Functional convergence and optimality of plug-in estimators for stationary densities of moving average processes

By Anton Schick[1] and Wolfgang Wefelmeyer

Binghamton University and University of Cologne

We give new results, under mild assumptions, on convergence rates in L_1 and L_2 for residual-based kernel estimators of the innovation density of moving average processes. Exploiting the convolution representation of the stationary density of moving average processes, these estimators can be used to obtain $n^{1/2}$-consistent plug-in estimators for this stationary density. Here we derive functional weak convergence results in L_1 and $C_0(\mathbb{R})$ for these plug-in estimators. If efficient estimators for the finite-dimensional parameters of the process are used in our construction, semiparametric efficiency of our plug-in estimators is obtained.

1. Introduction. Smooth functionals of appropriate density estimators and regression function estimators are known to converge at the parametric rate $n^{-1/2}$, even though the function estimators themselves converge only at slower rates, depending on the smoothness of the estimated function. Analogous results hold for functionals of derivatives of densities and regression functions.

For *nonparametric* models and i.i.d. observations, there is by now a considerable literature on such "plug-in" estimators in which the parametric rate is obtained, the influence function is calculated, and the estimators are shown to be asymptotically efficient in the sense of having minimal asymptotic variance among regular estimators. Of particular interest have been nonlinear integral functionals of a density f and its derivatives $f^{(k)}$. For $\int f^{(k)}(x)^2\,dx$ see Hall and Marron (1987) and Bickel and Ritov (1988); for generalizations $\int \phi(f(x),x)\,dx$ and $\int \phi(f(x),\dots,f^{(k)}(x),x)\,dx$ see Laurent (1996) and Birgé and Massart (1995). The Shannon entropy $-\int f(x)\log f(x)\,dx$ is considered in Dudewicz and van der Meulen (1981), Tsybakov and van der Meulen (1996), and Eggermont and LaRiccia (1999). Abramson and Goldstein (1991) study the equidistribution functional $2\int f(x)g(x)/(f(x)+g(x))\,dx$ of two densities. Frees (1994) treats the density of a symmetric function $h(X_1,\dots,X_m)$ of $m>1$ independent and identically distributed random variables at a point. His result generalizes to non-identically distributed random variables. This covers in particular convolution densities $g(x) = \int f(x-\vartheta y)f(y)\,dy$ at a fixed point x and for known scale parameter ϑ, considered by Saavedra and Cao (2000); such densities arise as stationary densities of

Received

AMS 2000 subject classifications. 62G07, 62M09, 62M10

Key words and phrases. Plug-in estimator, time series, semiparametric model, functional central limit theorem, efficient estimator, least dispersed estimator.

[1]Supported in part by NSF Grant DMS 0072174

first-order moving average processes $X_t = \varepsilon_t + \vartheta\varepsilon_{t-1}$ with innovation density f and known ϑ. This also covers densities of functions $u_i(X_1) + \cdots + u_m(X_m)$ at a point. Schick and Wefelmeyer (2004b) obtain functional central limit theorems for appropriate plug-in estimators of such densities, viewed as elements of the function spaces $C_0(\mathbb{R})$ and $L_1(\mathbb{R})$. For results on plug-in estimators of general functionals we refer to Goldstein and Khas'minskii (1995).

There are analogous nonparametric results on plug-in estimators based on i.i.d. observations $(X_1, Y_1), \dots, (X_n, Y_n)$ for functionals of the regression function $r(x) = E(Y|X = x)$ and the quantile regression function $q_\alpha(x) = \inf\{y : P(Y \le y|X = x) \ge \alpha\}$. Goldstein and Messer (1992) and Loh (1995) study $\int r(x)^2\,dx$; Efromovich and Samarov (2000) treat $\int r^{(k)}(x)^2\,dx$. Stoker (1991), Samarov (1991, 1993), and Li (1996) consider the average regression derivative $Er'(X)$. Doksum and Samarov (1995) introduce three estimators of a weighted version of Pearson's correlation ratio $\operatorname{Var} r(X)/\operatorname{Var} Y$. Chaudhuri, Doksum and Samarov (1996, 1997) estimate average weighted quantile regression derivatives $E[q'_\alpha(X)w(X)]$.

Suppose now that the model has additional structure. For example, in the regression model we might assume that the error is independent of the covariate and/or that we have a parametric model for the regression function. This complicates the calculation of the asymptotic variance bound and the construction of efficient plug-in estimators. There is much less literature on such problems. A well-studied degenerate case is the error variance $E\varepsilon^2$ in the nonparametric regression model $Y = r(X) + \varepsilon$, with error ε centered and independent of the covariate X. Hall and Marron (1990) estimate the error variance by the empirical variance of the residuals and calculate the asymptotic variance of this estimator. Müller, Schick and Wefelmeyer (2004a,b) show that the estimator is efficient, and adaptive with respect to the regression function. For other functionals of the error distribution, the empirical estimator is not adaptive but still efficient; see Akritas and Van Keilegom (2001) and Müller, Schick and Wefelmeyer (2004a,b). In the corresponding semiparametric model, with $r = r_\vartheta$ known up to some parameter ϑ, the empirical estimators can be improved; see Schick and Wefelmeyer (2002a) for an autoregressive version of such a result.

Here we are interested in $n^{1/2}$-consistent and efficient estimation of the stationary density of a moving average process. This model has structural features analogous to those mentioned in the previous paragraph: it is driven by independent innovations, and it is semiparametric. There is a rich literature on estimating stationary densities of stochastic processes by kernel estimators $\frac{1}{n}\sum_{j=1}^n k_b(x - X_j)$; see e.g. Chanda (1983), Yakowitz (1989), Hart and Vieu (1990), Tran (1992), Chan and Tran (1992), Hall and Tran (1996), and Honda (2000) who also discuss applications. Under appropriate conditions, these estimators have similar (nonparametric) rates as for i.i.d. observations. For continuous-time processes, parametric rates of kernel estimators are more common; see Castellana and Leadbetter (1986), Bosq (1993, 1995), Blanke and Bosq (1997), and Bosq, Merlevède and Peligrad (1999). However, such estimators do not exploit the specific structure of the process. For the case of the MA(1) model $X_t = \varepsilon_t + \vartheta\varepsilon_{t-1}$, Saavedra and Cao (1999) make use of the above-mentioned representation $g(x) = \int f(x - \vartheta y)f(y)\,dy$ of the stationary density at x and propose the plug-in estimator $\hat{g}(x) = \int \hat{f}(x - \hat{\vartheta}y)\hat{f}(y)\,dy$. Here $\hat{\vartheta}$ is $n^{1/2}$-consistent and $\hat{f}$ is a kernel estimator based on estimated innovations $\hat{\varepsilon}_j = \sum_{s=0}^{j-1} \hat{\vartheta}^s X_{j-s}$. Saavedra and Cao observe that the asymptotic variance of the plug-in estimator decreases as n^{-1}. Under rather mild conditions, Schick and Wefelmeyer (2004a) give sharper results, in the spirit of the above nonparametric references; they show in particular asymptotic linearity and discuss efficiency. Heuristically, the required stochastic expansion of $\hat{g}(x)$ is obtained by writing

$$\hat{g}(x) - g(x) \approx \int (\hat{f}(x - \vartheta y) - f(x - \vartheta y)) f(y)\,dy + \int f(x - \vartheta y)(\hat{f}(y) - f(y))\,dy - (\hat{\vartheta} - \vartheta) \int y f'(x - \vartheta y) f(y)\,dy.$$

The first two terms are of order $n^{-1/2}$ because they may be viewed as (centered) plug-in estimators; the last term is of order $n^{-1/2}$ if $\hat{\vartheta}$ is $n^{1/2}$-consistent. Related results exist for continuous-time processes: Efficient and $n^{1/2}$-consistent estimators for the stationary density of diffusion processes on a time interval $[0, n]$ with nonparametric drift are constructed in Kutoyants (1997a,b;1998;1999); for the derivative of the density see Dalalyan and Kutoyants (2003). We also refer to Chapter 4 in Kutoyants (2004).

The present paper extends the results on MA(1) models in two directions, at the same time weakening the conditions further. One extension is to moving average processes of (fixed) higher order. The other extension is that we do not consider the stationary density at a fixed point x only, but view g and $\hat{g}$ as elements of the function spaces L_1 or $C_0(\mathbb{R})$ and obtain that the process $n^{1/2}(\hat{g} - g)$ converges in distribution in these spaces to a centered Gaussian process. This seems to be the first non-local result on functional convergence of density estimators. The results can be used in a straightforward way for testing whether the time series is Gaussian, and for efficient estimation of various linear and nonlinear functionals of the stationary law. Schick and Wefelmeyer (2004c) have extended these results further, to invertible infinite-order linear processes, including ARMA models.

Specifically, we consider an MA(q) process

$$X_t = \varepsilon_t + \vartheta_1 \varepsilon_{t-1} + \cdots + \vartheta_q \varepsilon_{t-q},$$

where the ε_t are i.i.d. innovations with finite second moment and density f. We assume that the parameter $\vartheta = (\vartheta_1, \dots, \vartheta_q)^\top$ satisfies $\vartheta_q \neq 0$, and that the complex polynomial $p_\vartheta(z) = 1 + \vartheta_1 z + \cdots + \vartheta_q z^q$ has no roots in the unit disk. This assumption guarantees stationarity of the process. It also implies invertibility, i.e. a representation of the innovations in terms of the observations,

$$\varepsilon_t = \sum_{s=0}^{\infty} \alpha_s(\vartheta) X_{t-s},$$

where the $\alpha_s(\vartheta)$ are the coefficients in the series $1/p_\vartheta(z) = \sum_{s=0}^{\infty} \alpha_s(\vartheta) z^s$.

We suppose that we observe $X_1, \dots, X_n$ from this process. Our first goal is to study estimators of f. They are based on estimated innovations. Under the above assumptions, there exist $n^{1/2}$-consistent estimators $\hat\vartheta$ of ϑ. We use such an estimator to estimate the innovations by the truncated series

$$\hat\varepsilon_j = \sum_{s=0}^{j-1} \alpha_s(\hat\vartheta) X_{j-s}.$$

The estimators $\hat\varepsilon_j$ are good only for large values of j. Therefore, we will not use the first r estimated innovations $\hat\varepsilon_1, \dots, \hat\varepsilon_r$ to construct estimators for f, and estimate f by kernel estimators

$$\hat f(x) = \frac{1}{n-r} \sum_{j=r+1}^{n} k_b(x - \hat\varepsilon_j). \tag{1.1}$$

Here $k_b(u) = k(u/b)/b$ for some density k and some bandwidth b.

In Section 2 we derive rates of convergence in probability for $\hat f$ in the L_1 and L_2 norms. These rates are new. They are the same as those for kernel estimators based on the actual innovations $\varepsilon_{r+1}, \dots, \varepsilon_n$, i.e. kernel estimators based on i.i.d. observations. For the *sup*-norm, *strong* convergence rates of kernel estimators based on residuals have been obtained in similar models: Fazal (1977) and Li (1995) consider linear regression with fixed design; Liebscher (1999) treats nonlinear autoregressive models.

In Sections 3 and 4 we address estimation of the stationary density of the MA(q) process. It has the representation

$$g(x) = \int \cdots \int f\Big(x - \sum_{i=1}^{q} \vartheta_i y_i\Big) f(y_1) \cdots f(y_q)\, dy_1 \cdots dy_q, \quad x \in \mathbb{R}. \tag{1.2}$$

We estimate g by plugging in the estimator $\hat f$ of the innovation density. In Sections 3 and 4 we prove that $n^{1/2}(\hat g - g)$ converges in distribution, in L_1 and in $C_0(\mathbb{R})$, respectively, to a centered Gaussian process. In Section 5 we show that $\hat g$ is efficient in those function spaces if an efficient estimator for ϑ is used. Efficiency is understood in the semiparametric sense discussed in Bickel, Klaassen, Ritov and Wellner (1998) for the i.i.d. case. Recall that in nonparametric models like the ones mentioned above, all regular estimators are asymptotically equivalent, and therefore proving efficiency is straightforward, whereas in our semiparametric model, the calculations of the influence function of the estimator and of the asymptotic variance bound pose difficulties.

2. Estimation of the innovation density. We study the estimator $\hat f$ introduced in (1.1). For notational convenience we assume that the observations are $X_{-r+1}, \dots, X_n$. Then we can write

$$\hat\varepsilon_j = \sum_{s=0}^{j+r-1} \alpha_s(\hat\vartheta) X_{j-s}, \quad j = 1, \dots, n,$$

$$\hat f(x) = \frac{1}{n} \sum_{j=1}^{n} k_b(x - \hat\varepsilon_j).$$

We will let r tend to infinity slowly. As a first approximation to $\hat\varepsilon_j$ we use

$$\tilde\varepsilon_j = \varepsilon_j + (\hat\vartheta - \vartheta)^\top \dot\varepsilon_j,$$

where

$$\dot\varepsilon_j = \sum_{s=0}^{\infty} \dot\alpha_s(\vartheta) X_{j-s},$$

with $\dot\alpha_s(\vartheta)$ denoting the gradient of $\alpha(\vartheta)$ with respect to ϑ.

LEMMA 1. *Suppose* $\hat\vartheta$ *is* $n^{1/2}$*-consistent for* ϑ*, and* $r/\log n \to \infty$ *and* $r/(\log n)^2 \to 0$*. Then*

$$\sum_{j=1}^{n} |\hat\varepsilon_j - \tilde\varepsilon_j| = O_p(1). \tag{2.1}$$

PROOF. Recall that $\varepsilon_j = \sum_{s=0}^{\infty} \alpha_s(\vartheta) X_{j-s}$. We can bound the left-hand side of (2.1) by $T_1 + T_2 + \|\hat\vartheta - \vartheta\| T_3$ with

$$T_1 = \sum_{j=1}^{n} \sum_{s=0}^{j+r-1} |\alpha_s(\hat\vartheta) - \alpha_s(\vartheta) - (\hat\vartheta - \vartheta)^\top \dot\alpha(\vartheta)|\, |X_{j-s}|,$$

$$T_2 = \sum_{j=1}^{n} \sum_{s=j+r}^{\infty} |\alpha_s(\vartheta) X_{j-s}|,$$

$$T_3 = \sum_{j=1}^{n} \sum_{s=j+r}^{\infty} \|\dot\alpha_s(\vartheta)\|\, |X_{j-s}|.$$

It is well known that α_s and its derivatives decay exponentially locally uniformly. Thus there are $\eta > 0$, $\rho < 1$, and a constant C such that

$$\sup_{\|\tau-\vartheta\|\le\eta} |\alpha_s(\tau)| + \|\dot\alpha_s(\tau)\| + \|\ddot\alpha_s(\tau)\| \le C\rho^s.$$

Hence

$$ET_2 = \sum_{j=1}^{n} \sum_{s=j+r}^{\infty} |\alpha_s(\vartheta)|\, E|X_0| = O\Big(\sum_{j=1}^{n} \rho^{j+r}\Big) = O(\rho^r),$$

$$ET_3 = \sum_{j=1}^{n} \sum_{s=j+r}^{\infty} \|\dot\alpha_s(\vartheta)\|\, E|X_0| = O\Big(\sum_{j=1}^{n} \rho^{j+r}\Big) = O(\rho^r).$$

Since $\hat\vartheta$ is $n^{1/2}$-consistent, the probability of $\|\hat\vartheta - \vartheta\| > \eta$ tends to zero, and

$$T_1 \le \|\hat\vartheta - \vartheta\|^2 \sum_{j=1}^{n} \sum_{s=0}^{j+r-1} \sup_{\|\tau-\vartheta\|\le\eta} \|\ddot\alpha_s(\tau)\|\, |X_{j-s}| + o_p(1) = O_p(1).$$

In the last step we have used $n^{1/2}$-consistency and

$$E\sum_{j=1}^{n}\sum_{s=0}^{j+r-1}\rho^s|X_{j-s}| = O(n).$$

□

We use Lemma 1 and smoothness of f to obtain rates of convergence of $\hat{f}$ in the L_p-norms for $p = 1, 2$. We say a function h is *L_p-Lipschitz* if there is a constant C such that

$$\int |h(x+t) - h(x)|^p \, dx \le C^p |t|^p, \quad t \in \mathbb{R}.$$

We call C the L_p-Lipschitz constant.

LEMMA 2. *Let h be absolutely continuous with almost everywhere derivative h' in L_p for some $p \in [1,\infty)$. Then h is L_p-Lipschitz with L_p-Lipschitz constant $C = \|h\|_p$. Moreover, for every random variable Y with $E[|Y|^p]$ finite, we have*

$$\int \big|E[h(x+tY] - h(x) - tE[Y]h'(x)\big|^p \, dx = o(|t|^p).$$

If h' is L_p-Lipschitz and $E[Y^{2p}]$ is finite, then we even have

$$\int \big|E[h(x+tY] - h(x) - tE[Y]h'(x)\big|^p \, dx \le \|h''\|_p^p E[Y^{2p}]|t|^{2p}.$$

PROOF. By absolute continuity, $h(x+t) - h(x) = t\int_0^1 h'(x+ut)\,du$, and the moment inequality gives

$$\int |h(x+t) - h(x)|^p \, dx \le |t|^p \iint_0^1 |h'(x+ut)|^p \, du \, dx = |t|^p \int |h'(x)|^p \, dx.$$

Similarly,

$$\int \big|E[h(x+tY) - h(x) - tYh'(x)]\big|^p \, dx \le |t|^p E\Big[|Y|^p \iint_0^1 |h'(x+utY) - h'(x)|^p \, du \, dx\Big].$$

This bound is of order $o(|t|^p)$ by the L_p-continuity of translations, see e.g. Theorem 9.5 in Rudin (1974), the Lebesgue dominated convergence theorem, and the bound $\int |h(x-s) - h(x)|^p \, dx \le s^p \|h'\|_p^p$. It is of order $o(t^{2p})$ if h' is L_p-Lipschitz and $E[Y^{2p}]$ is finite. □

REMARK 1. Recall that the density f has *finite Fisher information* (for location) if f is absolutely continuous and

$$J_f = \int \Big(\frac{f'(x)}{f(x)}\Big)^2 f(x)\,dx < \infty. \tag{2.2}$$

In this case, we have $\|f\|_\infty \le \|f'\|_1 \le J_f^{1/2}$ and $\|f'\|_2^2 \le \|f\|_\infty J_f \le J_f^{3/2}$. Thus, if f has finite Fisher information for location, then f is L_1- and L_2-Lipschitz. □

THEOREM 1. *Suppose f has finite second moment and is L_1-Lipschitz. Suppose k has finite second moment and is twice continuously differentiable, $b \to 0$ as $n \to \infty$, and the integrals $\int (1+|u|)^2 |k'(u)|\, du$ and $\int |k''(u)|\, du$ are finite. Let $\hat{\vartheta}$ be a $n^{1/2}$-consistent estimator of ϑ. Then*

$$\|\hat{f} - f\|_1 = O_p(n^{-1}b^{-2}) + O(b) + O_p(n^{-1/2}b^{-1/2})$$

In particular, if $b \sim n^{-1/3}$, we get $\|\hat{f} - f\|_1 = O_p(n^{-1/3})$.

PROOF. Let

$$\tilde{f}(x) = \frac{1}{n}\sum_{j=1}^n k_b(x - \tilde{\varepsilon}_j) \quad \text{and} \quad \bar{f}(x) = \frac{1}{n}\sum_{j=1}^n k_b(x - \varepsilon_j).$$

Then, using Lemma 1,

$$\begin{aligned}\|\hat{f} - \tilde{f}\|_1 &\le \frac{1}{n}\sum_{j=1}^n \int |k_b(x - \hat{\varepsilon}_j) - k_b(x - \tilde{\varepsilon}_j)|\, dx \\ &= \frac{1}{n}\sum_{j=1}^n \int \left|(\hat{\varepsilon}_j - \tilde{\varepsilon}_j)\int_0^1 k_b'(x - \tilde{\varepsilon}_j - u(\hat{\varepsilon}_j - \tilde{\varepsilon}_j))\, du\right| dx \\ &\le \frac{1}{n}\sum_{j=1}^n |\hat{\varepsilon}_j - \tilde{\varepsilon}_j|\, \|k_b'\|_1 = O_p(n^{-1}b^{-1}).\end{aligned}$$

By the L_1-Lipschitz property of f and the moment assumptions on k,

$$\|f * k_b - f\|_1 \le \int\int |f(x - bu) - f(x)|\, dx\, k(u)\, du \le \int C|bu|k(u)\, du = O(b).$$

A similar argument, using $\int k'(u)\, du = 0$, yields

$$|k_b' * f\|_1 = b^{-1}\int \left|\int (f(x - bu) - f(x))k'(u)\, du\right| dx \le \int C|uk'(u)|\, du = O(1). \tag{2.3}$$

Since f and k have finite second moment and k is bounded, it follows from Lemma 2 in Devroye (1992) that

$$\|\bar{f} - f * k_b\|_1 = O_p(n^{-1/2}b^{-1/2}).$$

Thus it remains to show that

$$\|\tilde{f} - \bar{f}\|_1 = O_p(n^{-1}b^{-2}) + O_p(n^{-1/2}). \tag{2.4}$$

For this let

$$\hat{h}(x) = \frac{1}{n}\sum_{j=1}^n \dot{\varepsilon}_j k_b'(x - \varepsilon_j), \quad x \in \mathbb{R}.$$

A Taylor expansion gives

$$\|\tilde{f} - \bar{f} + (\hat{\vartheta} - \vartheta)^\top \hat{h}\|_1 \le \|\hat{\vartheta} - \vartheta\|^2 \frac{1}{n}\sum_{j=1}^n \|\dot{\varepsilon}_j\|^2 \|k_b''\|_1 = O_p(n^{-1}b^{-2}). \tag{2.5}$$

Let

$$\bar{h}(x) = \frac{1}{n}\sum_{j=1}^{n} \dot{\varepsilon}_j E[k_b'(x-\varepsilon_j)] = \frac{1}{n}\sum_{j=1}^{n} \dot{\varepsilon}_j k_b' * f(x), \quad x \in \mathbb{R}.$$

Then $\hat{h}(x) - \bar{h}(x)$ is a martingale, and

$$nE[\|\hat{h}(x) - \bar{h}(x)\|^2] \le E[\|\dot{\varepsilon}_1\|^2]E[k_b'(x-\varepsilon_1)^2] = E[\|\dot{\varepsilon}_1\|^2]\int k_b'(x-y)^2 f(y)\,dy.$$

Using this bound we can show that $\|\hat{h}(x) - \bar{h}(x)\|_1 = O_p(n^{-1/2}b^{-3/2})$. Indeed, with $r_n = n^{-1/2}b^{-3/2}E[\|\dot{\varepsilon}_1\|^2]^{1/2}$, we can bound

$$\begin{aligned} E[\|\hat{h}(x) - \bar{h}(x)\|_1] &\le \int E[\|\hat{h}(x) - \bar{h}(x)\|^2]^{1/2}\,dx \\ &\le r_n \int \Big(\int f(x-by)k'(y)^2\,dy\Big)^{1/2} dx \\ &\le r_n \Big(\int \frac{dx}{1+x^2} \iint (1+x^2) f(x-by)k'(y)^2\,dx\,dy\Big)^{1/2} \\ &= O(r_n). \end{aligned}$$

In view of (2.3) we have $\int \|\bar{h}(x)\|\,dx = O_p(1)$. Together with $n^{1/2}$-consistency of $\hat{\vartheta}$, the above shows that $\|(\hat{\vartheta}-\vartheta)^\top \hat{h}\|_1 = O_p(n^{-1/2} + n^{-1}b^{-3/2})$. This and (2.5) yield (2.4). This completes the proof. □

REMARK 2. Other results are possible under different assumptions on f. If f is absolutely continuous with an integrable almost everywhere derivative f' and k has mean zero, then Lemma 2 shows that $\|f * k_b - f\|_1 = o(b)$. In this case we can use $b \sim n^{-1/4}$ and get $\|\hat{f} - f\|_1 = o_p(n^{-1/4})$. If f' is also L_1-Lipschitz, then one can show that $\|f * k_b - f\|_1 = O(b^2)$, and the choice $b \sim n^{-1/5}$ yields $\|\hat{f} - f\|_1 = O_p(n^{-2/5})$. Faster rates are possible under additional smoothness on f and if higher order kernels are employed. □

COROLLARY 1. *Under the assumptions of Theorem 1 we have*

$$\int |x|\,|\hat{f}(x) - f(x)|\,dx = O_p(\|\hat{f} - f\|_1^{1/2}). \tag{2.6}$$

PROOF. Use the Cauchy–Schwarz inequality and $|\hat{f}(x) - f(x)| \le \hat{f}(x) + f(x)$ to bound the square of the left-hand side of (2.6) by

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

The desired result follows from this bound and

$$\int x^2 \hat{f}(x)\,dx = \frac{1}{n}\sum_{j=1}^{n}\int (\hat{\varepsilon}_j + bu)^2 k(u)\,du \le \frac{1}{n}\sum_{j=1}^{n} 2\hat{\varepsilon}_j^2 + 2b^2\int u^2 k(u)\,du = O_p(1).$$

In the last step we have used that $\max_{1\le j\le n}|\hat{\varepsilon}_j - \varepsilon_j| = O_p(1)$, which is a consequence of Lemma 1. □

THEOREM 2. *Suppose f has finite fourth moment and is L_2-Lipschitz. Suppose k has finite second moment and is three times continuously differentiable, and the integrals $\int(1+v^2)|k'(v)|\,dv$, $\int(1+v^2)|k''(v)|\,dv$, and $\int|k'''(v)|\,dv$ are finite. Let $\hat\vartheta$ be a $n^{1/2}$-consistent estimator of ϑ. Let $b \to 0$ as $n \to \infty$. Then*

$$\|\hat f - f\|_2 \le O_p(n^{-1}b^{-3/2}) + O_p(n^{-1/2}b^{-1/2}) + O(b) + O_p(n^{-3/2}b^{-7/2}).$$

In particular, if $b \sim n^{-1/3}$, we get $\|\hat f - f\|_2 = O_p(n^{-1/3})$.

PROOF. Let $\tilde f$ and $\bar f$ be as in the proof of Theorem 1. Using the Cauchy–Schwarz inequality in the form $(\sum a_j b_j)^2 \le \sum |a_j| \sum |a_j| b_j^2$, Fubini's theorem, and Lemma 1, we obtain

$$\begin{aligned}\|\hat f - \tilde f\|_2^2 &= \int \Big| \frac{1}{n}\sum_{j=1}^n (\hat\varepsilon_j - \tilde\varepsilon_j) \int_0^1 k_b'(x - \tilde\varepsilon_j - u(\hat\varepsilon_j - \tilde\varepsilon_j))\,du \Big|^2 dx \\ &\le \frac{1}{n}\sum_{j=1}^n |\hat\varepsilon_j - \tilde\varepsilon_j| \int \frac{1}{n}\sum_{j=1}^n |\hat\varepsilon_j - \tilde\varepsilon_j| \int_0^1 |k_b'(x - \tilde\varepsilon_j - u(\hat\varepsilon_j - \tilde\varepsilon_j))|^2\,du\,dx \\ &= \Big(\frac{1}{n}\sum_{j=1}^n |\hat\varepsilon_j - \tilde\varepsilon_j|\Big)^2 \|k_b'\|_2^2 = O_p(n^{-2}b^{-3}).\end{aligned}$$

It is well known that

$$E[\|\bar f - f * k_b\|_2^2] \le n^{-1}\|k_b\|_2^2 = O(n^{-1}b^{-1}).$$

Thus $\|\bar f - f * k_b\|_2 = O_p(n^{-1/2}b^{-1/2})$. From the L_2-Lipschitz property of f and the moment assumptions on k we derive

$$\|f * k_b - f\|_2^2 \le \iint (f(x-bu) - f(x))^2 k(u)\,du\,dx \le C \int b^2u^2k(u)\,du = O(b^2).$$

Since $\int k'(u)\,du = 0$ and $\int k''(u)\,du = 0$, we can use a similar argument to conclude that

$$\|k_b' * f\|_2 = O(1) \quad\text{and}\quad \|k_b'' * f\|_2 = O(b^{-1}). \tag{2.7}$$

For example,

$$\begin{aligned}\|k_b' * f\|_2^2 &= b^{-2}\int \Big(\int (f(x-bu) - f(x))k'(u)\,du\Big)^2 dx \\ &\le b^{-2}\int\int (f(x-bu)-f(x))^2|k'(u)|\,du \int |k'(u)|\,du\,dx = O(1).\end{aligned}$$

To complete the proof we shall now show that

$$\|\tilde f - \bar f\|_2^2 = O_p(n^{-3}b^{-7}) + O_p(n^{-2}b^{-3}) + O_p(n^{-1}). \tag{2.8}$$

Let $\hat h$ and $\bar h$ be as in the proof of Theorem 1, and set

$$\hat H(x) = \frac{1}{n}\sum_{j=1}^n \dot\varepsilon_j\dot\varepsilon_j^\top k_b''(x - \varepsilon_j) \quad\text{and}\quad \bar H(x) = \frac{1}{n}\sum_{j=1}^n \dot\varepsilon_j\dot\varepsilon_j^\top k_b'' * f(x), \quad x \in \mathbb{R}.$$

Then a Taylor expansion yields

$$\left\|\tilde f-\bar f-(\hat\vartheta-\vartheta)^\top\hat h-\frac12(\hat\vartheta-\vartheta)^\top\hat H(\hat\vartheta-\vartheta)\right\|_2^2\le\int\Big|\frac1n\sum_{j=1}^n|(\hat\vartheta-\vartheta)^\top\dot\varepsilon_j|^3I_j(x,b)\Big|^2\,dx,$$

where

$$I_j(x,b)=\int_0^1\int_0^1\int_0^1 vw^2\big|k_b'''(x-\varepsilon_j-uvw(\hat\vartheta-\vartheta)^\top\dot\varepsilon_j)\big|\,du\,dv\,dw.$$

Now apply the Cauchy–Schwarz inequality as at the beginning of this proof and use the fact that $\int I_j^2(x,b)\,dx\le\int(k_b'''(x))^2\,dx=b^{-7}\|k'''\|_2^2$ to see that

$$\int\Big|\frac1n\sum_{j=1}^n|(\hat\vartheta-\vartheta)^\top\dot\varepsilon_j|^3I_j(x,b)\Big|^2\,dx\le\Big(\frac1n\sum_{j=1}^n|(\hat\vartheta-\vartheta)^\top\dot\varepsilon_j|^3\Big)^2b^{-7}\|k'''\|_2^2.$$

Since f has finite fourth moment, we obtain that $E[\|\dot\varepsilon_1\|^4]<\infty$. It is now easy to see that

$$\left\|\tilde f-\bar f-(\hat\vartheta-\vartheta)^\top\hat h-\frac12(\hat\vartheta-\vartheta)^\top\hat H(\hat\vartheta-\vartheta)\right\|_2^2=O_p(n^{-3}b^{-7}).$$

Next we have

$$\int E[\|\hat h(x)-\bar h(x)\|^2]\,dx\le n^{-1}E[\|\dot\varepsilon_1\|^2]\iint k_b'(x-y)^2f(y)\,dy\,dx=O(n^{-1}b^{-3})$$

and similarly,

$$\int E[\|\hat H(x)-\bar H(x)\|^2]\,dx\le n^{-1}E[\|\dot\varepsilon_1\|^4]\iint k_b''(x-y)^2f(y)\,dy\,dx=O(n^{-1}b^{-5}).$$

In view of (2.7) we also have

$$\int\|\bar h(x)\|_2^2\,dx=O_p(1)\quad\text{and}\quad\int\|\bar H(x)\|_2^2\,dx=O_p(b^{-2}).$$

From the above and the $n^{1/2}$-consistency of $\hat\vartheta$ we obtain

$$\|(\hat\vartheta-\vartheta)^\top\hat h\|_2^2=O_p(n^{-2}b^{-3}+n^{-1}),$$
$$\|(\hat\vartheta-\vartheta)^\top\hat H(\hat\vartheta-\vartheta)\|_2^2=O_p(n^{-3}b^{-5}+n^{-2}b^{-2}).$$

Combining the above we obtain the desired (2.8). □

REMARK 3. Other results are possible under different assumptions on f. If f is absolutely continuous with an integrable almost everywhere derivative f' and k has mean zero, then Lemma 2 shows that $\|f*k_b-f\|_2=o(b)$. In this case we can use $b\sim n^{-1/4}$ and get $\|\hat f-f\|_2=o_p(n^{-1/4})$. If f' is also L_2-Lipschitz, then Lemma 2 shows that $\|f*k_b-f\|_2=O(b^2)$, and the choice $b\sim n^{-1/5}$ yields $\|\hat f-f\|_2=O_p(n^{-2/5})$. □

COROLLARY 2. *In addition to the assumptions of Theorem 2, let f be bounded and let k have a finite fourth moment. Then*

$$\int x^2(\hat f(x)-f(x))^2\,dx=O_p(b^{-1/2}\|\hat f-f\|_2).\tag{2.9}$$

PROOF. It is easy to see that $\|\hat{f}\|_\infty \le \|k\|_\infty b^{-1}$. As in the proof of Corollary 1 one verifies that

$$\int x^4 \hat{f}(x)\,dx = \frac{1}{n}\sum_{j=1}^{n}(\hat{\varepsilon}_j + bu)^4 k(u)\,du \le \frac{1}{n}\sum_{j=1}^{n} 8\hat{\varepsilon}_j^4 + 8b^4 \int u^4 k(u)\,du = O_p(1).$$

Now use the Cauchy–Schwarz inequality to bound the square of the left-hand side of (2.9) by $\|\hat{f}-f\|_2^2 \int x^4(\hat{f}(x)-f(x))^2\,dx$ and thus also by the larger term

$$\|\hat{f}-f\|_2^2 \int x^4(\hat{f}(x)+f(x))\,dx\,(\|\hat{f}\|_\infty + \|f\|_\infty) = O_p(b^{-1})\|\hat{f}-f\|_2^2.$$

This is the desired result. □

We conclude this section with a technical result about scaling which will be used later.

LEMMA 3. *Let h be an integrable function that is absolutely continuous with an almost everywhere derivative h' that satisfies $\int |x||h'(x)|\,dx < \infty$. Then, for $s \ne 0$, and as $t \to s$,*

$$\int \Big|\frac{1}{|t|}h\Big(\frac{x}{t}\Big) - \frac{1}{|s|}h\Big(\frac{x}{s}\Big) + \frac{t-s}{s|s|}\Big[h\Big(\frac{x}{s}\Big) + \frac{x}{s}h'\Big(\frac{x}{s}\Big)\Big]\Big|\,dx = o(|t-s|).$$

PROOF. Substituting $u = x/s$ and letting $a = s/t$ (which we may and do assume to be positive), one simplifies the left-hand side to

$$\int \Big|ah(au) - h(u) - \frac{a-1}{a}(h(u) - uh'(u))\Big|\,du.$$

Let $\tilde{h}(y) = e^y h(e^y)$, $y \in \mathbb{R}$. Then $\tilde{h}$ is absolutely continuous with integrable almost everywhere derivative $\tilde{h}'(y) = e^y h(e^y) = e^{2y}h'(e^y)$. Thus the previous lemma with $Y = 1$ and the fact that $e^{-b}(e^b - 1) = b + o(b)$ as $b \to 0$, yield

$$\int \big|\tilde{h}(y+b) - \tilde{h}(y) - e^{-b}(e^b-1)\tilde{h}'(y)\big|\,dy = o(b).$$

Letting $b = \log(a)$, the substitution $y = e^u$ gives

$$\int_0^\infty \Big|ah(au) - h(u) - \frac{a-1}{a}(h(u) - uh'(u))\Big|\,du.$$

A similar argument yields the result for integration over negative u. □

3. Convergence in L_1 of estimators for the stationary density. We now address the estimation of the stationary density g viewed as an element of the function space L_1. Recall that g has the representation (1.2):

$$g(x) = \int \cdots \int f\Big(x - \sum_{i=1}^{q}\vartheta_i y_i\Big) f(y_1)\cdots f(y_q)\,dy_1\cdots dy_q, \quad x \in \mathbb{R}.$$

Alternatively, we can write g as the convolution $g = f * f_{\tau_1} * \cdots * f_{\tau_m}$ of the densities $f, f_{\tau_1}, \ldots, f_{\tau_m}$, where $\tau_1, \ldots, \tau_m$ are the non-zero components of ϑ and f_τ denotes the density of $\tau\varepsilon_1$ for non-zero τ, so that $f_\tau(x) = f(x/\tau)/|\tau|$, $x \in \mathbb{R}$. Since $\vartheta_q \neq 0$, we can thus write

$$g(x) = \int \cdots \int f * f_{\vartheta_q}\Big(x - \sum_{i=1}^{q-1} \vartheta_i y_i\Big) f(y_1) \cdots f(y_{q-1})\, dy_1 \cdots dy_{q-1}, \quad x \in \mathbb{R}.$$

In what follows we assume that f is absolutely continuous with an integrable almost everywhere derivative f'. This implies that f is bounded and L_1-Lipschitz. It also implies that $f * f_\tau$ is continuously differentiable with a derivative $f' * f_\tau$ that is absolutely continuous with almost everywhere derivative $f' * (f_\tau)'$. We have $f' * (f_\tau)'(x) = \frac{1}{\tau} \int f'(x - \tau y) f'(y)\, dy$. From this we immediately see that the density g is continuously differentiable with integrable derivative g' given by

$$g'(x) = \int \cdots \int f'\Big(x - \sum_{i=1}^{q} \vartheta_i y_i\Big) f(y_1) \cdots f(y_q)\, dy_1 \cdots dy_q, \quad x \in \mathbb{R},$$

and that g' is absolutely continuous with integrable almost everywhere derivative g'' given by

$$g''(x) = \frac{1}{\vartheta_q} \int \cdots \int f'\Big(x - \sum_{i=1}^{q} \vartheta_i y_i\Big) f(y_1) \cdots f(y_{q-1}) f'(y_q)\, dy_1 \cdots dy_q, \quad x \in \mathbb{R}.$$

If we require in addition that $\int |x f'(x)|\, dx < \infty$, then the gradient $\dot{g}(x) = (\dot{g}_1(x), \dots, \dot{g}_q(x))^\top$ of $g(x)$ with respect to the parameter ϑ exists for every x and has ν-th component given by

$$\dot{g}_\nu(x) = -\int \cdots \int y_\nu f'\Big(x - \sum_{i=1}^{q} \vartheta_i y_i\Big) f(y_1) \cdots f(y_q)\, dy_1 \cdots dy_q.$$

Actually differentiability holds uniformly in x:

$$\sup_{x \in \mathbb{R}} |g_{\vartheta+\delta}(x) - g(x) - \delta^\top \dot{g}(x)| = o(\|\delta\|), \tag{3.1}$$

and in the L_1-sense:

$$\|g_{\vartheta+\delta} - g - \delta^\top \dot{g}\|_1 = o(\|\delta\|), \tag{3.2}$$

where $g_{\vartheta+\delta}$ denotes the stationary density for the parameter value $\vartheta + \delta$. These are verified with the help of Lemmas 2 and 3.

We estimate g by the plug-in estimator

$$\hat{g}(x) = \int \cdots \int \hat{f}\Big(x - \sum_{i=1}^{q} \hat{\vartheta}_i y_i\Big) \hat{f}(y_1) \cdots \hat{f}(y_q)\, dy_1 \cdots dy_q, \quad x \in \mathbb{R},$$

where $\hat{f}$ is as in Section 2. We view $\hat{g}$ as an element of L_1 and show that under mild additional assumptions, the process $n^{1/2}(\hat{g} - g)$ converges in distribution in the space L_1 to a centered Gaussian process.

To describe this result, it will be convenient to set $\vartheta_0 = 1$. Note that g is the density of $Y = \varepsilon_0 + \vartheta_1\varepsilon_1 + \cdots + \vartheta_q\varepsilon_q = \sum_{i=0}^q \vartheta_i\varepsilon_i$. Now let p_i denote the density of $Y - \vartheta_i\varepsilon_i$ for $i = 0, \dots q$. Then

$$p_0(x) = \int \cdots \int f_{\vartheta_q}\Big(x - \sum_{i=1}^{q-1} \vartheta_i y_i\Big) f(y_1)\cdots f(y_{q-1})\, dy_1 \cdots dy_{q-1}, \quad x \in \mathbb{R},$$

and for $i = 1, \dots, q$,

$$p_i(x) = \int \cdots \int f\Big(x - \sum_{j:j\neq i} \vartheta_j y_j\Big) \prod_{j:j\neq i} f(y_j)\, dy_j, \quad x \in \mathbb{R}.$$

We have for $i = 0, \dots, q$,

$$\int p_i(x - \vartheta_i y) f(y)\, dy = g(x), \quad x \in \mathbb{R}.$$

Now define L_1-valued processes $\mathbb{H}_{n,0}, \dots, \mathbb{H}_{n,q}$ by

$$\mathbb{H}_{n,i}(x) = \frac{1}{n}\sum_{j=1}^n \Big(p_i(x - \vartheta_i\varepsilon_j) - \int p_i(x - \vartheta_i y) f(y)\, dy\Big), \quad x \in \mathbb{R}.$$

With $\mathbb{F}$ denoting the empirical distribution function of the innovations $\varepsilon_1, \dots, \varepsilon_n$ and F the distribution function with density f, we can write

$$\mathbb{H}_{n,i}(x) = \int p_i(x - \vartheta_i y)\, d(\mathbb{F}(y) - F(y)), \quad x \in \mathbb{R}.$$

By the assumptions on f, the densities $p_0, \dots, p_q$ are absolutely continuous with integrable almost everywhere derivatives. Thus we have the representation

$$\mathbb{H}_{n,i}(x) = \vartheta_i \int p_i'(x - \vartheta_i y)(\mathbb{F}(y) - F(y))\, dy, \quad x \in \mathbb{R}. \tag{3.3}$$

We shall now use this representation to establish tightness of $n^{1/2}\mathbb{H}_{n,i}$ in L_1. For this we also need the following characterization of compact sets in L_1, which is known as the Fréchet–Kolmogorov theorem; see Yosida (1980, p. 275).

LEMMA 4. *A closed subset H of L_1 is compact if and only if*

$$\sup_{h\in H} \|h\|_1 < \infty,$$

$$\lim_{t\to 0} \sup_{h\in H} \int |h(x-t) - h(x)|\, dx = 0,$$

$$\lim_{K\uparrow\infty} \sup_{h\in H} \int_{|x|>K} |h(x)|\, dx = 0.$$

Let $\Delta = n^{1/2}(\mathbb{F} - F)$ denote the empirical process and assume that the function $\psi = (1-F)^{1/2}F^{1/2}$ is integrable. Then

$$E[\|\Delta\|_1] = \int E[|\Delta(x)|]\, dx \le \int E[\Delta^2(x)]^{1/2}\, dx = \|\psi\|_1 < \infty. \tag{3.4}$$

We also find that

$$E \int_{|x| \geq K} |\Delta(x)| \, dx \leq \int_{|x| \geq K} \psi(x) \, dx \to 0 \quad \text{as } K \to \infty. \tag{3.5}$$

Moreover, for positive $t \in \mathbb{R}$ and finite K,

$$\|n^{1/2}\mathbb{H}_{n,i}\|_1 \leq |\vartheta_i| \|p_i'\|_1 \|\Delta\|_1, \tag{3.6}$$

$$\int n^{1/2} |\mathbb{H}_{n,i}(x+t) - \mathbb{H}_{n,i}(x)| \, dx \leq |\vartheta_i| \|\Delta\|_1 \int |p_i'(x+t) - p_i'(x)| \, dx, \tag{3.7}$$

$$\begin{aligned} \int_{|x|>2K} |n^{1/2}\mathbb{H}_{n,i}(x)| \, dx \leq |\vartheta_i| \|p_i'\|_1 &\int_{|\vartheta_i y|>K} |\Delta(y)| \, dy \\ &+ |\vartheta_i| \|\Delta\|_1 \int_{|x|>K} |p_i'(x)| \, dx. \end{aligned} \tag{3.8}$$

By the integrability of p_i',

$$\int |p_i'(z+t) - p_i'(z)| \, dz \to 0 \quad \text{as } t \to 0. \tag{3.9}$$

Applying (3.4), (3.5), and (3.9) to the inequalities (3.6) to (3.8) and using Lemma 4, we see that $n^{1/2}\mathbb{H}_{n,i}$ is tight in L_1. Consequently, $\mathbb{H}_{n,1} + \cdots + \mathbb{H}_{n,m}$ converges in distribution in the space L_1 to a centered Gaussian process.

For $\alpha > 1$ we have

$$\|\psi\|_1^2 \leq \int (1+|x|)^{-\alpha} \, dx \int (1+|x|)^{\alpha} (1-F(x))F(x) \, dx.$$

This shows that integrability of ψ is implied if f has a finite moment of order $\beta > 2$.

We are now ready to state the main result of this section. Let V denote the map defined by

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

and set

$$\mu = (1 + \vartheta_1 + \cdots + \vartheta_q) E[\dot{\varepsilon}_1].$$

THEOREM 3. *Suppose f has finite moment of order $\beta > 2$ and is absolutely continuous with an almost everywhere derivative f' which satisfies $\|f'V\|_1 < \infty$. Let the kernel k be as in Theorem 1 and have mean zero. Let the bandwidth b satisfy $nb^4 \to 0$ and $nb^{8/3} \to 0$. Let $\hat{\vartheta}$ be a $n^{1/2}$-consistent estimator of ϑ. Then*

$$\|\hat{g} - g - (\mathbb{H}_{n,0} + \cdots + \mathbb{H}_{n,q}) - (\hat{\vartheta} - \vartheta)^\top (\dot{g} - \mu g')\|_1 = o_p(n^{-1/2}).$$

Moreover, $n^{1/2}(\mathbb{H}_{n,0} + \cdots + \mathbb{H}_{n,q})$ converges in distribution in the space L_1 to a centered Gaussian process.

Theorem 3 is a simple consequence of the following two lemmas. To state them, we introduce the bounded linear operator A_i from L_1 to L_1 which maps an integrable function h to the integrable function $A_i h$ defined by

$$A_i h(x) = \int p_i(x - \vartheta_i y) h(y) \, dy, \quad x \in \mathbb{R}.$$

Abbreviate $f * k_b$ by f_*.

LEMMA 5. *Under the assumptions of Theorem 3 we have*

$$\|\hat{g} - g - (A_0(\hat{f} - f_*) + \cdots + A_q(\hat{f} - f_*)) - (\hat{\vartheta} - \vartheta)^\top \dot{g}\|_1 = o_p(n^{-1/2}).$$

LEMMA 6. *Under the assumptions of Theorem 3 we have for* $i = 0, \ldots, q$,

$$\|A_i(\hat{f} - f_*) - \mathbb{H}_{n,i} + (\hat{\vartheta} - \vartheta)^\top E[\dot{\varepsilon}_1]\vartheta_i g'\|_1 = o_p(n^{-1/2}).$$

PROOF OF LEMMA 5. It follows from $nb^{8/3} \to \infty$ that $n^{-1}b^{-2} + n^{-1/2}b^{-1/2} = o(n^{-1/4})$. Thus, in view of the proofs of Theorem 1 and Corollary 1 it holds that

$$\|\hat{f} - f_*\|_1 = o_p(n^{-1/4}), \tag{3.10}$$

$$\|(\hat{f} - f_*)V\|_1 = o_p(1). \tag{3.11}$$

For $t \in \mathbb{R}^q$ and integrable $h_0, \ldots, h_q$, let $L(t, h_0, \ldots, h_q)$ denote the integrable function defined by

$$L(t, h_0, \ldots, h_q)(x) = \int \cdots \int h_0\Big(x - \sum_{i=1}^q t_i y_i\Big) h_1(y_1) \cdots h_q(y_q)\, dy_1 \cdots dy_q, \quad x \in \mathbb{R}.$$

It is easy to see that

$$\|L(t, h_0, \ldots, h_q)\|_1 \le \|h_0\|_1 \cdots \|h_q\|_1. \tag{3.12}$$

Moreover, if h_0 is absolutely continuous with integrable almost everywhere derivative, then

$$\|L(t, h_0, \ldots, h_q) - L(s, h_0, \ldots, h_q)\|_1 \le q^{1/2}\|h_0'\|_1\|t - s\|\, \|h_1 V\|_1 \cdots \|h_q V\|_1. \tag{3.13}$$

Indeed we can bound the left-hand side of (3.13) by

$$\|h_0'\|_1 \sum_{i=1}^q |t_i - s_i| \int \cdots \int |y_i h_1(y_1) \cdots h_q(y_q)|\, dy_1 \cdots dy_q.$$

We can write $\hat{g} = L(\hat{\vartheta}, \hat{f}, \ldots, \hat{f})$, $g = L(\vartheta, f, \ldots, f)$ and $g_t = L(t, f, \ldots, f)$. It follows from (3.2) and the $n^{1/2}$-consistency of $\hat{\vartheta}$ that

$$\|g_{\hat{\vartheta}} - g - (\hat{\vartheta} - \vartheta)^\top \dot{g}\|_1 = o_p(n^{-1/2}).$$

Set $g_t^* = L(t, f_*, \ldots, f_*)$. We can express g_t^* as

$$g_t^*(x) = \int g_t(x - bu)\tilde{k}_t(u)\, du$$

with $\tilde{k}_t = L(t, k, \ldots, k)$ a density with mean zero and finite variance which is $1 + \|t\|^2$ times the variance of k. Thus, by Lemma 2,

$$\|g_{\hat{\vartheta}}^* - g_{\hat{\vartheta}}\|_1 \le \|g_{\hat{\vartheta}}''\|_1 b^2 \int u^2 \tilde{k}_{\hat{\vartheta}}(u)\, du = O_p(b^2) = o_p(n^{-1/2}). \tag{3.14}$$

For a subset A of $\{0, \dots, q\}$ we set

$$\gamma(t, A) = L(t, h_0, \dots, h_q) \quad \text{with} \quad h_i = \begin{cases} \hat{f} - f_*, & i \in A, \\ f_*, & i \neq A, \end{cases}$$

and for $r = 0, \dots, q$ we set

$$\Gamma(r, t) = \sum_{|A|=r} \gamma(t, A).$$

Since $\hat{f} = f_* + \hat{f} - f_*$ and $h_i \mapsto L(t, h_0, \dots, h_q)$ is linear for each $i = 0, \dots, q$, we obtain the expansion

$$\hat{g} = \Gamma(0, \hat{\vartheta}) + \cdots + \Gamma(q + 1, \hat{\vartheta}).$$

We obtain from (3.12) that $\|\Gamma(r, \hat{\vartheta})\|_1 \le \sum_{|A|=r} \|\gamma(\hat{\vartheta}, A)\|_1 \le \binom{q+1}{r} \|\hat{f} - f_*\|_1^r$. Thus, by (3.10),

$$\|\hat{g} - \Gamma(0, \hat{\vartheta}) - \Gamma(1, \hat{\vartheta})\|_1 \le \sum_{r=2}^{q+1} \|\Gamma(r, \hat{\vartheta})\|_1 = o_p(n^{-1/2}).$$

Note that $\Gamma(0, \hat{\vartheta}) = g_{\hat{\vartheta}}$. Thus to finish the proof it suffices to show that

$$\|\Gamma(1, \hat{\vartheta}) - \Gamma(1, \vartheta)\|_1 \le \sum_{i=0}^{q} \|\gamma(\hat{\vartheta}, i) - \gamma(\vartheta, i)\|_1 = o_p(n^{-1/2}), \tag{3.15}$$

where $\gamma(\vartheta, i) = \gamma(\vartheta, \{i\})$. We obtain from (3.13), (3.10) and the $n^{1/2}$-consistency of $\hat{\vartheta}$ that

$$\begin{aligned} \|\gamma(\hat{\vartheta}, 1) - \gamma(\vartheta, 1)\|_1 &= \|L(\hat{\vartheta}, f_*, \hat{f} - f_*, f_*, \dots, f_*) - L(\vartheta, f_*, \hat{f} - f_*, f_*, \dots, f_*)\|_1 \\ &\le q^{1/2} \|f_*'\|_1 \|\hat{\vartheta} - \vartheta\| \, \|(\hat{f} - f_*)V\|_1 \|f_* V\|_1^{q-1} = o_p(n^{-1/2}). \end{aligned}$$

Here we used that $\|f_*'\|_1 = O(1)$ and $\|f_* V\|_1 = O(1)$. The former was shown in the proof of Theorem 1, and the latter follows from direct calculations. Similarly, for $i = 2, \dots, q$ we have $\|\gamma(\hat{\vartheta}, i) - \gamma(\vartheta, i)\|_1 = o_p(n^{-1/2})$. For the case $i = 0$ we let $\hat{\vartheta}_* = (\hat{\vartheta}_1, \dots, \hat{\vartheta}_{q-1}, 1)^\top$ and $\vartheta_* = (\vartheta_1, \dots, \vartheta_{q-1}, 1)^\top$ and set $\hat{\phi} = f_{\hat{\vartheta}_q} * k_{\hat{\vartheta}_q b}$ and $\phi = f_{\vartheta_q} * k_{\vartheta_q b}$. Then we use the commutativity of convolutions to derive that

$$\begin{aligned} \gamma(\hat{\vartheta}, 0) &= L(\hat{\vartheta}, \hat{f} - f_*, f_*, \dots, f_*) = L(\hat{\vartheta}_*, \hat{f} - f_*, f_*, \dots, f_*, \hat{\phi}) \\ &= L(\hat{\vartheta}_*, \hat{\phi}, f_*, \dots, f_*, \hat{f} - f_*) \end{aligned}$$

and $\gamma(\vartheta, 0) = L(\vartheta_*, \phi, f_*, \dots, f_*, \hat{f} - f_*)$. By linearity of $L(t, h_0, \dots, h_q)$ in h_0 we can write

$$\begin{aligned} \gamma(\hat{\vartheta}, 0) - \gamma(\vartheta, 0) &= L(\hat{\vartheta}_*, \hat{\phi} - \phi, f_*, \dots, f_*, \hat{f} - f_*) \\ &\quad + L(\hat{\vartheta}_*, \phi, f_*, \dots, f_*, \hat{f} - f_*) - L(\vartheta_*, \phi, f_*, \dots, f_*, \hat{f} - f_*). \end{aligned}$$

The arguments for the case $i = 1$ above yield

$$\|L(\hat{\vartheta}_*, \phi, f_*, \dots, f_*, \hat{f} - f_*) - L(\vartheta_*, \phi, f_*, \dots, f_*, \hat{f} - f_*)\|_1 = o_p(n^{-1/2}).$$

Next, it follows from inequality (3.12) that

$$\|L(\hat\vartheta_*, \hat\phi - \phi, f_*, \dots, f_*, \hat f - f_*)\|_1 \le \|\hat\phi - \phi\|_1 \|\hat f - f_*\|_1.$$

Lemma 3 and the $n^{1/2}$-consistency of $\hat\vartheta$ yield

$$\|f_{\hat\vartheta_q} * k_{\hat\vartheta_q b} - f_{\vartheta_q} * k_{\hat\vartheta_q b}\|_1 \le \|f_{\hat\vartheta_q} - f_{\vartheta_q}\|_1 = O_p(n^{-1/2}).$$

Since f_{ϑ_q} is L_1-Lipschitz, we obtain

$$\|f_{\vartheta_q} * k_{\hat\vartheta_q b} - f_{\vartheta_q} * k_{\vartheta_q b}\|_1 = O_p(b|\hat\vartheta_q - \vartheta_q|) = o_p(n^{-1/2}).$$

Thus we get $\|\hat\phi - \phi\|_1 = O_p(n^{-1/2})$ and hence $\|\gamma(\hat\vartheta, 0) - \gamma(\vartheta, 0)\|_1 = o_p(n^{-1/2})$. This completes the proof of (3.15). The desired result now follows since $\gamma(\vartheta, i) = A_i(\hat f - f_*)$ for $i = 0, \dots, q$. □

Proof of Lemma 6. Fix $i \in \{0, \dots, q\}$. Let $\tilde f$ and $\bar f$ be as in the proof of Theorem 1. Then we can express

$$A_i(\hat f - f_*) = A_i(\hat f - \tilde f) + A_i(\tilde f - \bar f) + A_i(\bar f - f_*).$$

Since $\|\hat f - \tilde f\|_1 = O_p(n^{-1}b^{-1})$ as shown in the proof of Theorem 1, we obtain that

$$\|A_i(\hat f - \tilde f)\|_1 \le \|\hat f - \tilde f\|_1 = o_p(n^{-1/2}).$$

It is easy to see that

$$A_i(\bar f - f_*)(x) = \int \mathbb{H}_{n,i}(x - \vartheta_i bu) k(u)\, du, \quad x \in \mathbb{R}. \tag{3.16}$$

Since $\int |h_n(x - a_n u) - h(x)|\, dx \to 0$ as $\|h_n - h\|_1 \to 0$ and $a_n \to 0$, and since $n^{1/2}\mathbb{H}_{n,i}$ converges in distribution in the space L_1, we obtain from Rubin's Theorem (see e.g. Theorem 5.5 in Billingsley, 1968) that

$$\|A_i(\bar f - f_*) - \mathbb{H}_{n,i}\|_1 = o_p(n^{-1/2}).$$

The desired result will thus follow if we show that

$$\|A_i(\tilde f - \bar f) + (\hat\vartheta - \vartheta)^\top E[\dot\varepsilon_1]\vartheta_i g'\|_1 = o_p(n^{-1/2}). \tag{3.17}$$

For this, note first that

$$A_i(\tilde f)(x) = \frac{1}{n}\sum_{j=1}^n \int p_i(x - \vartheta_i(\varepsilon_j + (\hat\vartheta - \vartheta)^\top \dot\varepsilon_j + bu))k(u)\, du,$$

$$A_i(\bar f)(x) = \frac{1}{n}\sum_{j=1}^n \int p_i(x - \vartheta_i(\varepsilon_j + bu))k(u)\, du.$$

It follows from the properties of f that p_i is absolutely continuous with integrable almost everywhere derivative p_i'. Now set

$$\chi_i(x) = \frac{1}{n}\sum_{j=1}^n \dot\varepsilon_j \int p_i'(x - \vartheta_i(\varepsilon_j + bu))k(u)\, du, \quad x \in \mathbb{R}.$$

It is easy to see that we can bound $D_i = \|A_i(\tilde f - \bar f) + \vartheta_i(\hat\vartheta - \vartheta)^\top \chi_i\|_1$ by

$$\begin{aligned} D_i &\le \frac{1}{n}\sum_{j=1}^n \int |p_i(x - \vartheta_i(\hat\vartheta - \vartheta)\dot\varepsilon_j) - p_i(x) + \vartheta_i(\hat\vartheta - \vartheta)^\top \dot\varepsilon_j p_i'(x)|\,dx \\ &\le \frac{1}{n}\sum_{j=1}^n |\vartheta_i|\,\|\hat\vartheta - \vartheta\|\,\|\dot\varepsilon_j\| \int\int_0^1 |p_i'(x - s\vartheta_i(\hat\vartheta - \vartheta)\dot\varepsilon_j) - p_i'(x)|\,ds\,dx \\ &\le |\vartheta_i|\,\|\hat\vartheta - \vartheta\| \frac{1}{n}\sum_{j=1}^n \|\dot\varepsilon_j\| \sup_{|t|\le\xi_n} \int |p_i'(x-t) - p_i'(x)|\,dx, \end{aligned}$$

where $\xi_n = \max_{1\le j\le n} |(\hat\vartheta - \vartheta)^\top \dot\varepsilon_j|$. Since $\hat\vartheta$ is $n^{1/2}$-consistent and

$$P\Big(\max_{1\le j\le n} n^{-1/2}\|\dot\varepsilon_j\| > \eta\Big) \le nP(\|\dot\varepsilon_1\| > \eta n^{1/2}) \le E\big[\|\dot\varepsilon_1\|^2 \mathbf{1}[\|\dot\varepsilon_1\| > \eta n^{1/2}]\big] \to 0,$$

we can conclude that $\xi_n = o_p(1)$. This and (3.9) yield

$$\|A_i(\tilde f - \bar f) + \vartheta_i(\hat\vartheta - \vartheta)^\top \chi_i\|_1 = o_p(n^{-1/2}).$$

It is easy to check that

$$\|\chi_i - E[\dot\varepsilon_1]g'\|_1 = o_p(1).$$

This completes the proof of (3.17) □

Remark 4. Under the conditions of Theorem 3, the optimal bandwidth rate for estimating f is $b \sim n^{-1/3}$. The requirement $nb^4 \to 0$ allows us to over-smooth the kernel estimates of f. The condition $nb^4 \to 0$ is used to conclude (3.14). It cannot be relaxed even if we impose additional smoothness on f as long as we insist on using kernels of order two. For higher order kernels, however, it can be relaxed. For example, if k is a kernel of order four and f'' is integrable, then we can weaken the requirement $nb^4 \to 0$ to $nb^8 \to 0$. Indeed, we then have $\|g^*_{\hat\vartheta} - g_{\hat\vartheta}\|_1 = O(b^4)$. The latter even holds without the additional smoothness assumption on f for such kernels as long as three of the coefficients of ϑ are non-zero. More generally, one can show that if m of the coefficients of ϑ are non-zero and one uses a kernel of order $m+1$, then $\|g^*_{\hat\vartheta} - g_{\hat\vartheta}\|_1 = O(b^{m+1})$. For $m = 2$ and a kernel of order three we can take $b \sim n^{-1/5}$, although this choice of bandwidth over-smoothes the kernel estimator $\hat f$. □

4. Convergence in $C_0(\mathbb{R})$ of estimators for the stationary density. In the previous section we have studied estimation of g in the function space L_1. This is a natural space when dealing with densities. However, we are sometimes interested in other norms, in particular in the sup-norm. In this case it is more convenient to view g as an element of $C_0(\mathbb{R})$, the set of (uniformly) continuous functions h from $\mathbb{R}$ to $\mathbb{R}$ which vanish at infinity in the sense of the one-point compactification: $\lim_{K\to\infty} \sup_{|x|>K} |h(x)| = 0$. Endowed with the sup-norm, $C_0(\mathbb{R})$ becomes a separable Banach space.

The goal of this section is to translate the results obtained in the L_1-norm to the sup-norm. More precisely, we shall prove a sup-norm version of the expansion given in Theorem 3, and then conclude that the process $n^{1/2}(\hat{g} - g)$ converges in distribution in the space $C_0(\mathbb{R})$ to some centered Gaussian process.

Note that integrable uniformly continuous functions belong to $C_0(\mathbb{R})$. Assume now that f is absolutely continuous with an integrable almost everywhere derivate f'. Then the densities g, $p_0, \dots, p_q$ are also absolutely continuous with integrable almost everywhere derivatives and hence belong to $C_0(\mathbb{R})$. From this we immediately obtain that the processes $\mathbb{H}_{n,0}, \dots, \mathbb{H}_{n,q}$ introduced in the previous section have sample paths in $C_0(\mathbb{R})$ and hence are $C_0(\mathbb{R})$-valued random elements. Let us now show that $n^{1/2}\mathbb{H}_{n,i}$ is tight in $C_0(\mathbb{R})$ for each $i = 0, \dots, q$. To obtain the latter, we recall the following characterization of compact subsets of $C_0(\mathbb{R})$. For a proof see Schick and Wefelmeyer (2004b).

LEMMA 7. *A closed subset H of $C_0(\mathbb{R})$ is compact if and only if*

$$\lim_{\delta \downarrow 0} \sup_{h \in H} \sup_{|z-y| \le \delta} |h(z) - h(y)| = 0,$$

$$\lim_{K \to \infty} \sup_{h \in H} \sup_{|x| \ge K} |h(x)| = 0.$$

To obtain tightness of $n^{1/2}\mathbb{H}_{n,i}$, first recall the representation

$$n^{1/2}\mathbb{H}_{n,i}(x) = \vartheta_i \int p_i'(x - \vartheta_i y)\Delta(y)\, dy, \quad x \in \mathbb{R}.$$

Thus we may assume that $\vartheta_i \neq 0$ and get the bounds

$$\sup_{|z-y| \le \delta} |n^{1/2}(\mathbb{H}_{n,i}(z) - \mathbb{H}_{n,i}(y))| \le \|\Delta\|_\infty \sup_{|t| \le \delta} \int |p_i'(u+t) - p_i'(u)|\, dt,$$

$$\sup_{|x| > 2M} |n^{1/2}\mathbb{H}_{n,i}(x)| \le \sup_{|\vartheta_i y| > M} |\Delta(y)|\, \|p_i'\|_1 + \|\Delta\|_\infty \int_{|u| > M} |p_i'(u)|\, du.$$

These bounds, relation (3.9), and well-known properties of the empirical process give, in view of the above lemma, the desired tightness of the process $n^{1/2}\mathbb{H}_{n,i}$. It is now easy to check that $\mathbb{H}_{n,0} + \dots + \mathbb{H}_{n,q}$ converges in distribution in the space $C_0(\mathbb{R})$ to a centered Gaussian process.

We are now ready to state the main result of this section. Recall that $V(x) = 1 + |x|$ and $\mu = (1 + \vartheta_1 + \dots + \vartheta_q)E[\dot{\varepsilon}_1]$.

THEOREM 4. *Suppose f has finite fourth moment and is absolutely continuous with an almost everywhere derivative f' that satisfies $\|f'V\|_1 < \infty$ and $\|f'V\|_2 < \infty$. Let the kernel k be as in Theorem 2 and have mean zero and finite fourth moment. Let the bandwidth b satisfy $nb^4 \to 0$ and $nb^{14/5} \to \infty$. Let $\hat{\vartheta}$ be a $n^{1/2}$-consistent estimator of ϑ. Then*

$$\|\hat{g} - g - (\mathbb{H}_{n,0} + \dots + \mathbb{H}_{n,q}) - (\hat{\vartheta} - \vartheta)^\top (\dot{g} - \mu g')\|_\infty = o_p(n^{-1/2}).$$

Moreover, $n^{1/2}(\mathbb{H}_{n,0} + \dots + \mathbb{H}_{n,q})$ converges in distribution in the space $C_0(\mathbb{R})$ to a centered Gaussian process.

PROOF. Since the density p_i is uniformly continuous, the range of the operator A_i contains only integrable and uniformly continuous functions and is thus a subset of $C_0(\mathbb{R})$. Actually, A_i is also a bounded linear operator from L_1 into $C_0(\mathbb{R})$, as

$$\|A_i h\|_\infty \le \|p_i\|_\infty \|h\|_1. \tag{4.1}$$

Moreover, if h is also square-integrable and $\vartheta_i \neq 0$, we obtain the alternative bound

$$\|A_i h\|_\infty \le |\vartheta_i|^{-1/2} \|p_i\|_2 \|h\|_2. \tag{4.2}$$

Indeed, an application of the Cauchy–Schwarz inequality shows that

$$|A_i h(x)| \le \int p_i(x - \vartheta_i y)^2 \, dy \int h(y)^2 \, dy \le \frac{1}{|\vartheta_i|} \|p_i\|_2^2 \|h\|_2^2, \quad x \in \mathbb{R}. \tag{4.3}$$

Since f' is square-integrable, we find that g'' is bounded. More precisely, we have, for t close to ϑ,

$$\|g_t''\|_\infty \le \|f' * f_{t_q}'\|_\infty \le \|f'\|_2 \|f_{t_q}'\|_2 \le |t_q|^{-3/2} \|f'\|_2^2.$$

From this and the fact that k has mean zero and finite variance we obtain by a standard argument that

$$\|g_{\hat\vartheta}^* - g_{\hat\vartheta}\|_1 \le \|g_{\hat\vartheta}''\|_\infty b^2 \int u^2 \tilde{k}_{\hat\vartheta}(u) \, du = O_p(b^2) = o_p(n^{-1/2}). \tag{4.4}$$

It follows from proofs of Theorems 1 and 2, Corollaries 1 and 2, and the choice of bandwidth that

$$\|\hat f - f_*\|_i = o_p(n^{-1/4}) \quad \text{and} \quad \|(\hat f - f_*)V\|_i = o_p(1), \quad i = 1, 2. \tag{4.5}$$

It suffices to prove

$$\|\hat g - g - (A_0(\hat f - f_*) + \cdots + A_q(\hat f - f_*)) - (\hat\vartheta - \vartheta)^\top \dot g\|_\infty = o_p(n^{-1/2}) \tag{4.6}$$

and, for $i = 0, \dots, q$,

$$\|A_i(\hat f - f_*) - \mathbb{H}_{n,i} + (\hat\vartheta - \vartheta)^\top E[\dot\varepsilon_1] \vartheta_i g'\|_\infty = o_p(n^{-1/2}). \tag{4.7}$$

The proof of (4.6) is as the proof of Lemma 5; but now use (3.1) instead of (3.2), use (4.4) instead of (3.14), apply the bound

$$\|L(t, h_0, \dots, h_q)\|_\infty \le |t_q|^{-1} \|h_0\|_2 \|h_q\|_2 \|h_1\|_1 \dots \|h_{q-1}\|_1 \tag{4.8}$$

instead of the bound (3.12), and replace (3.13) with the bound

$$\begin{aligned} &\|L(t, h_0, \dots, h_q) - L(s, h_0, \dots, h_q)\|_\infty \\ &\quad \le q^{1/2} \|t - s\| \int_0^1 |s_q + v(t_q - s_q)|^{-1/2} \, dv \|h_0'\|_2 \|h_q V\|_2 \|h_1 V\|_1 \cdots \|h_{q-1} V\|_1, \end{aligned}$$

valid for absolutely continuous h_0 with square-integrable almost everywhere derivative. To prove this last inequality, bound its left-hand side by the supremum over x of

$$\int\cdots\int \Big| \int_0^1 h_0'\Big(x - \sum_{\nu=1}^q (s_\nu + v(t_\nu - s_\nu))y_\nu\Big)\,dv \sum_{i=1}^q (t_i - s_i)y_i \prod_{j=1}^q h_j(y_j)\Big| dy_1 \dots dy_q$$

$$\le \sum_{i=1}^q (t_i - s_i) \int_0^1 \int\cdots\int \Big|h_0'\Big(x - \sum_{\nu=1}^q (s_\nu + v(t_\nu - s_\nu))y_\nu\Big)\Big| \prod_{j=1}^q |V(y_j)h_j(y_j)|\,dy_j\,dv,$$

and then argue as in (4.3) above.

It remains to verify (4.7). Fix $i \in \{0, \dots, q\}$. If $\vartheta_i = 0$, then (4.7) holds as its left-hand side equals zero. Now assume that $\vartheta_i \neq 0$. Let $\tilde{f}$ and $\bar{f}$ be as in the proof of Theorem 1. We shall show that

$$\|A_i(\hat{f} - \tilde{f})\|_\infty = O_p(n^{-1}b^{-1}), \tag{4.9}$$

$$\|A_i(\bar{f} - f_*) - \mathbb{H}_{n,i}\|_\infty = o_p(n^{-1/2}), \tag{4.10}$$

$$\|A_i(\tilde{f} - \bar{f}) + (\hat{\vartheta} - \vartheta)^\top E[\dot{\varepsilon}_1]\vartheta_i g'\|_\infty = o_p(n^{-1/2}). \tag{4.11}$$

Since $\|\hat{f} - \tilde{f}\|_1 = O_p(n^{-1}b^{-1})$ as shown in the proof of Theorem 2, we obtain (4.9) from (4.1). It is easy to see that

$$\sup_{x\in\mathbb{R}} \Big| \int (h_n(x - a_n u) - h(x))k(u)\,du \Big| \to 0$$

if $\|h_n - h\|_\infty \to 0$, $a_n \to 0$ and $h \in C_0(\mathbb{R})$. In view of this, representation (3.16), and the weak convergence of $n^{1/2}\mathbb{H}_{n,i}$ in $C_0(\mathbb{R})$, we derive (4.10) from Rubin's Theorem. Since g'' is bounded, we have that $\|g' * k_b - g'\|_\infty = O(b)$. Thus it suffices to verify (4.11) with g' replaced by $g' * k_b$. In other words: We need to show that $\|D_i\|_\infty = o_p(n^{-1/2})$, where

$$D_i = A_i(\tilde{f} - \bar{f}) + (\hat{\vartheta} - \vartheta)^\top E[\dot{\varepsilon}_1]\vartheta_i g' * k_b.$$

Let $\tilde{\mathbb{F}}$ be the empirical distribution function based on $\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_n$. Using representation (3.3) we find that

$$D_i(x) = \vartheta_i \iint p_i'(x - \vartheta_i(y + bu))\big(\tilde{\mathbb{F}}(y) - \mathbb{F}(y) + (\hat{\vartheta} - \vartheta)^\top E[\dot{\varepsilon}_1]f(y)\big)\,dy\,k(u)\,du.$$

With the argument used to establish (4.3), we now derive the bound

$$\|D_i\|_\infty \le |\vartheta_i|^{1/2}\|p_i'\|_2\|\tilde{\mathbb{F}} - \mathbb{F} + (\hat{\vartheta} - \vartheta)^\top E[\dot{\varepsilon}_1]f\|_2.$$

Thus it suffices to show that

$$n^{1/2}\|\tilde{\mathbb{F}} - \mathbb{F} + (\hat{\vartheta} - \vartheta)^\top E[\dot{\varepsilon}_1]f\|_2 = o_p(1). \tag{4.12}$$

A similar result for the sup-norm was obtained by Koul (1996) in the context of nonlinear autoregression models. He required the density f to be positive. We follow his approach in establishing (4.12), but will not need f to be positive.

Let us define, for $x \in \mathbb{R}$ and $t \in \mathbb{R}^q$,

$$H(x,t) = n^{-1/2} \sum_{j=1}^{n} \Big(\mathbf{1}[\varepsilon_j + n^{-1/2} t^\top \dot{\varepsilon}_j \le x] - F(x - n^{-1/2} t^\top \dot{\varepsilon}_j) \Big),$$

$$U(x,t) = \sum_{j=1}^{n} \Big(F(x - n^{-1/2} t^\top \dot{\varepsilon}_j) - F(x) + n^{-1/2} t^\top \dot{\varepsilon}_j f(x) \Big)^2.$$

Then, with $\hat{t} = n^{1/2}(\hat{\vartheta} - \vartheta)$, we can bound the square of the left-hand side of (4.12) by

$$3 \int (H(x,\hat{t}) - H(x,0))^2 \, dx + 3 \int U(x,\hat{t}) \, dx + 3\|f\|_2^2 \|\hat{t}\|^2 \Big\| \frac{1}{n} \sum_{j=1}^{n} (\dot{\varepsilon}_j - E[\dot{\varepsilon}_1]) \Big\|^2.$$

Since $\|f'V\|_1$ is finite, f is bounded and hence square integrable. Thus the last term tends to zero in probability by the ergodic theorem. Since f is L_2-Lipschitz in view of $\|f'V\|_2 < \infty$ and $\xi_n = n^{-1/2} \max_{1 \le j \le n} |\hat{t}^\top \dot{\varepsilon}_j| = o_p(1)$ as shown in the proof of Lemma 6, we obtain $\int U(x,\hat{t})\, dx = o_p(1)$. Thus, the desired (4.12) follows if we show that for all positive integers M,

$$\sup_{\|t\| \le M} \int (H(x,t) - H(x,0))^2 \, dx = o_p(1). \tag{4.13}$$

Fix such an M and set $\mathcal{S} = \{-1,1\}^q$. For $j = 1, \dots, n$, let S_j be the $\mathcal{S}$-valued random vector whose i-th coordinate equals the sign of the i-th coordinate of $\dot{\varepsilon}_j$. For $\sigma \in \mathcal{S}$, let

$$H_\sigma(x,t) = n^{-1/2} \sum_{j=1}^{n} \mathbf{1}[S_j = \sigma] \Big(\mathbf{1}[\varepsilon_j + n^{-1/2} t^\top \dot{\varepsilon}_j \le x] - F(x - n^{-1/2} t^\top \dot{\varepsilon}_j) \Big).$$

Then $H(x,t) = \sum_{\sigma \in \mathcal{S}} H_\sigma(x,t)$, and (4.13) follows if we show that for all $\sigma \in \mathcal{S}$,

$$\sup_{\|t\| \le M} \int (H_\sigma(x,t) - H_\sigma(x,0))^2 \, dx = o_p(1). \tag{4.14}$$

Now fix $\sigma \in \mathcal{S}$ and a large integer K. Partition the cube $[-M,M]^q$ into $(2MK)^q$ cubes of equal volume. Let $\mathcal{C}$ denote the collection of these cubes. For each $C \in \mathcal{C}$ there exist vertices t_C and T_C of C such that $t_C^\top \dot{\varepsilon}_j \le t^\top \dot{\varepsilon}_j \le T_C^\top \dot{\varepsilon}_j$ for all $t \in C$ and for all $\dot{\varepsilon}_j$ with $S_j = \sigma$. Using this and monotonicity we can now show that $H_\sigma(x,T_C) - R_C(x) \le H_\sigma(x,t) \le H_\sigma(x,t_C) + R_C(x)$ for all $t \in C$, where

$$R_C(x) = n^{-1/2} \sum_{j=1}^{n} \mathbf{1}[S_j = \sigma] \, \Big(F(x - n^{-1/2} t_C^\top \dot{\varepsilon}_j) - F(x - n^{-1/2} T_C^\top \dot{\varepsilon}_j) \Big).$$

It is now easy to see that the left-hand side of (4.14) is bounded by

$$3 \max_{C \in \mathcal{C}} \int \Big((H_\sigma(x,t_C) - H_\sigma(x,0))^2 + (H_\sigma(x,t_C) - H_\sigma(x,T_C))^2 + R_C^2(x) \Big) \, dx.$$

Since $H_\sigma(x,t) - H_\sigma(x,s)$ is a martingale, we find, utilizing also stationarity, that

$$E[(H_\sigma(x,t) - H_\sigma(x,s))^2] \le E[F(x - n^{-1/2} t^\top \dot{\varepsilon}_1) - F(x - n^{-1/2} s^\top \dot{\varepsilon}_1)].$$

Since F is L_1-Lipschitz, we thus obtain

$$\int E[(H_\sigma(x,t) - H_\sigma(x,s))^2]\,dx \le \|f\|_1 E[\|\dot\varepsilon_1\|] n^{-1/2}\|t-s\|.$$

By the Cauchy–Schwarz inequality we have

$$R_C^2(x) \le \sum_{j=1}^n \mathbf{1}[S_j = \sigma]\Big(F(x - n^{-1/2}t_C^\top \dot\varepsilon_j) - F(x - n^{-1/2}T_C^\top \dot\varepsilon_j)\Big)^2.$$

Since f is square-integrable, F is L_2-Lipschitz, and

$$\int R_C^2(x)\,dx \le \|f\|_2^2 \frac{1}{n}\sum_{j=1}^n \|\dot\varepsilon_j\|^2\|T_C - t_C\|^2 \le \|f\|_2^2\frac{1}{n}\sum_{j=1}^n \|\dot\varepsilon_j\|^2 qK^{-2}.$$

Combining the above shows that the expected value of the left-hand side of (4.14) is bounded by

$$\begin{aligned}3\sum_{C\in\mathcal{C}} & E[\|\dot\varepsilon_1\|]n^{-1/2}(\|t_C\| + \|T_C - t_C\|) + 3\|f\|_2^2 E[\|\dot\varepsilon_1\|^2]qK^{-2} \\ &\le 3E[\|\dot\varepsilon_1\|](2MK)^q n^{-1/2}q^{1/2}(M + K^{-1}) + 3\|f\|_2^2 E[\|\dot\varepsilon_1\|^2]qK^{-2}.\end{aligned}$$

Since this is valid for all integers K, relation (4.14) holds. This completes the proof. □

5. Efficiency of estimators for the stationary density. We show that $\hat g$ is efficient if an efficient estimator for ϑ is used. This is a straightforward generalization of the efficiency result for MA(1) processes in Schick and Wefelmeyer (2004a), and we will be brief. Fix true parameters ϑ and f. Introduce a local model by perturbing ϑ as $\vartheta_{nc} = \vartheta + n^{-1/2}c$ with $c \in \mathbb{R}^q$, and f as f_{nh} with

$$\int\left(f_{nh}(x)^{1/2} - f(x)^{1/2} - n^{-1/2}\frac{1}{2}h(x)f(x)^{1/2}\right)^2 dx = o(n^{-1}).$$

The Hellinger derivative h is in $L_{2,0}(f) = \{h \in L_2(f) : \int h(x)f(x)\,dx = 0\}$. For technical convenience we choose f_{nh} such that, in addition, $\|f_{nh} - f\|_\infty \to 0$. Assume that f has finite Fisher information $J_f = \int \ell^2(x)f(x)\,dx$, with $\ell = f'/f$, in the sense of (2.2). Since f has a finite second moment, one obtains from an application of the Cauchy–Schwarz inequality that $\|f'V\|_1$ is finite. Write P_n and P_{nch} for the joint distribution of $(X_{-r+1},\dots,X_n)$ under (ϑ, f) and (ϑ_{nc}, f_{nh}), respectively, and set $\varepsilon = \varepsilon_1$. We have local asymptotic normality (LAN),

$$\log\frac{dP_{nch}}{dP_n} = n^{-1/2}\sum_{j=1}^n \big(c^\top \dot\varepsilon_j \ell(\varepsilon_j) + h(\varepsilon_j)\big) - \frac{1}{2}\|(c,h)\|_{\mathrm{LAN}}^2 + o_p(1), \tag{5.1}$$

with squared *LAN norm*

$$\|(c,h)\|_{\mathrm{LAN}}^2 = J_f c^\top E[\dot\varepsilon\dot\varepsilon^\top]c + 2c^\top E[\dot\varepsilon]E[\ell(\varepsilon)h(\varepsilon)] + E[h(\varepsilon)^2].$$

LAN for MA(q) processes follows from known results for more general time series. For $E[\varepsilon] = 0$ and fixed f, see Kreiss (1987), Jeganathan (1995), and Drost, Klaassen and Werker (1997); for varying f see Koul and Schick (1997). The *LAN inner product* induced by the LAN norm is

$$((c,h),(d,k))_{\mathrm{LAN}} = J_f c^\top E[\dot\varepsilon\dot\varepsilon^\top]d + c^\top E[\dot\varepsilon]E[\ell(\varepsilon)k(\varepsilon)] \\ + d^\top E[\dot\varepsilon]E[\ell(\varepsilon)h(\varepsilon)] + E[h(\varepsilon)k(\varepsilon)].$$

Now consider a real-valued functional κ of (ϑ, f) that is *differentiable* at the true (ϑ, f) in the (usual) sense that there exist $c_* \in \mathbb{R}^q$ and $h_* \in L_{2,0}(f)$ such that for all $c \in \mathbb{R}^q$ and $h \in L_{2,0}(f)$,

$$n^{1/2}\big(\kappa(\vartheta_{nc}, f_{nh}) - \kappa(\vartheta, f)\big) \to c_*^\top c + E[h_*(\varepsilon)h(\varepsilon)]. \tag{5.2}$$

The convolution theorem characterizes efficient estimators of κ in terms of the gradient of κ in the LAN inner product. This *LAN gradient* is the pair (c_κ, h_κ) with $c_\kappa \in \mathbb{R}^q$ and $h_\kappa \in L_{2,0}(f)$ such that

$$c_*^\top c + E[h_*(\varepsilon)h(\varepsilon)] = ((c_\kappa, h_\kappa),(c,h))_{\mathrm{LAN}} \quad \text{for all } c \in \mathbb{R}^q, h \in L_{2,0}(f).$$

Setting first $c = 0$ and then $h = 0$, one obtains

$$c_\kappa = J_f^{-1}\mathrm{Cov}[\dot\varepsilon]^{-1}(c_* - E[\dot\varepsilon]E[\ell(\varepsilon)h_*(\varepsilon)]), \quad h_\kappa = h_* - c_\kappa^\top E[\dot\varepsilon]\ell(\varepsilon). \tag{5.3}$$

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

$$n^{1/2}(\hat\kappa - \kappa(\vartheta_{nc}, f_{nh})) \Rightarrow L \quad \text{under } P_{nch} \quad \text{for all } c \in \mathbb{R}^q, h \in L_{2,0}(f).$$

The convolution theorem says that L is the convolution of some random variable and a normal random variable with mean zero and variance $\|(c_\kappa, h_\kappa)\|^2_{\mathrm{LAN}}$. This justifies calling $\hat\kappa$ *efficient* at (ϑ, f) if L is distributed as this normal random variable. It also follows from the convolution theorem that $\hat\kappa$ is regular and efficient if and only if

$$n^{1/2}(\hat\kappa - \kappa(\vartheta, f)) = n^{-1/2}\sum_{j=1}^n \big(c_\kappa^\top \dot\varepsilon_j \ell(\varepsilon_j) + h_\kappa(\varepsilon_j)\big) + o_p(1). \tag{5.4}$$

We apply this characterization to $g(x)$ and to the components of ϑ, interpreted as functionals of (ϑ, f). First we calculate the LAN gradient of

$$\kappa_x(\vartheta, f) = g(x) = \int \cdots \int f\Big(x - \sum_{i=1}^q \vartheta_i y_i\Big) f(y_1)\cdots f(y_q)\, dy_1 \cdots dy_q.$$

Recall that $\mu = (1 + \vartheta_1 + \cdots + \vartheta_q)E[\dot\varepsilon]$.

Lemma 8. *If f has finite Fisher information J_f, then the functional κ_x is differentiable at (ϑ, f) with LAN gradient (c_x, h_x) given by*

$$c_x = J_f^{-1}\mathrm{Cov}[\dot\varepsilon]^{-1}(\dot g(x) - \mu g'(x)), \quad h_x = \psi_x - c_x^\top E[\dot\varepsilon]\ell$$

with

$$\psi_x(y) = \sum_{i=0}^q \Big(p_i(x - \vartheta_i y) - \int p_i(x - \vartheta_i z) f(z)\, dz\Big).$$

PROOF. It is straightforward to check that the functional κ_x is differentiable in terms of the usual inner product for (c, h):

$$n^{1/2}\big(\kappa_x(\vartheta_{nc}, f_{nh}) - \kappa_x(\vartheta, f)\big) \to c^\top \dot g(x) + E[h(\varepsilon)\psi_x(\varepsilon)].$$

This is differentiability (5.2) with $c_* = \dot g(x)$ and $h_* = \psi_x$. The LAN gradient (c_x, h_x) is now obtained from (5.3), using $E[\ell(\varepsilon)\psi_x(\varepsilon)] = (1 + \vartheta_1 + \cdots + \vartheta_q)g'(x)$. □

By characterization (5.4), an estimator $\hat\kappa_x$ is regular and efficient for $g(x)$ if and only if

$$\begin{aligned} &n^{1/2}(\hat\kappa_x - g(x)) \\ &\quad = n^{-1/2}\sum_{j=1}^n \Big(\psi_x(\varepsilon_j) + (\dot g(x) - \mu g'(x))^\top \mathrm{Cov}[\dot\varepsilon]^{-1}(\dot\varepsilon_j - E[\dot\varepsilon])J_f^{-1}\ell(\varepsilon_j)\Big) + o_p(1). \end{aligned}$$

Note that $\frac{1}{n}\sum_{j=1}^n \psi_x(\varepsilon_j) = \mathbb{H}_{n,0} + \cdots + \mathbb{H}_{n,q}$. Comparing with Theorems 3 and 4, we see that our estimator $\hat g(x)$ is efficient if

$$n^{1/2}(\hat\vartheta - \vartheta) = n^{-1/2}\sum_{j=1}^n \mathrm{Cov}[\dot\varepsilon]^{-1}(\dot\varepsilon_j - E[\dot\varepsilon])J_f^{-1}\ell(\varepsilon_j) + o_p(1).$$

This is the characterization (5.4) of a (componentwise) efficient estimator of ϑ. Indeed, the functional $\kappa(\vartheta, f) = \vartheta_i$ is differentiable in the sense of (5.2) with $c_* = e_i$, the i-th q-dimensional unit vector, and $h_* = 0$. Hence by Lemma 8 its LAN gradient is (c_i, h_i) with

$$c_i = J_f^{-1}\mathrm{Cov}[\dot\varepsilon]^{-1}e_i, \quad h_i = -c_i^\top E[\dot\varepsilon]\ell.$$

Efficient estimators of ϑ were constructed in Kreiss (1987) under the assumption of symmetry, and in Drost, Klaassen and Werker (1997), Koul and Schick (1997), and Schick and Wefelmeyer (2002b) under the assumption that $E\varepsilon = 0$. These constructions can be adapted to our slightly more general situation, see Schick and Wefelmeyer (2004a) for the case $q = 1$.

Since $\hat g(x)$ is efficient for $g(x)$ whatever x, it follows that $(\hat g(x_1), \ldots, \hat g(x_k))$ is efficient for $(g(x_1), \ldots, g(x_k))$ for any $x_1, \ldots, x_k$ and any k. As an immediate consequence, under the assumptions of Sections 3 and 4, our estimator $\hat g$ is efficient for g in the spaces L_1 and $C_0(\mathbb{R})$.

REFERENCES

Abramson, I. and Goldstein, L. (1991). Efficient nonparametric testing by functional estimation. *J. Theoret. Probab.* 4, 137–159.

Akritas, M. G. and Van Keilegom, I. (2001). Non-parametric estimation of the residual distribution. *Scand. J. Statist.* 28, 549–567.

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

Bickel, P. J. and Ritov, Y. (1988). Estimating integrated squared density derivatives: Sharp best order of convergence estimates. *Sankhyā Ser. A* 50, 381–393.

Billingsley, P. (1968). *Convergence of Probability Measures.* Wiley, New York.

Birgé, L. and Massart, P. (1995). Estimation of integral functionals of a density. *Ann. Statist.* 23, 11–29.

Blanke, D. and Bosq, D. (1997). Accurate rates of density estimators for continuous-time processes. *Statist. Probab. Lett.* 33, 185–191.

Bosq, D. (1993). Vitesses optimales et superoptimales des estimateurs fonctionnels pour les processus à temps continu. *C. R. Acad. Sci. Paris Sér. I. Math.* 317, 1075–1078.

Bosq, D. (1995). Sur le comportement exotique de l'estimateur à noyau de la densité marginale d'un processus à temps continu. *C. R. Acad. Sci. Paris Sér. I. Math.* 320, 369–372.

Bosq, D., Merlevède, F. and Peligrad, M. (1999). Asymptotic normality for density kernel estimators in discrete and continuous time. *J. Multivariate Anal.* 68, 78–95.

Castellana, J. V. and Leadbetter, M. R. (1986). On smoothed probability density estimation for stationary processes. *Stochastic Process. Appl.* 21, 179–193.

Chan, N. H. and Tran, L. T. (1992). Nonparametric tests for serial dependence. *J. Time Ser. Anal.* 13 19–28.

Chanda, K. C. (1983). Density estimation for linear processes. *Ann. Inst. Statist. Math.* 35, 439–446.

Chaudhuri, P., Doksum, K. and Samarov, A. (1996). Nonparametric estimation of global functionals of conditional quantiles. In: *Robust Statistics, Data Analysis, and Computer Intensive Methods* (H. Rieder, ed.), 63–78, Lecture Notes in Statistics 109, Springer, New York.

Chaudhuri, P., Doksum, K. and Samarov, A. (1997). On average derivative quantile regression. *Ann. Statist.* 25, 715–744.

Dalalyan, A. S. and Kutoyants, Y. A. (2003). Asymptotically efficient estimation of the derivative of the invariant density. *Stat. Inference Stoch. Process.* 6, 89–107.

Devroye, L. (1992). A note on the usefulness of superkernels in density estimation. *Ann. Statist.* 20, 2037–2056.

Doksum, K. and Samarov, A. (1995). Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. *Ann. Statist.* 23, 1443–1473.

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

Dudewicz, E. J. and van der Meulen, E. C. (1981). Entropy-based tests of uniformity. *J. Amer. Statist. Assoc.* 76, 967–974.

Efromovich, S. and Samarov, A. (2000). Adaptive estimation of the integral of squared regression derivatives. *Scand. J. Statist.* 27, 335–351.

Eggermont, P. P. B. and LaRiccia, V. N. (1999). Best asymptotic normality of the kernel density entropy estimator for smooth densities. *IEEE Trans. Inform. Theory* 45, 1321–1326.

Fazal, S. S. (1977). On estimating the density function of error variables in regression. *Sankhyā Ser. A* 39, 378–386.

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

Goldstein, L. and Khas'minskii, R. (1995). On efficient estimation of smooth functionals. *Theory Probab. Appl.* 40, 151–156.

Goldstein, L. and Messer, K. (1992). Optimal plug-in estimators for nonparametric functional estimation. *Ann. Statist.* 20, 1306–1328.

Hall, P. and Marron, J. S. (1987). Estimation of integrated squared density derivatives. *Statist. Probab. Lett.* 6, 109–115.

Hall, P. and Marron, J. S. (1990). On variance estimation in nonparametric regression. *Biometrika* 77, 415–419.

Hallin, M. and Tran, L. T. (1996). Kernel density estimation for linear processes: Asymptotic normality and optimal bandwidth derivation. *Ann. Inst. Statist. Math.* 48, 429–449.

Hart, J. D. and Vieu, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data. *Ann. Statist.* 18, 873–890.

Honda, T. (2000). Nonparametric density estimation for a long-range dependent linear process. *Ann. Inst. Statist. Math.* 52, 599–611.

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

Koul, H. L. (1996). Asymptotics of some estimators and sequential empiricals in non-linear time series. *Ann. Statist.* 24, 380–404.

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

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

Kutoyants, Y. A. (1997a). Some problems of nonparametric estimation by observations of ergodic diffusion process. *Statist. Probab. Lett.* 32, 311–320.

Kutoyants, Y. A. (1997b). On unbiased density estimation for ergodic diffusion. *Statist. Probab. Lett.* 34, 133–140.

Kutoyants, Yu. A. (1998). On density estimation by the observations of ergodic diffusion processes. In: *Statistics and Control of Stochastic Processes* (Y. M. Kabanov, B. L. Rozovskii and A. N. Shiryaev, eds.), 253–274, World Scientific, Singapore.

Kutoyants, Yu. A. (1999). Efficient density estimation for ergodic diffusion processes. *Stat. Inference Stoch. Process.* 1, 131–155.

Kutoyants, Yu. A. (2004). *Statistical Inference for Ergodic Diffusion Processes.* Springer Series in Statistics, Springer, London.

Laurent, B. (1997). Estimation of integral functionals of a density and its derivatives. *Bernoulli* 3, 181–211.

Li, W. (1996). Asymptotic equivalence of estimators of average derivatives. *Econom. Lett.* 52, 241–245.

Li, Z. (1995). A study of nonparametric regression of error distribution in linear model based on L_1-norm. *Hiroshima Math. J.* 25, 171–205.

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

Loh, W.-L. (1997). Estimating the integral of squared regression function with Latin hypercube sampling. *Statist. Probab. Lett.* 31, 339–349.

Müller, U. U., Schick, A. and Wefelmeyer, W. (2004a). Estimating linear functionals of the error distribution in nonparametric regression. *Scand. J. Statist.* 31, 63–78.

Müller, U. U., Schick, A. and Wefelmeyer, W. (2004b). Estimating functionals of the error distribution in parametric and nonparametric regression. To appear in: *J. Nonparametr. Statist.*

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

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.

Samarov, A. (1991). On asymptotic efficiency of average derivative estimates. In: *Nonparametric Functional Estimation and Related Topics*, (G. G. Roussas, ed.), NATO ASI Series C 335, 167–172, Kluwer, Dordrecht.

Samarov, A. (1993). Exploring regression structure using nonparametric functional estimation. *J. Amer. Statist. Assoc.* 88, 836–847.

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

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

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. To appear in: *J. Nonparametr. Statist.*

Schick, A. and Wefelmeyer, W. (2004c). Root n consistent density estimators for invertible linear processes. Technical Report, Department of Mathematical Sciences, Binghamton University. http://math.binghamton.edu/anton/preprint.html.

Stoker, T. M. (1991). Equivalence of direct, indirect, and slope estimators of average derivatives. In: *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, (W. A. Barnett, J. Powell and G. Tauchen, eds.), International Symposia in Economic Theory and Econometrics, 99–118, Cambridge University Press.

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

Tsybakov, A. B. and van der Meulen, E. C. (1996). Root-n consistent estimators of entropy for densities with unbounded support. *Scand. J. Statist.* 23, 75–83.

Yakowitz, S. (1989). Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *J. Multivariate Anal.* 30, 124–136.

Yosida, K. (1980). *Functional Analysis*. 6th ed. Grundlehren der mathematischen Wissenschaften 123. Springer, Berlin.

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

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