \le \phi_0 < 4$. Thus $\phi_0 > 3$ and $nb_n^{\phi_0} \to \infty$ implies $nb_n^3 \to \infty$. Consequently we obtain from Lemma 3.3 that $\|\hat{\gamma} - f'E[\dot{r}_\vartheta(\mathbf{X})]\|_V = o_p(1)$. This and Lemma 3.2 give the desired result. $\square$

Lemmas 2.2, 2.4, 3.1 and Corollary 3.1 give the following convergence rate for the residual-based density estimator $\hat{\hat{f}}$ in the V-norm.

Theorem 3.1. *Suppose Conditions K, R, S, V1 and V2 hold and f is V-smooth. If $nb_n^{\phi_1} \to \infty$ with $\phi_1 = (40+15q)/(14+5q)$, then*

$$\|\hat{\hat{f}} - f\|_V^2 = O_p(n^{-1}b_n^{-1}) + o(b_n^2) + O_p\left(n^{-1}b_n^{-2}\sum_{i=1}^n(\hat{\varepsilon}_i - \varepsilon_i^*)^2\right) = o_p(n^{-1}b_n^{-2} + b_n^2).$$

Remark 3.2. Suppose the assumptions of Theorem 3.1 are met. Then $\|\hat{\hat{f}}-f\|_V = o_p(1)$. Mimicking the proof of Corollary 2.1 yields the stronger $\|\hat{\hat{f}} - f\|_{V_*} = o_p(1)$. If $b_n \sim n^{-1/4}$, then we even have $\|\hat{\hat{f}} - f\|_V = o_p(n^{-1/4})$. Better rates are available if we impose additional smoothness assumptions on f or if we require (3.1) to hold with $o_p(1)$ replaced by $O_p(n^{-1/3})$. In this latter case we have $\sum_{i=1}^n (\hat{\varepsilon}_i - \varepsilon_i^*)^2 = O_p(n^{-1/3})$. Then $\|\hat{\hat{f}} - f\|_V = O_p(n^{-1/3})$ if $b_n \sim n^{-1/3}$. A sufficient condition for the strengthened version of (3.1) is a Hölder condition with exponent $1/3$; see Remark 3.1.

If Condition R holds with $o_p(1)$ replaced by $O_p(n^{-2/3})$, then $\sum_{i=1}^n (\hat{\varepsilon}_i - \varepsilon_i^*)^2 = O_p(n^{-2/3})$. Thus Lemma 3.1 and Corollary 3.2 give the following result.

Theorem 3.2. *Suppose Conditions K, S, V1 and V2 hold, f is V-smooth, and Condition R holds with $o_p(1)$ replaced by $O_p(n^{-2/3})$. Let $nb_n^{\phi_0} \to \infty$ with $\phi_0 = (50 + 20q)/(14 + 5q)$, then*

$$\|\hat{\hat{f}} - \hat{f} - f' E[\dot{r}_\vartheta(\mathbf{X})]^\top (\hat{\vartheta} - \vartheta)\|_V = o_p(n^{-1/2}).$$

Since f is assumed to have mean zero, it is more natural to choose an estimator that also has this property. For this purpose we now consider a weighted kernel estimator $\hat{\hat{f}}_w(y) = \frac{1}{n} \sum_{i=1}^n \hat{w}_i k_{b_n}(y - \hat{\varepsilon}_i)$ with positive weights $\hat{w}_i = 1/(1 + \hat{\lambda}\hat{\varepsilon}_i)$, where $\hat{\lambda}$ is chosen such that $\sum_{i=1}^n \hat{w}_i \hat{\varepsilon}_i = 0$. As shown by Owen (1988, 2001) such a $\hat{\lambda}$ exists on the event $\min_{1 \le i \le n} \hat{\varepsilon}_i < 0 < \max_{1 \le i \le n} \hat{\varepsilon}_i$. We choose $\hat{\lambda} = 0$ otherwise. The above event has probability tending to one by (3.3), and since the innovations are centered. In view of (3.3) we have

$$(3.15) \qquad S = \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^2 + o_p(1) = \sigma^2 + o_p(1),$$

$$(3.16) \qquad Z_* = \max_{1 \le i \le n} |\hat{\varepsilon}_i| = \max_{1 \le i \le n} |\varepsilon_i| + o_p(1) = o_p(n^{1/2}).$$

Also, in view of (3.4) and the ergodic theorem,

$$\frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i = \frac{1}{n} \sum_{i=1}^n \varepsilon_i^* + o_p(n^{-1/2}) = \frac{1}{n} \sum_{i=1}^n \varepsilon_i - E[\dot{r}_\vartheta(\mathbf{X})]^\top (\hat{\vartheta} - \vartheta) + o_p(n^{-1/2}) = O_p(n^{-1/2}).$$

We can then proceed as in Owen (2001, pp. 219–221), with his $X_i - \mu_0$ replaced by $\hat{\varepsilon}_i$, to conclude that

$$(3.17) \qquad \hat{\lambda} = S^{-1} \frac{1}{n} \sum_{i=1}^n \hat{\varepsilon}_i + o_p(n^{-1/2})$$

and therefore

$$(3.18) \qquad \hat{\lambda} = \sigma^{-2} \Big(\frac{1}{n} \sum_{i=1}^n \varepsilon_i - E[\dot{r}_\vartheta(\mathbf{X})]^\top (\hat{\vartheta} - \vartheta) \Big) + o_p(n^{-1/2}).$$

It is now easy to see that

$$(3.19) \qquad \hat{w}_* = \max_{1 \le i \le n} |\hat{w}_i - 1| = o_p(1).$$

We are ready to compare $\hat{\hat{f}}_w$ with $\hat{\hat{f}}$ in the V-norm.

Lemma 3.4. *Suppose Conditions K, R, S, V1 and V2 hold and f is V-smooth. Let $nb_n^{\phi_1} \to \infty$ with $\phi_1 = (40 + 15q)/(14 + 5q)$. Then, with $\psi(y) = yf(y)$, $\|\hat{\hat{f}}_w - \hat{\hat{f}} + \hat{\lambda}\,\psi\|_V = o_p(n^{-1/2})$.*

Proof. The proof is similar to that of Lemma 2.5. But now we use (3.18) and (3.19) instead of (2.11) and (2.12). To prove the analogue of $\|\hat{\psi}_w - \hat{\psi}\|_V = o_p(1)$, we now use (3.7) and

$$\frac{1}{n}\sum_{i=1}^{n}|\hat{\varepsilon}_i|V(\varepsilon_i) \le \left(\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i^2\,\frac{1}{n}\sum_{i=1}^{n}V^2(\varepsilon_i)\right)^{1/2} = O_p(1),$$

which follows from (3.15). We also have $\|\hat{\hat{f}}\|_V = O_p(1)$ and $\|\hat{\hat{f}} - f\|_{V_*} = o_p(1)$ by Theorem 3.1 and Remark 3.2. $\qquad\square$

Lemma 3.4, Theorem 3.2 and relation (3.18) give the following expansion for $\hat{\hat{f}}_w$.

Theorem 3.3. *Suppose Conditions K, S, V1 and V2 hold, Condition R holds with $o_p(1)$ replaced by $O_p(n^{-2/3})$, and f is V-smooth. Let $nb_n^{(50+20q)/(14+5q)} \to \infty$. Then, with $\psi(y) = yf(y)$,*

$$\left\|\hat{\hat{f}}_w - \hat{f} + \sigma^{-2}\psi\,\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i - (\sigma^{-2}\psi + f')E[\dot{r}_\vartheta(\mathbf{X})]^\top(\hat{\vartheta} - \vartheta)\right\|_V = o_p(n^{-1/2}).$$

A sufficient condition for the strengthened version of (3.1) is a Hölder condition with exponent $2/3$, see Remark 3.1. If $E[\dot{r}_\vartheta(\mathbf{X})] = 0$, as is the case in the classical linear autoregressive model of order p, we have the simpler conclusion $\|\hat{\hat{f}}_w - \hat{f} + \sigma^{-2}\psi\,\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\|_V = o_p(n^{-1/2})$.

4. A SMOOTHED AND WEIGHTED EMPIRICAL DISTRIBUTION FUNCTION

As a first application of the previous results, we consider estimation of the innovation distribution function F. From now on we assume the following.

(A1) The kernel k is a symmetric twice continuously differentiable density with compact support.

(A2) The bandwidth b_n satisfies $nb_n^4 \to 0$ and $nb_n^{(50+20q)/(14+5q)} \to \infty$.

(A3) Condition R holds with $o_p(1)$ replaced by $O_p(n^{-2/3})$.

Our estimator is the smoothed and weighted residual-based empirical distribution function

$$\hat{\hat{F}}_{sw}(t) = \int_{-\infty}^{t}\hat{\hat{f}}_w(u)\,du = \frac{1}{n}\sum_{i=1}^{n}\hat{w}_iK\left(\frac{t - \hat{\varepsilon}_i}{b_n}\right), \quad t \in \mathbb{R},$$

where K is the distribution function of the kernel k. The smoothed empirical distribution function (based on the actual innovations) is

$$\hat{F}_s(t) = \int_{-\infty}^{t}\hat{f}(u)\,du = \frac{1}{n}\sum_{i=1}^{n}K\left(\frac{t - \varepsilon_i}{b_n}\right), \quad t \in \mathbb{R}.$$

We can use Theorem 3.3 with $V = 1$ to obtain the following expansion for the difference of these estimators.

Corollary 4.1. *Suppose (A1)–(A3) hold and f is absolutely continuous with $\|f'\|_1 < \infty$. Then*

$$\sup_{t\in\mathbb{R}}\left|\hat{\hat{F}}_{sw}(t) - \hat{F}_s(t) + c_t\frac{1}{n}\sum_{i=1}^n \varepsilon_i - (c_t + f(t))E[\dot{r}_\vartheta(\mathbf{X})]^\top(\hat\vartheta - \vartheta)\right| = o_p(n^{-1/2}),$$

where $c_t = \sigma^{-2}E[\varepsilon\mathbf{1}[\varepsilon \le t]] = \int_{-\infty}^t yf(y)\,dy\big/\int y^2 f(y)\,dy.$

Let $\hat{F}(t) = \frac{1}{n}\sum_{i=1}^n \mathbf{1}[\varepsilon_i \le t]$ denote the empirical distribution based on the actual innovations. Since the empirical process $\nu_n = n^{1/2}(\hat{F} - F)$ satisfies $\sup_{|u|\le b_n}\sup_{t\in\mathbb{R}}|\nu_n(t-u) - \nu_n(t)| = o_p(1)$, we obtain that $\sup_{t\in\mathbb{R}}|\hat{F}_s(t) - \hat{F}(t) - \int(F(t+b_n u) - F(t))k(u)\,du| = o_p(n^{-1/2})$. As k has mean zero, we see that $\Delta = \sup_{t\in\mathbb{R}}|\int(F(t + b_n u) - F(t))k(u)\,du| \le \sup_{|u|\le b_n}\sup_{t\in\mathbb{R}}|F(t-u) - F(t) + uf(t)|$. If f is Lipschitz, then $\Delta = O(b_n^2)$ and f is absolutely continuous. Thus we have the following expansion for the difference between the smoothed and weighted residual-based empirical distribution function and the empirical distribution function based on the true innovations.

Theorem 4.1. *Suppose (A1)–(A3) hold. Let f be Lipschitz and $\|f'\|_1 < \infty$. Then*

$$\sup_{t\in\mathbb{R}}\left|\hat{\hat{F}}_{sw}(t) - \hat{F}(t) + c_t\frac{1}{n}\sum_{i=1}^n \varepsilon_i - (c_t + f(t))E[\dot{r}_\vartheta(\mathbf{X})]^\top(\hat\vartheta - \vartheta)\right| = o_p(n^{-1/2}).$$

The terms involving c_t come from weighting the kernel estimator. Weighting usually leads to a smaller asymptotic variance. The gains can be considerable as the next example shows.

Example 4.1. Consider the classical autoregressive process $X_t = \vartheta X_{t-1} + \varepsilon_t$ of order one with $|\vartheta| < 1$. In this case $E[\dot{r}_\vartheta(\mathbf{X})] = E[X_0] = 0$ and $\sup_{t\in\mathbb{R}}|\hat{\hat{F}}_{sw}(t) - \hat{F}(t) + c_t\frac{1}{n}\sum_{i=1}^n \varepsilon_i| = o_p(n^{-1/2})$ for any $n^{1/2}$-consistent estimator of ϑ. The smoothed but unweighted estimator $\hat{\hat{F}}_s$, which is the distribution function of $\hat{\hat{f}}$, satisfies $\sup_{t\in\mathbb{R}}|\hat{\hat{F}}_s(t) - \hat{F}(t)| = o_p(n^{-1/2})$. The asymptotic variances of the estimators $\hat{\hat{F}}_{sw}(t)$ and $\hat{\hat{F}}_s(t)$ for a fixed t are $F(t)(1 - F(t)) - c_t^2\sigma^2$ and $F(t)(1 - F(t))$. For the standard normal distribution F and $t = 0$ we calculate these asymptotic variances as $1/4 - 1/(2\pi)$ and $1/4$. Thus using the weighted estimator yields a reduction of the asymptotic variance of about 64 percent.

To address efficiency issues we assume from now on that f has finite Fisher information for location. This means that f is absolutely continuous with a.e. derivative f', and $E[\ell^2(\varepsilon)] = \int \ell^2(y)f(y)\,dy$ is finite, where $\ell = -f'/f$ is the score function for location. Then f is V-smooth for each f-square integrable V. This follows from the inequality

$$\|f'\|_V^2 = \left(\int V(y)|\ell(y)|f(y)\,dy\right)^2 \le \int V^2(y)f(y)\,dy\int\ell^2(y)f(y)\,dy.$$

There exists a rich literature on efficient estimators of the finite-dimensional parameter in related semiparametric time series models; see for example Kreiss (1987a, b), Linton (1993), Jeganathan (1995), Drost and Klaassen (1997) and Drost, Klaassen and Werker (1997), and Schick and Wefelmeyer (2002b). Koul and Schick (1997) have characterized and constructed efficient

estimators for ϑ in nonlinear autoregression with mean zero innovations as needed here. Such an efficient estimator $\hat{\vartheta}$ satisfies $\hat{\vartheta} = \vartheta + \Lambda^{-1}\frac{1}{n}\sum_{i=1}^{n} S(\mathbf{X}_{i-1}, \varepsilon_i) + o_p(n^{-1/2})$, where $S(\mathbf{X}, \varepsilon) = \left(\dot{r}_\vartheta(\mathbf{X}) - E[\dot{r}_\vartheta(\mathbf{X})]\right)\ell(\varepsilon) + \sigma^{-2}E[\dot{r}_\vartheta(\mathbf{X})]\,\varepsilon$, and Λ is the covariance matrix of $S(\mathbf{X}, \varepsilon)$. For such an estimator we have

$$\hat{\hat{F}}_{sw}(t) = \hat{F}(t) - c_t \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i + (c_t + f(t))E[\dot{r}_\vartheta(\mathbf{X})]^\top \Lambda^{-1}\frac{1}{n}\sum_{i=1}^{n} S(\mathbf{X}_{i-1}, \varepsilon_i) + o_p(n^{-1/2}).$$

This is the characterization of an efficient estimator for $F(t)$, see Schick and Wefelmeyer (2002a).

Alternative estimators to $\hat{\hat{F}}_{sw}(t)$ are $\hat{\hat{F}}_w(t) = \frac{1}{n}\sum_{i=1}^{n}\hat{w}_i\mathbf{1}[\hat{\varepsilon}_i \le t]$ and $\hat{\hat{F}}(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}[\hat{\varepsilon}_i \le t]$. If f is uniformly continuous, and Condition R is strengthened to

$$\max_{1 \le i \le n}\sup_{|\tau-\vartheta| \le Cn^{-1/2}}\left|r_\tau(\mathbf{X}_{i-1}) - r_\vartheta(\mathbf{X}_{i-1}) - \dot{r}_\vartheta(\mathbf{X}_{i-1})^\top(\tau-\vartheta)\right| = o_p(n^{-1/2}),$$

then Schick and Wefelmeyer (2002a) show that

$$\sup_{t \in \mathbb{R}}|\hat{\hat{F}}(t) - \hat{F}(t) - f(t)E[\dot{r}_\vartheta(\mathbf{X})]^\top(\hat{\vartheta} - \vartheta)| = o_p(n^{-1/2}) \quad \text{and} \quad \sup_{t \in \mathbb{R}}|\hat{c}_t - c_t| = o_p(1),$$

with $\hat{c}_t = \sum_{i=1}^{n}\hat{\varepsilon}_i\mathbf{1}[\hat{\varepsilon}_i \le t]\Big/\sum_{i=1}^{n}\hat{\varepsilon}_i^2$. Because of the identity $\hat{w}_i = 1 - \hat{\lambda}\hat{w}_i\hat{\varepsilon}_i$, we get

$$\sup_{t \in \mathbb{R}}\left|\hat{\hat{F}}_w(t) - \hat{\hat{F}}(t) + \hat{\lambda}\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i\mathbf{1}[\hat{\varepsilon}_i \le t]\right| = o_p(n^{-1/2}).$$

Thus $\hat{\hat{F}}_w$ has the same uniform stochastic expansion as the smoothed version $\hat{\hat{F}}_{sw}$, under weaker assumptions on f. The estimators $\hat{\hat{F}}_w$ and $\hat{\hat{F}}_{sw}$ have the advantage that they are true distribution functions. The alternative efficient estimator of Schick and Wefelmeyer (2002a), $\tilde{F}(t) = \hat{\hat{F}}(t) - \hat{c}_t\frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i$, is not a distribution function.

5. SMOOTHED AND WEIGHTED EMPIRICAL MOMENTS

Let m be an integer greater than 1. As a further application of our results on density estimators we consider estimation of the m-th moment $\mu_m = \int y^m f(y)\,dy$ of f. Our estimator is the smoothed and weighted empirical moment based on the residuals,

$$\hat{\hat{\mu}}_{m,w} = \int y^m \hat{\hat{f}}_w(y)\,dy.$$

Since $\hat{\mu}_m = \int y^m \hat{f}(y)\,dy = \frac{1}{n}\sum_{i=1}^{n}\int(\varepsilon_i + b_n u)^m k(u)\,du = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^m + O(b_n^2)$, an application of Theorem 3.3 with $V(y) = (1 + |y|)^m$ gives the following expansion.

Theorem 5.1. *Suppose (A1)–(A3) hold. Let f have a finite absolute moment of order greater than $2m+1$ and be V-smooth for $V(x) = (1 + |x|)^m$. Then*

$$\hat{\hat{\mu}}_{m,w} = \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i^m - \frac{\mu_{m+1}}{\sigma^2}\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i + \left(\frac{\mu_{m+1}}{\sigma^2} - m\mu_{m-1}\right)E[\dot{r}_\vartheta(\mathbf{X})]^\top(\hat{\vartheta} - \vartheta) + o_p(n^{-1/2}).$$

Let $\hat{\bar{f}}(y) = \frac{1}{n} \sum_{i=1}^{n} \hat{k}_{b_n}(y - \hat{\varepsilon}_i)$ denote the unweighted kernel estimator, and $\hat{\bar{\mu}}_m = \int y^m \hat{\bar{f}}(y) \, dy$ the unweighted (smoothed) empirical moment. The weighted (smoothed) empirical moment $\hat{\bar{\mu}}_{m,w}$ can have considerably smaller asymptotic variance. Consider for simplicity a *linear* autoregressive model. Then $E[\dot{r}_\vartheta(\mathbf{X})] = 0$ and

$$\hat{\bar{\mu}}_m = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^m + o_p(n^{-1/2}), \quad \hat{\bar{\mu}}_{m,w} = \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i^m - \frac{\mu_{m+1}}{\sigma^2} \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i + o_p(n^{-1/2}).$$

The asymptotic variances of $\hat{\bar{\mu}}_m$ and $\hat{\bar{\mu}}_{m,w}$ are μ_{2m} and $\mu_{2m} - \mu_{m+1}^2/\sigma^2$, respectively. For $m = 3$ and f normal these variances are $15\sigma^6$ and $(15 - 9)\sigma^6 = 6\sigma^6$, respectively.

If we also assume that f has finite Fisher information for location, then $\hat{\bar{\mu}}_{m,w}$ is efficient by Schick and Wefelmeyer (2002a) if $\hat{\vartheta}$ is. They construct an alternative efficient estimator

$$\tilde{\mu}_m = \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i^m - \frac{\sum_{i=1}^{n} \hat{\varepsilon}_i^{m+1}}{\sum_{i=1}^{n} \hat{\varepsilon}_i^2} \frac{1}{n} \sum_{i=1}^{n} \hat{\varepsilon}_i$$

under weaker assumptions; they only require Condition R and a finite moment of order $2m$. Another efficient estimator is given by

$$\frac{1}{n} \sum_{i=1}^{n} \hat{w}_i \hat{\varepsilon}_i^m.$$

Indeed, because of the identity $\hat{w}_i = 1 - \hat{\lambda} \hat{w}_i \hat{\varepsilon}_i$, this estimator is asymptotically equivalent to $\tilde{\mu}_n$ by (3.15), (3.17) and (3.19).

REFERENCES

Chan, K. S. and Tong, H. (1985). On the use of the deterministic Lyapunov function fro the ergodicity of stochastic difference equations. *Adv. Appl. Probab.* **17**, 666-678.

Chen, S. X. (1997). Empirical likelihood-based kernel density estimation. *Austral. J. Statist.* **39**, 47–56.

Devroye, L. (1992). A note on the usefulness of superkernels in density estimation. *Ann. Statist.* **20**, 2037–2058.

Drost, F. C. and Klaassen, C. A. J. (1997). Efficient estimation in semiparametric GARCH models, *J. Econometrics* **81**, 193–221.

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

Hall, P. and Presnell, B. (1999). Density estimation under constraints. *J. Comput. Graph. Statist.* **8**, 259–277.

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

Koul, H. L. and Schick, A. (1996). Adaptive estimation in a random coefficient autoregressive model. *Ann. Statist.* **24**, 1025–1052.

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

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

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

Liebscher, E. (1999). Estimating the density of the residuals in autoregressive models. *Stat. Inference Stoch. Process.* **2**, 105–117.

Linton, O. (1993). Adaptive estimation in ARCH models. *Econometric Theory* **9**, 539–569.

Müller, U. U., Schick, A. and Wefelmeyer, W. (2003). Efficient prediction for linear and nonlinear autoregressive models. Technical Report, Department of Mathematical Sciences, Binghamton University.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249.

Owen, A. B. (2001). *Empirical Likelihood.* Monographs on Statistics and Applied Probability **92**, Chapman & Hall / CRC, London.

Petrucelli, J. D. and Woolford, S. W. (1984). A threshold AR(1) model. *J. Appl. Probab.* **21**, 270-286.

Robinson, P. M. (1986). Nonparametric estimation from time series residuals. *Cahiers Centre Études Rech. Opér.* **28**, 197–202.

Robinson, P. M. (1987). Time series residuals with application to probability density estimation. *J. Time Ser. Anal.* **8**, 329–344.

Rudin, W. (1974). *Real and Complex Analysis. 2nd ed.* McGraw-Hill, New York.

Schick, A. and Wefelmeyer, W. (2002a). Estimating the innovation distribution in nonlinear autoregressive models. *Ann. Inst. Statist. Math.* **54**, 245–260.

Schick, A. and Wefelmeyer, W. (2002b). Efficient estimation in invertible linear processes. *Math. Meth. Statist.* **11**, 358–379.

Zhang, B. (1998). A note on kernel density estimation with auxiliary information. *Comm. Statist. Theory Methods* **27**, 1–11.

URSULA U. MÜLLER, UNIVERSITÄT BREMEN, FACHBEREICH 3: MATHEMATIK UND INFORMATIK, POSTFACH 330 440, 28334 BREMEN, GERMANY

ANTON SCHICK, BINGHAMTON UNIVERSITY, DEPARTMENT OF MATHEMATICAL SCIENCES, BINGHAMTON, NY 13902-6000, USA

WOLFGANG WEFELMEYER, UNIVERSITÄT ZU KÖLN, MATHEMATISCHES INSTITUT, WEYERTAL 86-90, 50931 KÖLN, GERMANY