PLUG-IN ESTIMATORS FOR HIGHER-ORDER TRANSITION DENSITIES IN AUTOREGRESSION [*]

Anton Schick[1] and Wolfgang Wefelmeyer[2]

Abstract. In this paper we obtain root-n consistency and functional central limit theorems in weighted L_1-spaces for plug-in estimators of the two-step transition density in the classical stationary linear autoregressive model of order one, assuming essentially only that the innovation density has bounded variation. We also show that plugging in a properly weighted residual-based kernel estimator for the unknown innovation density improves on plugging in an unweighted residual-based kernel estimator. These weights are chosen to exploit the fact that the innovations have mean zero. If an efficient estimator for the autoregression parameter is used, then the weighted plug-in estimator for the two-step transition density is efficient. Our approach generalizes to invertible linear processes.

1991 Mathematics Subject Classification. 62M05, 62M10.

...

Introduction

Suppose we observe $X_0, \ldots, X_n$ coming from a stationary linear autoregressive process of order one:

$$X_t = \rho X_{t-1} + \varepsilon_t, \quad t \in \mathbb{Z},$$

with $\rho \in (-1,1)$ and independent and identically distributed innovations $\{\varepsilon_t, t \in \mathbb{Z}\}$ with mean zero, finite variance σ^2 and a density f. We are interested in estimating the two-step transition density q of X_{t+2} given $X_t = x$ for some fixed x. We could estimate $q(z)$ nonparametrically by writing it as $h(x,z)/g(z)$ with g and h the stationary densities of X_t and (X_t, X_{t+2}), respectively, and plugging in kernel estimators for h and g; see Roussas (1969, 1988), Nguyen (1984) and Athreya and Atuncar (1998). The rates would be those for estimating a two-dimensional density. We could

use the Markov structure and write $q(z)$ as $\int s(x,y)s(y,z)\,dy$ with s the one-step transition density, and plugging in nonparametric estimators for s. This would give better rates, because the problem reduces to that of estimating one-dimensional functions, namely conditional expectations. This is work in progress. Functional central limit theorems cannot be expected (except locally, on intervals shrinking as the bandwidth) since for different values of z the estimators eventually use essentially different parts of the observations.

Using the autoregressive structure, we can express the two-step transition density given x as

$$q(z) = \int f(z - \rho y - \rho^2 x) f(y)\,dy, \quad z \in \mathbb{R}.$$

This is a smooth functional of f and ρ, and we shall show that plugging in appropriate estimators for f and ρ leads to a root-n consistent estimator of q. We must of course exclude the degenerate case $\rho = 0$.

First we discuss limitations and possible extensions of our approach. For the *one*-step transition density $f(\cdot - \rho x)$ of X_{t+1} given $X_t = x$, the rate of the estimator $\hat{f}(\cdot - \hat{\rho}x)$ equals the rate of $\hat{f}$, and we do not have root-n consistency. Our results generalize however to estimation of m-step transition densities of X_{t+m} given $X_t = x$, which can be expressed as

$$\int \cdots \int f(\cdot - \rho y_{m-1} - \cdots - \rho^{m-1} y_1 - \rho^m x) f(y_1) \ldots f(y_{m-1})\,dy_1 \ldots dy_{m-1}.$$

Our results also generalize to higher-order autoregression. To keep the paper readable, we consider only the simplest case.

Root-n consistent estimation of q is possible because q can be written as a convolution. This idea has already been used for other estimation problems in time series. Saavedra and Cao (1999) obtain pointwise root-n consistency of a plug-in estimator for the stationary density p of a *moving-average* process of order one, $X_t = \rho \varepsilon_{t-1} + \varepsilon_t$, which can be expressed as $p(y) = \int f(y - \rho x) f(x)\,dx$. Schick and Wefelmeyer (2004a) prove that versions of this estimator are asymptotically normal and efficient. Schick and Wefelmeyer (2004c) show that such estimators obey functional central limit theorems in L_1 and C_0; they also consider higher-order moving average processes. — Estimating q is similar to estimating the *stationary* density p of X_t, which can be expressed as $p(y) = \int f(y - \rho x) p(x)\,dx$. Root-$n$ consistent estimators for the stationary density of general invertible linear processes are derived Schick and Wefelmeyer (2007a, 2007c).

Similar results exist for *i.i.d.* observations $X_1, \ldots, X_n$ and functionals *not* involving unknown parameters like ρ. Frees (1994) shows that his local U-statistic estimators for densities of certain functions $q(X_1, \ldots, X_m)$ with $m \geq 2$ are pointwise root-n consistent. Saavedra and Cao (2000) consider the special case $X_1 + aX_2$ with a known. Schick and Wefelmeyer (2004b, 2007b) obtain functional central limit theorems for plug-in estimators of densities of $u_1(X_1) + \cdots + u_m(X_m)$ and $X_1 + X_2$ in L_1 and C_0. Giné and Mason (2007a, 2007b) prove functional central limit theorems and laws of the iterated logarithm in L_p, $1 \leq p \leq \infty$, for local U-statistic estimators of densities of functions $q(X_1, \ldots, X_m)$, under minimal conditions, and uniformly in the bandwidth.

To describe our plug-in estimator, let $\hat{\rho}$ denote the *least squares estimator*

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

Note that $\hat{\rho}$ has the martingale approximation

$$\hat{\rho} - \rho = \frac{\sum_{j=1}^{n} X_{j-1} \varepsilon_j}{\sum_{j=1}^{n} X_{j-1}^2} = \frac{1}{E[X_0^2]} \frac{1}{n} \sum_{j=1}^{n} X_{j-1} \varepsilon_j + o_p(n^{-1/2})$$

and $E[X_0^2] = \sigma^2/(1 - \rho^2)$. In particular, $\hat{\rho}$ is asymptotically normal with variance $1 - \rho^2$. We mimic the innovation ε_j by the residual

$$\hat{\varepsilon}_j = X_j - \hat{\rho} X_{j-1}, \quad j = 1, \dots, n.$$

We can then estimate the innovation density f by the kernel estimator based on these residuals,

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

where $k_b(y) = k(y/b)/b$ for some density k and some bandwidth $b = b_n$. Substituting these estimators for f and ρ in the expression for q yields the following plug-in estimator of q:

$$\hat{q}(z) = \int \hat{f}(z - \hat{\rho} y - \hat{\rho}^2 x) \hat{f}(y) \, dy, \quad z \in \mathbb{R}.$$

In place of the kernel estimator $\hat{f}$ we can use a *weighted* kernel estimator

$$\hat{f}_w(y) = \frac{1}{n} \sum_{j=1}^{n} w_j k_b(y - \hat{\varepsilon}_j), \quad y \in \mathbb{R}.$$

This results in the plug-in estimator

$$\hat{q}_w(z) = \int \hat{f}_w(z - \hat{\rho} y - \hat{\rho}^2 x) \hat{f}_w(y) \, dy, \quad z \in \mathbb{R}.$$

Since the innovations have mean zero, we choose weights for which the weighted empirical distribution of the residuals has mean zero. Motivated by Owen (1988, 2001), we take weights of the form

$$w_j = \frac{1}{1 + \hat{\lambda} \hat{\varepsilon}_j}, \tag{0.1}$$

where $\hat{\lambda}$ is chosen such that the weights $w_1, \dots, w_n$ are positive and $\sum_{j=1}^{n} w_j \hat{\varepsilon}_j = 0$. This is possible on the event $\{\min_{1 \le j \le n} \hat{\varepsilon}_j < 0 < \max_{1 \le j \le n} \hat{\varepsilon}_j\}$, which has probability tending to one; otherwise we set $\hat{\lambda} = 0$.

We shall show that the estimators $\hat{q}$ and $\hat{q}_w$ are root-n consistent in the L_1-norm and obey functional central limit theorems in L_1 with the weighted estimator resulting in a smaller asymptotic

variance structure. To describe these results let us define functions χ_0, χ_1 and χ_2 by

$$\chi_0(z) = \int f(z - \rho y) f(y) \, dy,$$

$$\chi_1(z) = \int f(z - \rho y) y f(y) \, dy,$$

$$\chi_2(z) = \int (z - \rho y) f(z - \rho y) f(y) \, dy, \quad z \in \mathbb{R}.$$

We shall require that f is of bounded variation. Then the first two of these functions will be shown to be absolutely continuous with integrable a.e. derivatives χ_0' and χ_1'. Now set $\chi = \chi_1 + \chi_2$ and

$$\dot{q}(z) = -2\rho x \chi_0'(z) - \chi_1'(z), \quad z \in \mathbb{R}.$$

In the following, let ε and X be distributed as ε_t and X_t. Let f_ρ denote the density of $\rho\varepsilon$, i.e. $f_\rho(y) = f(y/\rho)/|\rho|$. Define functions ψ and ψ_* by

$$\psi(y, z) = f_\rho(z - y) + f(z - \rho y) \quad \text{and} \quad \psi_*(y, z) = \psi(y, z) - \sigma^{-2} y \chi(z), \quad y, z \in \mathbb{R}.$$

These can be used to define processes

$$\Psi(z) = \frac{1}{n} \sum_{j=1}^{n} (\psi(\varepsilon_j, z) - E[\psi(\varepsilon, z)]), \quad z \in \mathbb{R},$$

$$\Psi_*(z) = \frac{1}{n} \sum_{j=1}^{n} (\psi_*(\varepsilon_j, z) - E[\psi_*(\varepsilon, z)]), \quad z \in \mathbb{R}.$$

Finally, for $t \in \mathbb{R}$, introduce the shift operator S_t which assigns to an integrable function h the shifted version $S_t h$ defined by $S_t h(z) = h(z - t)$.

Theorem 1. *Let $\rho \neq 0$. Suppose f has bounded variation and a finite moment of order greater than $16/7$. Let k be the standard normal density and $b \sim (n \log n)^{-1/4}$. Then*

$$\|\hat{q} - q - S_{\rho^2 x} \Psi - (\hat{\rho} - \rho) S_{\rho^2 x} \dot{q}\|_1 = o_p(n^{-1/2}),$$

$$\|\hat{q}_w - q - S_{\rho^2 x} \Psi_* - (\hat{\rho} - \rho) S_{\rho^2 x} \dot{q}\|_1 = o_p(n^{-1/2}).$$

Moreover, $n^{1/2}(\hat{q} - q)$ converges in distribution in L_1 to $S_{\rho^2 x}(\mathbb{G}_* + \sigma^{-1} Z_1 \chi + \sqrt{1 - \rho^2} Z_2 \dot{q})$, while $n^{1/2}(\hat{q}_w - q)$ converges in distribution in L_1 to $S_{\rho^2 x}(\mathbb{G}_* + \sqrt{1 - \rho^2} Z_2 \dot{q})$, where Z_1 and Z_2 are independent standard normal random variables independent of the centered L_1-valued Gaussian process $\mathbb{G}_*$ that has the same covariance structure as $\psi_*(\varepsilon, \cdot)$.

The rest of the paper is organized as follows. In Section 1 we describe our main result, a version of Theorem 1 for certain weighted L_1-norms. Theorem 1 then follows by taking the weight function to be constant. Section 2 gives conditions under which $\hat{q}_w$ is efficient for all finite-dimensional marginals if an efficient estimator $\hat{\rho}$ is used. Section 3 derives stochastic approximations for the residual-based density estimators $\hat{f}$ and $\hat{f}_w$ used for plug-in. The proof of the stochastic approximations (1.6) and (1.7) for $\hat{q}$ and $\hat{q}_w$ is in Section 4.

1. A VERSION OF THEOREM 1 IN WEIGHTED L_1-SPACES

Let V be a positive measurable function. The V-norm of a measurable function h is defined by

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

This is the usual L_1-norm if $V = 1$. Let L_V denote the space of all (equivalence classes of) measurable functions with finite V-norm. In other words, L_V is simply the L_1-space for the measure that has density V with respect to the Lebesgue measure. Throughout we always impose the following assumptions on the function V.

Assumption 1. The function V is continuous at 0 with $V(0) = 1$ and satisfies

$$V(s+t) \leq V(s)V(t), \quad s,t \in \mathbb{R}, \tag{1.1}$$

$$V(st) \leq V(t), \quad |s| \leq 1,\ t \in \mathbb{R}. \tag{1.2}$$

Examples of such functions are $V(y) = (1+|y|)^r$ and $V(y) = \exp(r|y|)$ for non-negative r. The class of functions satisfying Assumption 1 is closed under multiplication and under taking positive powers. Consequently, the following functions V_* and W_α share Assumption 1 with V,

$$V_*(y) = (1+|y|)V(y) \quad \text{and} \quad W_\alpha(y) = (1+|y|)^\alpha V^2(y), \quad y \in \mathbb{R},$$

with $\alpha \geq 0$.

Assumption 1 was introduced by Schick and Wefelmeyer (2007b). Property (1.1) is useful when dealing with shifts and convolutions, while (1.2) is useful when dealing with rescaling. More precisely, for functions g and h of finite V-norm we have

$$\|S_t h\|_V \leq V(t)\|h\|_V \quad \text{and} \quad \|g*h\|_V \leq \|g\|_V \|h\|_V$$

and, with $h_s(y) = h(y/s)/|s|$ for $y \in \mathbb{R}$,

$$\|h_s\|_V \leq \|h\|_V, \quad 0 < |s| \leq 1.$$

Furthermore, for each $\alpha > 1$ we have the inequality

$$\|h\|_V^2 \leq C_\alpha \|h^2\|_{W_\alpha} \tag{1.3}$$

with $C_\alpha = \int (1+|y|)^{-\alpha}\,dy$. See Schick and Wefelmeyer (2007b) for details.

To state our result for the space L_V we need to strengthen the assumption that f has bounded variation. Let h be an integrable function of bounded variation. Then there are finite measures μ_1 and μ_2 with equal mass $\mu_1(\mathbb{R}) = \mu_2(\mathbb{R})$ such that

$$h(x) = \mu_1((-\infty, x]) - \mu_2((-\infty, x])$$

for all but countably many x. This motivates the following definition. We say that an integrable function h has *bounded V-variation* if h has bounded variation and the measures μ_1 and μ_2 can be chosen such that $\int V \, d\mu_1$ and $\int V \, d\mu_2$ are finite.

In this section we no longer insist that $\hat{\rho}$ is the least squares estimator. Instead we consider more generally an estimator $\hat{\rho}$ that satisfies a martingale approximation

$$\hat{\rho} - \rho = \frac{1}{n} \sum_{j=1}^{n} X_{j-1} \phi(\varepsilon_j) + o_p(n^{-1/2}) \tag{1.4}$$

for some measurable function ϕ such that $E[\phi(\varepsilon)] = 0$ and $E[\phi^2(\varepsilon)]$ is finite. The least squares estimator satisfies (1.4) with $\phi(y) = y/E[X^2] = (1-\rho^2)y/\sigma^2$, $y \in \mathbb{R}$.

Theorem 2. *Let $\rho \neq 0$. Suppose the density k has mean zero and is four times continuously differentiable with bounded derivatives and k and its four derivatives have finite V_*^2-norms. Let $b \sim (n \log n)^{-1/4}$. Suppose f has bounded V-variation and*

$$\int \left((1+|y|)^\alpha V^2(y) + y^2 V(y) + |y|^\xi \right) f(y) \, dy < \infty \tag{1.5}$$

for some $\alpha > 1$ and some $\xi > 16/7$. Let (1.4) hold and let γ denote the standard deviation of $X_0\phi(\varepsilon_1)$. Then

$$\|\hat{q} - q - S_{\rho^2 x}\Psi - (\hat{\rho} - \rho)S_{\rho^2 x}\dot{q}\|_V = o_p(n^{-1/2}), \tag{1.6}$$

$$\|\hat{q}_w - q - S_{\rho^2 x}\Psi_* - (\hat{\rho} - \rho)S_{\rho^2 x}\dot{q}\|_V = o_p(n^{-1/2}). \tag{1.7}$$

Moreover, $n^{1/2}(\hat{q}-q)$ converges in distribution in L_V to $S_{\rho^2 x}(\mathbb{G}_* + \sigma^{-1}Z_1\chi + \gamma Z_2\dot{q})$, while $n^{1/2}(\hat{q}_w - q)$ converges in distribution in L_V to $S_{\rho^2 x}(\mathbb{G}_* + \gamma Z_2\dot{q})$, where Z_1 and Z_2 are independent standard normal random variables independent of the centered L_V-valued Gaussian process $\mathbb{G}_*$ that has the same covariance structure as $\psi_*(\varepsilon, \cdot)$.

We obtain Theorem 1 by taking $V = 1$, k the standard normal density, and $\hat{\rho}$ the least squares estimator, for which $\gamma = \sqrt{1-\rho^2}$. If $V(x) = (1+|x|)^r$ for some non-negative r, then (1.5) amounts to a moment condition. For example, for $r \geq 1$, (1.5) is equivalent to f having a finite moment of order greater than $1 + 2r$.

We shall defer the proof of the expansions (1.6) and (1.7) to Section 4. A proof of the weak convergence of $n^{1/2}(\hat{q} - q)$ and $n^{1/2}(\hat{q}_w - q)$ is given next. For this let Z_1, Z_2 and $\mathbb{G}_*$ be as in Theorem 2. In view of the expansions (1.6) and (1.7) and the continuity of shifts and addition in L_V, it suffices to prove that $\mathbb{A}_n = n^{1/2}(\Psi_*, \Psi - \Psi_*, (\hat{\rho} - \rho)\dot{q})$ converges in distribution in L_V^3 to $\mathbb{A} = (\mathbb{G}_*, \sigma^{-1}Z_1\chi, \gamma Z_2\dot{q})$. For this it is enough to show that the sequence $\mathbb{A}_n$ is tight in L_V^3 and that $M_h(\mathbb{A}_n)$ converges to $M_h(\mathbb{A})$ for every triplet $h = (h_1, h_2, h_3)$ of bounded measurable functions, where M_h is the operator from L_V^3 to $\mathbb{R}$ defined by

$$M_h(g) = \int \left(h_1(z)g_1(z) + h_2(z)g_2(z) + h_3(z)g_3(z) \right) dz, \quad g = (g_1, g_2, g_3) \in L_V^3.$$

Here we made use of the fact that the family M_h is a probability determining class on L_V^3. In view of (1.4) we obtain the expansion

$$M_h(\mathbb{A}_n) = n^{-1/2} \sum_{j=1}^{n} \left(a_1(\varepsilon_j) - E[a_1(\varepsilon)] - \frac{a_2 \varepsilon_j}{\sigma^2} + a_3 X_{j-1} \phi(\varepsilon_j) \right) + o_p(n^{-1/2})$$

with $a_1(y) = \int h_1(z) \psi_*(y, z) \, dz$, $a_2 = \int h_2(z) \chi(z) \, dz$ and $a_3 = \int h_3(z) \dot{q}(z) \, dz$. Note also that $E[X] = 0$ and $E[a_1(\varepsilon)\varepsilon] = 0$, the latter in view of (1.8). Thus the martingale central limit theorem yields that $M_h(\mathbb{A}_n)$ is asymptotically normal with mean zero and variance

$$\mathrm{Var}(a_1(\varepsilon)) + \frac{a_2^2}{\sigma^2} + a_3^2 \gamma^2.$$

This is also the variance of the centered normal random variable $M_h(\mathbb{A})$. Thus we have shown that $M_h(\mathbb{A}_n)$ converges to $M_h(\mathbb{A})$ for each triplet $h = (h_1, h_2, h_3)$ of bounded measurable functions. Hence we are left to establish tightness of $\mathbb{A}_n$. For this we recall the central limit theorem for the space L_V; see e.g. Ledoux and Talagrand (1991), Theorem 10.10.

Theorem 3. *Let $Y, Y_1, Y_2, \ldots$ be independent and identically distributed L_V-valued random variables with mean zero. Then $n^{-1/2} \sum_{j=1}^{n} Y_j$ converges in distribution in the space L_V to a centered Gaussian process (with the same covariance structure as Y) if and only if*

$$\lim_{t \to \infty} t^2 P(\|Y\|_V > t) = 0 \quad and \quad \int V(z) (E[Y^2(z)])^{1/2} \, dz < \infty.$$

A sufficient condition for this central limit theorem is that $E[Y^2]$ has finite W_α-norm for some $\alpha > 1$; see Schick and Wefelmeyer (2007b).

Abbreviate W_α by W for the α of Theorem 2. For each $z \in \mathbb{R}$, we calculate

$$E[\psi(\varepsilon, z)\varepsilon] = \int f_\rho(z - y) y f(y) \, dy + \chi_1(z) = \int (z - y) f(z - y) f_\rho(y) \, dy + \chi_1(z) = \chi(z) \qquad (1.8)$$

and thus find

$$E[\psi_*^2(\varepsilon, z)] = E[\psi^2(\varepsilon, z)] - \sigma^{-2} \chi^2(z) \le E[\psi^2(\varepsilon, z)] \le 2 f_\rho^2 * f(z) + 2 f^2 * f_\rho(z).$$

Since f has finite W-norm and is bounded under the assumptions of Theorem 2, we obtain that

$$\|E[\psi_*^2(\varepsilon, \cdot)]\|_W \le \|E[\psi^2(\varepsilon, \cdot)]\|_W \le 2\|f_\rho^2 * f\|_W + 2\|f^2 * f_\rho\|_W$$
$$\le 2\|f_\rho^2\|_W \|f\|_W + 2\|f^2\|_W \|f_\rho\|_W \le 4D\|f\|_W^2$$

for some constant D. Thus $n^{1/2}\Psi$ and $n^{1/2}\Psi_*$ obey a central limit theorem in L_V. Since $n^{1/2}(\hat{\rho} - \rho)$ is tight, so is $n^{1/2}(\hat{\rho} - \rho)\dot{q}$ in view of the continuity of the map $(s, h) \mapsto sh$ from $\mathbb{R} \times L_V$ into L_V. This shows that the sequence $\mathbb{A}_n$ is tight in L_V^3.

We end this section by stating some expansions in L_V. We say a function h is L_V-*Lipschitz* if there is a constant C such that

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

It was shown in Schick and Wefelmeyer (2007b), Lemma 8, that functions of bounded V-variation are L_V-Lipschitz.

Let $\mathcal{A}$ denote the set of functions g with finite V-norm that are absolutely continuous and whose a.e. derivatives g' have finite V-norms. Let $\mathcal{A}_L = \{g \in \mathcal{A} : g' \text{ is } L_V\text{-Lipschitz}\}$ and $\mathcal{A}_2 = \{g \in \mathcal{A} : g' \in \mathcal{A}\}$.

Lemma 1. *Let K be a measurable function with finite V_*-norm. Then $\nu_i = \int u^i K(u)\, du$ is finite for $i = 0, 1$, and the following are true.*

  (1) *If g is L_V-Lipschitz, then $\|g * K_b - \nu_0 g\|_V = O(b)$.*
  (2) *If g belongs to $\mathcal{A}$, then $\|g * K_b - \nu_0 g + b\nu_1 g'\|_V = o(b)$.*
  (3) *If g belongs to $\mathcal{A}_L$ and $\int V(u)u^2|K(u)|\, du$ is finite, then $\|g * K_b - \nu_0 g + b\nu_1 g'\|_V = O(b^2)$.*

Proof. The first assertion follows from Lemma 7(1) of Schick and Wefelmeyer (2007b) with $\gamma = 1$; the last two assertions follow from their Lemma 6. $\qquad\square$

2. Efficiency

In this section we recall a characterization of efficient estimators in autoregressive models and use it to prove that $\hat{q}_w$ is efficient if an efficient estimator $\hat{\rho}$ is used. Fix $\rho \in (-1, 1)$ and an innovation density f. Assume that f is absolutely continuous with a.e. derivative f', and that the *Fisher information for location,* $J = E[\ell^2(\varepsilon)]$, is finite, where $\ell = -f'/f$. Introduce perturbations $\rho_{nr} = \rho + n^{-1/2}r$, $r \in \mathbb{R}$, and $f_{nh} = f(1 + n^{-1/2}h)$ with h in the space H of bounded measurable functions such that $E[h(\varepsilon)] = E[\varepsilon h(\varepsilon)] = 0$. Let P_{n+1} and $P_{n+1,rh}$ denote the joint stationary law of $(X_0, \ldots, X_n)$ under (ρ, f) and (ρ_{nr}, f_{nh}), respectively. Then we have *local asymptotic normality* under P_{n+1},

$$\log \frac{dP_{n+1,rh}}{dP_{n+1}} = n^{-1/2} \sum_{j=1}^{n} \big(rX_{j-1}\ell(\varepsilon_j) + h(\varepsilon_j)\big) - \frac{1}{2}\big(r^2 E[X^2]J + E[h^2(\varepsilon)]\big) + o_p(1);$$

see e.g. Koul and Schick (1997). Let $\bar{H}$ denote the closure of H in $L_2(f)$. The squared norm on the right-hand side defines an inner product on $\mathbb{R} \times \bar{H}$. A real-valued functional κ of (ρ, f) is called *differentiable* at (ρ, f) with *gradient* $(r_\kappa, h_\kappa) \in \mathbb{R} \times \bar{H}$ if

$$n^{1/2}(\kappa(\rho_{nr}, f_{nh}) - \kappa(\rho, f)) \to r_\kappa r E[X^2]J + E[h_\kappa(\varepsilon)h(\varepsilon)], \quad (r, h) \in \mathbb{R} \times H.$$

An estimator $\hat{\kappa}$ of κ is called *regular* at (ρ, f) with *limit* L if L is a random variable such that

$$n^{1/2}(\hat{\kappa} - \kappa(\rho_{nr}, f_{nh})) \Rightarrow L \text{ under } P_{n+1,rh}, \quad (r, h) \in \mathbb{R} \times H.$$

The convolution theorem of Hájek and LeCam says that L is distributed as the convolution of some random variable with a normal random variable N that has mean 0 and variance $r_\kappa^2 E[X^2]J + E[h_\kappa^2(\varepsilon)]$. This justifies calling $\hat{\kappa}$ *efficient* if L is distributed as N.

An estimator $\hat{\kappa}$ of κ is called *asymptotically linear* at (ρ, f) with *influence function* g if $g \in L_2(P_2)$ with $E(g(X_0, X_1)|X_0) = 0$ and

$$n^{1/2}(\hat{\kappa} - \kappa(\rho, f)) = n^{-1/2} \sum_{j=1}^{n} g(X_{j-1}, X_j) + o_p(1).$$

It follows from a version of the convolution theorem that an estimator $\hat{\kappa}$ is regular and efficient if and only if it is asymptotically linear with influence function $g(X_0, X_1) = r_\kappa X_0 \ell(\varepsilon_1) + h_\kappa(\varepsilon_1)$. We refer to Bickel et al. (1998) for these results.

We read off immediately the well-known result that an estimator for ρ is regular and efficient if (and only if) it is asymptotically linear with influence function $r_\rho X_0 \ell(\varepsilon_1)$, where $1/r_\rho = E[X^2]J$, i.e. expansion (1.4) holds with $\phi(y) = r_\rho \ell(y)$. Efficient estimators for parameters of time series are constructed in Kreiss (1987a), (1987b), Jeganathan (1995), Drost et al. (1997) and Koul and Schick (1997). If we use an efficient estimator $\hat{\rho}$ for $\hat{q}_w$, then, by Theorem 2, $\hat{q}_w(z)$ has influence function $r_z X_0 \ell(\varepsilon_1) + S_{\rho^2 x}(\psi_*(\varepsilon_1, z) - E[\psi_*(\varepsilon, z)])$ with $r_z = r_\rho S_{\rho^2 x} \dot{q}(z)$. To prove that $\hat{q}_w(z)$ is efficient, we need only check that $(r_z, S_{\rho^2 x}(\psi_*(\varepsilon_1, z) - E[\psi_*(\varepsilon, z)]))$ is the gradient of $\kappa(\rho, f) = q(z)$. Note first that for $(r, h) \in \mathbb{R} \times H$ we can write

$$n^{1/2}(\kappa(\rho_{nr}, f_{nh}) - \kappa(\rho, f))$$
$$= n^{1/2}\left(\int f_{nh}(z - \rho_{nr} y - \rho_{nr}^2 x) f_{nh}(y)\, dy - \int f(z - \rho y - \rho^2 x) f(y)\, dy \right).$$

By Taylor expansion this converges to

$$-r 2\rho x \int f(z - \rho y - \rho^2 x) \ell(z - \rho y - \rho^2 x) f(y)\, dy$$
$$-r \int f(z - \rho y - \rho^2 x) y \ell(z - \rho y - \rho^2 x) f(y)\, dy$$
$$+ \int f(z - \rho y - \rho^2 x) h(z - \rho y - \rho^2 x) f(y)\, dy$$
$$+ \int f(z - \rho y - \rho^2 x) h(y) f(y)\, dy$$

which can be rewritten as

$$-r\left(2\rho x S_{\rho^2 x} \chi_0'(z) + S_{\rho^2 x} \chi_1'(z) \right)$$
$$+ S_{\rho^2 x} \int f_\rho(z - y) h(y) f(y)\, dy + S_{\rho^2 x} \int f(z - \rho y) h(y) f(y)\, dy$$
$$= r S_{\rho^2 x} \dot{q}(z) + E[S_{\rho^2 x} \psi(\varepsilon, z) h(\varepsilon)]$$
$$= r_z r E[X^2] J + E[S_{\rho^2 x}(\psi_*(\varepsilon, z) - E[\psi_*(\varepsilon, z)]) h(\varepsilon)].$$

This shows that $\kappa(\rho, f) = q(z)$ has the desired gradient. In the last step we have used that $\psi_*(\varepsilon, z) - E[\psi_*(\varepsilon, z)]$ is the projection of $\psi(\varepsilon, z)$ onto $\bar{H}$.

3. The behavior of the density estimators

In this section we derive convergence rates and stochastic expansions for residual-based density estimators. Throughout the section we let K denote a measurable function with finite V-norm and set

$$\nu_K = \int K(y)\, dy \quad \text{and} \quad K_b(y) = K(y/b)/b, \quad y \in \mathbb{R}.$$

For $z \in \mathbb{R}$ let

$$\hat{H}_w(z) = \frac{1}{n} \sum_{j=1}^n w_j K_b(z - \hat{\varepsilon}_j), \qquad \hat{H}(z) = \frac{1}{n} \sum_{j=1}^n K_b(z - \hat{\varepsilon}_j),$$

$$\tilde{H}(z) = \frac{1}{n} \sum_{j=1}^n K_b(z - \varepsilon_j), \qquad \bar{H}(z) = E[\tilde{H}(z)] = K_b * f(z).$$

Now fix an α with $1 < \alpha \le 2$ and set $W = W_\alpha$. Since f has a finite second moment, so do the stationary random variables X_t, and this yields

$$\max_{1 \le j \le n} |X_{j-1}| = o_p(n^{1/2}). \tag{3.1}$$

The root-n consistency of $\hat{\rho}$ then implies that

$$\tau_n := \max_{1 \le j \le n} |\hat{\varepsilon}_j - \varepsilon_j| = |\hat{\rho} - \rho| \max_{1 \le j \le n} |X_{j-1}| = o_p(1). \tag{3.2}$$

From Lemma 1 we obtain the following result.

Lemma 2. *Let f have finite V-norm and be L_V-Lipschitz. Let K have finite V_*-norm. Then*

$$\|\bar{H} - \nu_K f\|_V = O(b).$$

Lemma 3. *Let f and K^2 have finite W-norms. Then*

$$\|\tilde{H} - \bar{H}\|_V = O_p(n^{-1/2} b^{-1/2}).$$

Proof. We calculate

$$n \operatorname{Var}(\tilde{H}(y)) \le E[K_b^2(y - \varepsilon)] = K_b^2 * f(y), \quad y \in \mathbb{R},$$

and thus obtain in view of (1.3) the bound

$$nE[\|\tilde{H} - \bar{H}\|_V^2] \le C_\alpha \|K_b^2 * f\|_W \le C_\alpha \|f\|_W \|K_b^2\|_W.$$

This yields the desired result as $\|K_b^2\|_W \le b^{-1} \|K^2\|_W$ for $b \le 1$. $\qquad\square$

Lemma 4. *Suppose that f is L_V-Lipschitz and has finite W-norm. Let K be twice continuously differentiable, let $(K')^2$ and $(K'')^2$ have finite W-norms, and let K' and K'' have finite V_*-norms. Let $nb^4 \to 0$. Then*

$$\|\hat{H} - \tilde{H}\|_V = O_p(n^{-1} b^{-2}).$$

Proof. We may assume that $b \leq 1$. In view of $V \leq V_* \leq W$, the functions f, K' and K'' have finite V-norms. Let K'_b and K''_b denote the first and second derivative of K_b. Then $K'_b(x) = K'(x/b)/b^2$ and $K''_b(x) = K''(x/b)/b^3$. Using this it is easy to see that

$$\|K'_b\|_V = O(b^{-1}), \quad \|K''_b\|_V = O(b^{-2}), \quad \|(K'_b)^2\|_W = O(b^{-3}), \quad \|(K''_b)^2\|_W = O(b^{-5}).$$

Since f is L_V-Lipschitz and K' and K'' have finite V_*-norms, and since the integrals $\int K'(u)\,du$ and $\int K''(u)\,du$ equal zero, part (i) of Lemma 1 implies that

$$\|f * K'_b\|_V = O(1) \quad \text{and} \quad \|f * K''_b\|_V = O(b^{-1}). \tag{3.3}$$

Set

$$\Gamma(y) = \frac{1}{n}\sum_{j=1}^{n} X_{j-1}K'_b(y - \varepsilon_j), \qquad \bar{\Gamma}(y) = \frac{1}{n}\sum_{j=1}^{n} X_{j-1}K'_b * f(y), \quad y \in \mathbb{R}.$$

Then we have

$$\|\bar{\Gamma}\|_V = O_p\Big(\frac{1}{n}\sum_{j=1}^{n} X_{j-1}\Big) = O_p(n^{-1/2}) = o_p(n^{-1}b^{-2}).$$

Note that $nE[(\Gamma(y) - \bar{\Gamma}(y))^2] = E[X^2](K'_b)^2 * f(y)$. Using this and (1.3) we obtain that

$$nE[\|\Gamma - \bar{\Gamma}\|_V^2] \leq C_\alpha E[X^2]\|(K'_b)^2 * f\|_W \leq C_\alpha E[X^2]\|f\|_W\|(K'_b)^2\|_W = O(b^{-3}).$$

The above shows that

$$\|(\hat{\rho} - \rho)\Gamma\|_V = O_p(n^{-1}b^{-3/2}).$$

Thus it suffices to show that

$$\|\hat{H} - \tilde{H} - (\hat{\rho} - \rho)\Gamma\|_V = O_p(n^{-1}b^{-2}).$$

A Taylor expansion shows that $\hat{H}(y) - \tilde{H}(y) - (\hat{\rho} - \rho)\Gamma(y)$ equals

$$\frac{1}{n}\sum_{j=1}^{n}(\hat{\rho} - \rho)^2 X_{j-1}^2 \int_0^1 K''_b\big(y - \varepsilon_j + u(\hat{\rho} - \rho)X_{j-1}\big)(1 - u)\,du.$$

Using $\|S_t h\|_V \leq V(t)\|h\|_V$ and $V(s + t) \leq V(s)V(t)$ we obtain

$$\|\hat{H} - \tilde{H} - (\hat{\rho} - \rho)\Gamma\|_V \leq \frac{1}{n}\sum_{j=1}^{n}(\hat{\rho} - \rho)^2 X_{j-1}^2 V(\varepsilon_j)V(\tau_n)\|K''_b\|_V.$$

This gives the desired result in view of the bounds

$$(\hat{\rho} - \rho)^2\|K''_b\|_V = O_p(n^{-1}b^{-2}), \quad V(\tau_n) = 1 + o_p(1), \quad \frac{1}{n}\sum_{j=1}^{n} X_{j-1}^2 V(\varepsilon_j) = O_p(1),$$

the latter by the finiteness of $E[X_0^2 V(\varepsilon_1)]$. $\qquad \square$

A stronger result is possible under an additional moment assumption.

Lemma 5. *Let the assumptions of Lemma 4 hold with $b \sim (n \log n)^{-1/4}$. Let f have a finite moment of order $\xi > 16/7$. Then*

$$\|\hat{H} - \tilde{H}\|_V = o_p(n^{-1/2}).$$

Proof. We may assume that $b \leq 1$ and $\xi < 4$. It follows from the moment assumption on f that the stationary random variables X_t also have finite moments of order ξ. This implies that

$$\max_{1 \leq j \leq n} |X_{j-1}| = o_p(n^{1/\xi}). \tag{3.4}$$

In view of the results in the proof of Lemma 4 it suffices to show that

$$\|\hat{H} - \tilde{H} - (\hat{\rho} - \rho)\Gamma\|_V = o_p(n^{-1/2}).$$

Set $r_{nj} = n^{-1/2}X_{j-1}$ and $\hat{\Delta} = n^{1/2}(\hat{\rho} - \rho)$. Then $\hat{\varepsilon}_j - \varepsilon_j = -(\hat{\rho} - \rho)X_{j-1} = -\hat{\Delta}r_{nj}$. This allows us to write $\hat{H}(y) - \tilde{H}(y) - (\hat{\rho} - \rho)\Gamma(y)$ as $R_{\hat{\Delta}}(y)$, where

$$R_\Delta(y) = \frac{1}{n}\sum_{j=1}^n \left(K_b(y - \varepsilon_j + \Delta r_{nj}) - K_b(y - \varepsilon_i) - \Delta r_{nj}K_b'(y - \varepsilon_j) \right).$$

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

$$\sup_{|\Delta| \leq C} \|R_\Delta\|_V = o_p(n^{-1/2}). \tag{3.5}$$

Now fix such a C. A Taylor expansion shows that

$$R_\Delta(y) = \frac{1}{n}\sum_{j=1}^n (\Delta r_{nj})^2 \int_0^1 K_b''(y - \varepsilon_j + u\Delta r_{nj})(1 - u)\, du$$

and that

$$R_{\Delta+\tilde{\Delta}}(y) - R_\Delta(y) = \frac{1}{n}\sum_{j=1}^n \tilde{\Delta}r_{nj} \int_0^1 (\Delta + v\tilde{\Delta})r_{nj} \int_0^1 K_b''\big(y - \varepsilon_j + u(\Delta + v\tilde{\Delta})r_{nj}\big)\, du\, dv.$$

Let $a = a_n$ be a sequence of positive numbers such that $C \geq a$ and $a \sim (\log n)^{-1}$. It follows from $\|S_t h\|_V \leq V(t)\|h\|_V$ and the properties of V that

$$\sup_{|\Delta| \leq C} \sup_{|\tilde{\Delta}| \leq a} \|R_{\Delta+\tilde{\Delta}} - R_\Delta\|_V \leq a(C+a)V\Big((C+a)\max_{1 \leq j \leq n}|r_{nj}|\Big)\|K_b''\|_V \frac{1}{n^2}\sum_{j=1}^n X_{j-1}^2 V(\varepsilon_j).$$

Since $\max_{1 \leq j \leq n}|r_{nj}| = o_p(1)$ and $E[X_{j-1}^2 V(\varepsilon_j)] = E[X^2]\|f\|_V$, we see that

$$\sup_{|\Delta| \leq C} \sup_{|\tilde{\Delta}| \leq a} \|R_{\Delta+\tilde{\Delta}} - R_\Delta\|_V = O_p(an^{-1}b^{-2}) = O_p((n \log n)^{-1/2}). \tag{3.6}$$

Let now $R_\Delta^*(y)$ be defined as $R_\Delta(y)$, but with r_{nj} replaced by $r_{nj}^* = r_{nj}\mathbf{1}[|X_{j-1}| \leq n^{1/\xi}]$. Then $R_\Delta^*(y)$ and $R_\Delta(y)$ can differ only on the event $\{\max_{1 \leq j \leq n}|X_{j-1}| > n^{1/\xi}\}$ which has probability

tending to zero. This shows that

$$\sup_{|\Delta| \leq C} \|R_\Delta^* - R_\Delta\|_V = o_p(n^{-1/2}). \tag{3.7}$$

Now set

$$\bar{R}_\Delta^*(y) = \frac{1}{n} \sum_{j=1}^n (\Delta r_{nj}^*)^2 \int_0^1 \int K_{b_n}''\big(y - z + u\Delta r_{nj}^*\big) f(z)\, dz\, (1 - u)\, du.$$

In view of (3.3) and (1.1) we obtain

$$\sup_{|\Delta| \leq C} \|\bar{R}_\Delta^*\|_V \leq \frac{1}{n} \sum_{j=1}^n \big(\Delta r_{nj}^*\big)^2 V\Big(C \max_{1 \leq j \leq n} |r_{nj}|\Big) \|K_b'' * f\|_V = O_p(n^{-1}b^{-1}) = o_p(n^{-1/2}). \tag{3.8}$$

Next,

$$E[\|(R_\Delta^* - \bar{R}_\Delta^*)^2\|_W] \leq \int_0^1 \int W(y) E[\Gamma_\Delta^2(y, u)]\, dy\, du,$$

with

$$\Gamma_\Delta(y, u) = \frac{1}{n} \sum_{j=1}^n (\Delta r_{nj}^*)^2 \Big(K_{b_n}''\big(y - \varepsilon_j + u\Delta r_{nj}^*\big) - \int K_{b_n}''\big(y - z + u\Delta r_{nj}^*\big) f(z)\, dz\Big)$$

a martingale. Since

$$E[\Gamma_\Delta^2(y, u)] \leq n^{-3}|\Delta|^4 E[|X_0|^4 \mathbf{1}[|X_0| \leq n^{1/\xi}](K_{b_n}''(y - \varepsilon_1 + u\Delta n^{-1/2} X_0))^2],$$

we obtain from (1.1) with W in place of V that

$$n \sup_{|\Delta| \leq C} E[\|(R_\Delta^* - \bar{R}_\Delta^*)^2\|_W] \leq W(Cn^{-1/2+1/\xi}) C^4 n^{-2+(4-\xi)/\xi} b_n^{-5} E[|X|^\xi] \|f\|_W \|(K'')^2\|_W.$$

Since $-2 + (4 - \xi)/\xi + 5/4 < 0$ in view of $\xi > 16/7$, we arrive at the rate

$$n \sup_{|\Delta| \leq C} E[\|(R_\Delta^* - \bar{R}_\Delta^*)^2\|_W] = O_p(n^{-\zeta})$$

for some $\zeta > 0$. In view of (1.3) we then have, for every finite subset D_n of the interval $(-C, C)$ with M_n elements,

$$\begin{aligned} P\Big(\max_{\Delta \in D_n} n^{1/2}\|R_\Delta^* - \bar{R}_\Delta^*\|_V > \eta\Big) &\leq \sum_{\Delta \in D_n} P(n^{1/2}\|R_\Delta^* - \bar{R}_\Delta^*\|_V > \eta) \\ &\leq \sum_{\Delta \in D_n} P(nC_\alpha \|(R_\Delta^* - \bar{R}_\Delta^*)^2\|_W > \eta^2) \\ &\leq \sum_{\Delta \in D_n} \frac{nC_\alpha}{\eta^2} E[\|(R_\Delta^* - \bar{R}_\Delta^*)^2\|_W], \quad \eta > 0. \end{aligned}$$

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

$$P\Big(\max_{\Delta \in D_n} n^{1/2}\|R_\Delta^* - \bar{R}_\Delta^*\|_V > \eta\Big) = O(M_n n^{-\zeta}). \tag{3.9}$$

Now take D_n to be such that the intervals of length a_n centered at elements of D_n cover the interval $[-C, C]$. We can choose D_n such that $M_n = O(a_n^{-1}) = O(\log n)$. The desired (3.5) follows from (3.6)–(3.9). $\qquad\square$

Lemma 6. *Let the assumptions on f and K of Lemmas 2 to 4 hold. Let $nb^4 \to 0$ and $nb^{8/3} \to \infty$. Then we have*

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

In addition, if K has finite W-norm and $\int u^2 |K(u)| \, du$ is finite, then

$$\|\hat{H} - \nu_K f\|_{V_*} = o_p(n^{-1/16}).$$

If also $\int y^2 V(y) f(y) \, dy$ is finite, then

$$\|\hat{H} - \nu_K f\|_{V_*} = o_p(n^{-1/8}).$$

Proof. The first conclusion is a consequence of Lemmas 2 to 4. Since $\alpha > 1$, we have

$$(1 + |y|) V^{3/2}(y) \le W(y), \quad y \in \mathbb{R},$$

and hence the inequality

$$\|h\|_{V_*}^2 = \int (1 + |y|) V^{3/2}(y) |h(y)| \, dy \int (1 + |y|) V^{1/2}(y) |h(y)| \, dy \le \|h\|_W \|h\|_U^{1/2} \|h\|_V^{1/2}$$

where $U(y) = (1 + |y|)^2$, $y \in \mathbb{R}$. The second conclusion now follows from the facts that $\|f\|_W$ and $\|f\|_U$ are finite and that $\|\hat{H}\|_W = O_p(1)$ and $\|\hat{H}\|_U = O_p(1)$. For example, using $\|S_t h\|_W \le W(t) \|h\|_W$ and $W(x + y) \le W(x) W(y)$, we find

$$\|\hat{H}\|_W \le \frac{1}{n} \sum_{j=1}^n \int |K_b(z - \hat{\varepsilon}_j)| W(z) \, dz \le \|K_b\|_W \frac{1}{n} \sum_{j=1}^n W(\hat{\varepsilon}_j) \le \|K\|_W W(\tau_n) \frac{1}{n} \sum_{j=1}^n W(\varepsilon_j)$$

for $b \le 1$ and thus $\|\hat{H}\|_W = O_p(1)$ in view of (3.2) and finiteness of $E[W(\varepsilon)] = \|f\|_W$.

The third conclusion is proved similarly using instead the inequality

$$\|h\|_{V_*}^2 \le \int (1 + |y|)^2 V(y) |h(y)| \, dy \int V(y) |h(y)| \, dy$$

$\qquad\square$

Let ι denote the identity map on $\mathbb{R}$.

Lemma 7. *Under all the assumptions of Lemma 6 except the very last, we have*

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

Proof. It was shown in Müller et al. (2005) that

$$\hat{\lambda} = \sigma^{-2} \frac{1}{n} \sum_{j=1}^{n} \varepsilon_j + o_p(n^{-1/2}) = O_p(n^{-1/2}), \tag{3.10}$$

$$w_* = \max_{1 \le j \le n} |w_j - 1| = o_p(1), \tag{3.11}$$

and that $w_j - 1 = -\hat{\lambda}\hat{\varepsilon}_j w_j$ for $j = 1, \ldots, n$. Thus we obtain the identity

$$\hat{H}_w(y) - \hat{H}(y) + \hat{\lambda} y f(y) = -\hat{\lambda} y (\hat{H}_w(y) - f(y)) + b\hat{\lambda} \frac{1}{n} \sum_{j=1}^{n} w_j (\iota K)_b(y - \hat{\varepsilon}_j).$$

Since ιK has finite V-norm, we obtain

$$\left\| \frac{1}{n} \sum_{j=1}^{n} w_j (\iota K)_b(\cdot - \hat{\varepsilon}_j) \right\|_V \le \frac{1}{n} \sum_{j=1}^{n} w_j V(\hat{\varepsilon}_j) \|\iota K\|_V = O_p(1).$$

Thus it suffices to show that $\|\hat{H}_w - \nu_K f\|_{V_*} = o_p(1)$. But this follows from

$$\|\hat{H}_w - \hat{H}\|_{V_*} \le w_* \frac{1}{n} \sum_{j=1}^{n} V_*(\hat{\varepsilon}_j) \|K\|_{V_*} = o_p(1)$$

and $\|\hat{H} - \nu_K f\|_{V_*} = o_p(1)$ in view of Lemma 6. $\qquad \square$

All the assumptions on K appearing in the previous lemmas are met if K is twice continuously differentiable with bounded derivatives and if K and its derivatives K' and K'' have finite V_*^2-norms.

Now let $\tilde{f}$ be the kernel estimator based on the actual observations,

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

and set $\bar{f} = f * k_b$. If we take $K = k$, then we have $\nu_K = 1$, $\hat{H}_w = \hat{f}_w$, $\hat{H} = \hat{f}$, $\tilde{H} = \tilde{f}$ and $\bar{H} = f * k_b = \bar{f}$; if we take $K = k'$, then $\nu_K = 0$, $\hat{H}_w = b\hat{f}'_w$, $\hat{H} = b\hat{f}'$, $\tilde{H} = b\tilde{f}'$ and $\bar{H} = b\bar{f}'$; and if we take $K = k''$ we have $\nu_K = 0$, $\hat{H}_w = b^2 \hat{f}''_w$, $\hat{H} = b^2 \hat{f}''$, $\tilde{H} = b^2 \tilde{f}''$ and $\bar{H} = b^2 \bar{f}''$. Thus we have the following results.

Proposition 1. *Suppose the density k is four times continuously differentiable with bounded derivatives and k and its four derivatives have finite V_*^2-norms. Let $b \sim (n \log n)^{-1/4}$. Suppose f has finite W-norm and a finite moment of order greater than $16/7$ and is L_V-Lipschitz, and $\iota^2 f$ has finite V-norm. Then*

$$\|\hat{f}_w - f\|_V = o_p(n^{-1/4}), \qquad \|\hat{f}_w - f\|_{V_*} = o_p(n^{-1/8}),$$
$$\|\hat{f}_w - \tilde{f} + \hat{\lambda}\iota f\|_V = o_p(n^{-1/2}), \qquad \|\hat{f}_w - \tilde{f}\|_V = o_p(n^{-1/2} \log n),$$
$$\|\hat{f}'_w - \bar{f}'\|_V = o_p(n^{-1/8} \log n), \qquad \|\hat{f}''_w - \bar{f}''\|_V = o_p(n^{1/8} \log n).$$

Moreover,

$$\|\hat{f}_w\|_V + \|\iota \hat{f}_w\|_V + \|\iota^2 \hat{f}_w\|_V = O_p(1) \quad \text{and} \quad \|\hat{f}''_w\|_V = o_p(n^{1/4} \log n).$$

4. PROOFS OF (1.6) AND (1.7)

For the proofs of (1.6) and (1.7) we need some smoothness results for convolutions. For r in $(-1, 1)$ and integrable functions g and h, let us denote by $B_r(g, h)$ the integrable function defined by

$$B_r(g, h)(z) = \int g(z - ry)h(y)\, dy, \quad z \in \mathbb{R}.$$

It is easy to check that if g and h have finite V-norms, so does $B_r(g, h)$, and

$$\|B_r(g, h)\|_V \leq \|g\|_V \|h\|_V.$$

Since $B_r(\bar{g}, \bar{h}) - B_r(g, h) = B_r(\bar{g} - g, \bar{h}) + B_r(g, \bar{h} - h)$, we find that

$$\|B_r(\bar{g}, \bar{h}) - B_r(g, h)\|_V \leq \|\bar{g} - g\|_V \|\bar{h}\|_V + \|g\|_V \|\bar{h} - h\|_V,$$

for $\bar{g}, g, \bar{h}, h$ with finite V-norms.

Lemma 8. *Let g and h have finite V-norms and $|r|, |\rho| < 1$ with $\rho \neq 0$. Then the following are true.*

(1) *If g has bounded V-variation, then $B_\rho(g, h)$ belongs to $\mathcal{A}$.*

(2) *If h has bounded V-variation, then $B_\rho(g, h)$ belongs to $\mathcal{A}$.*

(3) *If g and h have bounded V-variation, then $B_\rho(g, h)$ belongs to $\mathcal{A}_L$.*

(4) *If g belongs to $\mathcal{A}$ and ιh has finite V-norm, then, for $|r - \rho| \leq 1 - |\rho|$,*

$$\|B_r(g, h) - B_\rho(g, h)\|_V \leq |r - \rho| \|g'\|_V \|\iota h\|_V.$$

(5) *If g belongs to $\mathcal{A}_2$, then*

$$\|S_t B_r(g, h) - B_r(g, h) - t B_r(-g', h)\|_V \leq t^2 V(t) \|g''\|_V \|h\|_V$$

and, if also $\iota^2 h$ has a finite V-norm, then, for $|r - \rho| \leq |\rho|$,

$$\|B_r(g, h) - B_\rho(g, h) - (r - \rho) B_r(-g', \iota h)\|_V \leq (r - \rho)^2 \|g''\|_V \|\iota^2 h\|_V.$$

Proof. The proof of (1) is contained in the proof of Lemma 8 in Schick and Wefelmeyer (2007b). The proof of (2) is essentially the same. To prove (3), recall first that a function of bounded V-variation is L_V-Lipschitz. Then (3) follows from Lemma 4.6 in Schick and Wefelmeyer (2006). To prove (4), write

$$B_r(g, h)(z) - B_\rho(g, h)(z) = -(r - \rho) \int \int_0^1 g'(z - \rho y - t(r - \rho)y)\, dt\, yh(y)\, dy$$

to obtain

$$\|B_r(g, h) - B_\rho(g, h)\|_V \leq |r - \rho| \int V(z) \int \int_0^1 |g'(z - \rho y - u(r - \rho)y)|\, du\, |yh(y)|\, dy\, dz$$

$$= |r - \rho| \int_0^1 \int \int V(z + \rho y + u(r - \rho)y)|g'(z)|\, dz\, |yh(y)|\, dy\, du.$$

Then use $V(s+t) \le V(s)V(t)$. The proof of (5) is similar. $\square$

We are now ready to prove (1.7). The proof of (1.6) is similar. Set $\hat{\Delta} = (\hat{\rho}^2 - \rho^2)x$. We have the identities

$$\hat{q}_w = S_{\hat{\rho}^2 x} B_{\hat{\rho}}(\hat{f}_w, \hat{f}_w) \quad \text{and} \quad q = S_{\rho^2 x} B_\rho(f, f).$$

Since $\|S_t h\|_V \le V(t)\|h\|_V$ and $S_{t+u} = S_t S_u$, it suffices to prove

$$\|S_{\hat{\Delta}} B_{\hat{\rho}}(\hat{f}_w, \hat{f}_w) - B_\rho(f, f) - \Psi_* - (\hat{\rho} - \rho)\dot{q}\|_V = o_p(n^{-1/2}).$$

Let us set

$$\mathbb{A}_{n,1}(z) = \frac{1}{n} \sum_{j=1}^n (f_\rho(z - \varepsilon_j) - E[f_\rho(z - \varepsilon)]),$$

$$\mathbb{A}_{n,2}(z) = \frac{1}{n} \sum_{j=1}^n (f(z - \rho\varepsilon_j) - E[f(z - \rho\varepsilon)]), \quad z \in \mathbb{R}.$$

Then we have

$$\Psi_* = \mathbb{A}_{n,1} + \mathbb{A}_{n,2} - \sigma^{-2}\frac{1}{n}\sum_{j=1}^n \varepsilon_j \chi.$$

We can show that $\mathbb{A}_{n,1} * k_b = B_\rho(\tilde{f} - \bar{f}, f)$ and $\mathbb{A}_{n,2} * k_b = B_\rho(f, \tilde{f} - \bar{f})$. By the tightness of $n^{1/2}\mathbb{A}_{n,i}$, we obtain $\|n^{1/2}A_{n,i} * k_b - n^{1/2}\mathbb{A}_{n,i}\|_V = o_p(1)$ for $i = 1, 2$. In view of this and (3.10) it is enough to show that

$$\|S_{\hat{\Delta}} B_{\hat{\rho}}(\hat{f}_w, \hat{f}_w) - B_\rho(f, f) - B_\rho(\tilde{f} - \bar{f}, f) - B_\rho(f, \tilde{f} - \bar{f}) - (\hat{\rho} - \rho)\dot{q} + \hat{\lambda}\chi\|_V = o_p(n^{-1/2}).$$

We have the identities

$$\dot{q} = 2\rho x B_\rho(-f, f)' + B_\rho(-f, \iota f)' \quad \text{and} \quad \chi = B_\rho(f, \iota f) + B_\rho(\iota f, f).$$

Thus the desired result will follow if show

$$\|S_{\hat{\Delta}} B_{\hat{\rho}}(\hat{f}_w, \hat{f}_w) - B_{\hat{\rho}}(\hat{f}_w, \hat{f}_w) - (\hat{\rho} - \rho)2\rho x B_\rho(-f, f)'\|_V = o_p(n^{-1/2}), \tag{4.1}$$

$$\|B_{\hat{\rho}}(\hat{f}_w, \hat{f}_w) - B_\rho(\hat{f}_w, \hat{f}_w) - (\hat{\rho} - \rho)B_\rho(-f, \iota f)'\|_V = o_p(n^{-1/2}), \tag{4.2}$$

$$\|B_\rho(\hat{f}_w, \hat{f}_w) - B_\rho(\tilde{f}, \hat{f}_w) + \hat{\lambda} B_\rho(\iota f, f)\|_V = o_p(n^{-1/2}), \tag{4.3}$$

$$\|B_\rho(\tilde{f}, \hat{f}_w) - B_\rho(\tilde{f}, \tilde{f}) + \hat{\lambda} B_\rho(f, \iota f)\|_V = o_p(n^{-1/2}), \tag{4.4}$$

$$\|B_\rho(\tilde{f}, \tilde{f}) - B_\rho(\bar{f}, \bar{f}) - B_\rho(\tilde{f} - \bar{f}, f) - B_\rho(f, \tilde{f} - \bar{f})\|_V = o_p(n^{-1/2}), \tag{4.5}$$

$$\|B_\rho(\bar{f}, \bar{f}) - B_\rho(f, f)\|_V = O(b^2) = o_p(n^{-1/2}). \tag{4.6}$$

Since $\int k'(u)\,du = 0$ and $\int uk'(u)\,du = -1$, Lemma 1 yields $\|g * k'_b\|_V = O(1)$ if g is L_V-Lipschitz and $\|g * k'_b - g'\|_V = o(1)$ if g belongs to $\mathcal{A}$. In view of Lemma 8 and Proposition 1 we have

$$\|S_{\hat{\Delta}} B_{\hat{\rho}}(\hat{f}_w, \hat{f}_w) - B_{\hat{\rho}}(\hat{f}_w, \hat{f}_w) - \hat{\Delta} B_{\hat{\rho}}(-\hat{f}'_w, \hat{f}_w)\|_V \leq \hat{\Delta}^2 V(\hat{\Delta}) \|\hat{f}''_w\|_V \|\hat{f}_w\|_V = o_p(n^{-1/2}),$$

$$\|B_{\hat{\rho}}(\hat{f}'_w, \hat{f}_w) - B_\rho(\hat{f}'_w, \hat{f}_w)\|_V \leq |\hat{\rho} - \rho| \|\hat{f}''_w\|_V \|\iota \hat{f}_w\|_V = o_p(1),$$

$$\|B_\rho(\hat{f}'_w, \hat{f}_w) - B_\rho(\bar{f}', f)\|_V \leq \|\hat{f}'_w - \bar{f}'\|_V \|\hat{f}_w\|_V + \|\bar{f}'\|_V \|\hat{f}_w - f\|_V = o_p(1),$$

$$\|B_\rho(\bar{f}', f) - B_\rho(f, f)'\|_V = \|B_\rho(f, f) * k'_b - B_\rho(f, f)'\|_V \rightarrow 0.$$

These four bounds and the fact that $\hat{\Delta} = (\hat{\rho} - \rho)2\rho x + o_p(n^{-1/2})$ show that (4.1) holds.

Using again Lemma 8 and Proposition 1, we find

$$\|B_{\hat{\rho}}(\hat{f}_w, \hat{f}_w) - B_\rho(\hat{f}_w, \hat{f}_w) - (\hat{\rho} - \rho) B_\rho(-\hat{f}'_w, \iota \hat{f}_w)\|_V \leq (\hat{\rho} - \rho)^2 \|\hat{f}''_w\|_V \|\iota^2 \hat{f}_w\|_V = o_p(n^{-1/2}),$$

$$\|B_\rho(-\hat{f}'_w, \iota \hat{f}_w) - B_\rho(-\bar{f}', \iota f)\|_V \leq \|\hat{f}'_w - \bar{f}'\|_V \|\iota \hat{f}_w\|_V + \|\bar{f}'\|_V \|\iota(\hat{f}_w - f)\|_V = o_p(1),$$

$$\|B_\rho(-\bar{f}', \iota f) - B_\rho(-f, \iota f)'\|_V = \|B_\rho(-f, \iota f) * k'_b - B_\rho(-f, \iota f)'\|_V \rightarrow 0.$$

These three bounds show that (4.2) holds.

Since

$$\|B_\rho(\hat{f}_w, \hat{f}_w) - B_\rho(\tilde{f}, \hat{f}_w) - B_\rho(\hat{f}_w - \tilde{f}, f)\|_V \leq \|\hat{f}_w - \tilde{f}\|_V \|\hat{f}_w - f\|_V = o_p(n^{-1/2}),$$

$$\|B_\rho(\hat{f}_w - \tilde{f}, f) + \hat{\lambda} B_\rho(\iota f, f)\|_V \leq \|f\|_V \|\hat{f}_w - \tilde{f} + \hat{\lambda} \iota f\|_V = o_p(n^{-1/2}),$$

we obtain (4.3). A similar argument yields (4.4).

The inequalities

$$\|B_\rho(\tilde{f}, \tilde{f}) - B_\rho(\bar{f}, \tilde{f}) - B_\rho(\tilde{f} - \bar{f}, f)\|_V \leq \|\tilde{f} - \bar{f}\|_V \|\tilde{f} - f\|_V = o_p(n^{-1/2}),$$

$$\|B_\rho(\bar{f}, \tilde{f}) - B_\rho(\bar{f}, \bar{f}) - B_\rho(f, \tilde{f} - \bar{f})\|_V \leq \|\tilde{f} - \bar{f}\|_V \|\tilde{f} - f\|_V = o_p(n^{-1/2})$$

imply (4.5).

It is easy to check that $B_\rho(\bar{f}, \bar{f}) = B_\rho(f, f) * k_b * k_{\rho b} = B_\rho(f, f) * (k * k_\rho)_b$. Since $B_\rho(f, f)$ belongs to $\mathcal{A}_L$ by Lemma 8 and $K = k * k_\rho$ has finite V_*^2-norm, part (iii) of Lemma 1 gives (4.6).

References

[1] Athreya, K. B. and Atuncar, G. S. (1998). Kernel estimation for real-valued Markov chains. *Sankhyā Ser. A* **60**, 1–17.

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

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

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

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

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

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

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

[9] Kreiss, J.-P. (1987a). On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **1**, 112–133.

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

[11] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces. Isoperimetry and Processes.* Ergebnisse der Mathematik und ihrer Grenzgebiete (3) 23, Springer, Berlin.

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

[13] Nguyen, H. T. (1984). Recursive nonparametric estimation in stationary Markov processes. *Publ. Inst. Statist. Univ. Paris* **29**, 65–84.

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

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

[16] Roussas, G. G. (1969). Nonparametric estimation in Markov processes. *Ann. Inst. Statist. Math.* **21**, 73–87.

[17] Roussas, G. G. (1988). Nonparametric estimation in mixing sequences of random variables. *J. Statist. Plann. Inference* **18**, 135–149.

[18] 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.

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

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

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

[22] 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.

[23] Schick, A. and Wefelmeyer, W. (2006). Convergence rates in weighted L_1-spaces of kernel density estimators for linear processes. Technical Report, Department of Mathematical Sciences, Binghamton University.

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

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

[26] Schick, A. and Wefelmeyer, W. (2007c). Root-n consistency in weighted L_1-spaces for density estimators of invertible linear processes. To appear in: *Stat. Inference Stoch. Process.*