Electronic Journal of Statistics
Vol. 17 (2023) 2880–2911
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2166

Efficient density estimation in an AR(1) model

Anton Schick

Department of Mathematics and Statistics, Binghamton University, Binghamton, NY 13902-6000, USA
e-mail: anton@math.binghamton.edu

Wolfgang Wefelmeyer

Abteilung Mathematik, Universität zu Köln, Weyertal 86-90, 50931 Köln, Germany
e-mail: wefelm@math.uni-koeln.de

Abstract: This paper studies a class of plug-in estimators of the stationary density of an autoregressive model with autoregression parameter $0 < \varrho < 1$. These use two types of estimator of the innovation density, a standard kernel estimator and a weighted kernel estimator with weights chosen to mimic the condition that the innovation density has mean zero. Bahadur expansions are obtained for this class of estimators in L_1, the space of integrable functions. These stochastic expansions establish root-n consistency in the L_1-norm. It is shown that the density estimators based on the weighted kernel estimators are asymptotically efficient if an asymptotically efficient estimator of the autoregression parameter is used. Here asymptotic efficiency is understood in the sense of the Hájek–Le Cam convolution theorem.

MSC2020 subject classifications: Primary 62G07, 62G20; secondary 62M10.
Keywords and phrases: Bahadur expansions in L_1, Hadamard differentiability, kernel density estimation.

Received March 2023.

1. Introduction

Consider observations $X_0, \dots, X_n$ from a stationary autoregressive process of order 1,

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

with unknown autoregression parameter ϱ in the open interval $(0, 1)$. The interval $(0, 1)$ is chosen for notational convenience. The following carries over to ϱ in the open interval $(-1, 0)$. The innovations ε_t, $t \in \mathbb{Z}$, are i.i.d. with a common density f, mean zero and finite variance σ^2, and $\{X_s, s \le t\}$ and $\{\varepsilon_r, r > t\}$ are independent. Then X_t has the infinite series representation

$$X_t = \varepsilon_t + \sum_{j=1}^{\infty} \varrho^j \varepsilon_{t-j}.$$

This is a semiparametric model with parameters ϱ and f. We are interested in estimating the stationary density g of the process. The usual density estimators based on the observations $X_0, \dots, X_n$ are generally developed for nonparametric Markov chains or more general time series models. They do not use the autoregressive structure of the data. See, for example, [4, 25, 7, 22, 3, 6, 8, 24, 2] and [20].

For the autoregressive process, the stationary density g satisfies the equation

$$g(x) = \int_{-\infty}^{\infty} f(x - \varrho y) g(y)\, dy, \quad x \in \mathbb{R}.$$

Thus a natural estimator of g is given by the plug-in estimator

$$\hat{g}_0(x) = \int_{-\infty}^{\infty} \hat{f}(x - \hat{\varrho} y) \hat{g}(y)\, dy, \quad x \in \mathbb{R},$$

with $\hat{\varrho}$ a root-n consistent estimator of ϱ, $\hat{f}$ an estimator of f based on the residuals $\hat{\varepsilon}_j = X_j - \hat{\varrho} X_{j-1}$, $j = 1, \dots, n$, and $\hat{g}$ a kernel estimator of g based on the observations $X_0, \dots, X_n$. We view our estimators as members of L_1, the set of measurable functions h from $\mathbb{R}$ to $\mathbb{R}$ with finite L_1-norm

$$\|h\|_1 = \int_{-\infty}^{\infty} |h(x)|\, dx.$$

It can be deduced from [21] that the plug-in estimator $\hat{g}_0$ is root-n consistent in L_1 under mild assumptions. Similar results for moving average processes are in [17, 18].

We can repeat the above plug-in procedure with $\hat{g}_0$ replacing $\hat{g}$. This leads to the estimator

$$\hat{g}_1(x) = \int_{-\infty}^{\infty} \hat{f}(x - \hat{\varrho} y) \hat{g}_0(y)\, dy, \quad x \in \mathbb{R}.$$

One expects the estimator $\hat{g}_1$ to be better than $\hat{g}_0$ as it uses a better initial estimator of g. Proceeding in this way one recursively defines new estimators

$$\hat{g}_{k+1}(x) = \int_{-\infty}^{\infty} \hat{f}(x - \hat{\varrho} y) \hat{g}_k(y)\, dy, \quad x \in \mathbb{R},$$

for positive integers k. It is easy to check that $\hat{g}_k$ has the representation

$$\hat{g}_k(x) = \int_{\mathbb{R}^{k+1}} \hat{f}\Big(x - \sum_{i=1}^{k} \hat{\varrho}^i y_i - \hat{\varrho}^{k+1} z\Big) \prod_{j=1}^{k} \hat{f}(y_j)\, dy_j\ \hat{g}(z)\, dz \tag{1.1}$$

for nonnegative k.

In this paper we study the estimator $\hat{g}_{k_n}$ where k_n is a sequence of integers that grow to infinity slowly. We derive root-n consistency of $\hat{g}_{k_n}$ in the L_1-norm through a Bahadur expansion, see Theorem 2.1. We establish Hadamard differentiability of the stationary density in Theorem 2.2. Using these results we show that for proper choices of $\hat f$, $\hat\varrho$ and k_n the estimator $\hat g_{k_n}$ is asymptotically efficient. By this we mean that $\int_{-\infty}^{\infty}\phi(x)\hat g_{k_n}(x)\,dx$ is a least dispersed regular estimator of $\int_{-\infty}^{\infty}\phi(x)g(x)\,dx$ for each bounded and measurable function ϕ in the sense of a semiparametric version of the Hájek–Le Cam convolution theorem.

2. Results

We study the asymptotic behavior of $\hat g_{k_n}$ under the following assumptions.

(A1) The density f has finite Fisher information for location.
(A2) The estimator $\hat\varrho$ satisfies the stochastic expansion

$$\hat\varrho = \varrho + \frac{1}{n}\sum_{j=1}^{n}\psi(X_{j-1},\varepsilon_j) + o_P(n^{-1/2})$$

for a function ψ satisfying $\int_{-\infty}^{\infty}\psi(x,y)f(y)\,dy = 0$ and

$$\Psi = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\psi^2(x,y)f(y)\,dy\,g(x)\,dx < \infty.$$

Recall that the density f has finite Fisher information for location if f is absolutely continuous and the integral

$$J_f = \int_{-\infty}^{\infty}\frac{(f'(x))^2}{f(x)}\,dx$$

is finite, where f' denotes the almost everywhere derivative of f. In this case we let $\ell_f = -f'/f$ denote the score function for location. Assumption (A1) implies that f' is integrable with L_1-norm $\|f'\|_1 = \|\ell_f f\|_1 \le J_f^{1/2}$. This allows the representation

$$f(x) = \int_{-\infty}^{x} f'(t)\,dt, \quad x\in\mathbb{R},$$

and shows that f is bounded by $\|f'\|_1$. Furthermore, the moment assumptions on f, assumption (A1) and an application of the Cauchy–Schwarz inequality show that the integral

$$\int_{-\infty}^{\infty}(1+|x|)|f'(x)|\,dx = \int_{-\infty}^{\infty}|\ell_f(x)|(1+|x|)f(x)\,dx$$

is finite.

It follows from (A2) that $n^{1/2}(\hat\varrho-\varrho)$ converges in distribution to a normal random variable with mean zero and variance Ψ. The sample autocorrelation coefficient

$$\frac{\frac{1}{n}\sum_{i=1}^{n}X_{i-1}X_i}{\frac{1}{n}\sum_{i=1}^{n}X_{i-1}^2}$$

meets this requirement with $\psi(x,y) = xy/E[X_0^2] = xy(1-\varrho^2)/\sigma^2$. An asymptotically efficient estimator of ρ is characterized by (A2) with

$$\psi(x,y) = \frac{x\ell_f(y)}{E[X_0^2]J_f} = \frac{x\ell_f(y)(1-\varrho^2)}{\sigma^2 J_f}.$$

Such an estimator was constructed in [11].

We shall work with two estimators of f. The first one is the usual kernel density estimator

$$\hat{f}_1(x) = \frac{1}{n}\sum_{j=1}^{n} K_b(x - \hat{\varepsilon}_j), \quad x \in \mathbb{R},$$

based on the residuals. Here $K_b(x) = (1/b)K(x/b)$ for a density K and a bandwidth b. For asymptotic efficiency, it is necessary to exploit that the innovation density f has mean zero. Our second estimator mimics the mean zero property using weights stemming from an empirical likelihood approach for mean zero observations,

$$\hat{f}_2(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{1+\hat{\lambda}\hat{\varepsilon}_j} K_b(x - \hat{\varepsilon}_j), \quad x \in \mathbb{R},$$

where $\hat{\lambda}$ is chosen such that $1+\hat{\lambda}\hat{\varepsilon}_1, \dots, 1+\hat{\lambda}\hat{\varepsilon}_n$ are positive and

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

on the event $\{\min_{1\le j\le n} \hat{\varepsilon}_j < 0 < \max_{1\le j\le n} \hat{\varepsilon}_j\}$ and is taken to be zero otherwise. The second estimator satisfies

$$\int_{-\infty}^{\infty} y\hat{f}_2(y)\,dy = 0$$

on this event and thus mimics that f has mean zero. Rates of convergence in the L_1-norm of these two estimators were derived in [13] in the more general setting of nonlinear autoregressive models. We shall improve these results for the present model in later sections. Both density estimators have the same rates of convergence in the L_1-norm, but the estimator $\hat{f}_2$ performs better as plug-in estimator for linear functionals of f. This was observed in [13] in the context of estimating the innovation distribution function and further exploited in [14] in the prediction for autoregressive models.

To state our first result we introduce some notation. We start with the random variables

$$\dot{X}_0 = \sum_{j=1}^{\infty} j\varrho^{j-1}\varepsilon_{-j} \quad \text{and} \quad Y_j = X_0 - \varrho^j\varepsilon_{-j} = \sum_{i=0}^{\infty} \mathbf{1}[i \ne j]\varrho^i\varepsilon_{-i}, \quad j \ge 0.$$

Let $\dot{g}$ denote the function defined by

$$\dot{g}(x) = E[-\dot{X}_0 f'(x - \rho X_{-1})], \quad x \in \mathbb{R}.$$

This function is integrable with L_1-norm

$$\|\dot g\|_1 \le \|f'\|_1 E[|\dot X_0|] \le \|f'\|_1 \frac{E[|\varepsilon_0|]}{(1-\varrho)^2} = \frac{\|f'\|_1 \|\iota_{\mathbb{R}} f\|_1}{(1-\varrho)^2}, \tag{2.1}$$

where $\iota_{\mathbb{R}}$ denotes the identity map on $\mathbb{R}$. For $j = 0, 1, 2, \dots,$ let γ_j denote the density of Y_j. Then we have the following representation of the stationary density,

$$g(x) = \int_{-\infty}^{\infty} \gamma_j(x - \varrho^j y) f(y)\, dy, \quad x \in \mathbb{R}, \tag{2.2}$$

for each such j. Now introduce functions γ and γ^* by

$$\gamma(x,y) = \sum_{j=0}^{\infty} \big(\gamma_j(x - \varrho^j y) - g(x)\big), \quad x, y \in \mathbb{R},$$

and

$$\gamma^*(x,y) = \gamma(x,y) - \int_{-\infty}^{\infty} \gamma(x,z) z f(z)\, dz \frac{y}{\sigma^2}, \quad x, y \in \mathbb{R}.$$

These functions satisfy the integrability conditions

$$\Big(\int_{\mathbb{R}^2} |\gamma(x,y)|\, dx f(y)\, dy\Big)^2 \le \pi \int_{\mathbb{R}^2} (1+x^2)|\gamma(x,y)|^2\, dx f(y)\, dy < \infty \tag{2.3}$$

and

$$\Big(\int_{\mathbb{R}^2} |\gamma^*(x,y)|\, dx f(y)\, dy\Big)^2 \le \pi \int_{\mathbb{R}^2} (1+x^2)|\gamma^*(x,y)|^2\, dx f(y)\, dy < \infty \tag{2.4}$$

as will be shown in Section 3. Finally we introduce the average

$$\bar\Psi_n = \frac{1}{n} \sum_{i=1}^{n} \psi(X_{i-1}, \varepsilon_i)$$

and assume that the kernel estimator $\hat g$ also uses the kernel K and the bandwidth b,

$$\hat g(x) = \frac{1}{n+1} \sum_{j=0}^{n} K_b(x - X_j), \quad x \in \mathbb{R}.$$

The kernel K will be taken to be a symmetric density satisfying some smoothness and integrability conditions as stated in the next theorem. We do not require K to have compact support.

Theorem 2.1. *Suppose* (A1) *and* (A2) *are met, the kernel K is a symmetric density with finite variance and is twice continuously differentiable with $\|(1+\iota_{\mathbb{R}}^2)K'\|_1$ and $\|(1+\iota_{\mathbb{R}}^2)(K'')^2\|_1$ finite, and the sequence k_n and the bandwidth $b = b_n$ are chosen to satisfy*

$$\frac{k_n}{\log(n)} \to \infty, \quad k_n^4 b_n^4 n \to 0 \quad \textit{and} \quad \frac{k_n^2}{n b_n^3} \to 0.$$

Then, for the choice $\hat{f} = \hat{f}_1$*, the estimator* $\hat{g}_{k_n}$ *satisfies the* L_1*-Bahadur expansion*

$$\int_{-\infty}^{\infty} \Big| \hat{g}_{k_n}(x) - g(x) - \bar{\Psi}_n \dot{g}(x) - \frac{1}{n}\sum_{i=1}^{n} \gamma(x, \varepsilon_i) \Big| \, dx = o_P(n^{-1/2})$$

while, for the choice $\hat{f} = \hat{f}_2$*, it satisfies the* L_1*-Bahadur expansion*

$$\int_{-\infty}^{\infty} \Big| \hat{g}_{k_n}(x) - g(x) - \bar{\Psi}_n \dot{g}(x) - \frac{1}{n}\sum_{j=1}^{n} \gamma^*(x, \varepsilon_i) \Big| \, dx = o_P(n^{-1/2}).$$

The proof of Theorem 2.1 is in Section 6. The assumptions on k_n and b_n are met by taking $k_n = (\log n)^\alpha$ for some $\alpha > 1$ and $b_n = n^{-\beta}$ for some β in the open interval $(1/4, 1/3)$. The standard normal density is a possible choice for K.

Inspecting the proof of Theorem 2.1 reveals that the theorem remains valid if we omit integration with respect to z in the formula (1.1), resulting in the estimator

$$\hat{p}(x) = \int_{\mathbb{R}^{k_n}} \hat{f}\Big(x - \sum_{i=1}^{k_n} \hat{\varrho}^i y_i\Big) \prod_{j=1}^{k_n} \hat{f}(y_j)\, dy_j, \quad x \in \mathbb{R}.$$

In view of the identity

$$\int_{-\infty}^{\infty} h(y)\hat{f}_1(y)\, dy = \frac{1}{n}\sum_{j=1}^{n} \int_{-\infty}^{\infty} h(\hat{\varepsilon}_j - u) K_b(u)\, du,$$

valid for any bounded measurable function h, this estimator with $\hat{f} = \hat{f}_1$ can be written as a V-statistic

$$\hat{p}(x) = \frac{1}{n^{k_n+1}} \sum_{j_0=1}^{n} \cdots \sum_{j_{k_n}=1}^{n} \mathbb{K}_n\Big(x - \sum_{i=0}^{k_n} \hat{\varrho}^i \hat{\varepsilon}_{j_i}\Big), \quad x \in \mathbb{R},$$

with $\mathbb{K}_n$ the convolution of the densities $K_b, K_{\hat{\varrho}b}, \dots, K_{\hat{\varrho}^{k_n}b}$. For the estimator $\hat{f} = \hat{f}_2$ we can write $\hat{p}(x)$ as a *weighted* V-statistic

$$\hat{p}(x) = \frac{1}{n^{k_n+1}} \sum_{j_0=1}^{n} \cdots \sum_{j_{k_n}=1}^{n} \frac{\mathbb{K}_n(x - \sum_{i=0}^{k_n} \hat{\varrho}^i \hat{\varepsilon}_{j_i})}{\prod_{l=0}^{k_n}(1 + \hat{\lambda}\hat{\varepsilon}_{j_l})}, \quad x \in \mathbb{R}.$$

If we take for K the standard normal density, then $\mathbb{K}_n$ equals the normal density with mean zero and variance $b^2 \sum_{i=0}^{k_n} \hat{\varrho}^{2i}$. This allows for a straightforward computation of the estimator $\hat{p}$ for both, $\hat{f}_1$ and $\hat{f}_2$.

It follows from the integrability conditions on γ and γ^* that the CLT in L_1 applies both to the L_1-valued random variables $Z_j = \gamma(\cdot, \varepsilon_j)$, $j = 1, 2, \dots,$ and to $Z_j^* = \gamma^*(\cdot, \varepsilon_j)$, $j = 1, 2 \dots,$ and yields that

$$n^{-1/2}\sum_{j=1}^{n} Z_j = n^{-1/2}\sum_{j=1}^{n} \gamma(\cdot, \varepsilon_j) \quad \text{and} \quad n^{-1/2}\sum_{j=1}^{n} Z_j^* = n^{-1/2}\sum_{j=1}^{n} \gamma^*(\cdot, \varepsilon_j)$$

converge in distribution to centered Gaussian processes. Indeed, as shown in Lemma 3 of [19], the integrability conditions imply the necessary and sufficient conditions of the CLT in L_1; see [12], Theorem 10.10, or [23], p. 92.

Our next result gives Hadamard differentiability of the stationary density which will be crucial in the characterization of asymptotically efficient estimators of g in L_1. For this we write $g_{\varrho,f}$ for g to stress the dependence of g on the parameters ϱ and f. Let $\mathcal{H}$ denote the set of all measurable functions h which satisfy

$$\int_{-\infty}^{\infty} h(y)f(y)\,dy = 0, \quad \int_{-\infty}^{\infty} yh(y)f(y)\,dy = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} h^2(y)f(y)\,dy < \infty.$$

This set is the tangent space at f of the set $\mathcal{F}$ of all densities with mean zero, finite variance and finite Fisher information. Indeed, one can show that for each h in $\mathcal{H}$ there is a sequence f_n of densities with finite Fisher information satisfying

$$\int_{-\infty}^{\infty} \Big(n^{1/2}\big(\sqrt{f_n(x)} - \sqrt{f(x)}\big) - \frac{1}{2}h(x)\sqrt{f(x)}\Big)^2\,dx \to 0, \tag{2.5}$$

$$\int_{-\infty}^{\infty} x f_n(x)\,dx = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} x^2\big|f_n(x) - f(x)\big|\,dx \to 0. \tag{2.6}$$

The proof of the next theorem, establishing Hadamard differentiability, is in Section 7.

Theorem 2.2. *Suppose f has finite Fisher information for location. Let h belong to $\mathcal{H}$. Let f_n be a sequence of densities satisfying (2.5) and (2.6) and let ϱ_n be a sequence in $(0,1)$ satisfying $n^{1/2}(\varrho_n - \varrho) \to t$ for some real t. Then g_{ϱ_n,f_n} satisfies*

$$\int_{-\infty}^{\infty} \big|n^{1/2}\big(g_{\varrho_n,f_n}(x) - g_{\varrho,f}(x)\big) - Ah(x) - \dot{g}(x)t\big|\,dx = o_P(n^{-1/2})$$

with

$$Ah(x) = \int_{-\infty}^{\infty} \gamma^*(x,y)h(y)f(y)\,dy, \quad x \in \mathbb{R}.$$

We will now show that the density estimator $\hat{g}_{k_n}$ is asymptotically efficient for g if we use $\hat{f}_2$ and an asymptotically efficient estimator $\hat{\varrho}$ for ϱ. More specifically we consider linear functionals of the density g of the form $\Phi(g) = \int_{-\infty}^{\infty} \phi(y)g(y)\,dy$ with ϕ bounded and measurable, i.e., $\phi \in L_\infty$, and show that $\Phi(\hat{g}_{k_n})$ is asymptotically efficient for the functional $\Phi(g)$. We follow efficiency proofs for other models and functionals and will be brief. See [11, 5, 10], and [16]. Write $P_{n\varrho f}$ for the joint distribution of $(X_0, \dots, X_n)$ when ϱ and f are true. Choose ϱ_n and f_n close to ϱ and f as in Theorem 2.2. Under the above assumptions it follows from [10] that the *local log-likelihood ratio* admits the stochastic expansion

$$\log \frac{dP_{n\varrho_n f_n}}{dP_{n\varrho f}}(X_0, \dots, X_n) = n^{-1/2} \sum_{j=1}^{n} \big[tX_{j-1}\ell_f(\varepsilon_j) + h(\varepsilon_j)\big] - \frac{1}{2}\Lambda(t,h) + o_P(1)$$

with $\Lambda(t,h) = E[(tX_0\ell_f(\varepsilon_1) + h(\varepsilon_1))^2]$.

In other words, the model is *locally asymptotically normal* (LAN) with *central sequence* $n^{-1/2}\sum_{j=1}^n tX_{j-1}\ell_f(\varepsilon_j) + n^{-1/2}\sum_{j=1}^n h(\varepsilon_j)$. It follows from Theorem 2.1 and the above characterization of asymptotically efficient estimators for ϱ that

$$n^{1/2}(\Phi(\hat{g}_{k_n}) - \Phi(g))$$

is approximated stochastically by an expression of the form of a central sequence. This implies that $\Phi(\hat{g}_{k_n})$ is asymptotically efficient for $\Phi(g)$ (and also regular and asymptotically linear).

This asymptotic efficiency result is an instance of the plug-in principle formulated in [9] in the i.i.d. case. In order to see this, fix ϱ, write $\varepsilon_j(\varrho) = \varepsilon_j = X_j - \varrho X_{j-1}$, and set

$$\tilde{f}_2(x,\varrho) = \frac{1}{n}\sum_{j=1}^n \frac{1}{1+\tilde{\lambda}(\varrho)\varepsilon_j(\varrho)} K_b(x - \varepsilon_j(\varrho)), \quad x \in \mathbb{R},$$

where $\tilde{\lambda}(\varrho)$ is chosen such that $1+\tilde{\lambda}(\varrho)\varepsilon_1(\varrho), \dots, 1+\tilde{\lambda}(\varrho)\varepsilon_n(\varrho)$ are positive and

$$\frac{1}{n}\sum_{j=1}^n \frac{\varepsilon_j(\varrho)}{1+\tilde{\lambda}(\varrho)\varepsilon_j(\varrho)} = 0$$

on the event $\{\min_{1\le j\le n}\varepsilon_j(\varrho) < 0 < \max_{1\le j\le n}\varepsilon_j(\varrho)\}$ and is taken to be zero otherwise. Set

$$\tilde{g}_k(x,\varrho) = \int_{\mathbb{R}^{k+1}} \tilde{f}_2\Big(x - \sum_{i=1}^k \varrho^i y_i - \varrho^{k+1} z, \varrho\Big) \prod_{j=1}^k \tilde{f}_2(y_j,\varrho)\,dy_j\; \hat{g}(z)\,dz, \quad x \in \mathbb{R}.$$

Then for k_n as in Theorem 2.1 the estimator $\Phi(\tilde{g}_{k_n}(\cdot,\varrho)) = \int_{-\infty}^{\infty}\phi(x)\tilde{g}_{k_n}(x,\varrho)\,dx$ is asymptotically efficient for $\Phi(g) = \int_{-\infty}^{\infty}\phi(y)g(y)\,dy$ when ϱ is fixed. Plugging in an asymptotically efficient estimator $\hat{\varrho}$ for ϱ, we obtain an asymptotically efficient estimator

$$\Phi(\tilde{g}_{k_n}(\cdot,\hat{\varrho})) = \int_{-\infty}^{\infty}\phi(x)\tilde{g}_{k_n}(x,\hat{\varrho})\,dx = \int_{-\infty}^{\infty}\phi(x)\hat{g}_{k_n}(x)\,dx$$

when ϱ is unknown.

3. Some auxiliary lemmas

In this section we collect some lemmas that will be used in the proofs of our theorems. We start with three inequalities.

Lemma 3.1. *For numbers r and s in the interval $(0,1)$, we have the inequalities*

$$\sum_{j=1}^{\infty} |r^j - s^j| \le \frac{|r-s|}{(1-\max\{r,s\})^2}, \tag{3.1}$$

$$\sum_{j=1}^{\infty} |r^j - s^j|^2 \le \frac{|r-s|^2}{(1-\max\{r,s\})^3}, \tag{3.2}$$

$$\sum_{j=1}^{\infty} |r^j - s^j - js^{j-1}(r-s)| \le \frac{|r-s|^2}{(1-\max\{r,s\})^3}. \tag{3.3}$$

Proof. Recall the infinite series

$$\sum_{j=1}^{\infty} jt^{j-1} = \frac{1}{(1-t)^2} \quad \text{and} \quad \sum_{j=1}^{\infty} j(j-1)t^{j-2} = \frac{2}{(1-t)^3}, \qquad |t| < 1.$$

Using the inequality $|r^j - s^j| \le |r-s| j \max\{r,s\}^{j-1}$ and the first infinite series, we obtain (3.1). Using $|r^j - s^j|^2 \le \frac{1}{2}(r-s)^2 2j(2j-1)\max\{r,s\}^{2j-2}$ and the second infinite series, we obtain (3.2). Using

$$|r^j - s^j - js^{j-1}(r-s)| \le \frac{1}{2}(r-s)^2 j(j-1)\max\{r,s\}^{j-2}$$

and the second infinite series, we obtain (3.3). □

Lemma 3.2. *Let h be a measurable function. Then we have the inequality*

$$\|h\|_1^2 \le \int_{-\infty}^{\infty} \pi(1+x^2)h^2(x)\,dx.$$

Proof. Let us set $w(x) = \pi(1+x^2)$, $x \in \mathbb{R}$. Then $1/w$ is the Cauchy density. We calculate

$$\|h\|_1^2 = \|\sqrt{w}h/\sqrt{w}\|_1^2 \le \|wh^2\|_1 \|1/w\|_1 = \|wh^2\|_1$$

which is the desired result. □

Lemma 3.3. *Let p and q be two integrable functions with $\|\iota_{\mathbb{R}}^2 p\|_1$ and $\|\iota_{\mathbb{R}}^2 q\|_1$ finite. Then the inequality $\|\iota_{\mathbb{R}} p - \iota_{\mathbb{R}} q\|_1^2 \le (\|\iota_{\mathbb{R}}^2 p\|_1 + \|\iota_{\mathbb{R}}^2 q\|_1)\|p-q\|_1$ holds.*

Proof. Bound $|\iota_{\mathbb{R}} p - \iota_{\mathbb{R}} q|$ by $|p-q|^{1/2}(|p|+|q|)^{1/2}|\iota_{\mathbb{R}}|$ and then use the Cauchy–Schwarz inequality. □

Let h be an integrable function. Then the L_1-*modulus of continuity* of h is the map w_h defined by

$$w_h(t) = \sup_{|u| \le t} \int_{-\infty}^{\infty} |h(x-u) - h(x)|\,dx, \quad t \ge 0.$$

The map w_h is bounded by $2\|h\|_1$ and continuous at 0; see [15], Theorem 9.5, for the latter. We say h is L_1-*Lipschitz* if there is a constant Λ such that

$$\int_{-\infty}^{\infty} |h(x-t) - h(x)|\,dx \le \Lambda |t|, \quad t \in \mathbb{R}.$$

In this case the inequality $w_h(t) \le \Lambda t$ holds for all $t \ge 0$.

Lemma 3.4. *Let h be an integrable function and $T, T_1, T_2, \dots$ be random variables such that $E[|T_n - T|] \to 0$. Then*

$$\int_{-\infty}^{\infty} \big|E[h(x - T_n)] - E[h(x - T)]\big|\, dx \to 0.$$

Proof. In view of the inequality $|E[X]| \le E[|X|]$ and Fubini's theorem, the integral is bounded by $E[w_h(|T_n - T|)]$, and the desired result follows from the dominated convergence theorem. □

Let $\mathbb{H}_1$ denote the set of all integrable functions of the form

$$h(x) = \int_{-\infty}^{x} h'(u)\, du$$

for some integrable function h' and let $\mathbb{H}_2$ denote set of all h in $\mathbb{H}_1$ with h' in $\mathbb{H}_1$. We write h'' for $(h')'$. If h belongs to $\mathbb{H}_1$, then h is bounded by $\|h'\|_1$ and uniformly continuous. More precisely, we have

$$|h(y) - h(x)| = \Big| \int_{-\infty}^{y} h'(s)\, ds - \int_{-\infty}^{y} h'(s - (y - x))\, ds \Big| \le w_{h'}(|y - x|)$$

for all real x and y. As an integrable and uniformly continuous function, $h(x) \to 0$ as $x \to \infty$. This implies that h' integrates to zero,

$$\int_{-\infty}^{\infty} h'(x)\, dx = 0,$$

and this gives the alternative representation

$$h(x) = -\int_{x}^{\infty} h'(u)\, du.$$

Using this we can show that $\|h\|_1 \le \|\iota_{\mathbb{R}} h'\|_1$. Indeed, the left-hand side is bounded by

$$\int_{-\infty}^{0} \int_{-\infty}^{x} |h'(t)|\, dt\, dx + \int_{0}^{\infty} \int_{x}^{\infty} |h'(t)|\, dt\, dx \le \int_{-\infty}^{0} |th'(t)|\, dt + \int_{0}^{\infty} |th'(t)|\, dt.$$

In view of this inequality, we conclude that a continuously differentiable function h belongs to $\mathbb{H}_1$ if $\lim_{|x|\to\infty} h(x) = 0$ and $\|(1 + |\iota_{\mathbb{R}}|)h'\|_1$ is finite.

Assumption (A1) implies that the density f belongs to $\mathbb{H}_1$. The next two results are easily verified.

Lemma 3.5. *Let h belong to $\mathbb{H}_1$ and let t be a positive number. Then the function h_t defined by*

$$h_t(x) = h(x/t)/t, \quad x \in \mathbb{R},$$

belongs to $\mathbb{H}_1$, and we can take

$$h_t'(x) = h'(x/t)/t^2, \quad x \in \mathbb{R}.$$

Thus $\|h_t'\|_1 = \|h'\|_1/t$.

Lemma 3.6. *Let $h = h_1 * h_2$ denote the convolution of the integrable functions h_1 and h_2. Then the following are true.*

1. *If h_1 is L_1-Lipschitz with constant Λ, then h is L_1-Lipschitz with constant $\Lambda\|h_2\|_1$.*
2. *If h_1 belongs to $\mathbb{H}_1$, then h belongs to $\mathbb{H}_1$ with $h' = h_1' * h_2$.*
3. *If h_1 and h_2 belong to $\mathbb{H}_1$, then h belongs to $\mathbb{H}_2$ with $h'' = h_1' * h_2'$.*

Lemma 3.7. *Let h belong to $\mathbb{H}_1$. Then h is L_1-Lipschitz with constant $\|h'\|_1$. Moreover, we have the inequality*

$$\int_{-\infty}^{\infty} \big|h(x-t) - h(x) + th'(x)\big|\, dx \le |t| w_{h'}(|t|), \quad t \in \mathbb{R}.$$

In particular, if h' is L_1-Lipschitz with constant Λ, then we have

$$\int_{-\infty}^{\infty} \big|h(x-t) - h(x) + th'(x)\big|\, dx \le t^2\Lambda, \quad t \in \mathbb{R}.$$

Proof. Fix $t \in \mathbb{R}$. Then we have the identity

$$h(x-t) - h(x) = -t \int_0^1 h'(x - st)\, ds$$

and consequently the bounds

$$\int_{-\infty}^{\infty} |h(x-t) - h(x)|\, dx \le |t| \int_{-\infty}^{\infty} \int_0^1 |h'(x-st)|\, ds\, dx = |t|\|h'\|_1$$

and

$$\int_{-\infty}^{\infty} \big|h(x-t) - h(x) + th'(x)\big|\, dx \le \int_{-\infty}^{\infty} |t| \int_0^1 |h'(x-st) - h'(x)|\, ds\, dx$$
$$\le |t| w_{h'}(|t|).$$

If h' is L_1-Lipschitz with constant Λ, then $|t|w_{h'}(|t|) \le \Lambda t^2$. □

Lemma 3.8. *Let h belong to $\mathbb{H}_1$ and let T, U and V be random variables. If T and U have finite means, then we have the inequalities*

$$\int_{-\infty}^{\infty} \big|E[h(x-V-T)] - E[h(x-V)]\big|\, dx \le \|h'\|_1 E[|T|]$$

and

$$\int_{-\infty}^{\infty} \big|E[h(x-V-T)] - E[h(x-V)] + E[Uh'(x-V)]\big|\, dx$$
$$\le E[|T|w_{h'}(|T|)] + \|h'\|_1 E[|T-U|]$$
$$\le E[|U|w_{h'}(|T|)] + 3\|h'\|_1 E[|T-U|].$$

Proof. Using the formula $|E[X]| \le E[|X|]$, Fubini's theorem and then the substitution $u = x - V$ we obtain that the left-hand side of the first inequality is bounded by

$$E\int_{-\infty}^{\infty} |h(u-T) - h(u)|\,du,$$

and of the second inequality by

$$E\int_{-\infty}^{\infty} \big|h(u-T) - h(u) + Th'(u) - (T-U)h'(u)\big|\,du.$$

The desired result then follows from the previous lemma and the fact that $w_{h'}$ is bounded by $2\|h'\|_1$. □

Corollary 3.1. *Let f be in $\mathbb{H}_1$. Then, as $r \to \varrho$,*

$$\big\|g_{r,f} - g_{\varrho,f} - (r-\varrho)\dot{g}\big\|_1 = o(|r-\varrho|). \tag{3.4}$$

Proof. For $0 < r < 1$, let $S_r = \sum_{j=1}^{\infty} r^j \varepsilon_{-j}$. Recall $\dot{X}_0 = \sum_{j=1}^{\infty} j\varrho^{j-1}\varepsilon_{-j}$. It follows from Lemma 3.1 that

$$E[|S_r - S_\varrho|] \le \frac{|r-\varrho|\|\iota_{\mathbb{R}} f\|_1}{(1-\max\{r,\varrho\})^2}$$

and

$$E\big[|S_r - S_\varrho - (r-\varrho)\dot{X}_0|\big] \le \frac{|r-\varrho|^2\|\iota_{\mathbb{R}} f\|_1}{(1-\max\{r,\varrho\})^3}.$$

Note the identity

$$g_{r,f}(x) = E[f(x - S_r)].$$

Applying Lemma 3.8 with $h = f$, $V = S_\varrho = \varrho X_{-1}$, $T = S_r - S_\varrho$ and $U = (r-\varrho)\dot{X}_0$ shows that the left-hand side of (3.4) is bounded by

$$|r-\varrho| E\big[|\dot{X}_0| w_{f'}(|S_r - S_\varrho|)\big] + 3\|f'\|_1\|\iota_{\mathbb{R}} f\|_1 \frac{|r-\varrho|^2}{(1-\max\{r,\varrho\})^3}.$$

The desired result now follows from the dominated convergence theorem. □

For integrable functions p and q and $t > 0$, we denote by $B_t(p,q)$ the integrable function defined by

$$B_t(p,q)(x) = \int_{-\infty}^{\infty} p(x-ty)q(y)\,dy, \quad x \in \mathbb{R}.$$

The integrability of $B_t(p,q)$ follows from the inequality

$$\|B_t(p,q)\|_1 \le \|p\|_1\|q\|_1.$$

We can view B_t as a bilinear operator from $L_1 \times L_1$ to L_1. Since $B_t(p,q)$ is the convolution of p and q_t, where $q_t(x) = q(x/t)/t$, $x \in \mathbb{R}$, the following lemma is an immediate consequence of Lemmas 3.5 and 3.6.

Lemma 3.9. *Let p and q be integrable functions and t be a positive number. Then the following hold.*

1. *If p is L_1-Lipschitz with constant Λ, then $B_t(p,q)$ is L_1-Lipschitz with constant $\Lambda\|q\|_1$.*
2. *If p belongs to $\mathbb{H}_1$, then $B_t(p,q)$ belongs to $\mathbb{H}_1$ with $B_t(p,q)' = B_t(p',q)$.*
3. *If q belongs to $\mathbb{H}_1$, then $B_t(p,q)$ belongs to $\mathbb{H}_1$ with $B_t(p,q)' = B_t(p,q')/t$.*
4. *If p and q belong to $\mathbb{H}_1$, then $B_t(p,q)$ belongs to $\mathbb{H}_2$ with $B_t(p,q)'' = B_t(p',q')/t$.*

The next two results are consequences of Lemmas 3.7 and 3.8.

Lemma 3.10. *Let p and q be integrable functions with $\|\iota_{\mathbb{R}}q\|_1$ finite and p being L_1-Lipschitz with constant Λ. Then we have the inequality*

$$\left\|B_t(p,q) - p\int_{-\infty}^{\infty} q(y)\,dy\right\|_1 \le \Lambda\|\iota_{\mathbb{R}}q\|_1 t, \quad t > 0.$$

In particular, if the integral of q is zero, we have

$$\|B_t(p,q)\|_1 \le \Lambda\|\iota_{\mathbb{R}}q\|_1 t, \quad t > 0.$$

Lemma 3.11. *Let p belong to $\mathbb{H}_1$ and let q be a density. If q has finite mean, then we have the inequality*

$$\|B_t(p,q) - p\|_1 \le \|p'\|_1\|\iota_{\mathbb{R}}q\|_1 t, \quad t > 0.$$

If p' is L_1-Lipschitz with constant Λ and q has mean zero and finite variance, then we have the inequality

$$\|B_t(p,q) - p\|_1 \le \Lambda\|\iota_{\mathbb{R}}^2 q\|_1 t^2, \quad t > 0.$$

Let v be the function defined by

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

This function satisfies the inequality

$$v(x+y) \le v(x)v(y), \quad x, y \in \mathbb{R}.$$

If vh is integrable, then we have

$$\int_{-\infty}^{\infty} v(x)|h(x-t)|\,dx = \int_{-\infty}^{\infty} v(x+t)|h(x)|\,dx \le v(t)\|vh\|_1, \quad t \in \mathbb{R}.$$

From this we immediately derive the following result.

Lemma 3.12. *Let vp and vq be integrable. Then, for every $0 < t \le 1$, $vB_t(p,q)$ is integrable with*

$$\|vB_t(p,q)\|_1 \le \|vp\|_1 \int_{-\infty}^{\infty} v(ty)|q(y)|\,dy \le \|vp\|_1\|vq\|_1.$$

Lemma 3.13. *Let h belong to $\mathbb{H}_1$ with vh' integrable. Then $\|vh\|_\infty \le \|vh'\|_1$.*

Proof. For negative x we have the bound

$$v(x)|h(x)| \le v(x)\int_{-\infty}^{x} |h'(u)|\,du \le \int_{-\infty}^{x} v(u)|h'(u)|\,du,$$

while for positive x we have the inequality

$$v(x)|h(x)| \le v(x)\int_{x}^{\infty} |h'(u)|\,du \le \int_{x}^{\infty} v(u)|h'(u)|\,du.$$

These inequalities imply $\|vh\|_\infty \le \|vh'\|_1$. □

We have seen that (A1) implies that $\|vf'\|_1$ is finite. The stationary density g equals $B_\varrho(f,g)$ and therefore belongs to $\mathbb{H}_2$ with $g' = B_\varrho(f',g)$ and $g'' = B_\varrho(f',g')/\varrho$, yielding

$$\|g'\|_1 \le \|f'\|_1, \quad \|vg'\|_1 \le \|vf'\|_1\|vg\|_1 \quad \text{and} \quad \|g''\|_1 \le \|f'\|_1^2/\varrho.$$

Recall that γ_j denotes the density of $Y_j = X_0 - \varrho^j\varepsilon_{-j} = \sum_{i=0}^\infty \mathbf{1}[i \ne j]\varrho^i\varepsilon_{-i}$ for nonnegative integers j. Since Y_0 equals ϱX_{-1}, we have $\gamma_0(x) = g(x/\varrho)/\varrho$. Thus the density γ_0 belongs to $\mathbb{H}_2$ with $\gamma_0' = g'(x/\varrho)/\varrho^2$ and $\gamma_0''(x) = g''(x/\varrho)/\varrho^3$ yielding the bounds

$$\|\gamma_0'\|_1 \le \|f'\|_1/\varrho, \quad \|v\gamma_0'\|_1 \le \|vf'\|_1\|vg\|_1/\varrho \quad \text{and} \quad \|\gamma_0''\|_1 \le \|f'\|_1^2/\varrho^3.$$

For $j \ge 1$, the density γ_j equals $B_\varrho(f,\gamma_{j-1})$ as is easily checked and thus belongs to $\mathbb{H}_2$ with $\gamma_j' = B_\varrho(f',\gamma_{j-1})$ and $\gamma_j'' = B_\varrho(f',\gamma_{j-1}')/\varrho$ giving the bounds

$$\|\gamma_j'\|_1 \le \|f'\|_1, \quad \|v\gamma_j'\|_1 \le \|vf'\|_1\|v\gamma_{j-1}\|_1 \quad \text{and} \quad \|\gamma_j''\|_1 \le \|f'\|_1\|\gamma_{j-1}'\|_1/\varrho.$$

Note also the bounds $\|v\gamma_j\|_1 = E[1+|Y_j|] \le 1 + \sum_{i=0}^\infty \varrho^i\|\iota_{\mathbb{R}} f\| \le \|vf\|_1/(1-\rho)$ and similarly $\|vg\|_1 \le \|vf\|_1/(1-\varrho)$. Consequently, we have the following result.

Corollary 3.2. *Suppose (A1) holds. Then there are constants C_1, C_2 and C_3 such that the inequalities $\|\gamma_j'\|_1 \le C_1$, $\|v\gamma_j'\|_1 \le C_2$ and $\|\gamma_j''\|_1 \le C_3$ hold for all $j \ge 0$.*

Let us now verify the integrability condition (2.3). The first inequality follows from the moment inequality and Lemma 3.2. The finiteness of the integral

$$I_2 = \int_{\mathbb{R}^2} (1+x^2)\gamma^2(x,y)f(y)\,dy\,dx$$

follows if we verify the inequality

$$\tau_j = \int_{\mathbb{R}^2} (1+x^2)(\gamma_j(x-\varrho^j y) - g(x))^2 f(y)\,dx\,dy \le M\varrho^j, \quad j \ge 0,$$

for some finite constant M. Indeed, we first bound I_2 by

$$\sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\int_{\mathbb{R}^2}(1+x^2)|\gamma_i(x-\varrho^i y)-g(x)||\gamma_j(x-\varrho^j y)-g(x)|f(y)\,dy\,dx$$

and then use the Cauchy–Schwarz inequality to obtain

$$I_2 \le \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\sqrt{\tau_i\tau_j} = \Big(\sum_{i=0}^{\infty}\sqrt{\tau_i}\Big)^2 \le \frac{M}{(1-\sqrt{\varrho})^2}.$$

The formula (2.2) yields the identity

$$\int_{-\infty}^{\infty}(\gamma_j(x-\varrho^j y)-g(x))^2 f(y)\,dy = \int_{-\infty}^{\infty}\gamma_j^2(x-\varrho^j y)f(y)\,dy - g^2(x), \quad x\in\mathbb{R}.$$

Using the substitution $x=u+\varrho^j y$ and the fact that f has mean zero and finite variance σ^2, we calculate

$$\int_{\mathbb{R}^2}(1+x^2)\gamma_j^2(x-\varrho^j y)f(y)\,dy\,dx = \int_{-\infty}^{\infty}(1+u^2+\varrho^{2j}\sigma^2)\gamma_j^2(u)\,du$$

and then, utilizing the inequality $1+x^2\le(1+|x|)^2=v^2(x)$,

$$\begin{aligned}\tau_j &= \int_{-\infty}^{\infty}(1+x^2)(\gamma_j^2(x)-g^2(x))\,dx + \varrho^{2j}\sigma^2\int\gamma_j^2(x)\,dx\\ &\le (\|v\gamma_j\|_\infty+\|vg\|_\infty)\|v(\gamma_j-g)\|_1+\varrho^{2j}\sigma^2\|\gamma_j\|_\infty.\end{aligned}$$

Next we use the inequalities $\|\gamma_j\|_\infty\le\|v\gamma_j\|_\infty\le\|v\gamma_j'\|_1\le C_2$ and

$$\begin{aligned}\|v(g-\gamma_j)\|_1 &= \int_{-\infty}^{\infty}v(x)\Big|\int_{-\infty}^{\infty}\int_0^1\gamma_j'(x-s\varrho^j y)\varrho^j y\,ds f(y)\,dy\Big|\,dx\\ &\le \int_0^1\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}v(x-s\varrho^j y)|\gamma_j'(x-s\varrho^j y)|\,dx\,v(y)\varrho^j|y|f(y)\,dy\,ds\\ &\le \|v\gamma_j'\|_1\varrho^j\|v^2 f\|_1\end{aligned}$$

to conclude the exponential decay $\tau_j\le M\varrho^j$.

Let us now verify the integrability condition (2.4). The first inequality follows from the moment inequality and Lemma 3.2. The finiteness of the second integral follows from (2.3) and the inequality

$$\begin{aligned}\int(\gamma^*(x,y))^2 f(y)\,dy &= \int\gamma^2(x,y)f(y)\,dy - \Big(\int\gamma(x,y)yf(y)\,dy\Big)^2/\sigma^2\\ &\le \int\gamma^2(x,y)f(y)\,dy.\end{aligned}$$

4. Behavior of the innovation density estimators

Let $\tilde{f}$ denote the kernel density estimator

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

based on the actual innovations, and let

$$\bar{f}(x) = \int_{-\infty}^{\infty} K_b(x-y)f(y)\,dy = \int_{-\infty}^{\infty} f(x-bu)K(u)\,du, \quad x \in \mathbb{R},$$

denote its expectation. We have

$$\|\tilde{f} - \bar{f}\|_1 = O_P((nb)^{-1/2}). \tag{4.1}$$

Indeed we calculate, using Lemma 3.2 and the substitution $x = y + bu$,

$$\begin{aligned} nbE[\|\tilde{f} - \bar{f}\|_1^2] &\le \pi \int_{-\infty}^{\infty} (1+x^2)nbE[(\tilde{f}(x) - \bar{f}(x))^2]\,dx \\ &\le \pi \int_{-\infty}^{\infty} (1+x^2) \int_{-\infty}^{\infty} bK_b^2(x-y)f(y)\,dy\,dx \\ &= \pi \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (1+(y+bu)^2)f(y)K^2(u)\,dy\,du \\ &= \pi(1+\sigma^2)\|K^2\|_1 + \pi b^2\|(\iota_{\mathbb{R}}K)^2\|_1. \end{aligned}$$

The following result was proved in [13] under stronger assumptions. In particular, we remove the assumption that the error density has a finite moment of order 5/2.

Theorem 4.1. *Suppose the bandwidth $b = b_n$ satisfies $nb_n^4 \to 0$ and $nb_n^3 \to \infty$, the kernel K is a symmetric density with finite variance and is continuously differentiable with a bounded derivative K' with $\int_{-\infty}^{\infty}(1+u^2)|K'(u)|\,du$ finite, the estimator $\hat{\varrho}$ is root-n consistent, i.e., $\sqrt{n}(\hat{\varrho} - \varrho) = O_P(1)$, and the density f is L_1-Lipschitz with constant Λ. Then we have the stochastic rates*

$$\int_{-\infty}^{\infty} \big|\hat{f}_1(x) - \tilde{f}(x)\big|\,dx = O_P\Big(\frac{1}{nb^{3/2}}\Big)$$

and

$$\int_{-\infty}^{\infty} \big|\hat{f}_2(x) - \tilde{f}(x) + xf(x)\hat{\lambda}\big|\,dx = O_P\Big(\frac{1}{nb^{3/2}}\Big).$$

Proof. The assumptions on K imply that K belongs to $\mathbb{H}_1$. Thus Lemma 3.9 implies that $\bar{f} = B_b(f,K)$ belongs to $\mathbb{H}_1$ and $\bar{f}' = B_b(f,K')/b$ is L_1-Lipschitz with constant $\Lambda\|K'\|_1/b$. It follows from Lemma 3.10 and $\int_{-\infty}^{\infty} K'(u)\,du = 0$ that $\|\bar{f}'\|_1 \le \Lambda\|\iota_{\mathbb{R}}K'\|_1$.

The residuals are of the form

$$\hat\varepsilon_j = X_j - \hat\varrho X_{j-1} = \varepsilon_j - (\hat\varrho - \varrho)X_{j-1}, \quad j = 1,\dots,n.$$

This representation, the root-n consistency of $\hat\varrho$ and the stochastic rates

$$\frac{1}{n}\sum_{j=1}^n X_{j-1} = O_P(n^{-1/2}) \quad \text{and} \quad \frac{1}{n}\sum_{j=1}^n \varepsilon_j X_{j-1} = O_P(n^{-1/2})$$

yield

$$\frac{1}{n}\sum_{j=1}^n \hat\varepsilon_j = \frac{1}{n}\sum_{j=1}^n \varepsilon_j + O_P(1/n)$$

and

$$\frac{1}{n}\sum_{j=1}^n \hat\varepsilon_j^2 = \frac{1}{n}\sum_{j=1}^n \varepsilon_j^2 + O_P(1/n).$$

We also need the following results already established in [13],

$$\hat\lambda = \frac{1}{n}\sum_{j=1}^n \frac{\varepsilon_j}{\sigma^2} + o_P(n^{-1/2}) = O_P(n^{-1/2}), \tag{4.2}$$

$$\max_{1\le j\le n} |\hat\varepsilon_j| = o_P(n^{1/2}). \tag{4.3}$$

A consequence of the above is the stochastic rate

$$\frac{1}{n}\sum_{j=1}^n \frac{\hat\lambda^2\hat\varepsilon_j^2}{1+\hat\lambda\hat\varepsilon_j} = O_P(n^{-1}) \tag{4.4}$$

which plays a role in comparing $\hat f_2$ with $\hat f_1$.

We shall establish the following stochastic rates.

$$\int_{-\infty}^{\infty} \big|\hat f_2(x) - \hat f_1(x) + \hat\lambda x \hat f_1(x)\big|\,dx = O_P(bn^{-1/2}), \tag{4.5}$$

$$\|\hat f_1 - \tilde f\|_1 = O_P(1/(nb^{3/2})), \tag{4.6}$$

$$\int_{-\infty}^{\infty} |x|\big|\hat f_1(x) - f(x)\big|\,dx = O_P(b^{1/2}). \tag{4.7}$$

These stochastic rates together with $\hat\lambda = O_P(n^{-1/2})$ and $nb^4 \to 0$ imply the desired results.

To verify (4.5), we write

$$\begin{aligned}\hat f_2(x) - \hat f_1(x) + \hat\lambda x \hat f_1(x) &= \frac{1}{n}\sum_{j=1}^n K_b(x-\hat\varepsilon_j)\Big(\frac{1}{1+\hat\lambda\hat\varepsilon_j} - 1 + \hat\lambda\hat\varepsilon_j + \hat\lambda(x-\hat\varepsilon_j)\Big)\\ &= \frac{1}{n}\sum_{j=1}^n K_b(x-\hat\varepsilon_j)\Big(\frac{\hat\lambda^2\hat\varepsilon_j^2}{1+\hat\lambda\hat\varepsilon_j} + \hat\lambda(x-\hat\varepsilon_j)\Big)\end{aligned}$$

and then find that the left-hand side of (4.5) is bounded by

$$\frac{1}{n}\sum_{j=1}^{n}\frac{\hat\lambda^2\hat\varepsilon_j^2}{1+\hat\lambda\hat\varepsilon_j}+|\hat\lambda|b\int_{-\infty}^{\infty}|u|K(u)\,du=O_P(1/n)+O_P(bn^{-1/2}),$$

where we used (4.2) and (4.4) in the last step. This proves (4.5).

Next we prove (4.6). For this we write

$$\begin{aligned}\hat f_1(x)-\tilde f(x)&=\frac{1}{n}\sum_{j=1}^{n}\big(K_b\big(x-\varepsilon_j+(\hat\varrho-\varrho)X_{j-1}\big)-K_b(x-\varepsilon_j)\big)\\&=H_{\sqrt{n}(\hat\varrho-\varrho)}(x)+D(x)+(\hat\varrho-\varrho)\frac{1}{n}\sum_{j=1}^{n}X_{j-1}\bar f'(x)\end{aligned}$$

with

$$H_t(x)=\frac{1}{n}\sum_{j=1}^{n}\Big(K_b\Big(x-\varepsilon_j+\frac{tX_{j-1}}{\sqrt n}\Big)-K_b(x-\varepsilon_j)-\bar f\Big(x+\frac{tX_{j-1}}{\sqrt n}\Big)+\bar f(x)\Big),$$

$$D(x)=\frac{1}{n}\sum_{j=1}^{n}\big(\bar f(x+(\hat\varrho-\varrho)X_{j-1})-\bar f(x)-(\hat\varrho-\varrho)X_{j-1}\bar f'(x)\big).$$

As $\bar f'$ has norm $\|\bar f'\|_1\le\Lambda\|\iota_{\mathbb R}K'\|_1$ and is L_1-Lipschitz with constant $\Lambda\|K'\|_1/b$, we derive, utilizing Lemma 3.7 and setting $B=\Lambda\max\{\|K'\|_1,\|\iota_{\mathbb R}K'\|_1\}$,

$$\begin{aligned}\|\hat f_1-\tilde f\|_1&\le\|H_{\sqrt n(\hat\varrho-\varrho)}\|_1+B\Big(\frac{1}{b}(\hat\varrho-\varrho)^2\frac{1}{n}\sum_{j=1}^{n}X_{j-1}^2+|\hat\varrho-\varrho|\Big|\frac{1}{n}\sum_{j=1}^{n}X_{j-1}\Big|\Big)\\&=\|H_{\sqrt n(\hat\varrho-\varrho)}\|_1+O_P(1/(nb)).\end{aligned}$$

Thus the desired (4.6) follows if we verify that the stochastic rate

$$\sup_{|t|\le C}\|H_t\|_1=O_P(1/(nb^{3/2}))\tag{4.8}$$

holds for every large constant C. Fix such a C. Since $\max_{1\le j\le n}|X_{j-1}|$ is of stochastic order $o_P(n^{1/2})$, we may replace H_t by $\bar H_t$ where

$$\bar H_t(x)=\frac{1}{n}\sum_{j=1}^{n}\Big(K_b\Big(x-\varepsilon_j+\frac{tX_{j-1}}{\sqrt n}\Big)-K_b(x-\varepsilon_j)-\bar f\Big(x+\frac{tX_{j-1}}{\sqrt n}\Big)+\bar f(x)\Big)W_{j-1}$$

with

$$W_{j-1}=\mathbf{1}[C|X_{j-1}|\le\sqrt n],\quad j=1,\dots,n.$$

We have $\bar H_0(x)=0$ for all x, and for s and t in $[-C,C]$ we have

$$\bar H_t(x)-\bar H_s(x)=\frac{1}{n}\sum_{j=1}^{n}(t-s)\frac{X_{j-1}}{\sqrt n}W_{j-1}V_j(x)$$

with

$$V_j(x) = \int_0^1 \Big(K_b'\Big(x-\varepsilon_j+(s+v(t-s))\frac{X_{j-1}}{\sqrt{n}}\Big)-\bar f'\Big(x+(s+v(t-s))\frac{X_{j-1}}{\sqrt{n}}\Big)\Big)\,dv.$$

A martingale argument yields

$$\begin{aligned} E[(\bar H_t(x)-\bar H_s(x))^2] &\le \frac{(t-s)^2}{n^2}E[X_0^2W_0V_1(x)^2] \\ &\le \frac{(t-s)^2}{n^2b^3}\int_0^1 E\Big[X_0^2W_0M_b\Big(x-\varepsilon_1+(s+v(t-s))\frac{X_0}{\sqrt{n}}\Big)\Big]\,dv \end{aligned}$$

with $M_b(x) = (1/b)M(x/b)$ and $M = (K')^2$. Lemma 3.2 yields the bound

$$|\|\bar H_t\|_1 - \|\bar H_s\|_1|^2 \le \|\bar H_t - \bar H_s\|_1^2 \le \pi\int_{-\infty}^{\infty}(1+x^2)(\bar H_t(x)-\bar H_s(x))^2\,dx.$$

Combining the above yields the inequality

$$\begin{aligned} n^2b^3E[|\|\bar H_t\|_1 - \|\bar H_s\|_1|^2] &\le \pi\int_{-\infty}^{\infty}(1+x^2)n^2b^3E[(\bar H_t(x)-\bar H_s(x))^2]\,dx \\ &\le \pi(t-s)^2\int_0^1 I(v)\,dv \end{aligned}$$

with

$$I(v) = \int_{-\infty}^{\infty}(1+x^2)E\Big[X_0^2\mathbf{1}[C|X_0|\le\sqrt{n}]M_b\Big(x-\varepsilon_1+(s+v(t-s))\frac{X_0}{\sqrt{n}}\Big)\Big]\,dx$$

which equals

$$\int_{-\infty}^{\infty}E\big[X_0^2\mathbf{1}[C|X_0|\le\sqrt{n}]\big(1+\big(\varepsilon_1-(s+v(t-s))X_0/\sqrt{n}+bu\big)^2\big)\big]M(u)\,du$$

and can be bounded as follows,

$$\begin{aligned} I(v) &\le \int_{-\infty}^{\infty}E[X_0^2\mathbf{1}[C|X_0|\le\sqrt{n}]](1+3\varepsilon_1^2+3b^2u^2+3)M(u)\,du \\ &\le E[X_0^2]\int_{-\infty}^{\infty}4(1+\sigma^2+b^2u^2)M(u)\,du \\ &\le 4(1+\sigma^2+b^2)E[X_0^2]\int_{-\infty}^{\infty}(1+u^2)M(u)\,du, \quad 0<v<1. \end{aligned}$$

In the first inequality we used the Cauchy–Schwarz inequality and the fact that $s+v(t-s)$ belongs to the interval $[-C,C]$ and is hence bounded by C. Note that $\|(1+\iota_{\mathbb{R}}^2)M\|_1$ is finite by the assumptions on K'. In view of Theorem 12.3 in [1], the process $\{\mathbb{X}_n(t) = nb^{3/2}\|\bar H_t\|_1, |t|\le C\}$ is tight and this implies (4.8).

We are left to verify (4.7). Since f is L_1-Lipschitz, the identity $\bar{f} = B_b(f, K)$ and Lemma 3.10 yield $\|\bar{f} - f\|_1 \le \Lambda b \|\iota_{\mathbb{R}} K\|_1$. This, the rate $nb^3 \to \infty$, (4.1) and (4.6) establish the stochastic rate

$$\|\hat{f}_1 - f\|_1 \le \|\hat{f}_1 - \tilde{f}\|_1 + \|\tilde{f} - \bar{f}\|_1 + \|\bar{f} - f\|_1 = O_P\Big(\frac{b^{3/2}}{nb^3} + \frac{b}{(nb^3)^{1/2}} + b\Big) = O_P(b).$$

In view of Lemma 3.3 the desired result follows from this, the stochastic rate

$$\int_{-\infty}^{\infty} x^2 \hat{f}_1(x)\,dx = \frac{1}{n}\sum_{j=1}^{n} \hat{\varepsilon}_j^2 + \int_{-\infty}^{\infty} b^2 u^2 K(u)\,du = O_P(1)$$

and the fact that f has a finite second moment. □

Corollary 4.1. *Under the assumptions of Theorem 4.1 the estimator $\hat{f}_1$ satisfies*

$$\|\hat{f}_1 - f\|_1 = O_P(b) \quad \textit{and} \quad \|\iota_{\mathbb{R}}(\hat{f}_1 - f)\|_1 = O_P(b^{1/2}),$$

and the estimator $\hat{f}_2$ satisfies

$$\|\hat{f}_2 - f\|_1 = O_P(b) \quad \textit{and} \quad \|\iota_{\mathbb{R}}(\hat{f}_2 - f)\|_1 = O_P(b^{1/2}).$$

Proof. The first two rates were established in the proof of (4.7). The third rate follows from the first one and $\|\hat{f}_2 - \hat{f}_1\|_1 = O_P(n^{-1}b^{-3/2}) + O_P(n^{-1/2}) = o_P(b^{3/2})$ which is a consequence of Theorem 4.1. From (4.2) and (4.3) we derive the bound

$$\int_{-\infty}^{\infty} x^2 \hat{f}_2(x)\,dx \le \max_{1 \le j \le n} \frac{1}{1 + \hat{\lambda}\varepsilon_j} \int_{-\infty}^{\infty} x^2 \hat{f}_1(x)\,dx = O_P(1).$$

The argument used to prove (4.7) now yields $\|\iota_{\mathbb{R}}(\hat{f}_2 - f)\|_1 = O_P(b^{1/2})$. □

5. Behavior of the derivatives of the innovation density estimators

In this section we assume that the density f belongs to $\mathbb{H}_1$ and show that the derivatives of our kernel estimators estimate f' consistently in the L_1-norm.

Theorem 5.1. *Suppose the bandwidth $b = b_n$ satisfies $nb_n^4 \to 0$ and $nb_n^3 \to \infty$, the kernel K is a symmetric density with finite variance and is twice continuously differentiable with $\|(1+\iota_{\mathbb{R}}^2)K'\|_1$ and $\|(1+\iota_{\mathbb{R}}^2)(K'')^2\|_1$ finite, the estimator $\hat{\varrho}$ is root-n consistent, and f belongs to $\mathbb{H}_1$. Then we have the stochastic rates*

$$\|\hat{f}_1' - f'\|_1 = o_P(1) \quad \textit{and} \quad \|\hat{f}_2' - f'\|_1 = o_P(1).$$

Proof. The assumptions on K imply that K belongs to $\mathbb{H}_2$ and meets the assumptions of Theorem 4.1. Since f' is integrable and $\bar{f}'$ equals $B_b(f', K)$, we have

$$\|\bar{f}' - f'\|_1 \le \int_{-\infty}^{\infty} w_{f'}(|bu|) K(u)\,du \to 0. \tag{5.1}$$

The desired result thus follows from the following stochastic rates:

$$\|\hat{f}_2' - \hat{f}_1'\|_1 = O_P(1/(bn^{1/2})), \tag{5.2}$$

$$\|\hat{f}_1' - \tilde{f}'\|_1 = O_P(1/(nb^{5/2})), \tag{5.3}$$

$$\|\tilde{f}' - \bar{f}'\|_1 = O_P(1/(nb^3)^{1/2}). \tag{5.4}$$

Let us first establish (5.2). In view of $\hat{\lambda} = O_P(n^{-1/2})$, it suffices to show the rates $\|\iota_{\mathbb{R}}\hat{f}_1'\|_1 = O_P(1/b)$ and $\|\hat{f}_2' - \hat{f}_1' + \hat{\lambda}\iota_{\mathbb{R}}\hat{f}_1'\|_1 = O_P(n^{-1/2})$. The former follows from the inequality

$$\begin{aligned}
\|\iota_{\mathbb{R}}\hat{f}_1'\|_1 &\le \frac{1}{n}\sum_{j=1}^n \int_{-\infty}^{\infty} |x||K_b'(x-\hat{\varepsilon}_j)|\,dx \\
&\le \frac{1}{n}\sum_{j=1}^n \int_{-\infty}^{\infty} |\hat{\varepsilon}_j + bu|\frac{|K'(u)|}{b}\,du \\
&\le \frac{\|K'\|_1}{b}\Big(\frac{1}{n}\sum_{j=1}^n |\varepsilon_j| + |\hat{\varrho}-\varrho|\frac{1}{n}\sum_{j=1}^n |X_{j-1}|\Big) + \|\iota_{\mathbb{R}}K'\|_1.
\end{aligned}$$

For the latter we use the identity

$$\begin{aligned}
\hat{f}_2'(x) - \hat{f}_1'(x) + \hat{\lambda}x\hat{f}_1'(x) &= \frac{1}{n}\sum_{j=1}^n K_b'(x-\hat{\varepsilon}_j)\Big(\frac{1}{1+\hat{\lambda}\hat{\varepsilon}_j} - 1 + \hat{\lambda}\hat{\varepsilon}_j + \hat{\lambda}(x-\hat{\varepsilon}_j)\Big) \\
&= \frac{1}{n}\sum_{j=1}^n K_b'(x-\hat{\varepsilon}_j)\Big(\frac{\hat{\lambda}^2\hat{\varepsilon}_j^2}{1+\hat{\lambda}\hat{\varepsilon}_j} + \hat{\lambda}(x-\hat{\varepsilon}_j)\Big)
\end{aligned}$$

to obtain the inequality

$$\|\hat{f}_2' - \hat{f}_1' - \hat{\lambda}\iota_{\mathbb{R}}\hat{f}_1'\|_1 \le \frac{1}{n}\sum_{j=1}^n \frac{\hat{\lambda}^2\hat{\varepsilon}_j^2}{1+\hat{\lambda}\hat{\varepsilon}_j}\frac{\|K'\|_1}{b} + |\hat{\lambda}|\|\iota_{\mathbb{R}}K'\|_1 = O_P\Big(\frac{1}{nb}\Big) + O_P\Big(\frac{1}{\sqrt{n}}\Big)$$

where we used (4.2) and (4.4) in the last step. This proves (5.2).

Let us now prove (5.3). We write

$$\begin{aligned}
\hat{f}_1'(x) - \tilde{f}'(x) &= \frac{1}{n}\sum_{j=1}^n \big(K_b'(x-\varepsilon_j + (\hat{\varrho}-\varrho)X_{j-1}) - K_b'(x-\varepsilon_j)\big) \\
&= H'_{\sqrt{n}(\hat{\varrho}-\varrho)}(x) + D'(x) + (\hat{\varrho}-\varrho)\frac{1}{n}\sum_{j=1}^n X_{j-1}\bar{f}''(x)
\end{aligned}$$

with

$$H_t'(x) = \frac{1}{n}\sum_{j=1}^n \Big(K_b'\Big(x-\varepsilon_j+\frac{tX_{j-1}}{\sqrt{n}}\Big) - K_b'(x-\varepsilon_j) - \bar{f}'\Big(x+\frac{tX_{j-1}}{\sqrt{n}}\Big) + \bar{f}'(x)\Big),$$

$$D'(x) = \frac{1}{n}\sum_{j=1}^{n}\big(\bar f'(x+(\hat\varrho-\varrho)X_{j-1}) - \bar f'(x) - (\hat\varrho-\varrho)X_{j-1}\bar f''(x)\big).$$

By Lemma 3.9, $\bar f$ belongs to $\mathbb{H}_2$ and $\bar f'' = B_b(f', K')/b$ has norm $\|\bar f''\|_1 \le \|f'\|_1\|K'\|_1/b$ and is L_1-Lipschitz with Lipschitz constant $\|f'\|_1\|K''\|_1/b^2$. Using this and Lemma 3.7 we obtain with $B = \|f'\|_1 \max\{\|K'\|_1, \|K''\|_1\}$,

$$\begin{aligned}\|\hat f_1' - \tilde f'\|_1 &\le \|H'_{\sqrt{n}(\hat\varrho-\varrho)}\|_1 + B\Big(\frac{1}{b^2}(\hat\varrho-\varrho)^2\frac{1}{n}\sum_{j=1}^{n}X_{j-1}^2 + \frac{1}{b}|\hat\varrho-\varrho|\Big|\frac{1}{n}\sum_{j=1}^{n}X_{j-1}\Big|\Big)\\ &= \|H'_{\sqrt{n}(\hat\varrho-\varrho)}\|_1 + O_P(1/(nb^2)).\end{aligned}$$

Thus the desired (5.3) follows if we verify that the stochastic rate

$$\sup_{|t|\le C}\|H_t'\|_1 = O_P(1/(nb^{5/2})) \tag{5.5}$$

holds for every large constant C. Fix such a C. Since $\max_{1\le j\le n}|X_{j-1}|$ is of stochastic order $o_P(n^{1/2})$, we may replace H_t' by $\bar H_t'$ where

$$\bar H_t'(x) = \frac{1}{n}\sum_{j=1}^{n}\Big(K_b'\Big(x-\varepsilon_j+\frac{tX_{j-1}}{\sqrt n}\Big) - K_b'(x-\varepsilon_j) - \bar f'\Big(x+\frac{tX_{j-1}}{\sqrt n}\Big) + \bar f'(x)\Big)W_{j-1}$$

with $W_{j-1} = \mathbf{1}[C|X_{j-1}| \le \sqrt n]$, $j = 1,\dots,n$, as in the proof of Theorem 4.1. We have $\bar H_0'(x) = 0$ for all x, and for s and t in $[-C, C]$ we have

$$\bar H_t'(x) - \bar H_s'(x) = \frac{1}{n}\sum_{j=1}^{n}(t-s)\frac{X_{j-1}}{\sqrt n}W_{j-1}V_j(x)$$

with

$$V_j(x) = \int_0^1 \Big(K_b''\Big(x-\varepsilon_j+(s+v(t-s))\frac{X_{j-1}}{\sqrt n}\Big) - \bar f''\Big(x+(s+v(t-s))\frac{X_{j-1}}{\sqrt n}\Big)\Big)\,dv.$$

A martingale argument yields

$$\begin{aligned}E[(\bar H_t'(x) - \bar H_s'(x))^2] &\le \frac{(t-s)^2}{n^2}E[X_0^2W_0V_1(x)^2]\\ &\le \frac{(t-s)^2}{n^2b^5}\int_0^1 E\Big[X_0^2W_0M_b\Big(x-\varepsilon_1+(s+v(t-s))\frac{X_0}{\sqrt n}\Big)\Big]\,dv\end{aligned}$$

with $M_b(x) = (1/b)M(x/b)$ and $M = (K'')^2$. Lemma 3.2 yields the bound

$$\big|\|\bar H_t'\|_1 - \|\bar H_s'\|_1\big|^2 \le \|\bar H_t' - \bar H_s'\|_1^2 \le \pi\int_{-\infty}^{\infty}(1+x^2)(\bar H_t'(x) - \bar H_s'(x))^2\,dx.$$

Combining the above yields the inequality

$$\begin{aligned}n^2b^5E\big[\big|\|\bar H_t'\|_1 - \|\bar H_s'\|_1\big|^2\big] &\le \pi\int_{-\infty}^{\infty}(1+x^2)n^3b^5E[(\bar H_t'(x) - \bar H_s'(x))^2]\,dx\\ &\le \pi(t-s)^2\int_0^1 I(v)\,dv\end{aligned}$$

with

$$I(v) = \int_{-\infty}^{\infty} (1+x^2) E[X_0^2 \mathbf{1}[C|X_0| \le \sqrt{n}]] M_b\Big(x - \varepsilon_1 + (s + v(t-s))\frac{X_0}{\sqrt{n}}\Big)\, dx$$
$$\le 4(1+\sigma^2+b^2)E[X_0^2] \int_{-\infty}^{\infty} (1+u^2) M(u)\, du, \quad 0 < v < 1,$$

where the inequality is obtained as in the proof of Theorem 4.1. In view of Theorem 12.3 in [1], the process $\{\mathbb{X}_n'(t) = nb^{5/2}\|\bar{H}_t'\|_1, |t| \le C\}$ is tight and this implies (5.5).

We are left to verify (5.4). Using Lemma 3.2 we calculate

$$\begin{aligned} nb^3 E[\|\tilde{f}' - \bar{f}'\|_1^2] &\le \pi \int_{-\infty}^{\infty} (1+x^2) nb^3 E[(\tilde{f}'(x) - \bar{f}'(x))^2]\, dx \\ &\le \pi \int_{-\infty}^{\infty} (1+x^2) \int_{-\infty}^{\infty} b^3 (K_b'(x-y))^2 f(y)\, dy\, dx \\ &= \pi \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (1+(y+bu)^2) f(y) (K'(u))^2\, dy\, du \\ &= \pi(1+\sigma^2)\|(K')^2\|_1 + \pi b^2 \|(\iota_{\mathbb{R}} K')^2\|_1. \end{aligned}$$

This shows $\|\tilde{f}' - \bar{f}'\|_1 = O_P(1/\sqrt{nb^3})$.

6. Proof of Theorem 2.1

Let $\hat{f}$ denote either the estimator $\hat{f}_1$ or the estimator $\hat{f}_2$. In view of the properties of k_n and $b = b_n$, Corollary 4.1 implies the stochastic rates

$$(k_n+1)\|\hat{f} - f\|_1 = o_P(n^{-1/4}) \tag{6.1}$$

and

$$\|\iota_{\mathbb{R}}(\hat{f} - f)\|_1 = O_P(b^{1/2}), \tag{6.2}$$

while Theorem 5.1 yields

$$\|\hat{f}' - f'\|_1 = o_P(1). \tag{6.3}$$

Recall that $\hat{g}_{k_n}$ can be expressed as

$$\hat{g}_{k_n}(x) = \int_{\mathbb{R}^{k_n+1}} \hat{f}\Big(x - \sum_{i=1}^{k_n} \hat{\varrho}^i y_i - \hat{\varrho}^{k_n+1} z\Big) \prod_{j=1}^{k_n} \hat{f}(y_j)\, dy_j\, \hat{g}(z)\, dz, \quad x \in \mathbb{R}. \tag{6.4}$$

Corollary 3.1, assumption (A2) and the bound (2.1) yield

$$\Big\| g_{\hat{\varrho},f} - g - \dot{g}\frac{1}{n}\sum_{j=1}^{n} \psi(X_{j-1}, \varepsilon_j)\Big\|_1 = o_P(n^{-1/2}).$$

Thus we need to compare $\hat{g}_{k_n}$ with $g_{\hat{\varrho},f}$. For this we represent $g_{\hat{\varrho},f}(x)$ as

$$g_{\hat{\varrho},f}(x) = \int_{\mathbb{R}^{k_n+1}} f\Big(x - \sum_{i=1}^{k_n} \hat{\varrho}^i y_i - \hat{\varrho}^{k_n+1} z\Big) \prod_{j=1}^{k_n} f(y_j)\, dy_j\; g_{\hat{\varrho},f}(z)\, dz. \tag{6.5}$$

Replacing $\hat{\varrho}^{k_n+1}$ by ϱ^{k_n+1} and $g_{\hat{\varrho},f}(z)$ by $g(z)$ yields

$$g^*_{\hat{\varrho},f}(x) = \int_{\mathbb{R}^{k_n+1}} f\Big(x - \sum_{i=1}^{k_n} \hat{\varrho}^i y_i - \varrho^{k_n+1} z\Big) \prod_{j=1}^{k_n} f(y_j)\, dy_j\; g(z)\, dz, \quad x \in \mathbb{R}.$$

Replacing $\hat{\varrho}^{k_n+1}$ by ϱ^{k_n+1} and $\hat{g}(z)$ by $g(z)$ in (6.4) yields

$$\hat{g}^*_{k_n}(x) = \int_{\mathbb{R}^{k_n+1}} \hat{f}\Big(x - \sum_{i=1}^{k_n} \hat{\varrho}^i y_i - \varrho^{k_n+1} z\Big) \prod_{j=1}^{k_n} \hat{f}(y_j)\, dy_j\; g(z)\, dz, \quad x \in \mathbb{R}.$$

We begin by establishing the rates

$$\|\hat{g}_{k_n} - \hat{g}^*_{k_n}\|_1 = o_P(n^{-1/2}) \tag{6.6}$$

and

$$\|g_{\hat{\varrho},f} - g^*_{\hat{\varrho},f}\|_1 = o_P(n^{-1/2}). \tag{6.7}$$

We have the identities $\hat{g}_{k_n} = B_{\hat{\varrho}^{k_n+1}}(\hat{p}, \hat{g})$ and $\hat{g}^*_{k_n} = B_{\varrho^{k_n+1}}(\hat{p}, g)$ with

$$\hat{p}(x) = \int_{\mathbb{R}^{k_n}} \hat{f}\Big(x - \sum_{i=1}^{k_n} \hat{\varrho}^i y_i\Big) \prod_{j=1}^{k_n} \hat{f}(y_j)\, dy_j, \quad x \in \mathbb{R}.$$

It is easy to check that $\hat{p}$ is $\mathbb{H}_1$-valued with

$$\hat{p}'(x) = \int_{\mathbb{R}^{k_n}} \hat{f}'\Big(x - \sum_{i=1}^{k_n} \hat{\varrho}^i y_i\Big) \prod_{j=1}^{k} \hat{f}(y_j)\, dy_j, \quad x \in \mathbb{R},$$

and that $\|\hat{p}'\|_1 \le \|\hat{f}'\|_1 = O_P(1)$ holds in view of (6.3). Using Lemma 3.11 we obtain

$$\|\hat{g}_{k_n} - \hat{g}^*_{k_n}\|_1 \le \|\hat{g}_{k_n} - \hat{p}\|_1 + \|\hat{g}^*_{k_n} - \hat{p}\|_1 \le \|\hat{p}'\|_1\big(\hat{\varrho}^{k_n+1}\|\iota_{\mathbb{R}}\hat{g}\|_1 + \varrho^{k_n+1}\|\iota_{\mathbb{R}} g\|_1\big).$$

The desired (6.6) now follows from $\hat{\varrho}^{k_n+1} = o_P(n^{-1/2})$ and $\|\iota_{\mathbb{R}}\hat{g}\|_1 = O_P(1)$. Indeed, the former follows from

$$n^{1/2}\hat{\varrho}^{k_n+1} = \exp\Big(-(k_n+1)\Big(\log(1/\hat{\varrho}) - \frac{\log(n)}{2(k_n+1)}\Big)\Big) = o_P(1)$$

and the latter from

$$E[\|\iota_{\mathbb{R}}\hat{g}\|_1] = \|\iota_{\mathbb{R}} E[\hat{g}]\|_1 = \|\iota_{\mathbb{R}} g * K_b\|_1 \le \|\iota_{\mathbb{R}} g\|_1 + \|\iota_{\mathbb{R}} K_b\|_1 = \|\iota_{\mathbb{R}} g\|_1 + O(b).$$

A similar argument yields (6.7).

We are left to compare $\hat{g}^*_{k_n}$ and $g^*_{\hat{\varrho},f}$. For this we express $\hat{g}^*_{k_n}$ as $L_{\hat{\varrho}}(\hat{f},\dots,\hat{f})$ and $g^*_{\hat{\varrho},f}$ as $L_{\hat{\varrho}}(f,\dots,f)$ where

$$L_r(h_0,\dots,h_{k_n})(x) = \int_{\mathbb{R}^{k_n+1}} h_0\Big(x - \sum_{i=1}^{k_n} r^i y_i - \varrho^{k_n+1} z\Big) \prod_{j=1}^{k_n} h_j(y_j)\,dy_j g(z)\,dz$$

for integrable functions $h_0,\dots,h_{k_n}$ and positive numbers r. One checks

$$\|L_r(h_0,\dots,h_{k_n})\|_1 \le \prod_{j=0}^{k_n} \|h_j\|_1. \tag{6.8}$$

To simplify notation, we set

$$\bar{L}_{r,h} = L_r(h,\dots,h)$$

and, for a subset A of $\{0,\dots,k_n\}$,

$$L_{r,A}(p,q) = L_r(h_0,\dots,h_{k_n}) \quad \text{with } h_i = \begin{cases} p, & i \in A, \\ q, & i \notin A, \end{cases} \quad i = 0,\dots,k_n.$$

We use the identity

$$\prod_{j=0}^{k_n}(a_j + b_j) = \sum_{A\subset\{0,1,\dots,k_n\}} \prod_{j\in A} a_j \prod_{j\in A^c} b_j$$

to conclude

$$\bar{L}_{r,p+q} = \sum_{A\subset\{0,1,\dots,k_n\}} L_{r,A}(p,q).$$

Applying this with $r = \hat{\varrho}$, $p = \hat{f} - f$ and $q = f$, and singling out the terms with A of cardinality at most one, we obtain

$$\bar{L}_{\hat{\varrho},\hat{f}} = \bar{L}_{\hat{\varrho},f} + \sum_{j=0}^{k_n} L_{\hat{\varrho},\{j\}}(\hat{f} - f, f) + R_1$$

with

$$R_1 = \sum_{\substack{A\subset\{0,\dots,k_n\} \\ \mathrm{card}(A)\ge 2}} L_{\hat{\varrho},A}(\hat{f} - f, f)$$

By (6.8), $\|L_{r,A}(p,q)\|_1$ is bounded by $\|p\|_1^a \|q\|_1^{k_n+1-a}$ with a the cardinality of A. Since there are $\binom{k_n+1}{a} \le (k_n+1)^a/a!$ subsets of cardinality a, we obtain

$$\|R_1\|_1 \le \sum_{a=2}^{k_n+1} \frac{(k_n+1)^a}{a!} \|\hat{f} - f\|_1^a \le ((k_n+1)\|\hat{f} - f\|_1)^2 e^{(k_n+1)\|\hat{f}-f\|_1}.$$

In view of the rate $(k_n+1)\|\hat{f}-f\|_1 = o_P(n^{-1/4})$ given in (6.1), this establishes

$$\Big\|\hat{g}^*_{k_n} - g^*_{\hat{\varrho},f} - \sum_{j=0}^{k_n} L_{\hat{\varrho},\{j\}}(\hat{f}-f,f)\Big\|_1 = o_P(n^{-1/2}). \tag{6.9}$$

Our next goal is to verify

$$\Big\|\sum_{j=0}^{k_n} L_{\hat{\varrho},\{j\}}(\hat{f}-f,f) - \sum_{j=0}^{k_n} L_{\varrho,\{j\}}(\hat{f}-f,f)\Big\|_1 = o_P(n^{-1/2}). \tag{6.10}$$

To achieve this we derive bounds on the terms

$$D_i(h) = \|L_{\hat{\varrho},\{i\}}(h,f) - L_{\varrho,\{i\}}(h,f)\|_1, \quad i=0,\dots,k_n,$$

for $h \in \mathbb{H}_1$ with $\|\iota_{\mathbb{R}}h\|_1$ finite. Using Lemma 3.7 we obtain

$$D_0(h) \le \|h'\|_1 \sum_{j=1}^{k_n} |\hat{\varrho}^j - \varrho^j| \|\iota_{\mathbb{R}} f\|_1$$

and

$$\begin{aligned} D_i(h) &\le \|f'\|_1 \sum_{j=1}^{k_n} |\hat{\varrho}^j - \varrho^j| \big(\mathbf{1}[j\ne i]\|h\|_1\|\iota_{\mathbb{R}} f\|_1 + \mathbf{1}[j=i]\|\iota_{\mathbb{R}} h\|_1\big) \\ &\le \|f'\|_1\|h\|_1\|\iota_{\mathbb{R}} f\|_1 \sum_{j=1}^{k_n} |\hat{\varrho}^j - \varrho^j| + \|f'\|_1\|\iota_{\mathbb{R}} h\|_1 |\hat{\varrho}^i - \varrho^i|, \quad i=1,\dots,k_n. \end{aligned}$$

Using the inequality (3.1) with $r=\hat{\varrho}$ and $s=\varrho$ and taking $h=\hat{f}-f$ we obtain that the left-hand side of (6.10) is bounded by

$$\frac{\|\hat{f}'-f'\|_1\|\iota_{\mathbb{R}} f\|_1 + \|f'\|_1\|\iota_{\mathbb{R}}(\hat{f}-f)\|_1 + k_n\|f'\|_1\|\hat{f}-f\|_1\|\iota_{\mathbb{R}} f\|_1}{(1-\max\{\hat{\varrho},\varrho\})^2}|\hat{\varrho}-\varrho|.$$

This bound is of order $o_P(n^{-1/2})$ because $\hat{\varrho}-\varrho$ is of order $O_P(n^{-1/2})$ and the terms $k_n\|\hat{f}-f\|_1$, $\|\iota_{\mathbb{R}}(\hat{f}-f)\|_1$ and $\|\hat{f}'-f'\|_1$ are of order $o_P(1)$ in view of (6.1)–(6.3).

Next we observe the identity

$$\sum_{j=0}^{k_n} L_{\varrho,j}(\hat{f}-f,f) = \sum_{j=0}^{k_n} B_{\varrho^j}(\gamma_j,\hat{f}-f) = \sum_{j=0}^{k_n} \Gamma_j(\hat{f}-f)$$

where, for an integrable function h, $\Gamma_j h = B_{\varrho^j}(\gamma_j,h)$ is the function defined by

$$\Gamma_j h(x) = \int_{-\infty}^{\infty} \gamma_j(x-\varrho^j y)h(y)\,dy, \quad x\in\mathbb{R}.$$

We have $\|\Gamma_j h\|_1 \le \|h\|_1$ for all integrable h. Using this and Theorem 4.1 we derive

$$\sum_{j=0}^{k_n} \|\Gamma_j(\hat{f}_1 - \tilde{f})\|_1 \le (k_n+1)\|\hat{f}_1 - \tilde{f}\|_1 = O_P(k_n/(nb^{3/2}))$$

and

$$\sum_{j=0}^{k_n} \|\Gamma_j(\hat{f}_2 - \tilde{f}) + \hat{\lambda}\Gamma_j(\iota_{\mathbb{R}} f))\|_1 \le (k_n+1)\|\hat{f}_2 - \tilde{f} + \hat{\lambda}\iota_{\mathbb{R}} f\|_1 = O_P(k_n/(nb^{3/2})).$$

Since f has mean zero and finite variance, Lemma 3.10 and Corollary 3.2 yield the inequalities

$$\|\Gamma_j(\iota_{\mathbb{R}} f)\|_1 \le \varrho^j \|\gamma_j'\|_1 \|\iota_{\mathbb{R}}^2 f\|_1 \le C_1 \sigma^2 \varrho^j, \quad j \ge 0.$$

This and the expansion (4.2) yield

$$\Big\| \hat{\lambda} \sum_{j=0}^{k_n} \Gamma_j(\iota_{\mathbb{R}} f) - \frac{1}{n}\sum_{i=1}^n \frac{\varepsilon_i}{\sigma^2} \sum_{j=0}^{\infty} \Gamma_j(\iota_{\mathbb{R}} f) \Big\|_1 = o_P(n^{-1/2}).$$

Since f has mean zero, we compute

$$\Gamma_j(\iota_{\mathbb{R}} f)(x) = \int_{-\infty}^{\infty} (\gamma_j(x - \varrho^j y) - g(x)) y f(y)\, dy, \quad x \in \mathbb{R},$$

and obtain the identity

$$\sum_{j=0}^{\infty} \Gamma_j(\iota_{\mathbb{R}} f)(x) = \int_{-\infty}^{\infty} \gamma(x,y) y f(y)\, dy, \quad x \in \mathbb{R}.$$

In view of $k_n^2/(nb^3) \to 0$, we obtain

$$\Big\| \sum_{j=0}^{k_n} L_{\varrho,\{j\}}(\hat{f}_1 - f) - \sum_{j=0}^{k_n} \Gamma_j(\tilde{f} - f) \Big\|_1 = o_P(n^{-1/2}) \tag{6.11}$$

and

$$\Big\| \sum_{j=0}^{k_n} L_{\varrho,\{j\}}(\hat{f}_2 - f) - \sum_{j=0}^{k_n} \Gamma_j(\tilde{f} - f) + \sum_{j=0}^{\infty} \Gamma_j(\iota_{\mathbb{R}} f) \frac{1}{n}\sum_{i=1}^n \frac{\varepsilon_i}{\sigma^2} \Big\|_1 = o_P(n^{-1/2}). \tag{6.12}$$

For $j = 0, 1, \dots$, we compute

$$\Gamma_j \tilde{f}(x) = \int \bar{\gamma}_j(x - \varrho^j b u) K(u)\, du, \quad x \in \mathbb{R},$$

with

$$\bar{\gamma}_j(x) = \frac{1}{n}\sum_{i=1}^n \gamma_j(x - \varrho^j \varepsilon_i), \quad x \in \mathbb{R}.$$

By Corollary 3.2, γ_j belongs to $\mathbb{H}_2$ with $\|\gamma_j''\|_1 \le C_3$. This implies that $\bar\gamma_j$ is $\mathbb{H}_2$-valued and $\|\bar\gamma_j''\|_1 \le C_3$. Lemma 3.11, applied with $t = b\varrho^j$, $p = \bar\gamma_j$ and $q = K$, yields the bound

$$\|\Gamma_j\tilde f - \bar\gamma_j\|_1 \le C_3 b^2 \varrho^{2j} \|\iota_{\mathbb{R}}^2 K\|_1.$$

Since $\Gamma_j f = g$, we derive

$$\Big\|\sum_{j=0}^{k_n}\big(\Gamma_j(\tilde f - f) - (\bar\gamma_j - g)\big)\Big\|_1 \le \sum_{j=0}^{k_n}\|\Gamma_j\tilde f - \bar\gamma_j\|_1 = O_P(b^2) = o_P(n^{-1/2}).$$

Finally, using Corollary 3.2 we obtain the rate

$$E\Big[\Big\|\sum_{j=k_n+1}^{\infty}(\bar\gamma_j - g)\Big\|_1\Big] \le \sum_{j=k_n+1}^{\infty}\|\gamma_j'\|_1\varrho^j E[|\varepsilon_1 - \varepsilon_2|] = O(\varrho^{k_n+1}) = o_P(n^{-1/2}).$$

From this we conclude

$$\Big\|\sum_{j=0}^{k_n}\Gamma_j(\tilde f - f) - \bar\Gamma_n\Big\|_1 = o_P(n^{-1/2}) \tag{6.13}$$

with

$$\bar\Gamma_n(x) = \frac{1}{n}\sum_{i=1}^{n}\gamma(x,\varepsilon_i), \quad x \in \mathbb{R}.$$

The first L_1-Bahadur expansion follows from (6.6), (6.7), (6.9), (6.10), (6.11) and (6.13), while the second follows from (6.6), (6.7), (6.9), (6.10), (6.12) and (6.13).

7. Proof of Theorem 2.2

Corollary 3.1 and $n^{1/2}(\varrho_n - \varrho) \to t$ imply

$$\|n^{1/2}(g_{\varrho_n,f} - g_{\varrho,f}) - t\dot g\|_1 \to 0.$$

Thus we are left to show

$$\|n^{1/2}(g_{\varrho_n,f_n} - g_{\varrho_n,f}) - Ah\|_1 \to 0. \tag{7.1}$$

Recall that hf and $\iota_{\mathbb{R}}hf$ integrate to zero, i.e.,

$$\int_{-\infty}^{\infty} h(y)f(y)\,dy = 0 \quad \text{and} \quad \int_{-\infty}^{\infty} yh(y)f(y)\,dy = 0.$$

The second integral condition allows us to replace the function γ^* in the definition of Ah by γ, and this leads to the representation

$$Ah = \sum_{j=0}^{\infty}\Gamma_j(hf)$$

in view of the definition of γ and the first integral condition, which gives

$$\int_{-\infty}^{\infty} \big(\gamma_j(x-\varrho^j y) - g(x)\big) h(y) f(y)\, dy = \int_{-\infty}^{\infty} \gamma_j(x-\varrho^j y) h(y) f(y)\, dy = \Gamma_j(hf)(x).$$

It follows from Corollary 3.2 and Lemma 3.10 that $\|\Gamma_j(hf)\|_1 \le C_1 \varrho^j \|\iota_{\mathbb{R}} hf\|_1$. Using the definition of γ_0, we have the identity $\Gamma_0(hf) = B_\varrho(hf,g)$. Let k_n be a sequence of positive integers such that $k_n/(\log(n))^2 \to 1$. This implies that $n^{1/2}\varrho^{k_n} \to 0$ and therefore $\sum_{j=k_n+1}^{\infty} \|\Gamma_j(hf)\|_1 = o(n^{-1/2})$. Thus we achieve (7.1) by verifying

$$\|n^{1/2}(g_{\varrho_n,f_n} - \bar{g}_n) - B_\varrho(hf,g)\|_1 \to 0 \tag{7.2}$$

and

$$\Big\| n^{1/2}(\bar{g}_n - g_{\varrho_n,f}) - \sum_{j=1}^{k_n} \Gamma_j(hf) \Big\|_1 \to 0 \tag{7.3}$$

with

$$\bar{g}_n(x) = \int_{\mathbb{R}^{k_n+1}} f\Big(x - \sum_{j=1}^{k_n} \varrho_n^j y_j - \varrho^{k_n+1} z\Big) \prod_{j=1}^{k_n} f_n(y_j)\, dy_j g(z)\, dz, \quad x \in \mathbb{R}.$$

Let us set

$$\Delta_n = n^{1/2}(f_n - f) - hf.$$

We begin by showing that (2.5) and (2.6) imply

$$\|\Delta_n\|_1 \to 0 \quad \text{and} \quad \|\iota_{\mathbb{R}} \Delta_n\|_1 \to 0. \tag{7.4}$$

For this we set $\tau = \frac{1}{2} h\sqrt{f}$ and write

$$\begin{aligned} n^{1/2}(f_n - f) - hf &= n^{1/2}(\sqrt{f_n} - \sqrt{f})(\sqrt{f_n} + \sqrt{f}) - 2\tau\sqrt{f} \\ &= \big(n^{1/2}(\sqrt{f_n} - \sqrt{f}) - \tau\big)(\sqrt{f_n} + \sqrt{f}) + \tau(\sqrt{f_n} - \sqrt{f}). \end{aligned}$$

This shows that $\|\Delta_n\|_1$ is bounded by

$$\|n^{1/2}(\sqrt{f_n} - \sqrt{f}) - \tau\|_2 \|\sqrt{f_n} + \sqrt{f}\|_2 + \|\tau\|_2 \|\sqrt{f_n} - \sqrt{f}\|_2$$

and $\|\iota_{\mathbb{R}} \Delta_n\|_1$ is bounded by

$$\|n^{1/2}(\sqrt{f_n} - \sqrt{f}) - \tau\|_2 \|\iota_{\mathbb{R}}(\sqrt{f_n} + \sqrt{f})\|_2 + \|\tau\|_2 \|\iota_{\mathbb{R}}(\sqrt{f_n} - \sqrt{f})\|_2.$$

These bounds converge to 0 because (2.5) implies that $n^{1/2}(\sqrt{f_n} - \sqrt{f})$ converges in L_2 to τ and $\sqrt{f_n} - \sqrt{f}$ converges in L_2 to 0, and because (2.6) implies that $\|\iota_{\mathbb{R}}(\sqrt{f_n} - \sqrt{f}\|_2^2 \le \|\iota_{\mathbb{R}}^2 |f_n - f|\|_1 \to 0$.

As a consequence of (7.4), the bilinearity of B_{ϱ_n} and $\|B_{\varrho_n}(p,q)\|_1 \le \|p\|_1\|q\|_1$ we obtain

$$\|n^{1/2}(B_{\varrho_n}(f_n,g_n) - B_{\varrho_n}(f,g_n)) - B_{\varrho_n}(hf,g_n)\|_1 \le \|\Delta_n\|_1 \to 0 \tag{7.5}$$

with $g_n = g_{\varrho_n,f_n}$. For this g_n we have

$$B_{\varrho_n}(f,g_n)(x) = \int_{\mathbb{R}^{k_n+1}} f\Big(x - \sum_{i=1}^{k_n} \varrho_n^i y_i - \varrho_n^{k_n+1} z\Big) \prod_{j=1}^{k_n} f_n(y_j)\,dy_j g_n(z)\,dz$$

and obtain by an argument similar to the one used to derive (6.6),

$$\|n^{1/2}(B_{\varrho_n}(f,g_n) - \bar{g}_n)\|_1 \le \|f'\|_1 n^{1/2}\big(\varrho_n^{k_n+1}\|\iota_{\mathbb{R}} g_n\|_1 + \varrho^{k_n+1}\|\iota_{\mathbb{R}} g\|_1\big) \to 0. \tag{7.6}$$

Now we use the representations

$$B_{\varrho_n}(hf,g_n)(x) = E[(hf)(x - S_n)], \quad x \in \mathbb{R},$$

and

$$B_{\varrho}(hf,g)(x) = E[(hf)(x - S)], \quad x \in \mathbb{R},$$

with $S_n = \sum_{j=1}^\infty \varrho_n^j F_n^{-1}(U_j))$ and $S = \sum_{j=1}^\infty \varrho^j F^{-1}(U_j))$, where $U_1, U_2, \dots$ are independent uniform random variables and F_n^{-1} and F^{-1} are the left-inverses of the distribution functions F_n and F with respective densities f_n and f. We verify

$$E[|S_n - S|] \le \sum_{j=1}^\infty |\varrho_n^j - \varrho^j|\|\iota_{\mathbb{R}} f_n\|_1 + \sum_{j=1}^\infty \varrho^j \int_0^1 |F_n^{-1}(u) - F^{-1}(u)|\,du \to 0$$

with the help of Lemma 3.1, properties (2.6) and the inequality

$$\begin{aligned}\int_0^1 |F_n^{-1}(u) - F^{-1}(u)|\,du &= \int_{-\infty}^\infty |F_n(x) - F(x)|\,dx \\ &\le \int_{-\infty}^\infty |y||f_n(y) - f(y)|\,dy \to 0\end{aligned}$$

where the convergence to 0 follows from (7.4). An application of Lemma 3.4 yields

$$\|B_{\varrho_n}(hf,g_n) - B_{\varrho}(hf,g)\|_1 \to 0. \tag{7.7}$$

The desired (7.2) follows from (7.5)–(7.7).

To verify (7.3) we begin by observing the identity $\bar{g}_n = L_{\varrho_n,\{0\}}(f,f_n)$. An argument similar to the one that produced (6.9) yields

$$n^{1/2}\Big\|\bar{g}_n - \sum_{j=1}^{k_n} L_{\varrho_n,\{j\}}(f_n - f, f)\Big\|_1 \to 0,$$

now using $k_n\|f_n - f\|_1 = O(k_n n^{-1/2})$. Next, mimicking the argument that led to (6.10) yields

$$n^{1/2}\Big\|\sum_{j=1}^{k_n} L_{\varrho_n,\{j\}}(f_n - f, f) - \sum_{j=1}^{k_n} L_{\varrho,\{j\}}(f_n - f, f)\Big\|_1 \le \frac{C_n n^{1/2}|\varrho_n - \varrho|}{(1-\max\{\varrho_n,\varrho\})^2} \to 0$$

with $C_n = \|f'\|_1\big(\|\iota_{\mathbb{R}}(f_n - f)\|_1 + k_n\|f_n - f\|_1\|\iota_{\mathbb{R}} f\|_1\big) \to 0$. Finally, we have

$$n^{1/2}L_{\varrho,\{j\}}(f_n - f, f) - \Gamma_j(hf) = n^{1/2}B_{\varrho^j}(\gamma_j, f_n - f) - B_{\varrho^j}(\gamma_j, hf) = B_{\varrho^j}(\gamma_j, \Delta_n),$$

and Lemma 3.10, Corollary 3.2 and (7.4) imply

$$\Big\|\sum_{j=1}^{k_n} B_{\varrho^j}(\gamma_j, \Delta_n)\Big\|_1 \le \sum_{j=1}^{k_n} \|\gamma_j'\|_1 \varrho^j \|\iota_{\mathbb{R}}\Delta_n\|_1 \to 0.$$

Here we used the fact that hf integrates to zero. This completes the proof of (7.3).

References

[1] Billingsley, P. (1968). *Convergence of Probability Measures.* John Wiley & Sons. MR0233396

[2] Bryk, A. and Mielniczuk, J. (2005). Asymptotic properties of density estimates for linear processes: application of projection method. *J. Nonparametr. Stat.* **17** 121–133. MR2112690

[3] Chan, N. H. and Tran, L. T. (1992). Nonparametric tests for serial dependence. *J. Time Ser. Anal.* **13** 19–28. MR1149268

[4] Chanda, K. C. (1983). Density estimation for linear processes. *Ann. Inst. Statist. Math.* **35** 439–446. MR0739385

[5] Drost, F. C., Klaassen, C. A. J. and Werker, B. J. M. (1997). Adaptive estimation in time-series models. *Ann. Statist.* **25** 786–817. MR1439324

[6] 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. MR1424774

[7] Hart, J. D. and Vieu, P. (1990). Data-driven bandwidth choice for density estimation based on dependent data. *Ann. Statist.* **18** 873–890. MR1056341

[8] Honda, T. (2000). Nonparametric density estimation for a long-range dependent linear process. *Ann. Inst. Statist. Math.* **52** 599–611. MR1820739

[9] Klaassen, C. A. J. and Putter, H. (2005). Efficient estimation of Banach parameters in semiparametric models. *Ann. Statist.* **33** 307–346. MR2157805

[10] Koul, H. L. and Schick, A. (1997). Efficient estimation in nonlinear autoregressive time-series models. *Bernoulli* **3** 247–277. MR1468305

[11] Kreiss, J.-P. (1987). On adaptive estimation in autoregressive models when there are nuisance functions. *Statist. Decisions* **5** 59–76. MR0886878

[12] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces. Isoperimetry and Processes*. Ergebnisse der Mathematik und ihrer Grenzgebiete (3) **23**. Springer-Verlag, Berlin. MR1102015

[13] Müller, U. U., Schick, A. and Wefelmeyer, W. (2005). Weighted residual-based density estimators for nonlinear autoregressive models. *Statist. Sinica* **15** 177–195. MR2125727

[14] Müller, U. U., Schick, A. and Wefelmeyer, W. (2006). Efficient prediction for linear and nonlinear autoregressive models. *Ann. Statist.* **34** 2496–2533. MR2291508

[15] Rudin, W. (1974). *Real and Complex Analysis*, 2nd ed. McGraw-Hill, New York. MR0344043

[16] Schick, A. and Wefelmeyer, W. (2002). Efficient estimation in invertible linear processes. *Math. Meth. Statist.* **11** 358–379. MR1964452

[17] Schick, A. and Wefelmeyer, W. (2004a). Root n consistent and optimal density estimators for moving average processes. *Scand. J. Statist.* **31** 63–78. MR2042599

[18] Schick, A. and Wefelmeyer, W. (2004b). Functional convergence and optimality of plug-in estimators for stationary densities of moving average processes. *Bernoulli* **10** 889–917. MR2093616

[19] Schick, A. and Wefelmeyer, W. (2007). Root-n consistent density estimators of convolutions in weighted L_1-norms. *J. Statist. Plann. Inference* **137** 1765–1774. MR2323861

[20] Schick, A. and Wefelmeyer, W. (2008a). Convergence rates in weighted L_1-spaces of kernel density estimators for linear processes. *ALEA* **4** 117–129. MR2413090

[21] Schick, A. and Wefelmeyer, W. (2008b) Root-n consistency in weighted L_1-spaces for density estimators of invertible linear processes. *Stat. Inference Stoch. Process.* **11** 281–310. MR2438498

[22] Tran, L. T. (1992). Kernel density estimation for linear processes. *Stochastic Process. Appl.* **41** 281–296. MR1164181

[23] van der Vaart, A.W. and Wellner, J.A. (1996). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer Series in Statistics. Springer-Verlag, New York. MR1385671

[24] Wu, W. B. and Mielniczuk, J. (2002). Kernel density estimation for linear processes. *Ann. Statist.* **30** 1441–1459. MR1936325

[25] Yakowitz, S. (1989). Nonparametric density and regression estimation for Markov sequences without mixing assumptions. *J. Multivariate Anal.* **30** 124–136. MR1003712