# IMPROVED DENSITY ESTIMATORS FOR INVERTIBLE LINEAR PROCESSES

Anton Schick

Department of Mathematical Sciences

Binghamton University

Binghamton, NY 13902-6000, USA

anton@math.binghamton.edu

Wolfgang Wefelmeyer

Mathematisches Institut

Universität zu Köln

Weyertal 86-90

50931 Köln, Germany

wefelm@math.uni-koeln.de

ABSTRACT

The stationary density of a centered invertible linear processes can be represented as a convolution of innovation-based densities, and it can be estimated at the parametric rate by plugging residual-based kernel estimators into the convolution representation. We have shown elsewhere that a functional central limit theorem holds both in the space of continuous functions vanishing at infinity, and in weighted $L_1$-spaces. Here we show that we can improve the plug-in estimator considerably, exploiting the information that the innovations are centered, and replacing the kernel estimators by weighted versions, using the empirical likelihood approach.

## 1   INTRODUCTION

For independent and identically distributed (i.i.d.) observations $X_1, \ldots, X_n$, Frees (1994) has shown that densities of certain functions $q(X_1, \ldots, X_m)$ with $m \geq 2$ can be estimated pointwise at the parametric rate $n^{-1/2}$ by local U-statistics

$$\frac{1}{\binom{n}{m}} \sum_{1 \leq i_1 < \cdots < i_m \leq n} k_{b_n}(x - q(X_{i_1}, \ldots, X_{i_m})),$$

where $k_{b_n} = k(x/b_n)/b_n$ for some kernel $k$ (an integrable function that integrates to 1) and some bandwidth $b_n$ (tending to 0). Giné and Mason (2007a,b) prove functional central limit theorems and laws of the iterated logarithm in $L_p$ for $p \in [1, \infty]$.

Independently of Frees (1994), Saavedra and Cao (2000) have considered the special case $q(X_1, X_2) = X_1 + aX_2$ and obtained the pointwise rate $n^{-1/2}$ for the convolution $\int \hat{f}(x - ay)\hat{f}(y)\,dy$ of kernel estimators, i.e. by a plug-in estimator for the functional $\int f(x - ay)f(y)\,dy$ of the density $f$ of the observations. Such a plug-in estimator is asymptotically equivalent to the corresponding local U-statistic with kernel of the form $K(x) = \int k(x - ay)k(y)\,dy$. Schick and Wefelmeyer (2004b, 2007a) prove functional central limit theorems for plug-in estimators of the density of $q(X_1, \ldots, X_m) = u_1(X_1) + \cdots + u_m(X_m)$ and $q(X_1, X_2) = X_1 + X_2$ in $L_1$ and in the space $C_0$ of continuous functions vanishing at infinity. — Du and Schick (2007) obtain similar results for derivatives of convolutions. Schick and Wefelmeyer (2008b,c) show that the rate $n^{-1/2}$ may not be attained for plug-in estimators of densities of $q(X_1, X_2) = |X_1|^\nu + |X_2|^\nu$ with $\nu \geq 2$.

The results on density estimation for sums of independent random variables carry over to estimation of the stationary density of linear processes. Saavedra and Cao (1999) consider pointwise convergence of plug-in estimators for the stationary density of moving average processes of order one. Schick and Wefelmeyer (2004a) obtain pointwise asymptotic normality and efficiency in this case, and Schick and Wefelmeyer (2004b) generalize the result to higher order moving average processes and strengthen it to functional convergence in $L_1$ and $C_0$. Schick and Wefelmeyer (2007b, 2008a) consider observations $X_1, \ldots, X_n$ of a stationary linear process with infinite-order moving average representation

$$X_t = \varepsilon_t + \sum_{s=1}^{\infty} \varphi_s \varepsilon_{t-s}, \tag{1.1}$$

where the coefficients $\varphi_s$ are summable and the innovations $\varepsilon_t$ are i.i.d. with density $f$, mean zero and finite variance. They base their estimator for the stationary density on the convolution representation $X_t = \varepsilon_t + Y_t$. In order to obtain $n^{1/2}$-consistency, $Y_t$ must have a density, say $g$. Hence the case $Y_t = 0$ must be excluded by assuming that at least one of the moving average coefficients $\varphi_s$ is nonzero. Then the stationary density has the convolution representation $f * g$ and is estimated by plugging in residual-based kernel estimators $\hat{f}$ and $\hat{g}$. Invertibility of the linear process is used to construct residuals, i.e. estimators for the innovations.

These kernel estimators do not use the information that the innovations $\varepsilon_t$, and therefore also the series $Y_t$, have mean zero. In the present paper we will show how to make use of this information. We do this by applying the empirical likelihood approach of Owen (1988, 2001) to density estimation as follows. Consider first the simplest case of a kernel estimator

2

$\hat{f}(x) = (1/n) \sum_{j=1}^n k_{b_n}(x - X_j)$ based on i.i.d. observations $X_1, \ldots, X_n$ with density $f$ having mean zero and finite variance. Then we can improve $\hat{f}(x)$ by a randomly weighted version, with weights chosen such that the weighted empirical distribution also has mean zero,

$$\hat{f}_w(x) = \frac{1}{n} \sum_{j=1}^n w_j k_{b_n}(x - X_j) \quad \text{with} \quad \frac{1}{n} \sum_{j=1}^n w_j X_j = 0.$$

Weighted kernel estimators are studied by Chen (1997), Zhang (1998), Hall and Presnell (1999) and Müller, Schick and Wefelmeyer (2005, Section 2).

We mention in passing that weighting a kernel estimator $\hat{f}$ is asymptotically equivalent to correcting it additively as

$$\hat{f}_a(x) = \hat{f}(x) - \hat{\sigma}^{-2} x \hat{f}(x) \frac{1}{n} \sum_{j=1}^n X_j,$$

with $\hat{\sigma}^2 = (1/n) \sum_{j=1}^n X_j^2$ an estimator of $\sigma^2 = E[X^2]$. Such additive corrections are easier to calculate than the weights $w_j$, but the disadvantage is that $\hat{f}_a$ is only approximately a probability density. Similar corrections for empirical estimators have been introduced by Levit (1975) and Haberman (1984). The heuristic motivation for the additive correction of density estimators is the same as for empirical estimators: The constraint $E[X] = 0$ implies that efficient influence functions must be orthogonal to $X$, so we project the influence function $k_{b_n}(x - X)$ of $\hat{f}(x)$ onto the space of random variables orthogonal to $X$ to obtain the improved influence function

$$k_{b_n}(x - X) - \sigma^{-2} E[X k_{b_n}(x - X)] X.$$

The correction term is approximately $\sigma^{-2} x f(x) X$. Finally we replace unknown quantities by estimators.

In Section 2 we apply the empirical likelihood approach to obtain residual-based weighted kernel estimators $\hat{f}_w$ and $\hat{g}_w$ for the densities $f$ and $g$ of $\varepsilon_t$ and $Y_t$, respectively. The kernel estimator $\hat{f}(x) = (1/(n - p_n)) \sum_{j=p_n+1}^n k_b(x - \hat{\varepsilon}_j)$ is based on asymptotically *independent* residuals $\hat{\varepsilon}_j = X_j - \sum_{i=1}^{p_n} \hat{\varrho}_i X_{j-i}$ with appropriate estimators $\hat{\varrho}_i$ for $\varrho_i$, and the weighted version is $\hat{f}_w(x) = (1/(n - p_n)) \sum_{j=p_n+1}^n w_j k_b(x - \hat{\varepsilon}_j)$ with weights $w_j$ fulfilling the one-dimensional constraint $\sum_{j=p_n+1}^n w_j \hat{\varepsilon}_j = 0$. On the other hand, the kernel estimator $\hat{g}(x) = (1/(n - p_n)) \sum_{j=p_n+1}^n k_b(x - \hat{Y}_j)$ is based on *dependent* $\hat{Y}_j = X_j - \hat{\varepsilon}_j$, and we use a weighted version

$$\frac{1}{n - p_n - l_n} \sum_{j=p_n+l_n+1}^n w_j^* k_b(x - \hat{Y}_j)$$

3

with weights fulfilling an increasing number $l_n \sim \log n$ of constraints, i.e. "infinitely many" constraints

$$\frac{1}{n - p_n - l_n} \sum_{j=p_n+l_n+1}^{n} w_j^* \hat{\varepsilon}_{j-t} = 0, \quad t = 1, \dots, l_n.$$

Our main result gives a stochastic expansion of the convolution estimator $\hat{f}_w * \hat{g}_w$ for the stationary density. It holds both in $C_0$ and in weighted versions of $L_1$ and implies a functional central limit theorem for $\hat{f}_w * \hat{g}_w$ in these two spaces. The result follows from Schick and Wefelmeyer (2007b) and (2008a), together with stochastic expansions of $\hat{f}_w$ and $\hat{g}_w$ around $\hat{f}$ and $\hat{g}$, respectively, which are consequences of the lemmas in Section 3.

## 2   RESULTS

We consider observations $X_1, \dots, X_n$ of a stationary linear process $X_t$, $t \in \mathbb{Z}$, with infinite-order moving average representation (1.1). We assume that the coefficients $\varphi_s$ are summable and the innovations $\varepsilon_t$ are i.i.d. with density $f$, mean zero and finite variance. Then $X_0$ has a density, say $h$. In order to construct a $n^{1/2}$-consistent estimator of $h$, we write $X_t = \varepsilon_t + Y_t$ with

$$Y_t = X_t - \varepsilon_t = \sum_{s=1}^{\infty} \varphi_s \varepsilon_{t-s}.$$

For $Y_t$ to have a density $g$, we must exclude the degenerate case that the observations are i.i.d.:

(C) *At least one of the moving average coefficients $\varphi_s$ is nonzero.*

Then we can express the density $h$ of $X_0$ as the convolution $h = f * g$ of $f$ and $g$ and obtain an estimator of $h$ as $\hat{f} * \hat{g}$, where $\hat{f}$ and $\hat{g}$ are estimators of $f$ and $g$. We base these estimators on estimators of the innovations. For this we require invertibility of the process.

(I) *The function $\phi(z) = 1 + \sum_{s=1}^{\infty} \varphi_s z^s$ is bounded and bounded away from zero on the complex unit disk $\{z \in \mathbb{C} : |z| \leq 1\}$.*

Then $\rho(z) = 1/\phi(z) = 1 - \sum_{s=1}^{\infty} \varrho_s z^s$ is also bounded and bounded away from zero on the complex unit disk. Hence the innovations have the infinite-order autoregressive representation

$$\varepsilon_t = X_t - \sum_{s=1}^{\infty} \varrho_s X_{t-s}. \tag{2.1}$$

Let $p_n$ be positive integers with $p_n/n \to 0$. For $j = p_n + 1, \dots, n$ we mimic the innovation $\varepsilon_j$ by the residual

$$\hat{\varepsilon}_j = X_j - \sum_{i=1}^{p_n} \hat{\varrho}_i X_{j-i},$$

4

where $\hat{\varrho}_i$ is an estimator of $\varrho_i$ for $i = 1, \ldots, p_n$. We then estimate the innovation density by a kernel estimator based on the residuals,

$$\hat{f}(x) = \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} k_{b_n}(x - \hat{\varepsilon}_j), \quad x \in \mathbb{R},$$

and we estimate the density $g$ by a kernel estimator based on the differences $\hat{Y}_j = X_j - \hat{\varepsilon}_j$,

$$\hat{g}(x) = \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} k_{b_n}(x - \hat{Y}_j), \quad x \in \mathbb{R}.$$

In addition to (C) and (I) we use the following assumptions.

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

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

$$\sum_{i=1}^{p_n} (\hat{\varrho}_i - \varrho_i)^2 = O_p(q_n n^{-1})$$

*for some* $q_n$ *with* $1 \le q_n \le p_n$.

The usual estimators of the autoregression coefficients are the least squares estimators $\hat{\varrho}_1, \ldots, \hat{\varrho}_{p_n}$ which minimize $\sum_{j=p_n+1}^{n}(X_j - \sum_{i=1}^{p_n} \varrho_i X_{j-i})^2$. By Lemma 1 in Schick and Wefelmeyer (2007b), they meet condition (R) with $q_n = p_n$ if in addition

$$n p_n \sum_{s>p_n} \varrho_s^2 \to 0 \tag{2.2}$$

holds.

**1. Results for the sup-norm.** We obtain stochastic expansions for estimators of $h$ both in $C_0$ and in weighted $L_1$-spaces. The following assumptions are used for results in $C_0$.

(S) *The moving average coefficients satisfy* $\sum_{s=1}^{\infty} s|\varphi_s| < \infty$.

(F) *The density* $f$ *has mean zero, a finite fourth moment, is absolutely continuous with a bounded and integrable (almost everywhere) derivative* $f'$, *and the function* $x \mapsto x f'(x)$ *is bounded and integrable.*

We denote the number of non-zero coefficients among $\{\varphi_s : s \ge 1\}$ by

$$N = \sum_{s \ge 1} 1[\varphi_s \ne 0].$$

5

Then we can express (C) as $N \geq 1$. Smoothness of $g$ and $h$ can be linked to the number $N$. Our main results will thus be formulated in terms of $N$. The following conditions on the kernel and the bandwidth are kept general to allow for various smoothness assumptions in terms of a natural number $m$ that will play the role of a (known) minimal size for $N$. Under (C), we know that $N \geq 1$ so that we can always take $m = 1$.

(B) *The sequences $b_n$, $p_n$ and $q_n$ and the exponent $\zeta$ fulfill $p_n q_n b_n^{-1} n^{-1/2} \to 0$, $n b_n^{2m+2} = O(1)$, $n^{1/4} s_n \to 0$, $n^{1/2} b_n s_n = O(1)$, where*

$$s_n = b_n^{-1/2} n^{-1/2} + p_n q_n b_n^{-5/2} n^{-1} + b_n^{-3/2} n^{-\zeta - 1/2}.$$

(K) *The kernel $k$ has bounded, continuous and integrable derivatives up to order two and is of type $(m + 1, 2)$ as defined below.*

Here a kernel $k$ is said to be *of type $(m, c)$* if $\int t^i k(t)\, dt = 0$ for $i = 1, \ldots, m$ and if $\int |t|^{mc} |k(t)|\, dt$ is finite.

A possible choice of bandwidth is $b_n \sim n^{-1/(2m+2)}$. Then (B) is met if $4(m+1)\zeta > 1$ and $p_n q_n n^{-(2m-1)/(4(m+1))} \to 0$ hold. In particular, $p_n = q_n \sim n^\beta$ requires $8(m+1)\beta < 2m - 1$.

Let $\mathbb{G}_n$, $\mathbb{F}_n$ and $\mathbb{H}_n$ denote the processes defined by

$$\mathbb{F}_n(x) = \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \big( f(x - Y_j) - E[f(x - Y_j)] \big),$$

$$\mathbb{G}_n(x) = \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \big( g(x - \varepsilon_j) - E[g(x - \varepsilon_j)] \big),$$

$$\mathbb{H}_n(x) = \sum_{i=1}^{p_n} (\hat{\varrho}_i - \varrho_i) E[X_0 k_{b_n}(x - Y_i)]$$

for $x \in \mathbb{R}$.

Let $N \geq m$ and suppose that (I), (Q), (R), (S), (F), (K) and (B) hold. For the sup-norm $\| \cdot \|_\infty$, Schick and Wefelmeyer (2007b) show that

$$\|\hat{f} * \hat{g} - h - \mathbb{F}_n - \mathbb{G}_n + f' * \mathbb{H}_n\|_\infty = o_p(n^{-1/2}) \tag{2.3}$$

and that $n^{1/2} \mathbb{G}_n$ and $n^{1/2} \mathbb{F}_n$ are tight in $C_0$.

If the least squares estimators are used and (2.2) holds, then $n^{1/2} f' * \mathbb{H}_n$ is also tight in $C_0$. Moreover, the random vector $\hat{\Delta} = (\hat{\varrho}_1 - \varrho_1, \ldots, \hat{\varrho}_{p_n} - \varrho_{p_n})^\top$ fulfills

$$\hat{\Delta} = W_n^{-1} \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \mathbf{X}_{j-1} \varepsilon_j + o_p(n^{-1/2}) \tag{2.4}$$

with $\mathbf{X}_{j-1} = (X_{j-1}, \ldots, X_{j-p_n})^\top$ and $W_n = E[\mathbf{X}_0\mathbf{X}_0^\top]$. Then (2.3) can be written

$$\|n^{1/2}(\hat{f} * \hat{g} - h) - \mathbb{Z}_n\|_\infty = o_p(1) \tag{2.5}$$

with

$$\mathbb{Z}_n(x) = n^{-1/2} \sum_{j=1}^n \left( g(x - \varepsilon_j) + f(x - Y_j) - 2h(x) + \varepsilon_j \mathbf{X}_{j-1}^\top W_n^{-1} E[\mathbf{X}_0 f'(x - Y_1)] \right).$$

It follows that the process $n^{1/2}(\hat{f} * \hat{g} - h)$ converges weakly in $C_0$ to a centered Gaussian process with covariance

$$\Gamma(s, t) = \lim_{n \to \infty} \mathrm{Cov}(\mathbb{Z}_n(s), \mathbb{Z}_n(t)), \qquad s, t \in \mathbb{R}.$$

Since $f$ and $g$ are known to have mean zero, we can replace $\hat{f}$ and $\hat{g}$ by estimators that are constructed using this information. An estimator of $f$ that uses this information is the weighted density estimator

$$\hat{f}_w(x) = \frac{1}{n - p_n} \sum_{j=p_n+1}^n \frac{k_b(x - \hat{\varepsilon}_j)}{1 + \lambda\hat{\varepsilon}_j}, \quad x \in \mathbb{R},$$

where $\lambda$ is chosen such that

$$\frac{1}{n - p_n} \sum_{j=p_n+1}^n \frac{\hat{\varepsilon}_j}{1 + \lambda\hat{\varepsilon}_j} = 0$$

and $1 + \lambda\hat{\varepsilon}_j > 0$ for $j = 1, \ldots, n$.

A weighted version of $g$ could be obtained by mimicking the above with $\hat{Y}_j$ replacing $\hat{\varepsilon}_j$. Since the random variables $Y_j$ are not independent, we choose an alternative method which incorporates the dependence structure. One possibility is the blockwise empirical likelihood of Kitamura (1997) for weakly dependent stationary processes. In view of the linear structure for our model, we instead work with a different approach that uses an increasing number of constraints which are correlated with the observations. Let $l_n$ denote an integer that tends to infinity slowly, say $l_n \sim \log n$, and set $\hat{Z}_j = (\hat{\varepsilon}_{j-1}, \ldots, \hat{\varepsilon}_{j-l_n})^\top$. We now work with the weighted kernel estimator

$$\hat{g}_w(x) = \frac{1}{n - p_n - l_n} \sum_{j=p_n+l_n+1}^n \frac{k_b(x - \hat{Y}_j)}{1 + \mu^\top \hat{Z}_j}, \quad x \in \mathbb{R},$$

where $\mu$ is chosen such that

$$\frac{1}{n - p_n - l_n} \sum_{j=p_n+l_n+1}^n \frac{\hat{Z}_j}{1 + \mu^\top \hat{Z}_j} = 0$$

and $1 + \mu^\top \hat{Z}_j > 0$ for $j = p_n + l_n + 1, \ldots, n$.

We shall show that for our weighted density estimators, the expansion (2.3) holds with $\mathbb{F}_n$ and $\mathbb{G}_n$ replaced by

$$\mathbb{F}_n^*(x) = \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \big(f(x - Y_j) - E[f(x - Y_j)] - \gamma(x)\sigma^{-2}\varepsilon_j\big),$$

$$\mathbb{G}_n^*(x) = \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \big(g(x - \varepsilon_j) - E[g(x - \varepsilon_j)] - \phi * g(x)\sigma^{-2}\varepsilon_j\big),$$

where $\sigma^2 = E[\varepsilon_0^2]$, $\phi(x) = xf(x)$ and

$$\gamma(x) = \sum_{s=1}^{\infty} E[\varepsilon_0 f(x - Y_s)], \quad x \in \mathbb{R}.$$

It is easy to check that

$$\gamma(x) = \lim_{n\to\infty} (n - p_n)E[\mathbb{F}_n(x)\bar{\varepsilon}_n] \tag{2.6}$$

is the asymptotic covariance of $\mathbb{F}_n(x)$ and the average

$$\bar{\varepsilon}_n = \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \varepsilon_j, \tag{2.7}$$

while

$$g * \phi(x) = E[\varepsilon_0 g(x - \varepsilon_0)] = (n - p_n)E[\mathbb{G}_n(x)\bar{\varepsilon}_n] \tag{2.8}$$

is the asymptotic covariance of $\mathbb{G}_n(x)$ and $\bar{\varepsilon}_n$.

**Theorem 2.1.** *Let $N \geq m$ and suppose that (I), (Q), (R), (S), (F), (K) (B) and $l_n \sim \log n$ hold. Then*

$$\|\hat{f}_w * \hat{g}_w - h - \mathbb{F}_n^* - \mathbb{G}_n^* + f' * \mathbb{H}_n\|_\infty = o_p(n^{-1/2}).$$

*Proof.* The proof is based on the lemmas given in Section 3. In view of (2.3) it suffices to prove

$$\|\hat{f}_w * \hat{g}_w - \hat{f} * \hat{g} + (\phi * g + \gamma)\sigma^{-2}\bar{\varepsilon}_n\|_\infty = o_p(n^{-1/2}). \tag{2.9}$$

By (F), the function $\phi * g$ is bounded. Thus, in view of Lemma 3.1 and part (a) of Lemma 3.6, it suffices to verify

$$\|\hat{f}_w * \hat{g}_w - \hat{f} * \hat{g} + \lambda\phi * g + f * (\mu^\top \hat{\Gamma})\|_\infty = o_p(n^{-1/2}), \tag{2.10}$$

with $\hat{\Gamma}$ defined in (3.16). Set $r_f = \hat{f}_w - \hat{f} + \lambda\phi$ and $r_g = \hat{g}_w - \hat{g} + \mu^\top \hat{\Gamma}$. Then we can express

$$\hat{f}_w * \hat{g}_w - \hat{f} * \hat{g} + \lambda\phi * g + f * (\mu^\top \hat{\Gamma})$$
$$= (\hat{f} - \lambda\phi + r_f) * (\hat{g} - \mu^\top \hat{\Gamma} + r_g) - \hat{f} * \hat{g} + \lambda\phi * g + f * (\mu^\top \hat{\Gamma})$$
$$= -(\hat{f} - f) * (\mu^\top \hat{\Gamma}) + \hat{f} * r_g - \lambda\phi * (\hat{g} - g) + \lambda\phi * (\mu^\top \hat{\Gamma}) - \lambda\phi * r_g + r_f * \hat{g}_w.$$

8

From Schick and Wefelmeyer (2007b) we obtain $\|\hat{f} - f\|_2 = o_p(n^{-1/4})$ and $\|\hat{g} - g\|_2 = o_p(n^{-1/4})$. Since $b_n^{-1} = O(n^{1/4})$ by condition (B), we derive from Lemmas 3.1, 3.2 and 3.3 that $\|r_f\|_2 = o_p(n^{-1/2})$ and from Lemma 3.5 that $\|r_g\|_2 = o_p(n^{-1/2})$. By Lemma 3.1 we have $\lambda = O_p(n^{-1/2})$. Furthermore, $\phi$ is in $L_2$ and $\|\mu^\top \hat{\Gamma}\|_2 = O_p(l_n n^{-1/2} b^{-1/2})$. Using

$$\|u * v\|_\infty \leq \|u\|_2 \|v\|_2$$

we see that (2.10) holds. $\qquad\square$

Under (F) the functions $\phi$ and $g$ belong to $L_2$ implying that their convolution $\phi * g$ belongs to $C_0$. Under (F) one also has $\gamma$ in $C_0$; for details see the proof of Lemma 3.6 below. If least squares estimators are used, then (2.9) implies

$$\|n^{1/2}(\hat{f}_w * \hat{g}_w - h) - \mathbb{Z}_n^w\|_\infty = o_p(1) \tag{2.11}$$

with

$$\mathbb{Z}_n^w(x) = \mathbb{Z}_n(x) - (\phi * g(x) + \gamma(x))\sigma^{-2} n^{-1/2} \sum_{j=1}^n \varepsilon_j.$$

It follows that the process $n^{1/2}(\hat{f}_w * \hat{g}_w - h)$ converges in $C_0$ to a centered Gaussian process with covariance function

$$\Gamma_w(s,t) = \lim_{n \to \infty} \mathrm{Cov}(\mathbb{Z}_n^w(s), \mathbb{Z}_n^w(t)), \qquad s, t \in \mathbb{R}.$$

Using the results (2.6) and (2.8) and the fact that $\varepsilon_i$ and $X_{j-1}\varepsilon_j$ are uncorrelated for all $i$ and $j$, we see that the covariance of $\mathbb{Z}_n^w(x)$ and $\mathbb{Z}_n(x) - \mathbb{Z}_n^w(x)$ converges to zero. Thus the decrease in the asymptotic variance at a point $x$ is

$$\Gamma(x,x) - \Gamma_w(x,x) = \mathrm{Var}(\mathbb{Z}_n(x) - \mathbb{Z}_n^w(x)) = (\phi * g(x) + \gamma(x))^2 \sigma^{-2}.$$

We should note that

$$\phi * g(x) + \gamma(x) = \sum_{s=0}^\infty \int y h_s(x - \varphi_s y) f(y)\, dy$$

where $h_s$ denotes the density of $X_s - \varphi_s \varepsilon_0$ for $s = 0, 1, 2, \ldots$ with $\varphi_0 = 1$. Note that $h_0 = g$.

**Example 1.** Let $X_t = \vartheta X_{t-1} + \varepsilon_t$ be an AR(1) process with $\vartheta \neq 0$ and $|\vartheta| < 1$. Then $E[X_0^2] = \sigma^2/(1 - \vartheta^2)$. Furthermore, $Y_t = X_t - \varepsilon_t = \vartheta X_{t-1}$ and therefore

$$\nu'(x) = E[X_0 f'(x - Y_1)] = E[X_0 f'(x - \vartheta X_0)]$$

9

and

$$c(x) = \phi * g(x) + \gamma(x) = E[\varepsilon_0 g(x - \varepsilon_0)] + \sum_{s=0}^{\infty} E[\varepsilon_0 f(x - \vartheta X_s)].$$

Here we can take $p_n = 1$. The least squares estimator for $\vartheta$ is

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

We have

$$\hat{\vartheta} = \sigma^{-2}(1 - \vartheta^2)\frac{1}{n}\sum_{j=1}^{n} X_{j-1}\varepsilon_j + o_p(n^{-1/2}).$$

Hence the stochastic expansion (2.5) of $\hat{f} * \hat{g}$ holds with

$$\mathbb{Z}_n(x) = n^{-1/2}\sum_{j=1}^{n}\left(g(x - \varepsilon_j) + f(x - \vartheta X_{j-1}) - 2h(x) + \sigma^{-2}(1 - \vartheta^2)\nu'(x)X_{j-1}\varepsilon_j\right).$$

For the weighted version $\hat{f}_w * \hat{g}_w$ we obtain from Theorem 2.1 that

$$\|n^{1/2}(\hat{f}_w * \hat{g}_w - h) - \mathbb{Z}_n^w\|_\infty = o_p(1)$$

with

$$\mathbb{Z}_n^w(x) = \mathbb{Z}_n(x) - c(x)\sigma^{-2}n^{-1/2}\sum_{j=1}^{n}\varepsilon_j.$$

A formula for the asymptotic variance of $\mathbb{Z}_n(x)$ is given by

$$\begin{aligned}
\Gamma(x, x) =\ & g^2 * f(x) - h^2(x) + (1 - \vartheta^2)(\nu'(x))^2 + f^2 * g(x) - h^2(x) \\
& + 2\sum_{r=1}^{\infty}\left(\int g(x - y)h_r(x - \vartheta^r y)f(y)\,dy - h^2(x)\right) \\
& + 2(1 - \vartheta^2)\sigma^{-2}\nu'(x)\sum_{r=1}^{\infty}\iint yz\bar{f}_{r-1}(x - \vartheta^r y - \vartheta^{r+1}z)f(y)h(z)\,dy\,dz \\
& + 2\sum_{r=1}^{\infty}\left(\int f(x - \vartheta z)f_r(x - \vartheta^{r+1}z)h(z)\,dz - h^2(x)\right)
\end{aligned}$$

where $f_r$ is the density of $\sum_{i=0}^{r}\vartheta^i\varepsilon_i$ and $h_r$ is the density of $X_r - \varphi_r\varepsilon_0$.

Now assume that $f$ is the standard normal density. Then the asymptotic variance can be calculated explicitly and $c(x)$ equals $(1 + \vartheta)xh(x)$. In this case the relative variance reduction

$J_\vartheta(x) = c^2(x)/\Gamma(x,x)$ is an even function of $x$. Graphs of this function over the interval $[0,10]$ are given in the next figure for four choices of the parameter, namely $\vartheta = .3, .5, .7, .9$.
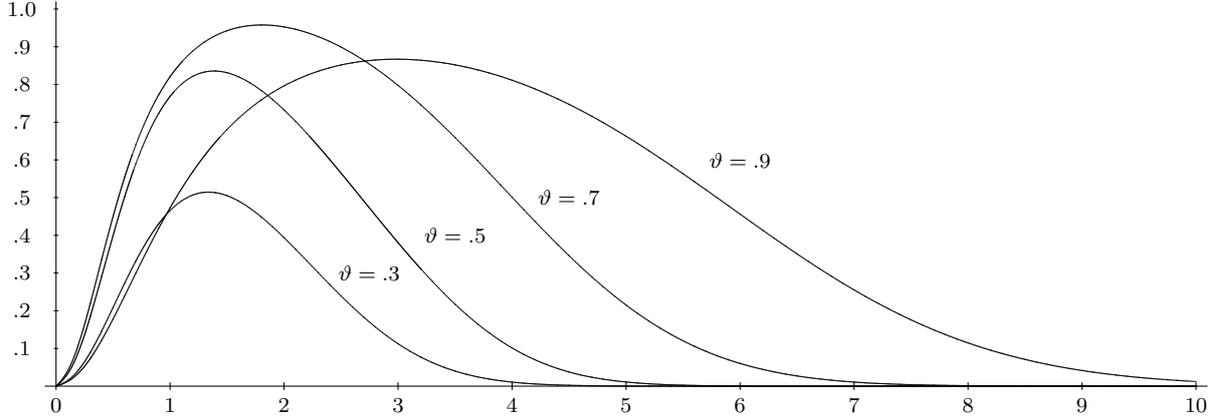


From the figure we see that variance reductions of over 90 percent are possible.

**Example 2.** Let $X_t = \varepsilon_t + \vartheta \varepsilon_{t-1}$ be an MA(1) process with $\vartheta \neq 0$ and $|\vartheta| < 1$. Then $Y_t = X_t - \varepsilon_t = \vartheta \varepsilon_{t-1}$ has density given by $g(x) = f_\vartheta(x) = f(x/\vartheta)/|\vartheta|$, and $\gamma(x) = E[\varepsilon_0 f(x - \vartheta \varepsilon_0)]$. Let $\hat\vartheta$ be a $n^{1/2}$-consistent estimator for $\vartheta$. As in Example 2 of Schick and Wefelmeyer (2007b) we obtain the stochastic expansion

$$\|n^{1/2}(\hat f * \hat g - h) - \mathbb{Z}_n\|_\infty = o_p(1)$$

with

$$\mathbb{Z}_n(x) = n^{-1/2} \sum_{j=1}^n \big( f_\vartheta(x - \varepsilon_j) + f(x - \vartheta \varepsilon_j) - 2h(x) \big)$$
$$- n^{1/2}(\hat\vartheta - \vartheta) E[\varepsilon_0 f'(x - \vartheta \varepsilon_0)].$$

We have $\phi * g(x) = E[\varepsilon_0 f_\vartheta(x - \varepsilon_0)]$. For the weighted version $\hat f_w * \hat g_w$ we obtain from Theorem 2.1 that

$$\|n^{1/2}(\hat f_w * \hat g_w - h) - \mathbb{Z}_n^w\|_\infty = o_p(1)$$

with

$$\mathbb{Z}_n^w(x) = \mathbb{Z}_n(x) - a(x) n^{-1/2} \sum_{j=1}^n \varepsilon_j,$$

where $a(x) = (E[\varepsilon_0 f_\vartheta(x - \varepsilon_0)] + E[\varepsilon_0 f(x - \vartheta \varepsilon_0)])\sigma^{-2}$. The weighted estimator $\hat f_w * \hat g_w(x)$ is asymptotically equivalent to the additively corrected estimator

$$\hat f * \hat g(x) - \hat a(x) \frac{1}{n - p_n} \sum_{j=p_n+1}^n \hat\varepsilon_j$$

11

of Schick and Wefelmeyer (2004a), with $\hat{a}(x)$ a consistent estimator for $a(x)$. Both estimators are efficient for $h(x)$ if an efficient estimator for $\vartheta$ is used. The asymptotic variance reduction of $\hat{f}_w * \hat{g}_w(x)$ over $\hat{f} * \hat{g}(x)$ is $a^2(x)\sigma^2 = (E[\varepsilon_0 f_\vartheta(x - \varepsilon_0)] + E[\varepsilon_0 f(x - \vartheta\varepsilon_0)])^2 \sigma^{-2}$.

**2. Results for weighted $L_1$-spaces.** Results analogous to those in $C_0$ can be obtained in weighted $L_1$-spaces. We denote by $L_V$ the space of functions $a$ with finite $V$-norm $\|a\|_V = \int V(x)|a(x)|\,dx$. In the following we consider the choice $V_r(x) = (1 + |x|)^r$ for some non-negative $r$. The case $r = 0$ or $V = 1$ corresponds to the natural distance between densities. The choice $V_r$ is useful if we want to estimate expectations $E[a(X_0)] = \int a(x)h(x)\,dx$ for functions $a$ bounded by a multiple of $V_r$. We need modified versions of Assumptions (S), (F), (K) and (B).

(SV) *The moving average coefficients satisfy $\sum_{s=1}^{\infty} s^\beta |\varphi_s| < \infty$ for some $\beta > 1$.*

We say that a function $a$ has *finite $V$-variation* if there are measures $\mu_1$ and $\mu_2$ of equal mass with $\int V\,d(\mu_1 + \mu_2)$ finite such that $a(x) = \mu_1((-\infty, x]) - \mu_2((-\infty, x])$ for Lebesgue almost all $x$. In this case, we call $\int V\,d(\mu_1 + \mu_2)$ the *$V$-variation of $a$*. Aside from moment conditions, we require only that $f$ has finite $V_{r+1}$-variation. In particular, $f$ need not be continuous.

(FV) *The density $f$ has mean zero, a finite moment of order $\xi > 2r + 3$ and finite $V_{r+1}$-variation.*

We formulate our assumptions on the kernel and the bandwidth in terms of the moment order $\xi$ and a natural number $m$ that plays again the role of a known minimal size for $N$.

(KV) *The kernel $k$ is twice continuously differentiable with bounded derivatives and $k$, $k'$ and $k''$ have finite $V_{2r+2}$-norms. Furthermore, $k$ has finite $V_{r+m+1}$-norm and satisfies $\int t^i k(t)\,dt = 0$ for $i = 1, \ldots, m$.*

(BV) *The bandwidth $b_n$ satisfies*

$$nb_n^{2m+2} \to 0, \quad n^{-1/4-\zeta}b_n^{-1} \to 0, \quad n^{-1/4}b_n^{-1/2} \to 0,$$

*and the sequences $b_n$, $q_n$ and $p_n$ satisfy*

$$n^{-3/4}p_n q_n b_n^{-2} \to 0, \quad n^{-1/2}p_n q_n b_n^{-1} \to 0 \quad \text{and} \quad p_n q_n n^{-1+2/\xi} = O(1).$$

Define

$$\nu_i(x) = E[X_0 f(x - Y_i)], \qquad x \in \mathbb{R}.$$

Let $N \geq m$ and suppose that (I), (Q), (R), (SV), (FV), (KV) and (BV) hold. Schick and Wefelmeyer (2008a) show that

$$\left\| \hat{f} * \hat{g} - h - \mathbb{F}_n - \mathbb{G}_n + \sum_{i=1}^{p_n} (\hat{\varrho}_i - \varrho_i) \nu_i' \right\|_{V_r} = o_p(n^{-1/2}). \qquad (2.12)$$

We have the following result for the weighted estimator $\hat{f}_w * \hat{g}_w$.

**Theorem 2.2.** *Let $N \geq m$ and suppose that (I), (Q), (R), (SV), (FV), (KV) (BV) and $l_n \sim \log n$ hold. Then*

$$\left\| \hat{f}_w * \hat{g}_w - h - \mathbb{F}_n^* - \mathbb{G}_n^* + \sum_{i=1}^{p_n} (\hat{\varrho}_i - \varrho_i) \nu_i' \right\|_{V_r} = o_p(n^{-1/2}).$$

*Proof.* In view of (2.12) it suffices to prove

$$\left\| \hat{f}_w * \hat{g}_w - \hat{f} * \hat{g} + (\phi * g + \gamma) \sigma^{-2} \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \varepsilon_j \right\|_{V_r} = o_p(n^{-1/2}).$$

The proof of this is parallel to that of Theorem 2.1, now using

$$\|u * v\|_{V_r} \leq \|u\|_{V_r} \|v\|_{V_r}.$$

$\square$

# 3    LEMMAS

In this section we always assume that the following condition is met.

(A) *The density $f$ has mean zero and a finite moment of order $\beta > 2$. The conditions (I), (Q) and (R) hold. The rates $p_n q_n n^{-1+2/\beta} = O(1)$ and $l_n \sim \log n$ hold.*

Note that the assumptions on $f$ are met under (F) and (FV). The boundedness of $p_n q_n n^{-1+2/\beta}$ is implied by (B) and (F) for the choice $\beta = 4$, and by (BV) and (FV) for the choice $\beta = \xi$.

Lemma 7 in Schick and Wefelmeyer (2007b) implies the following properties of the residuals:

$$\frac{1}{n - p_n} \sum_{j=p_n+1}^{n} (\hat{\varepsilon}_j - \varepsilon_j)^2 = O_p(p_n q_n n^{-1}) = O_p(n^{-2/\beta}), \qquad (3.1)$$

$$\max_{p_n < j \leq n} |\hat{\varepsilon}_j - \varepsilon_j| = O_p(n^{-\zeta}) + o_p(p_n^{1/2} q_n^{1/2} n^{-1/2+1/\beta}) = o_p(1), \qquad (3.2)$$

13

$$\frac{1}{n-p_n} \sum_{j=p_n+1}^{n} (\hat{\varepsilon}_j - \varepsilon_j) = O_p(n^{-1/2-\zeta}) + O_p(p_n^{1/2} q_n^{1/2} n^{-1}) = O_p(n^{-1/2-\tau}) \qquad (3.3)$$

with $\tau = \min\{\zeta, 1/\beta\} > 0$. The second property implies that

$$\frac{1}{n-p_n} \sum_{j=p_n+1}^{n} \hat{\varepsilon}_j^2 = \sigma^2 + o_p(1) \qquad (3.4)$$

and

$$\max_{p_n < j \le n} |\hat{\varepsilon}_j| \le \max_{p_n < j \le n} |\varepsilon_j| + \max_{p_n < j \le n} |\hat{\varepsilon}_j - \varepsilon_j| = o_p(n^{1/\beta}). \qquad (3.5)$$

Here $\max_{p_n < j \le n} |\varepsilon_j| = o_p(n^{1/\beta})$ follows from the fact that the innovations have a finite moment of order $\beta$.

Using (3.3), (3.4) and (3.5), we derive as in Owen (2001, pages 219–221) the following result.

**Lemma 3.1.** *Suppose (A) holds. Then*

$$\lambda = \sigma^{-2} \bar{\varepsilon}_n + o_p(n^{-1/2}).$$

Let $V$ be a continuous function on $\mathbb{R}$ satisfying $V(0) = 1$ and

$$V(x+y) \le V(x)V(y), \quad x, y \in \mathbb{R};$$
$$V(sx) \le V(x), \quad |s| \le 1, x \in \mathbb{R}.$$

Using the properties of $V$ and $\delta_n = \max_{p_n < j \le n} |\hat{\varepsilon}_j - \varepsilon_j|) = o_p(1)$, we obtain that

$$\frac{1}{n-p_n} \sum_{j=p_n+1}^{n} V(\hat{\varepsilon}_j) \le V(\delta_n) \frac{1}{n-p_n} \sum_{j=p_n+1}^{n} V(\varepsilon_j) = O_p(1) \qquad (3.6)$$

if $E[V(\varepsilon_0)] = \|f\|_V$ is finite, and

$$\frac{1}{n-p_n} \sum_{j=p_n+1}^{n} V(\hat{Y}_j) \le V(\delta_n) \frac{1}{n-p_n} \sum_{j=p_n+1}^{n} V(Y_j) = O_p(1) \qquad (3.7)$$

if $E[V(Y_1)] = \|g\|_V$ is finite. We set

$$V_*(x) = (1 + |x|)V(x), \qquad V_{**}(x) = (1 + |x|)^2 V(x), \quad x \in \mathbb{R},$$

and note that $V_*$ and $V_{**}$ have the same properties as $V$. Finally, we let

$$\hat{\phi}(x) = x\hat{f}(x), \quad x \in \mathbb{R}.$$

14

**Lemma 3.2.** *Suppose (A) holds and the bandwidth satisfies $b_n \to 0$ and $nb_n^2 \to \infty$. (a) If $\beta = 4$ and $\int (1+u^2)k^2(u)\,du$ is finite, then*

$$\|\hat{f}_w - \hat{f} + \lambda\hat{\phi}\|_2 = o_p(n^{-1/2}).$$

*(b) If $f$ and $k$ have finite $V_*$-norms, then*

$$\|\hat{f}_w - \hat{f} + \lambda\hat{\phi}\|_V = o_p(n^{-1/2}).$$

*Proof.* It follows from (3.5) and Lemma 3.1 that

$$\max_{p_n < j \le n} |\lambda\hat{\varepsilon}_j| = o_p(1). \tag{3.8}$$

Writing

$$\frac{1}{1 + \lambda\hat{\varepsilon}_j} = 1 - \lambda\hat{\varepsilon}_j + \frac{\lambda^2\hat{\varepsilon}_j^2}{1 + \lambda\hat{\varepsilon}_j},$$

we see that

$$\hat{f}_w(x) - \hat{f}(x) + \lambda x \hat{f}(x) = \lambda A(x) + \lambda^2 B(x)$$

with

$$A(x) = \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} (x - \hat{\varepsilon}_j)k_{b_n}(x - \hat{\varepsilon}_j),$$

$$B(x) = \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \frac{\hat{\varepsilon}_j^2}{1 + \lambda\hat{\varepsilon}_j} k_{b_n}(x - \hat{\varepsilon}_j).$$

An application of the Cauchy–Schwarz inequality and the substitution $x = \hat{\varepsilon}_j + b_n u$ show that

$$\|A\|_2^2 \le \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \int (x - \hat{\varepsilon}_j)^2 k_{b_n}^2(x - \hat{\varepsilon}_j)\,dx = b_n \int u^2 k^2(u)\,du.$$

Another application of the Cauchy–Schwarz inequality and the same substitution show that

$$b_n\|B\|_2^2 \le \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} \frac{\hat{\varepsilon}_j^4}{(1 + \lambda\hat{\varepsilon}_j)^2} \int b_n k_{b_n}^2(x - \hat{\varepsilon}_j)\,dx$$

$$\le \max_{p_n < j \le n} \frac{1}{(1 + \lambda\hat{\varepsilon}_j)^2} \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} (1 + |\hat{\varepsilon}_j|)^4 \int k^2(u)\,du.$$

Thus, if $f$ has a finite fourth moment and $\int (1+u^2)k^2(u)\,du$ is finite, then we have

$$n\|\hat{f}_w - \hat{f} + \lambda\hat{\phi}\|_2^2 \le 2n\lambda^2\|A\|_2^2 + 2n\lambda^4\|B\|_2^2 = o_p(1)$$

in view of $n\lambda^2 = O_p(1)$, (3.8), (3.6) with $V(x) = (1 + |x|)^4$, and the properties of $k$ and $b_n$.

15

For large $n$ such that $b_n < 1$ we have the bounds

$$\|A\|_V = \int V(x)\left|\frac{1}{n - p_n}\sum_{j=p_n+1}^{n}(x - \hat{\varepsilon}_j)k_{b_n}(x - \hat{\varepsilon}_j)\right|dx$$

$$\leq \frac{1}{n - p_n}\sum_{j=p_n+1}^{n}\int V(\hat{\varepsilon}_j + b_n u)|uk(u)|b_n\,du$$

$$\leq \frac{1}{n - p_n}\sum_{j=p_n+1}^{n}V(\hat{\varepsilon}_j)\int V(u)|uk(u)|b_n\,du$$

$$\leq \frac{1}{n - p_n}\sum_{j=p_n+1}^{n}V_*(\hat{\varepsilon}_j)b_n\|k\|_{V_*}$$

and

$$\|\lambda B\|_V = \int V(x)\left|\frac{1}{n - p_n}\sum_{j=p_n+1}^{n}\frac{\lambda\hat{\varepsilon}_j^2}{1 + \lambda\hat{\varepsilon}_j}k_{b_n}(x - \hat{\varepsilon}_j)\right|dx$$

$$\leq \frac{1}{n - p_n}\sum_{j=p_n+1}^{n}\frac{\lambda\hat{\varepsilon}_j^2}{1 + \lambda\hat{\varepsilon}_j}\int V(\hat{\varepsilon}_j + b_n u)|k(u)|\,du$$

$$\leq \frac{1}{n - p_n}\sum_{j=p_n+1}^{n}\frac{\lambda\hat{\varepsilon}_j^2}{1 + \lambda\hat{\varepsilon}_j}V(\hat{\varepsilon}_j)\int V(u)|k(u)|\,du.$$

$$\leq \max_{p_n < j \leq n}\frac{|\lambda\hat{\varepsilon}_j|}{|1 + \lambda\hat{\varepsilon}_j|}\frac{1}{n - p_n}\sum_{j=p_n+1}^{n}V_*(\hat{\varepsilon}_j)\|k\|_{V_*}.$$

Thus, if $f$ and $k$ have finite $V_*$-norms, then we have

$$\|\hat{f}_w - \hat{f} + \lambda\phi\|_V \leq |\lambda|(\|A\|_V + \|\lambda B\|_V) = o_p(n^{-1/2})$$

in view of $\lambda = O_p(n^{-1/2})$, (3.6) applied with $V_*$ in place of $V$, and (3.8). $\qquad\square$

**Lemma 3.3.** *Suppose (A) holds and the bandwidth satisfies $b_n \to 0$ and $nb_n^2 \to \infty$. (a) If $f$ and $|k|$ are bounded and have finite fourth moments, then*

$$\|\hat{\phi} - \phi\|_2^2 = O_p(b_n^{-1/2})\|\hat{f} - f\|_2.$$

*(b) If $f$ and $k$ have finite $V_{**}$-norms, then*

$$\|\hat{\phi} - \phi\|_V^2 = O_p(1)\|\hat{f} - f\|_V.$$

16

*Proof.* Without loss of generality we may assume that $b_n < 1$. An application of the Cauchy–Schwarz inequality yields

$$\|\hat{\phi} - \phi\|_2^4 = \left( \int x^2 (\hat{f}(x) - f(x))^2 \, dx \right)^2$$

$$\leq \int (1 + |x|)^4 (\hat{f}(x) - f(x))^2 \, dx \int (\hat{f}(x) - f(x))^2 \, dx$$

$$\leq \|\hat{f} - f\|_\infty (\|\hat{f}\|_{V_4} + \|f\|_{V_4}) \|\hat{f} - f\|_2^2.$$

By the properties of $f$ and $k$ we have $\|\hat{f} - f\|_\infty \leq \|k_{b_n}\|_\infty + \|f\|_\infty = O(b_n^{-1})$ and

$$\|\hat{f}\|_{V_4} \leq \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} V_4(\hat{\varepsilon}_j) \|k\|_{V_4}.$$

Thus part (a) is now immediate in view of (3.6).

In view of the properties of $f$ and $k$, part (b) follows from the inequalities

$$\|\hat{\phi} - \phi\|_V^2 \leq \|\hat{f} - f\|_{V_*}^2 \leq \|\hat{f} - f\|_{V_{**}} \|\hat{f} - f\|_V$$

and

$$\|\hat{f}\|_{V_{**}} \leq \frac{1}{n - p_n} \sum_{j=p_n+1}^{n} V_{**}(\hat{\varepsilon}_j) \|k\|_{V_{**}}$$

and (3.6). $\qquad\square$

Now let

$$Z_j = \left( \varepsilon_{j-1}, \dots, \varepsilon_{j-l_n} \right)^\top$$

and

$$\mathbf{1}_{l_n} = (1, \dots, 1)^\top \in \mathbb{R}^{l_n}.$$

We write $|v|$ for the euclidean norm of a vector $v$. We write $|A|_*$ for the operator norm of a matrix $A$ defined by

$$|A|_* = \sup_{|v|=1} |Av|.$$

For a symmetric $A$ one has $|A|_* = \sup_{|v|=1} |v^\top A v|$.

**Lemma 3.4.** *Suppose (A) holds. Then*

$$\mu = \sigma^{-2} \bar{\varepsilon}_n \mathbf{1}_{l_n} + o_p(n^{-1/2}).$$

17

*Proof.* For notational convenience we abbreviate the average operator

$$\frac{1}{n - p_n - l_n} \sum_{j=p_n+l_n+1}^{n} \quad \text{by} \quad \overline{\sum}_j.$$

Now we set

$$\hat{M}_n = \max_{p_n+l_n < j \le n} |\hat{Z}_j|, \quad M_n = \max_{p_n+l_n < j \le n} |Z_j|,$$

$$\hat{A}_n = \overline{\sum}_j \hat{Z}_j, \quad A_n = \overline{\sum}_j Z_j, \quad \hat{S}_n = \overline{\sum}_j \hat{Z}_j \hat{Z}_j^\top \quad \text{and} \quad S_n = \overline{\sum}_j Z_j Z_j^\top.$$

We derive from (3.1) to (3.3)

$$\overline{\sum}_j |\hat{Z}_j - Z_j|^2 = O_p(l_n n^{-2/\beta}), \tag{3.9}$$

$$\max_{p_n+l_n < j \le n} |\hat{Z}_j - Z_j| = o_p(l_n^{1/2}), \tag{3.10}$$

$$\hat{M}_n \le M_n + \max_{p_n+l_n < j \le n} |\hat{Z}_j - Z_j| = o_p(l_n^{1/2} n^{1/\beta}) + o_p(l_n^{1/2}), \tag{3.11}$$

$$\hat{A}_n = A_n + O_p(l_n^{1/2} n^{-1/2-\tau}) = O_p(l_n^{1/2} n^{-1/2}). \tag{3.12}$$

Let us now show that

$$|\hat{S}_n - \sigma^2 I_{l_n}|_* = O_p(l_n^{1/2} n^{-1/\beta} + l_n^3 n^{-\delta}), \tag{3.13}$$

where $I_{l_n}$ is the $l_n \times l_n$ identity matrix and $\delta = \min\{1/2, 1 - 2/\beta\}$. Since the matrices $\hat{S}_n$, $S_n$ and $I_{l_n}$ are symmetric, this follows if we show

$$|\hat{S}_n - S_n|_* = \sup_{|v|=1} \left| \overline{\sum}_j \left[ (v^\top \hat{Z}_j)^2 - (v^\top Z_j)^2 \right] \right| = O_p(l_n^{1/2} n^{-1/\beta}) \tag{3.14}$$

and

$$|S_n - \sigma^2 I_{l_n}|_* = \sup_{|v|=1} \left| \overline{\sum}_j (v^\top Z_j)^2 - \sigma^2 \right| = O_p(l_n^3 n^{-\delta}). \tag{3.15}$$

Since $(v^\top \hat{Z}_j)^2 = (v^\top Z_j)^2 + 2 v^\top Z_j v^\top (\hat{Z}_j - Z_j) + (v^\top (\hat{Z}_j - Z_j))^2$, (3.14) follows from (3.15), (3.9) and an application of the Cauchy–Schwarz inequality. Now set

$$c_n(i) = \overline{\sum}_j \varepsilon_j \varepsilon_{j-i}, \qquad i = 0, 1, 2, \ldots$$

For $1 \le i_1, i_2 \le l_n$, we have

$$\left| \overline{\sum}_j \varepsilon_{j-i_1} \varepsilon_{j-i_2} - c_n(|i_1 - i_2|) \right| \le \max_{p_n < j \le n} |\varepsilon_j|^2 \frac{l_n}{n - p_n - l_n}$$

18

and
$$\Big(E\Big[\sum_{i=1}^{l_n}|c_n(i)|\Big]\Big)^2 \le l_n \sum_{i=1}^{l_n} E[c_n^2(i)] = \frac{l_n^2\sigma^4}{n-p_n-l_n} = O(l_n^2 n^{-1}).$$

Thus for a unit vector $v$ we have

$$\overline{\sum}_j (v^\top Z_j)^2 - \sigma^2 = c_n(0) - \sigma^2 + R_n$$

where

$$|R_n| \le \max_{p_n < j \le n} |\varepsilon_j|^2 \frac{l_n^3}{n-p_n-l_n} + l_n^2 \sum_{a=1}^{l_n} |c_n(i)|$$

and hence

$$R_n = O_p(l_n^3 n^{2/\beta-1} + l_n^3 n^{-1/2}) = O_p(l_n^3 n^{-\delta}).$$

For $\beta \ge 4$, we have $c_n(0) - \sigma^2 = O_p(n^{-1/2}) = O_p(n^{-\delta})$. For $\beta < 4$, we have $c_n(0) - \sigma^2 = O_p(n^{-1+2/\beta}) = O_p(n^{-\delta})$. To see this write $c_n(0) - \sigma^2 = T_1 + T_2$, where

$$T_1 = \overline{\sum}_j \Big(\varepsilon_j^2 I[|\varepsilon_j| \le n^{1/\beta}] - E[\varepsilon_j^2 I[|\varepsilon_j| \le n^{1/\beta}]]\Big),$$
$$T_2 = \overline{\sum}_j \Big(\varepsilon_j^2 I[|\varepsilon_j| > n^{1/\beta}] - E[\varepsilon_j^2 I[|\varepsilon_j| > n^{1/\beta}]]\Big),$$

and verify that

$$(n-p_n-l_n)E[T_1^2] = E[\varepsilon^4 I[\varepsilon \le n^{1/\beta}]] \le n^{(4-\beta)/\beta} E[|\varepsilon|^\beta]$$

and

$$E[|T_2|] \le 2E[\varepsilon^2 I[|\varepsilon| > n^{1/\beta}]] \le 2n^{-(\beta-2)/\beta} E[|\varepsilon|^\beta].$$

Using the statements (3.11)–(3.13) one can first show that the probability of the event that the origin is an interior point of the convex hull of the random vectors $\hat{Z}_{p_n+l_n+1}, \ldots, \hat{Z}_n$ tends to one (we omit the elaborate argument) and then use this and arguments of Owen (2001, pages 220–221) to conclude

$$|\mu - \hat{S}_n^{-1}\hat{A}_n| = O_p(\hat{M}_n|\hat{A}_n|^2).$$

Thus in view of (3.11) to (3.13) one obtains

$$\mu = \sigma^{-2}A_n + o_p(n^{-1/2}) = \sigma^{-2}\bar{\varepsilon}_n \mathbf{1}_{l_n} + o_p(n^{-1/2}),$$

which is the desired result. □

Now set

$$\hat{\Gamma}(x) = \frac{1}{n-p_n-l_n} \sum_{j=p_n+l_n+1}^{n} \hat{Z}_j k_{b_n}(x - \hat{Y}_j), \quad x \in \mathbb{R}. \tag{3.16}$$

19

**Lemma 3.5.** *Suppose (A) holds and the bandwidth satisfies $b_n \to 0$ and $nb_n^2 \to \infty$. (a) If $\beta = 4$ and $\int k^2(u)\,du$ is finite, then*

$$\|\hat{g}_w - \hat{g} + \mu^\top \hat{\Gamma}\|_2 = o_p(n^{-1/2}).$$

*(b) If $g$ has finite $V^2$-norm and $k$ has finite $V$-norm, then*

$$\|\hat{g}_w - \hat{g} + \mu^\top \hat{\Gamma}\|_V = o_p(n^{-1/2}).$$

*Proof.* It follows from Lemma 3.4 and (3.2) that

$$\max_{p_n + l_n < j \le n} |\mu^\top \hat{Z}_j| = o_p(l_n n^{-1/2 + 1/\beta})$$

and thus

$$\Delta_n = \max_{p_n + l_n < j \le n} \frac{|\mu| |\mu^\top \hat{Z}_j|}{|1 + \mu^\top \hat{Z}_j|} = o_p(l_n^{3/2} n^{-1 + 1/\beta}).$$

Writing

$$\frac{1}{1 + \mu^\top \hat{Z}_j} = 1 - \mu^\top \hat{Z}_j + \frac{(\mu^\top \hat{Z}_j)^2}{1 + \mu^\top \hat{Z}_j},$$

we see that

$$\hat{g}_w(x) = \tilde{g}(x) - \mu^\top \hat{\Gamma}(x) + D(x)$$

with

$$\tilde{g}(x) = \overline{\sum}_j k_{b_n}(x - \hat{Y}_j),$$

$$D(x) = \overline{\sum}_j \frac{(\mu^\top \hat{Z}_j)^2}{1 + \mu^\top \hat{Z}_j} k_{b_n}(x - \hat{Y}_j).$$

Note that

$$\hat{g}(x) - \tilde{g}(x) = \left(1 - \frac{n - p_n}{n - p_n - l_n}\right)\hat{g}(x) + \frac{1}{n - p_n - l_n}\sum_{j = p_n + 1}^{p_n + l_n} k_{b_n}(x - \hat{Y}_j).$$

From this one derives that

$$\|\hat{g} - \tilde{g}\|_2 \le \frac{2l_n}{n - p_n - l_n}\|k_{b_n}\|_2 = \frac{2l_n}{(n - p_n - l_n)b_n^{1/2}}\|k\|_2 = o_p(n^{-1/2})$$

if $k$ has finite $L_2$-norm, and, in view of (3.7),

$$\|\hat{g} - \tilde{g}\|_V \le \frac{l_n \|k_{b_n}\|_V}{n - p_n - l_n}\left(\frac{1}{n - p_n}\sum_{j = p_n + 1}^{n} V(\hat{Y}_j) + \frac{1}{l_n}\sum_{j = p_n + 1}^{p_n + l_n} V(\hat{Y}_j)\right) = o_p(n^{-1/2})$$

20

if $k$ and $g$ have finite $V$-norms.

As in the proof of Lemma 3.2 we derive the bounds

$$b_n \|D\|_2^2 \leq \Delta_n^2 \overline{\sum}_j |\hat{Z}_j|^2 \int k^2(u)\,du,$$

$$\|D\|_V \leq \Delta_n \overline{\sum}_j |\hat{Z}_j| V(\hat{Y}_j) \|k\|_V.$$

It follows from (3.9) that

$$\overline{\sum}_j |\hat{Z}_j|^2 \leq 2\overline{\sum}_j |Z_j|^2 + 2\overline{\sum}_j |\hat{Z}_j - Z_j|^2 = O_p(l_n).$$

If $g$ has a finite $V^2$-norm, we obtain from (3.7) applied with $V^2$ instead of $V$ that

$$\overline{\sum}_j V^2(\hat{Y}_j) = O_p(1).$$

It is now easy to see that $\|\hat{g} - \tilde{g}\|_2 + \|D\|_2 = o_p(n^{-1/2})$ if $\beta = 4$ and $k$ is square-integrable, and that $\|\hat{g} - \tilde{g}\|_V + \|D\|_V = o_p(n^{-1/2})$ if $g$ has a finite $V^2$-norm and $k$ has a finite $V$-norm. This completes the proof. □

**Lemma 3.6.** *Suppose (A) holds and the bandwidth satisfies $b_n \to 0$ and $nb_n^2 \to \infty$. (a) Suppose (S) and (F) hold and $\|k\|_2 < \infty$. Then*

$$\|f * (\mu^\top \hat{\Gamma}) - \sigma^{-2}\gamma\bar{\varepsilon}_n\|_\infty = o_p(n^{-1/2}). \tag{3.17}$$

*(b) Suppose (S) and (FV) hold and $k$ has finite $V_r$-norm. Then*

$$\|f * (\mu^\top \hat{\Gamma}) - \sigma^{-2}\gamma\bar{\varepsilon}_n\|_{V_r} = o_p(n^{-1/2}). \tag{3.18}$$

*Proof.* For $x \in \mathbb{R}$, we set

$$\hat{U}_n(x) = \overline{\sum}_j \hat{Z}_j f(x - \hat{Y}_j) \quad \text{and} \quad U_n(x) = \overline{\sum}_j Z_j f(x - Y_j).$$

Then we can write $f * (\mu^\top \hat{\Gamma}) = (\mu^\top \hat{U}_n) * k_{b_n}$. Note also that

$$\bar{U}_n(x) = E[U_n(x)] = E[Z_1 f(x - Y_1)] = (\gamma_1(x), \dots, \gamma_{l_n}(x))^\top,$$

where, for a positive integer $j$,

$$\gamma_j(x) = E[\varepsilon_0 f(x - Y_j)], \quad x \in \mathbb{R}.$$

Let us first prove (a). It follows from (F) that $f$ is bounded, Lipschitz with Lipschitz constant $\|f'\|_\infty$, and $L_1$-Lipschitz with Lipschitz constant $\|f'\|_1$. Since $\varepsilon_0$ and $Y_j - \varphi_j \varepsilon_0$ are independent and $\varepsilon_0$ has mean zero, we see that

$$\gamma_j(x) = E[\varepsilon_0(f(x - Y_j) - f(x - Y_j + \varphi_j \varepsilon_0))], \quad x \in \mathbb{R},$$

and obtain

$$\|\gamma_j\|_\infty \le \|f'\|_\infty |\varphi_j| \sigma^2 \quad \text{and} \quad \|\gamma_j\|_1 \le \|f'\|_1 |\varphi_j| \sigma^2.$$

Thus we have the bounds

$$\|\gamma\|_\infty \le \sum_{j=1}^\infty \|\gamma_j\|_\infty \le \sigma^2 \|f'\|_\infty \sum_{j=1}^\infty |\varphi_j| < \infty,$$

$$\|\gamma\|_1 \le \sum_{j=1}^\infty \|\gamma_j\|_1 \le \sigma^2 \|f'\|_1 \sum_{j=1}^\infty |\varphi_j| < \infty,$$

and see that $\mathbf{1}_{l_n}^\top \bar{U}_n = \gamma_1 + \cdots + \gamma_{l_n}$ converges to $\gamma$ uniformly and in $L_1$.

Of course the function $\gamma_j$ is absolutely continuous with bounded almost everywhere derivative

$$\gamma_j'(x) = E[\varepsilon_0(f'(x - Y_j) - f'(x - Y_j + \varphi_j \varepsilon_0))], \quad x \in \mathbb{R}.$$

Let $i_* = \inf\{i : \varphi_i \ne 0\}$ and $\varphi = \varphi_{i_*}$. Then we can write $Y_j = \varphi \varepsilon_{j-i_*} + Y_j^*$ with $Y_j^* = \sum_{s > i_*} \varphi_s \varepsilon_{j-s}$. For $j > i_*$, we have

$$\gamma_j'(x) = E[\varepsilon_0(w(x - Y_j^*) - w(x - Y_j^* + \varphi_j \varepsilon_0))]$$

with $w = f' * f_\varphi = f'_\varphi * f$ and $f_\varphi(x) = f(x/\varphi)/|\varphi|$ the density of $\varphi \varepsilon_0$. Thus $w$ has a bounded and integrable derivative $w' = f' * f'_\varphi$. Consequently, for $j > i_*$, we derive

$$\|\gamma_j'\|_\infty \le |\varphi_j| \sigma^2 \|w'\|_\infty \quad \text{and} \quad \|\gamma_j'\|_1 \le |\varphi_j| \sigma^2 \|w'\|_1.$$

This shows that $\gamma$ has a bounded and integrable derivative $\gamma' = \sum_{j=1}^\infty \gamma_j'$. The above imply that $\gamma$ belongs to $C_0$.

In view of the inequalities $\|v * k_{b_n}\|_\infty \le \|v\|_\infty \|k\|_1$ and $\|v * k_{b_n}\|_\infty \le \|v\|_2 \|k_{b_n}\|_2 = \|v\|_2 \|k\|_2 b_n^{1/2}$, the desired statement (3.17) follows if we show that

$$\|\mu^\top(\hat{U}_n - U_n)\|_\infty = o_p(n^{-1/2}), \tag{3.19}$$

$$\|\mu^\top(U_n - \bar{U}_n)\|_2 = O_p(n^{-3/4}), \tag{3.20}$$

$$\|\mu^\top \bar{U}_n - \sigma^{-2} \gamma \bar{\varepsilon}_n\|_\infty = o_p(n^{-1/2}), \tag{3.21}$$

$$\|\gamma * k_{b_n} - \gamma\|_\infty \to 0. \tag{3.22}$$

Since $\gamma$ belongs to $C_0$, we obtain (3.22). Since $|\bar{U}_n| \le |\gamma_1| + \cdots + |\gamma_{l_n}|$, we obtain $\||\bar{U}_n|\|_\infty = O(1)$. This and Lemma 3.5 yield $\|(\mu - \sigma^{-2}\bar{\varepsilon}_n \mathbf{1}_{l_n})^\top \bar{U}_n\|_\infty = o_p(n^{-1/2})$. As $\mathbf{1}_{l_n}^\top \bar{U}_n$ converges to $\gamma$ uniformly and $\bar{\varepsilon}_n = O_p(n^{-1/2})$, one obtains (3.21). It is easy to check that

$$\|\mu^\top(\hat{U}_n - U_n)\|_\infty \le |\mu|\Big(\overline{\sum}_j |\hat{Z}_j - Z_j|\|f\|_\infty + \|f'\|_\infty \overline{\sum}_j |Z_j| \overline{\sum}_j |\hat{Y}_j - Y_j|\Big).$$

In view of $\mu = o_p(l_n^{1/2}n^{-1/2})$, (3.1), (3.9), and the identity $\hat{Y}_j - Y_j = \hat{\varepsilon}_j - \varepsilon_j$, we obtain (3.19). Note that the $i$-th coordinates of $U_n(x)$ and $\bar{U}_n(x)$ are

$$U_{ni}(x) = \overline{\sum}_j \varepsilon_{j-i} f(x - Y_j) \quad \text{and} \quad \bar{U}_{ni}(x) = \gamma_i(x).$$

For $i < i_*$, we find that

$$(n - p_n - l_n) \int \mathrm{Var}(U_{n,i}(x))\,dx \le \int \sigma^2 E[f^2(x - Y_1)]\,dx = \sigma^2\|f\|_2^2.$$

For $i \ge i_*$, it follows from Lemma 25 in Schick and Wefelmeyer (2008a), applied with $p = q = 0$, $h = f$ and $d_s = \varphi_{i_*+s}$, that

$$(n - p_n - l_n) \int \mathrm{Var}(U_{ni}(x))\,dx \le C\Big(1 + i - i_* + \sum_{s=0}^\infty (1 + s)\varphi_{i_*+s}\Big),$$

where

$$C = 8\max\{2, \|f'\|_1\}\|f\|_\infty 4^4(1 + (2c_*)^3 E[|\varepsilon_0|^3])^4$$

and

$$c_* = 1 + \sum_{j=1}^\infty |\varphi_j|.$$

This shows that

$$E[\||U_n - \bar{U}_n|\|_2^2] = \int \sum_{i=1}^{l_n} \mathrm{Var}(U_{ni}(x))\,dx = O_p(l_n^2 n^{-1})$$

from which we conclude that

$$\|\mu^\top(U_n - \bar{U}_n)\|_2 \le |\mu|\||U_n - \bar{U}_n|\|_2 = O_p(l_n^{3/2}n^{-1}) = o_p(n^{-3/4}).$$

Let us now prove part (b). It follows from Lemma 14 in Schick and Wefelmeyer (2008a) that

$$E[|Z_j|^2 V_s(Y_j)] \le l_n 2^{1+s}(1 + c_*^{2+s} E[|\varepsilon_0|^{2+s}]) \tag{3.23}$$

23

for $s \geq 0$ for which $E[|\varepsilon_0|^{2+s}]$ is finite. It follows from Lemmas 4 and 7 in Schick and Wefelmeyer (2008a) that (FV) implies that $V_{r+1}f$ is bounded and that $f$ is $V_s$-Lipschitz for every $0 \leq s \leq r+1$: There is a $\Lambda_s$ such that

$$\int V_s(x)|f(x-t) - f(x)|\, dx \leq \Lambda_s V_s(t)|t|, \quad t \in \mathbb{R}.$$

Now abbreviate $V_r$ by $V$. For this $V$, one also has $V(xy) \leq V(x)V(y)$. Using this, the properties of $V$ and the substitution $u = x - Y_j + \varphi_j \varepsilon_0$, we find

$$\|\gamma_j\|_V \leq \int V(x) E[|\varepsilon_0||f(x-Y_j) - f(x-Y_j+\varphi_j\varepsilon_0)|]\, dx$$

$$\leq E\Big[|\varepsilon_0|V(Y_j - \varphi_j\varepsilon_0)\int V(u)|f(u-\varphi_j\varepsilon_0) - f(u)|\, du\Big]$$

$$\leq \Lambda_r|\varphi_j|E[V(Y_j - \varphi_j\varepsilon_0)]E[\varepsilon_0^2 V(\varphi_j\varepsilon_0)]$$

$$\leq \Lambda_r|\varphi_j|V^2(\varphi_j)E[V(Y_j)]E[V(\varepsilon_0)]E[V_{r+2}(\varepsilon_0)].$$

This shows that $\|\gamma\|_V < \infty$ and that $\mathbf{1}_{l_n}^\top \bar{U}_n = \gamma_1 + \cdots + \gamma_{l_n}$ converges to $\gamma$ in the $V$-norm. In view of the inequality $\|u * v\|_V \leq \|u\|_V \|v\|_V$, the desired (3.18) follows if we show

$$\|\mu^\top \hat{U}_n - \sigma^{-2}\gamma\bar{\varepsilon}_n\|_V = o_p(n^{-1/2}) \tag{3.24}$$

and $\|\gamma * k_{b_n} - \gamma\|_V \to 0$. The latter follows as $\gamma$ and $k$ have finite $V$-norm; see Lemma 2 in Schick and Wefelmeyer (2008a). Thus we are left to verify (3.24).

We bound $\|\mu^\top(\hat{U}_n - U_n)\|_V$ by $|\mu|(\|f\|_V T_1 + L T_2)$, where

$$T_1 = \overline{\sum}_j |\hat{Z}_j - Z_j|V(\hat{Y}_j) \quad \text{and} \quad T_2 = \overline{\sum}_j |Z_j|V(Y_j)V(\hat{\varepsilon}_j - \varepsilon_j)|\hat{\varepsilon}_j - \varepsilon_j|.$$

Since $\delta_n = \max_{p_n < j \leq n} |\hat{\varepsilon}_j - \varepsilon_j| = o_p(1)$, we have

$$T_1^2 \leq V^2(\delta_n)\overline{\sum}_j |\hat{Z}_j - Z_j|^2 \overline{\sum}_j V^2(Y_j) = O_p(l_n n^{-2/\beta})$$

in view of (3.9) and $E[V^2(Y_1)] < \infty$, and

$$T_2^2 \leq V^2(\delta_n)\overline{\sum}_j |Z_j|^2 V^2(Y_j)\overline{\sum}_j |\hat{\varepsilon}_j - \varepsilon_j|^2 = O_p(l_n n^{-2/\beta})$$

in view of (3.1) and $E[|Z_1|^2 V^2(Y_1)] = O(l_n)$. This shows that

$$\|\mu^\top(\hat{U}_n - U_n)\|_V = o_p(n^{-1/2}). \tag{3.25}$$

Now set $W = V_{2r+\alpha}$ for some $\alpha \in (1,2]$ such that $2r + 2 + \alpha \leq \xi$. It follows from the Cauchy–Schwarz inequality that

$$\|a\|_V^2 \leq \int (1+|x|)^{-\alpha}\, dx \int W(x)a^2(x)\, dx. \tag{3.26}$$

24

For $i < i_*$, we have

$$(n - p_n - l_n) \int W(x) \operatorname{Var}(U_{ni}(x)) \, dx \leq \sigma^2 \int W(x) E[f^2(x - Y_1)] \, dx$$
$$= \sigma^2 E[W(Y_1)] \|f^2\|_W.$$

For $i \geq i_*$, it follows from Lemma 25 in Schick and Wefelmeyer (2008a), applied with $p = r + \alpha - 1 \leq q = r + 1$, $h = f$ and $d_s = \varphi_{i_*+s}$, that

$$(n - p_n - l_n) \int W(x) \operatorname{Var}(U_{ni}(x)) \, dx \leq C\Big(1 + i - i_* + \sum_{s=0}^{\infty} (1 + s)\varphi_{i_*+s}\Big),$$

where

$$C = 8 \max\{\Lambda_{r+1}, 2\|f\|_{V_{r+1}}\} \|V_{r+1} f\|_\infty 4^4 (1 + (2c_*)^{2r+2+\alpha} E[\varepsilon_0|^{2r+2+\alpha}])^4.$$

This shows that

$$\int W(x) E[|U_n(x) - \bar{U}_n(x)|^2] \, dx = O_p(l_n^2 n^{-1}).$$

Using this, (3.26) and $|\mu| = O_p(l_n^{1/2} n^{-1/2})$ we derive

$$\|\mu^\top (U_n - \bar{U}_n)\|_V = o_p(n^{-1/2}). \tag{3.27}$$

Since $|\bar{U}_n| \leq \sum_{j=1}^{l_n} |\gamma_j|$ and $\sum_{j=1}^{\infty} \|\gamma_j\|_V < \infty$, we see that

$$\|(\mu - \sigma^{-2}\bar{\varepsilon}_n \mathbf{1}_{l_n})^\top \bar{U}\|_V = o_p(n^{-1/2}) \tag{3.28}$$

in view of Lemma 3.5. Finally, since $\mathbf{1}_{l_n}^\top \bar{U}_n$ converges to $\gamma$ in the $V$-norm, we obtain

$$\|\sigma^{-2}\bar{\varepsilon}_n(\mathbf{1}_{l_n}^\top \bar{U} - \gamma)\|_V = o_p(n^{-1/2}). \tag{3.29}$$

Combining the above yields the desired (3.24). $\qquad\square$

## BIBLIOGRAPHY

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

Du, J. and Schick, A. (2007). Root-$n$ consistency and functional central limit theorems for estimators of derivatives of convolutions of densities. *Internat. J. Statist. Management Syst.*, **2**, 67–87.

Frees, E. W. (1994). Estimating densities of functions of observations. *J. Amer. Statist. Assoc.*, **89**, 517–525.

Giné, E. and Mason, D. (2007a). On local $U$-statistic processes and the estimation of densities of functions of several sample variables. *Ann. Statist.*, **35**, 1105–1145.

Giné, E. and Mason, D. (2007b). Laws of the iterated logarithm for the local U-statistic process. *J. Theoret. Probab.*, **20**, 457–485.

Haberman, S. J. (1984). Adjustment by minimum discriminant information. *Ann. Statist.*, **12**, 971–988. Erratum: **14** (1986), 358.

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

Kitamura, Y. (1997). Empirical likelihood methods with weakly dependent processes. *Ann. Statist.*, **25**, 2084–2102.

Levit, B. Y. (1975). Conditional estimation of linear functionals. *Problems Inform. Transmission*, **11**, 39–54.

Müller, U. U., Schick, A. and Wefelmeyer, W. (2005). Weighted residual-based density estimators for nonlinear autoregressive models. *Statist. Sinica*, **15**, 177–195.

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.

Saavedra, A. and Cao, R. (1999). Rate of convergence of a convolution-type estimator of the marginal density of an MA(1) process. *Stochastic Process. Appl.*, **80**, 129–155.

Saavedra, A. and Cao, R. (2000). On the estimation of the marginal density of a moving average process. *Canad. J. Statist.*, **28**, 799–815.

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

Schick, A. and Wefelmeyer, W. (2004b). Root $n$ consistent density estimators for sums of independent random variables. *J. Nonparametr. Statist.*, **16**, 925–935.

Schick, A. and Wefelmeyer, W. (2004c). Functional convergence and optimality of plug-in

estimators for stationary densities of moving average processes. *Bernoulli*, **10**, 889–917.

Schick, A. and Wefelmeyer, W. (2007a). Root-$n$ consistent density estimators of convolutions in weighted $L_1$-norms. *J. Statist. Plann. Inference*, **37**, 1765–1774.

Schick, A. and Wefelmeyer, W. (2007b). Uniformly root-$n$ consistent density estimators for weakly dependent invertible linear processes. *Ann. Statist.*, **35**, 815–843.

Schick, A. and Wefelmeyer, W. (2008a). Root-$n$ consistency in weighted $L_1$-spaces for density estimators of invertible linear processes. *Stat. Inference Stoch. Process.*, **11**, 281–310.

Schick, A. and Wefelmeyer, W. (2008b). Convergence rates of density estimators for sums of powers of observations. To appear in: *Metrika*.

Schick, A. and Wefelmeyer, W. (2008c). Non-standard behavior of density estimators for sums of squared observations. Technical Report, Department of Mathematical Sciences, Binghamton University.

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