
DOI: https://doi.org/10.4213/tvp1067
ASYMPTOTIC EFFICIENCY OF INVERSE ESTIMATORS

1. Introduction. General, abstract, inverse estimation concerns the recovery of an unknown input signal from blurred observations on a transformation of that signal. This transformation is known and we will assume that it is a bounded, injective, linear operator between Hilbert function spaces. Inverse estimation may also be referred to as indirect curve estimation and covers a rich variety of interesting practical problems ranging from imaging [4] and tomography [8] to all kinds of deconvolution or errors-in-variable models and physical systems like inverse heat conduction. An extensive summary of applications is given in [13].

Recently much attention has been focused on the convergence of the integrated mean square error (IMSE). For compact operators optimality of the convergence rate over smoothness classes has been established by

---

1) The first author was supported by NSF grant DMS 95-04485; the second author was supported by NSF grant DMS 95-04485 and a visitor's grant from the Nederlandse Stichting voor de Wiskunde.
Johnstone and Silverman [8], [9] at the level of constants, and for arbitrary bounded operators a lower bound for the rate has been derived in [14], although the precise constants were not obtained. Under suitable conditions, however, estimators were constructed that attained this optimal rate. Ibragimov and Hasminskii [6] focused on minimax risk for estimators of linear functionals in a direct curve estimation model.

In this paper we briefly review a construction method of inverse estimators that avoids preconditioning. Our main purpose is to address the question of optimality of these estimators related to weak convergence. More specifically we consider the question of asymptotic efficiency for indirect estimators of linear functionals in the sense of the Hájek-Le Cam «convolution theorem» [1], [17], [7], [2]. Although such a result boils down to application of an abstract version of this theorem, to the best of our knowledge such an application has not yet been carried out for the indirect curve estimation model in the literature, and we believe that it provides a nontrivial and relevant paradigm.

The construction of indirect curve estimators in Section 2 is based on the polar decomposition of the operator defining the model, along with Halmos' [5] version of the spectral theorem and a suitably defined generalized empirical characteristic function.

There is a dense linear submanifold of functionals for which the estimators are asymptotically normal at the $\sqrt{n}$-rate with optimal variance and hence asymptotically efficient in the sense of the Hájek-Le Cam convolution theorem. In the spectral domain these functionals should in some sense be divisible by the spectral density of the operator. Asymptotic efficiency results from showing that the tangent space to the indirect estimation model has codimension 1 and by showing that the projection of the influence function of the functional on this tangent space has the right length. See Sections 3 and 4 for details.

In Section 5 we consider the convolution theorem for a deconvolution by way of an example. The same example is used to show that for certain functionals outside the linear submanifold mentioned in the previous paragraph asymptotic normality of the estimators may hold true at a rate slower than $\sqrt{n}$. This virtually implies in particular that weak convergence of the signal as a whole in the ambient Hilbert space already breaks down at the level of finite dimensional distributions. Since it is well known that even in a direct model curve estimators do not converge weakly [11], [16], [15], it will not come as a surprise that this is also the case in the less stable indirect model. Therefore we have not further discussed this phenomenon here.

Let us conclude this section with a specification of the model. The present formulation will be of sufficient generality to include both indirect density and indirect regression estimation as special cases. Throughout this
paper \( L^2(\mathcal{D}, \mathcal{D}, \mu) =: L^2(\mu) \) and \( L^2(\mathcal{E}, \mathcal{E}, \nu) =: L^2(\nu) \) are separable Hilbert spaces over the real numbers, and the measures \( \mu \) and \( \nu \) are \( \sigma \)-finite. The unknown signal \( \theta \) to be estimated satisfies

\[
p = K\theta, \quad \theta \in \Theta \subset L^2(\mu),
\]

where \( K: L^2(\mu) \to L^2(\nu) \) is a known bounded, linear and injective operator and \( \Theta \) is a given subset. Without confusion an integral operator and its kernel will be denoted by the same symbol. It will be assumed that the kernel \( K \) is real without further reference. The data consist of an i.i.d. sample \( X_1, \ldots, X_n \) of random elements in some measurable space that are directly related to \( p \) in a sense to be made clear below. It should be noted that the inverse operator \( K^{-1} \) is typically unbounded in this kind of problem which is the cause of the extra difficulty of ill-posedness. There are two special models of particular interest where the relation between \( p \) and the sample can be made explicit. These models will be specified now.

A. Indirect density estimation. In this case \( p \) is itself a density and the data \( X_1, \ldots, X_n \) are independent copies of a random variable \( X \) with density \( p \). In most cases \( \theta \) is also a density and hence we will assume that \( \Theta \) is the class of all densities that are square integrable with respect to \( \mu \). The model can be summarized in the form

\[
X \sim p = K\theta, \quad \theta \in \Theta \subset L^2(\mu).
\]

It has been tacitly assumed that in this model \( K \) maps densities in \( L^2(\mu) \) into densities in \( L^2(\nu) \).

B. Indirect regression estimation. In this model the data \( X_1, \ldots, X_n \) are independent copies of \( X := (Y, Z) \) satisfying

\[
Y = (K\theta)(Z) + \varepsilon,
\]

where \( Z \) is the random design variable with values in \( \mathcal{E} \), and \( \varepsilon \) the real valued error variable independent of \( Z \) with zero mean and finite variance. The design variable is supposed to have a density \( f \) on \( \mathcal{E} \) such that

\[
\inf_{y \in \mathcal{E}} f(y) > 0, \quad \int_{\mathcal{E}} f \, d\nu = 1,
\]

which we are free to choose. It should be noted that (1.4) can be only fulfilled if \( \nu(\mathcal{E}) < \infty \) as is often the case in the regression model. Here we take \( \Theta := L^2(\mu) \).

2. Construction of estimators and further preliminaries. The polar decomposition [12] entails the existence of a partial isometry \( V: L^2(\mu) \to L^2(\nu) \) such that \( K = VR \) and \( K^* = RV^* \), where \( R := (K^*K)^{1/2} \). Since \( K \) is injective, \( R \) is strictly positive Hermitian. According to Halmos' [5] version of the spectral theorem there exists a finite measure
space \((S, \mathcal{F}, \Sigma)\), a unitary operator \(U: L^2(\mu) \to L^2(\Sigma) := L^2(S, \mathcal{F}, \Sigma)\), and a function \(\rho \in L^\infty(\Sigma)\) with \(\rho > 0\) in this case such that

\[ R = U^{-1} M_\rho U, \tag{2.1} \]

where \(M_\rho: L^2(\Sigma) \to L^2(\Sigma)\) is the multiplication operator defined by \(M_\rho \varphi = \rho \cdot \varphi, \varphi \in L^2(\Sigma)\). Equation (1.1) can now be written as \(p = V U^{-1} M_\rho U \theta\) or, equivalently, as

\[ \chi := W p = \rho \cdot U \theta, \quad \theta \in \Theta \subset L^2(\mu), \quad \text{where } W := UV^*. \tag{2.2} \]

This equivalence is apparent from the fact that \(V^* V\) is equal to the identity on the range of \(R\). The function \(\chi\) on the left in (2.2) will be referred to as generalized characteristic function. The following assumption will be very convenient.

\textbf{Assumption A.} The operators \(K, U,\) and \(W\) are integral operators with bounded kernels, i.e.,

\[ K \in L^\infty(\nu \times \mu), \quad U \in L^\infty(\Sigma \times \mu), \quad UV^* =: W \in L^\infty(\Sigma \times \nu). \tag{2.3} \]

We don’t require \(V\) to be an integral operator because we don’t want to exclude simple cases where \(V\) is the identity. Our first concern is to show that in either model \(\chi\) admits a natural unbiased estimator. Let \(\eta\) and \(\zeta\) be measurable functions on the range of \(X\) with values in \(\mathbb{R}\) and \(\mathbb{E}\) respectively, so that they produce random elements \(\eta(X)\) and \(\zeta(X)\). It will be assumed that

\[ \mathbb{E} \eta^{2+\delta}(X) < \infty, \quad \text{for some } \delta > 0. \tag{2.4} \]

The generalized empirical characteristic function, which contains all the relevant randomness, is defined as

\[ \hat{\chi}(s) := \frac{1}{n} \sum_{k=1}^n \eta(X_k) W(s, \zeta(X_k)) =: \frac{1}{n} \sum_{k=1}^n \hat{\chi}_k(s), \quad s \in S. \tag{2.5} \]

\textbf{Lemma 2.1.} The generalized empirical characteristic function satisfies

\[ \hat{\chi} \in L^\infty(\Sigma); \quad 1_S \hat{\chi} \in L^2(\Sigma) \quad \forall S \in \mathcal{F}: \Sigma(S) < \infty. \tag{2.6} \]

Choosing

\[ \eta(x) := 1, \quad \zeta(x) := x, \quad x \in \mathbb{E}, \tag{2.7} \]

for Model A, and

\[ \eta(x) := \frac{y}{f(z)}, \quad \zeta(x) := z, \quad x := (y, z) \in \mathbb{R} \times \mathbb{E}, \tag{2.8} \]

for Model B, we have

\[ \mathbb{E} \hat{\chi}(s) = \chi(s), \quad s \in S. \tag{2.9} \]

In other words, \(\hat{\chi}\) is an unbiased estimator in either case.
Proof. Claim (2.6) is obvious. In the situation of Model \( A \) we have
\[
\mathbb{E} \hat{\chi}(s) = \mathbb{E} W(s, X) = \int_{X} W(s, x) p(x) \, d\mu(x) = (Wp)(s) = \chi(s). \tag{2.10}
\]
For Model \( B \) we have \( \mathbb{E}(Y \mid Z) = (K\theta)(Z) \) so that
\[
\mathbb{E} \hat{\chi}(s) = \mathbb{E} \mathbb{E} \left( \frac{Y}{f(Z)} W(s, Z) \mid Z \right) = \mathbb{E} \frac{W(s, Z)(K\theta)(Z)}{f(Z)}
= \int_{X} \frac{W(s, z)}{f(z)} p(z) f(z) \, d\mu(z) = (Wp)(s) = \chi(s), \tag{2.11}
\]
as before. Note that condition (1.4) is used in (2.11).

Solving \( \theta \) from (2.2) one obtains
\[
\theta = U^{-1} \frac{\chi}{\rho}. \tag{2.12}
\]
However, if \( \chi \) is only approximately known, as in our case, the inversion of the multiplication operator \( M_{\rho} \) requires regularization since this inverse will be typically unbounded. Before we can formally define the estimators, we need one further assumption.

Assumption B. The spectral function \( \rho \) satisfies
\[
\Sigma \{ s \in S : \rho(s) \geq \alpha \} < \infty \quad \forall \, \alpha > 0. \tag{2.13}
\]

The estimator for the parameter \( \theta \) of actual interest will now be based on the spectral cut-off regularization of the inverse of \( M_{\rho} \) (see [10]), and is defined by
\[
\hat{\theta}_{\alpha} = U^{-1} 1_{\{\rho \geq \alpha\}} \frac{\hat{\chi}}{\rho}, \quad \alpha > 0. \tag{2.14}
\]
It should be noted that due to (2.6) and (2.13) we may infer that \( 1_{\{\rho \geq \alpha\}} \hat{\chi}/\rho \in L^{2}(\Sigma) \), so that \( \hat{\theta}_{\alpha} \) is indeed well defined. Clearly, the expectation of the estimator equals
\[
\theta_{\alpha} := \mathbb{E} \hat{\theta}_{\alpha} = U^{-1} \left( 1_{\{\rho \geq \alpha\}} \frac{\chi}{\rho} \right), \quad \alpha > 0, \tag{2.15}
\]
because of (2.9). Although \( \theta \) is real valued by definition, it is not a priori clear that its estimator \( \hat{\theta}_{\alpha} \) is real as well. We have, however, the following lemma.

Lemma 2.2. Under the current assumptions, and more in particular because \( K \) has a real valued kernel, it follows that \( \hat{\theta}_{\alpha} \) is also real valued.

Proof. Let us introduce the strictly positive Hermitian operator \( T := K^{*}K = R^{2} \). We have \( T = U^{-1} M_{\tau} U \), where \( \tau := \rho^{2} \) and \( \tau : S \rightarrow [0, \|T\|] \). Its spectral cut-off type regularized inverse is given by
Asymptotic efficiency of inverse estimators

$T_{\alpha}^{-1} := U^{-1}M_{(1/\tau)1_{\{r \geq \alpha\}}}U$, $\alpha > 0$. For $n \in \mathbb{N}$, $n > 1/\alpha$, choose a continuous function $f_n: [0, \|T\|] \rightarrow [0, 1/\alpha]$ with

$$f_n(\xi) = \begin{cases} \frac{1}{\xi}, & \text{for } 0 < \xi \leq \|T\|, \\ 0, & \text{for } 0 < \xi \leq \alpha \frac{1}{n}, \end{cases} \quad (2.16)$$

and a real polynomial $P_n$ with

$$|P_n(\xi) - f_n(\xi)| \leq \frac{1}{n}, \quad 0 \leq \xi \leq \|T\|. \quad (2.17)$$

For every $n$, $P_n(T)$ is a linear map $L^2(\mu) \rightarrow L^2(\mu)$, and if $f$ is real then $(P_n(T))f$ is real. Hence, we are done if $P_n(T) \rightarrow T_{\alpha}^{-1}$ pointwise on $L^2(\mu)$, which is the case if $UP_n(T)U^{-1} \rightarrow UT_{\alpha}^{-1}U^{-1}$ pointwise on $L^2(\Sigma)$. Now we have

$$UT_{\alpha}^{-1}U^{-1} = M_{(1/\tau)1_{\{r \geq \alpha\}}}, \quad (2.18)$$

$$UP_n(T)U^{-1} = P_n(UTU^{-1}) = P_n(M_\tau) = MP_{n0}. \quad (2.19)$$

Hence it remains to show that for all $\varphi \in L^2(\Sigma)$ we have

$$(P_n \circ \tau) \cdot \varphi \rightarrow \frac{1}{\tau} 1_{\{r \geq \alpha\}} \varphi. \quad (2.20)$$

By our choice of $P_n$ we have $|P_n(\xi)| \leq \|T\| + 1/n$ for all $n$ and $\xi$, and $P_n(\xi) \rightarrow 1/\xi$ if $\xi \geq \alpha$, $P_n(\xi) \rightarrow 0$ if $\xi < \alpha$. Hence it follows that

$$\|P_n \circ \tau\|_\infty \leq \|T\| + 1, \quad \text{for all } n, \quad (2.21)$$

$$P_n \circ \tau \rightarrow \frac{1}{\tau} 1_{\{r \geq \alpha\}}, \quad \text{pointwise on } S. \quad (2.22)$$

Finally, this entails (2.20) by the dominated convergence theorem.

To conclude the proof let us construct estimators of $\theta$ in a different way, restricting ourselves to Model A. (We can deal with Model B in a similar way.) The random function

$$\hat{q} := \frac{1}{n} \sum_{k=1}^{n} K^*(\cdot, X_k) \quad (2.23)$$

is real and in $L^2(\mu)$ with $\mathbb{E} \|\hat{q}\|^2 < \infty$. Since it is an unbiased estimator of $q := K^*p = T\theta$, a reasonable estimator of $\theta$ seems to be

$$\tilde{\theta}_\alpha := T_{\alpha}^{-1}\hat{q}, \quad \alpha > 0. \quad (2.24)$$

It follows from what we have seen above that $\tilde{\theta}_\alpha$ is real. It is not hard to see that $\{\tilde{\theta}_\alpha: \alpha > 0\} = \{\hat{\theta}_\alpha: \alpha > 0\}$ so that also the $\hat{\theta}_\alpha$ must be real.
3. Weak convergence. For a sequence of random elements in a Hilbert space to converge weakly, convergence of the finite dimensional distributions is a necessary condition; see, e.g., [3]. For ordinary direct curve estimation, these finite dimensional distributions still do converge weakly, but the tightness breaks down. We shall see that in the present case of indirect curve estimation there will be a problem with the weak convergence of the finite dimensional distributions. The one-dimensional distributions, in particular, are the distributions of the inner products of the random element and the vectors in the Hilbert space. In the present situation, these inner products do not in general converge weakly for every vector in $L^2(\mu)$ at the same rate, if they converge at all. For vectors in the linear submanifold

$$\mathcal{L} := \{ f := U^{-1}(\rho \cdot \varphi) : \varphi \in L^1(\Sigma) \cap L^2(\Sigma) \}$$

however, this weak convergence can be established. Let us introduce

$$D(s, t) := \mathbb{E}\left\{ \eta(X) W(s, \zeta(X)) - \chi(s) \right\} \left\{ \eta(X) W(t, \zeta(X)) - \chi(t) \right\}$$

$$= \mathbb{E} \eta^2(X) W(s, \zeta(X)) W(t, \zeta(X)) - \chi(s) \chi(t), \quad s, t \in S,$$

and note that the covariance function of the generalized empirical characteristic process equals

$$\mathbb{E}\{ \tilde{\chi}(s) - \chi(s) \} \{ \tilde{\chi}(t) - \chi(t) \} = \frac{1}{n} D(s, t), \quad s, t \in S.$$ (3.3)

Conditions (2.3) and (2.4) entail that

$$\sup_{s, t \in S} |D(s, t)| < \infty.$$ (3.4)

For the finite dimensional distributions the vectors will be also restricted to real valued functions in $\mathcal{L}$ which amounts to functions in $\mathcal{L} \cap L^2(\mu)$. Given any finite number $f_1, \ldots, f_m \in \mathcal{L} \cap L^2(\mu)$ of such functions, let $\varphi_1, \ldots, \varphi_m \in L^1(\Sigma) \cap L^2(\Sigma)$ be the corresponding elements in the spectral domain (that may still be complex valued) and introduce the numbers

$$\Delta_{ij} := \int_S \int_S \overline{\varphi_i(s)} D(s, t) \varphi_j(t) \, d\Sigma(s) \, d\Sigma(t).$$ (3.5)

They determine a semi-definite positive Hermitian matrix $\Delta$, say, with real entries.

**Theorem 3.1.** For $f_j := U^{-1}(\rho \cdot \varphi_j) \in \mathcal{L} \cap L^2(\mu)$, $j = 1, \ldots, m$, in the situation of either Model A or B we have

$$\sqrt{n} \left( \langle \hat{\theta}_{\alpha(n)} - \theta_{\alpha(n)}, f_1 \rangle, \ldots, \langle \hat{\theta}_{\alpha(n)} - \theta_{\alpha(n)}, f_m \rangle \right) \xrightarrow{d} \mathcal{N}(0, \Delta),$$ (3.6)

as $n \to \infty$, provided that $\alpha(n) \to 0$, as $n \to \infty$. 

Proof. For arbitrary $\lambda \in \mathbb{R}^m$ we see from (2.14) and (3.4) that

$$n \mathbf{D} \sum_{j=1}^{m} \lambda_j \langle \hat{\theta}_{\alpha(n)}, f_j \rangle = n \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \mathbf{E} \langle \hat{\theta}_{\alpha(n)} - \theta_{\alpha(n)}, f_i \rangle \langle f_j, \hat{\theta}_{\alpha(n)} - \theta_{\alpha(n)} \rangle \lambda_j$$

$$= n \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \mathbf{E} \langle U(\hat{\theta}_{\alpha(n)} - \theta_{\alpha(n)}), U f_i \rangle \langle U f_j, U(\hat{\theta}_{\alpha(n)} - \theta_{\alpha(n)}) \rangle \lambda_j$$

$$= n \sum_{i=1}^{m} \sum_{j=1}^{m} \lambda_i \left[ \mathbf{E} \int_{\rho \geq \alpha} \int_{\rho \geq \alpha} \varphi_i(s) \{ \hat{\chi}(s) - \chi(s) \} \times \{ \hat{\chi}(t) - \chi(t) \} \varphi_j(t) d\Sigma(s) d\Sigma(t) \right] \lambda_j \rightarrow \lambda^* \Delta \lambda, \quad (3.7)$$

as $n \to \infty$. For $\hat{\chi}_k$ see (2.5); let us write

$$T_{n,k} := \frac{1}{\sqrt{n}} \left| U^{-1} \frac{1}{\rho} 1_{\{ \rho \geq \alpha \}}(\hat{\chi}_k - \chi), \sum_{j=1}^{m} \lambda_j f_j \right|$$

$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{m} \lambda_j \langle 1_{\{ \rho \geq \alpha \}}(\hat{\chi}_k - \chi), \varphi_j \rangle, \quad (3.8)$$

and note that

$$\sqrt{n} \sum_{j=1}^{m} \lambda_j \langle \hat{\theta}_{\alpha(n)} - \theta_{\alpha(n)}, f_j \rangle = \sum_{k=1}^{n} T_{n,k}, \quad (3.9)$$

where the $T_{n,k}$ are i.i.d. for each $n$.

Because of (2.9) it follows that in both models $|\chi(s)| \leq \mathbf{E} |\hat{\chi}_k(s)| \leq \{ \mathbf{E} |\eta(X_k)| \} M_W$, for all $s \in S$, where $M_W$ is the essential supremum of the kernel $W$. Consequently, we have, with $\delta$ as in (2.4) and $\varphi := \sum_{j=1}^{m} \lambda_j \varphi_j$,

$$\sum_{k=1}^{n} \mathbf{E} |T_{n,k}|^{2+\delta} = n^{-\delta/2} \mathbf{E} \left| \int_{\rho \geq \alpha} \{ \hat{\chi}_k(s) - \chi(s) \} \varphi(s) d\Sigma(s) \right|^{2+\delta}$$

$$\leq n^{-\delta/2} \mathbf{E} \left[ \{ |\eta(X_k)| + \mathbf{E} |\eta(X_k)| \}^{2+\delta} \right]^{2+\delta} \left\{ \int |\varphi| d\Sigma \right\}^{2+\delta}$$

$$= O(n^{-\delta/2}) \quad \text{as } n \to \infty. \quad (3.10)$$

On the other hand, we have

$$\left\{ \sum_{k=1}^{n} \mathbf{D} T_{n,k} \right\}^{1+\delta/2} \rightarrow (\lambda^* \Delta \lambda)^{1+\delta/2} \quad \text{as } n \to \infty. \quad (3.11)$$

Apparently, Lyapunov's condition is fulfilled for the $T_{n,k}$ provided that $\lambda^* \Delta \lambda > 0$ in which case

$$\sqrt{n} \sum_{j=1}^{m} \lambda_j \langle \hat{\theta}_{\alpha(n)} - \theta_{\alpha(n)}, f_j \rangle \rightarrow N(0, \lambda^* \Delta \lambda) \quad \text{as } n \to \infty. \quad (3.12)$$
If $\lambda^* \Delta \lambda = 0$ the same holds true when the limiting normal distribution is interpreted as the law which is degenerate at 0. Claim (3.6) is obtained from (3.12) by exploiting the Cramér–Wold device.

**Corollary 3.1.** Let us choose $f := U^{-1} \rho \cdot \varphi \in \mathcal{L} \cap L^2(\mu)$ and assume that

$$\sqrt{n} (\theta_{\alpha(n)} - \theta, f ) \longrightarrow 0 \quad \text{as } n \to \infty. \quad (3.13)$$

Then it follows that

$$\sqrt{n} (\hat{\theta}_{\alpha(n)} - \theta, f ) \overset{d}{\longrightarrow} \mathcal{N}(0, \Delta^2(f)) \quad \text{as } n \to \infty, \quad (3.14)$$

where

$$\Delta^2(f) := \int_{\mathcal{S}} \int_{\mathcal{S}} \varphi(s) D(s, t) \varphi(t) d\Sigma(s) d\Sigma(t). \quad (3.15)$$

At this point some remarks are in order. For more specific considerations that are possible in less abstract models, we refer to Section 5.

1. The order of $\sqrt{n}$ in the weak convergence is due to the circumstance that taking an inner product is performing a smoothing operation. It should be noted that weak convergence of $0_\alpha(n)(s)$ at a fixed given point $s \in \mathcal{S}$, (3.16) will be of a different order. The order is also due to the convergence of the variance in (3.7) which hinges on the special assumptions regarding the functions $f_j$.

2. For functions $f_j \in L^2(\mu) \setminus \mathcal{L}$ there may still be weak convergence but of a different order depending on the rate at which $\alpha(n) \to 0$ as $n \to \infty$.

3. In general $\sqrt{n} (\hat{\theta}_{\alpha(n)} - \theta_{\alpha(n)})$ will not converge weakly to a Gaussian measure in $L^2(\mu)$. In the case of direct curve estimation all the finite dimensional distributions converge weakly, but there is a problem with the tightness. In the present case of indirect curve estimation we are not even assured that all the finite dimensional distributions converge weakly since (3.6) is restricted to functions in $\mathcal{L}$.

4. The weak convergence in (3.14), where centering is at the actual signal of interest, requires condition (3.13). For this condition to be satisfied a speed of convergence to zero of the $\alpha(n)$ is needed that does not in general correspond to speeds required for optimal convergence rates of the IMSE over smoothness classes.

4. **The convolution theorem and asymptotic efficiency.** Throughout this section let us focus on Model A of indirect density estimation, where $\Theta$ is the set of all probability densities in $L^2(\mu)$. For Model B a similar result could be derived but the proof would be a little more technical. The discussion is further facilitated by assuming that $K$ is itself strictly positive Hermitian, so that $K = R, L^2(\mu) = L^2(\nu)$, and

$$\mathcal{R}_K \quad \text{is dense in } L^2(\mu), \quad (4.1)$$
Asymptotic efficiency of inverse estimators

where $\mathcal{R}_K$ is the range of $K$. (In fact, (4.1) is necessary and sufficient for a Hermitian operator $K$ to be injective.) Property (4.1) may occasionally be useful to verify one of the conditions of the next lemma. Let

$$\mathcal{P} := \{\sqrt{K}\theta, \theta \in \Theta\} \subset L^2(\mu). \tag{4.2}$$

The tangent cone at $\sqrt{p} \in \mathcal{P}$ is the collection of all tangents at $\sqrt{p}$ to smooth curves «starting» at $\sqrt{p}$, and the closed linear hull of this cone is the tangent space and will be denoted by $T_{\sqrt{p}}(\mathcal{P})$.

For $p > 0$ on $X$ let us introduce the classes of functions

$$\mathcal{F}^+ := \left\{ f \in L^1(\mu) \cap L^2(\mu) : f \geq 0, \frac{Kf}{\sqrt{p}} \in L^2(\mu) \right\}, \tag{4.3}$$
$$\mathcal{F} := \left\{ f_1 - f_2, f_1, f_2 \in \mathcal{F}^+ \right\}, \tag{4.4}$$
$$\mathcal{G} := \left\{ \frac{Kf}{\sqrt{p}}, f \in \mathcal{F} \right\}. \tag{4.5}$$

**Lemma 4.1.** If $p > 0$ on $X$ and $\mathcal{G}$ is dense in $L^2(\mu)$ we have

$$T_{\sqrt{p}}(\mathcal{P}) = [\sqrt{p}]^\perp, \tag{4.6}$$

where $[\sqrt{p}]$ is the one-dimensional linear subspace spanned by $\sqrt{p}$.

**Proof.** The inclusion $T_{\sqrt{p}}(\mathcal{P}) \subset [\sqrt{p}]^\perp$ is immediate so let us focus on the reverse inclusion. Choose $f \in \mathcal{F}^+$ with $\int f \, d\mu > 0$ so that $\eta := f(f \, d\mu)^{-1} \in \Theta$. A curve in $\Theta$ is now obtained by setting $\theta_t := (\theta + t\eta)/(1 + t)$, $t > 0$, where $\theta := K^{-1}p$. This yields a curve $\sqrt{p_t} := \sqrt{K\theta_t}$, $t > 0$, in $\mathcal{P}$. Note that $\theta_0 = \theta$ and write $p_0 = p$.

We claim that

$$\frac{1}{t} (\sqrt{p_t} - \sqrt{p}) \to \frac{Kf}{2\sqrt{p}} - \frac{1}{2} \left\langle \frac{Kf}{\sqrt{p}}, \sqrt{p} \right\rangle \sqrt{p} \text{ as } t \downarrow 0, \text{ in } L^2(\mu). \tag{4.7}$$

To see this observe that $K\theta_t \to K\theta$ pointwise as $t \downarrow 0$, and that

$$\frac{1}{t} (\sqrt{K\theta_t} - \sqrt{K\theta}) = \frac{K(\eta - \theta)}{2\sqrt{K\theta}} - \frac{1}{\sqrt{K\theta_t} + \sqrt{K\theta}} \frac{1}{1 + t} K(\eta - \theta)$$

$$\frac{- K(\eta - \theta)}{2\sqrt{K\theta}} = \frac{K(\eta - \theta)}{2\sqrt{K\theta}} \left( \frac{2\sqrt{K\theta}}{\sqrt{K\theta_t} + \sqrt{K\theta}} \frac{1}{1 + t} - 1 \right). \tag{4.8}$$

The first factor in the last expression on the right in (4.8) is in $L^2(\mu)$, and the second factor is in absolute value bounded by 3 and converges pointwise to 0 as $t \downarrow 0$. By dominated convergence it follows that

$$\frac{1}{t} (\sqrt{p_t} - \sqrt{p}) \to \frac{K(\eta - \theta)}{2\sqrt{p}} \text{ as } t \downarrow 0, \text{ in } L^2(\mu). \tag{4.9}$$

The claim follows by noticing that $(\frac{1}{2} \sqrt{p}) K(\eta - \theta) = (\frac{1}{2} \sqrt{p}) K\eta_{-1/2} \sqrt{p} = (\frac{1}{2} \sqrt{p}) K\eta - ((\frac{1}{2} \sqrt{p}) K\eta, \sqrt{p}) \sqrt{p}$, because $K\eta$ is also a density.
By definition the function on the right in (4.7) is an element of $T_{\sqrt{\rho}}(\mathcal{P})$. Since $T_{\sqrt{\rho}}(\mathcal{P})$ is a linear space we may infer that even

$$g - \left\langle g, \sqrt{\rho} \right\rangle \sqrt{\rho} \in T_{\sqrt{\rho}}(\mathcal{P}) \quad \forall g \in \mathcal{G}. \quad (4.10)$$

Let $\Pi$ denote the orthogonal projection onto the subspace $[\sqrt{\rho}]^\perp$. With $g - \left\langle g, \sqrt{\rho} \right\rangle \sqrt{\rho} = \Pi g$ we may reformulate (4.10) as

$$\Pi \mathcal{G} \subset T_{\sqrt{\rho}}(\mathcal{P}). \quad (4.11)$$

By assumption $\mathcal{G}$ is dense in $L^2(\mathcal{E})$ which entails that $\Pi \mathcal{G}$ must be dense in $[\sqrt{\rho}]^\perp$. Because $T_{\sqrt{\rho}}(\mathcal{P})$ is closed we finally obtain

$$\text{cl}(\Pi \mathcal{G}) = [\sqrt{\rho}]^\perp \subset T_{\sqrt{\rho}}(\mathcal{P}), \quad (4.12)$$

which proves the lemma.

The verification that $\mathcal{G}$ in (4.5) is dense in $L^2(\mu)$ does not seem to be easy in its full generality. A fairly general case, however, arises when $\mu(\mathcal{E}) < \infty$, a condition which is often fulfilled in Model B. Now it makes sense to impose the stronger condition

$$\text{ess inf } \rho > 0. \quad (4.13)$$

In this situation we obviously have $\mathcal{F} = L^2(\mu)$ and

$$\mathcal{G} = \mathcal{R}_K. \quad (4.14)$$

Combination with (4.1) yields the following result.

**Lemma 4.2.** If $\mu(\mathcal{E}) < \infty$ and (4.13) is satisfied, $\mathcal{G}$ is dense in $L^2(\mu)$ and (4.6) holds true.

Throughout Section 5 we will be concerned with a given convolution operator on $L^2(\mathbb{R})$. Because $\mathbb{R}$ has infinite Lebesgue measure it is impossible to have a density $\rho$ which is bounded away from 0. Still, with some effort, it will be shown that $\mathcal{G}$ is dense in $L^2(\mathbb{R})$. Although it seems a fair conjecture that this result may be extended to certain classes of convolution operators, we will not pursue that question here.

The functional that we are interested in is

$$T(\rho) := \left\langle K^{-1} \rho, f \right\rangle, \quad p \in K\Theta, \quad (4.15)$$

where we still restrict ourselves to real valued $f$, more precisely to

$$f := U^{-1} \rho \cdot \varphi, \quad \varphi \in \mathcal{L} \cap L^2(\mu). \quad (4.16)$$

**Lemma 4.3.** The functional $T: K\Theta \to \mathbb{R}$ is differentiable at $\sqrt{\rho} \in \mathcal{P}$, satisfying $\sqrt{\rho} > 0$ on $\mathcal{E}$, relative to $T_{\sqrt{\rho}}(\mathcal{P})$ with derivative

$$2\sqrt{\rho}(U^{-1}\varphi). \quad (4.17)$$
Proof. Let \( t \to \sqrt{p_t}, 0 \leq t < \varepsilon \), be a smooth curve in \( \mathcal{P} \) with \( \sqrt{p_0} = \sqrt{p} \) and with derivative \( g \). For this curve we have

\[
\frac{1}{t} \{ T(p_t) - T(p) \} = \frac{1}{t} \langle K^{-1}(p_t - p), f \rangle = \frac{1}{t} \langle U^{-1} M_1 / p U(p_t - p), f \rangle = \frac{1}{t} \langle pt - p, U^{-1} \frac{1}{p} U f \rangle = \frac{1}{t} \{ \sqrt{p_t} - \sqrt{p} \}, \{ \sqrt{p_t} + \sqrt{p} \} U^{-1} \varphi \]

\[
\to \langle g, 2\sqrt{p} U^{-1} \varphi \rangle \text{ as } t \downarrow 0,
\]

(4.18)

exploiting the continuity of the inner product and the uniform boundedness of the \( \sqrt{p_t} \); by Assumption A we have

\[
0 \leq p_t(x) = \int K(x, y) \theta_t(y) \, d\mu(y) \leq \text{ess sup}_{x,y} |K(x, y)| \int \theta_t(y) \, d\mu(y)
\]

\[
\leq \text{ess sup}_{x,y} |K(x, y)| =: M_K < \infty \text{ for all } t.
\]

(4.19)

This proves the lemma.

Theorem 4.1. Let Assumptions A and B be fulfilled and suppose that \( \sqrt{p} > 0 \) on \( \mathcal{E} \). Any regular sequence of estimators of the functional \( T \) in (4.15) has limiting distribution of the form

\[
\mathcal{D} * \mathcal{N}(0, \sqrt{\langle \sqrt{p} (U^{-1} \varphi), \sqrt{p} (U^{-1} \varphi) \rangle}) = (4.20)
\]

where \( \mathcal{D} \) is a distribution on the real line. Under the extra assumption (3.13) the limiting normal distribution on the right in (3.14) equals the normal component in (4.20). Hence the statistics \( \langle \hat{\theta}_{\alpha(n)}, f \rangle \) are asymptotically efficient estimators of the functional.

Proof. All the conditions for application of Theorem 2.9 in [17] are satisfied; see also [1], [7] or [2]. The influence function of \( T \) being given by \( 2\sqrt{p} (U^{-1} \varphi) \) according to Lemma 4.3, the variance of the normal component of the limiting distribution equals one fourth of the squared length of the projection of this function onto \( T_{\sqrt{p}}(\mathcal{P}) \). Since the orthogonal complement is the one-dimensional space \( [\sqrt{p}] \) these considerations lead at once to the variance in (4.20).

To prove the asymptotic efficiency we need to compare this number with \( \Delta^2(f) \) in (3.15). It will be convenient to write

\[
f_0 := U^{-1} \varphi.
\]

(4.21)

On the one hand, we have

\[
\| \sqrt{p}(U^{-1} \varphi) - (\sqrt{p} (U^{-1} \varphi), \sqrt{p}) \sqrt{p} \|^2 = \| \sqrt{p} \cdot f_0 \|^2 - |(\sqrt{p} \cdot f_0, \sqrt{p})|^2
\]

\[
= \int |f_0|^2 \cdot p \, d\mu \int f_0 \cdot p \, d\mu.
\]

(4.22)
On the other hand, in the present case of Model A with $K = R$ and hence $V = I$, we see that

\[
\Delta^2(f) = \int_{S} \int_{S} \overline{\varphi(s)} U(s, X) \overline{U(t, X)} \varphi(t) d\Sigma(s) d\Sigma(t) \\
- \int_{S} \overline{\varphi(s)} \{E U(s, X) \} \{E U(t, X) \} \varphi(t) d\Sigma(s) d\Sigma(t) \\
= \int_{E} \left| \int_{S} U(t, x) \varphi(t) d\Sigma(t) \right|^2 p(x) d\mu(x) \\
- \left| \int_{E} \int_{S} U(t, x) \varphi(t) d\Sigma(t) p(x) d\mu(x) \right|^2 \\
= \int_{E} |f_0(x)|^2 p(x) d\mu(x) - \int_{E} f_0(x) p(x) d\mu(x) \right|^2, \tag{4.23}
\]

which apparently equals (4.22).

5. **Example: the «convolution theorem» for a deconvolution.**

A. **Construction of the estimators.** We will consider a special instance of the results described in Section 4. Let $K$ denote convolution with the double exponential density

\[
r(x) := \frac{1}{2} e^{-|x|}, \quad x \in \mathbb{R}, \tag{5.1}
\]
i.e., for $\theta \in L^2(\mathbb{R})$ we consider the integral equation

\[
p(x) = \int_{-\infty}^{\infty} r(x - y) \theta(y) dy = (r \ast \theta)(x), \quad x \in \mathbb{R}. \tag{5.2}
\]

Because $r \in L^1(\mathbb{R})$ it follows that $p \in L^2(\mathbb{R})$. For $\theta \in \Theta$, the class of all square integrable densities on $\mathbb{R}$, the resulting $p$ is also a density. The symmetry of $r$ entails that $K$ is Hermitian.

For any function $f \in L^1(\mathbb{R})$ the characteristic function is defined as

\[
\hat{f}(t) := \int_{-\infty}^{\infty} e^{itx} f(x) dx, \quad t \in \mathbb{R}. \tag{5.3}
\]

The Fourier transform is the unitary mapping $\mathcal{F} : L^2(\mathbb{R}) \rightarrow L^2(\mathbb{R})$ determined by

\[
\mathcal{F} f = \frac{1}{\sqrt{2\pi}} \hat{f}, \quad f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}). \tag{5.4}
\]

The characteristic function of $r$ equals

\[
\hat{r}(t) = \frac{1}{1 + t^2}, \quad t \in \mathbb{R}, \tag{5.5}
\]
and the operator $K$ has the representation

$$K = F^{-1}M_F F.$$  \hfill (5.6)

This representation is of the general form in (2.1) with $L^2(\mu) = L^2(\Sigma) = L^2(\mathbb{R})$, $U = F$, and $p = \tilde{r}$. Because $\tilde{r} > 0$ on $\mathbb{R}$ it follows that $K$ is itself strictly positive Hermitian: $K = R$ and hence $V = I$.

Obviously, the kernels of $K$ and $U = W = F$ are bounded, and

$$\lambda\{t: \tilde{r}(t) \geq \alpha\} = 2\sqrt{\frac{1 - \alpha}{\alpha}} < \infty \quad \forall \alpha \in (0, 1),$$  \hfill (5.7)

so that Assumptions A and B are satisfied. It follows from (2.5) that

$$\hat{\chi}(t) = \frac{1}{n} \sum_{k=1}^{n} F(t, X_k) = \frac{1}{\sqrt{2\pi n}} \sum_{k=1}^{n} e^{itX_k}, \quad t \in \mathbb{R}. \hfill (5.8)$$

Writing

$$\{\tilde{r} \geq \alpha\} = [-A, A], \quad A = A(\alpha) = \sqrt{\frac{1 - \alpha}{\alpha}},$$  \hfill (5.9)

according to (2.14) the estimator equals

$$\hat{\theta}_A(x) = \left(F^{-1} \left(\frac{\hat{\chi}}{\tilde{r}} \mathbf{1}_{[-A, A]}\right)\right)(x) = \frac{1}{2\pi n} \sum_{k=1}^{n} \int_{-A}^{A} e^{i(x_k - x)t}(1 + t^2) dt$$

$$= \frac{1}{2\pi n} \sum_{k=1}^{n} \int_{-A}^{A} (1 + t^2) \{\cos(x - X_k) t\} dt. \hfill (5.10)$$

Due to the symmetry of $1/\tilde{r}$ the imaginary part cancels and $\hat{\theta}$ is a real valued estimator indeed, which turns out to be of kernel type.

B. Asymptotic efficiency. By definition

$$\chi := Fp = \frac{1}{\sqrt{2\pi}} \tilde{p},$$  \hfill (5.11)

so that the current definition of both the actual and the empirical characteristic function differ by a factor $\sqrt{2\pi}$ from how they are usually defined. Let us choose

$$\varphi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}), \quad \text{real valued and symmetric.} \hfill (5.12)$$

Then, by definition,

$$f := F^{-1}(\tilde{r} \cdot \varphi) \in \mathcal{L}, \quad f_0 := F^{-1} \varphi \quad \text{are real valued.} \hfill (5.13)$$

Application of Theorem 4.1 yields the asymptotic efficiency of the estimator $\langle \hat{\theta}_A(n), f \rangle$ with limiting distribution

$$\sqrt{n}(\hat{\theta}_A(n) - \theta, f) \xrightarrow{d} \mathcal{N}(0, \Delta^2(f)) \quad \text{as } n \to \infty, \hfill (5.14)$$
provided only that $A(n) \to \infty$ at such a rate that
\[
\sqrt{n} \left\langle \theta_{A(n)} - \theta, f \right\rangle \to 0 \quad \text{as } n \to \infty, \quad (5.15)
\]
that $p > 0$, and that (cf. (4.5))
\[
\mathcal{G} = \left\{ \frac{r \ast f}{\sqrt{p}} : f \in L^1(\mathbb{R}) \cap L^2(\mathbb{R}) \right\} \quad \text{is dense in } L^2(\mathbb{R}). \quad (5.16)
\]

To find $A(n) =: A$ such that (5.15) is fulfilled, note that the inner product on the left equals
\[
\sqrt{n} \left| \left\langle \frac{x}{r} 1_{[-A, A]} - \mathcal{F} \theta, \tilde{r} \cdot \varphi \right\rangle \right| \leq \sqrt{n} \left( \int_{-\infty}^{-A} + \int_{A}^{\infty} \right) \left| (\mathcal{F} \theta)(t) \tilde{r}(t) \varphi(t) \right| dt
\]
\[
\leq \sqrt{n} \left( \int_{-\infty}^{-A} + \int_{A}^{\infty} \right) \frac{1}{1 + t^2} |\varphi(t)| dt
\]
\[
\leq 2\sqrt{n} ||\varphi|| \left\{ \int_{A}^{\infty} \left( \frac{1}{1 + t^2} \right)^2 dt \right\}^{1/2} = O(n^{1/2}A^{-3}) \quad \text{as } n \to \infty. \quad (5.17)
\]

Hence, if we choose
\[
A = A(n) = n^\xi \quad \text{for some } \xi > \frac{1}{6}, \quad (5.18)
\]
condition (5.15) is fulfilled.

Because $r$ has a strictly positive minimum on each compact interval it is immediate that $p = r \ast \theta > 0$ on $\mathbb{R}$ for each $\theta \in \Theta$. To verify (5.16) for any such $p$ first note that it suffices to prove that the closure of $\mathcal{G}$ contains all positive continuous functions that have compact support.

Thus, let $g: \mathbb{R} \to [0, \infty)$ be continuous with $g = 0$ outside some compact interval $[A, B]$, and let $\varepsilon > 0$. There is a $\delta > 0$ such that $\sqrt{p} \geq \delta$ on $[A, B]$ and there exists a $C^2$-function $h: \mathbb{R} \to [0, \infty)$ with
\[
\begin{cases}
|h - g\sqrt{p}| < \varepsilon \delta & \text{on } [A, B], \\
h = 0 & \text{on } [A, B]^c.
\end{cases} \quad (5.19)
\]

For any $g \in C^2$ with compact support we have $r \ast (g - g'') = g$. This follows by taking Fourier transforms: $(\mathcal{F}(r \ast (g - g'')))(t) = (1 + t^2)^{-1}((\mathcal{F}g)(t) + t^2(\mathcal{F}g)(t)) = (\mathcal{F}g)(t)$, $t \in \mathbb{R}$. Application to $h$ yields
\[
h = r \ast (h - h'') = r \ast ((h - h')^+) - r \ast ((h - h'')^-). \quad (5.20)
\]

For $x > 0$ we have
\[
e^x p(x) \geq \frac{1}{2} \int_0^\infty e^{x-y} \theta(x - y) dy. \quad (5.21)
\]
A similar expression holds true for \( x < 0 \) so that there exist \( 0 < C_1 < \infty \) and \( 0 < C_2 < \infty \) such that

\[
p(x) > C_1 e^{-|x|} \quad \text{for} \ |x| > C_2.
\]

(5.22)

It can also be easily shown that there exists \( 0 < C_3 < \infty \) such that

\[
r \ast 1_{[-D,D]} \leq C_3 e^{-|x|}, \quad x \in \mathbb{R}.
\]

(5.23)

Combination of (5.22) and (5.23) yields that

\[
\frac{r \ast 1_{[-D,D]}}{\sqrt{p}} \in L^2(\mathbb{R}).
\]

(5.24)

But this entails that for any \( g: \mathbb{R} \rightarrow [0, \infty) \) with compact support we must have

\[
\frac{r \ast g}{\sqrt{p}} \in L^2(\mathbb{R}).
\]

(5.25)

In view of the representation of \( h \) in (5.20) this last result entails that

\[
\frac{h}{\sqrt{p}} \in \mathcal{G}.
\]

(5.26)

The proof of (5.16) is concluded by observing that we have

\[
\left\| \frac{h}{\sqrt{p}} - g \right\| \leq \varepsilon \sqrt{B - A},
\]

(5.27)

since \(|h(x)/\sqrt{p(x)} - g(x)| \leq \varepsilon 1_{[A,B]}(x), \ x \in \mathbb{R}.

C. Asymptotic normality at different rates.

Here we will consider asymptotic normality of \( \langle \hat{\theta}_A, f \rangle \), centered at \( \langle \theta, f \rangle \), for \( f \in L^2(\mu) \setminus \mathcal{L} \). The suitable scaling factor depends on the choice of \( f \) and underlying \( \theta \), so that general statements can be hardly expected. We therefore limit ourselves to a specific choice of \( f \) and \( \theta \) which on the one hand is mathematically tractable and on the other hand yields a scaling factor different from \( \sqrt{n} \). In fact for \( \theta \) itself we will also take the double exponential density

\[
\theta(x) = \frac{1}{2} e^{-|x|}, \quad x \in \mathbb{R}, \quad \text{with} \quad \tilde{\theta}(t) = \frac{1}{1+t^2}, \quad t \in \mathbb{R},
\]

(5.28)

and we choose \( f =: f_\nu \in L^2(\mathbb{R}) \) in such a way that

\[
(\mathcal{F} f_\nu)(t) = (1 + t^2)^{-\nu} \quad \text{for some} \quad \frac{1}{4} < \nu < \frac{1}{2}.
\]

(5.29)

Larger values of \( \nu \) will not be suitable for our purposes. Note that, indeed, for \( \nu \) as indicated the function \( (\mathcal{F} f_\nu)/\widetilde{p} \) is not integrable or square integrable so that

\[
f_\nu \in L^2(\mathbb{R}) \setminus \mathcal{L}, \quad \frac{1}{4} < \nu < \frac{1}{2}.
\]

(5.30)
Since for this type of function the factor \( \tilde{r} \) no longer cancels in the expressions for the covariances we should rather introduce the kernel

\[
\tilde{S}(s, t) = \frac{D(s, t)}{\tilde{r}(s) \tilde{r}(t)} = \frac{\tilde{p}(s - t) - \tilde{p}(s) \tilde{p}(-t)}{\tilde{r}(s) \tilde{r}(t)} = \frac{\tilde{r}(s - t) \tilde{\theta}(s - t)}{\tilde{r}(s) \tilde{r}(t)} - \tilde{\theta}(s) \tilde{\theta}(-t), \quad s, t \in \mathbb{R}. \tag{5.31}
\]

In the present special case we have

\[
\tilde{S}(s, t) = \frac{(1 + s^2)(1 + t^2)}{\{1 + (s - t)^2\}^2} - \frac{1}{(1 + s^2)(1 + t^2)}. \tag{5.32}
\]

Employing the inner product with \( f_\nu \) in (5.29) we obtain

\[
\mathbf{D} \langle \hat{\theta}_A, f_\nu \rangle = \frac{1}{n} \int_{-A}^{A} \int_{-A}^{A} \frac{(1 + s^2)^{1-\nu}(1 + t^2)^{1-\nu}}{\{1 + (s - t)^2\}^2} \, ds \, dt - \frac{1}{n} \left\{ \int_{-A}^{A} (1 + s^2)^{-1-\nu} \, ds \right\}^2. \tag{5.33}
\]

It is clear that the second expression on the right in (5.33) is \( o(1) \) as \( n \to \infty \), for any choice of \( A \) so that the asymptotic order of the variance is determined by the first term. For application of the Lyapunov condition a lower bound is required:

\[
\frac{1}{n} \int_{-A}^{A} \int_{-A}^{A} \frac{(1 + s^2)^{1-\nu}(1 + t^2)^{1-\nu}}{\{1 + (s - t)^2\}^2} \, ds \, dt \\
\geq \frac{1}{n} \int_{-A}^{A} \int_{-A}^{A} \frac{(1 + s^2)^{1-\nu}(1 + t^2)^{1-\nu}}{\{1 + (2A)^2\}^2} \, ds \, dt \\
= \frac{1}{n} \frac{1}{(1 + 4A^2)^2} \left\{ \int_{-A}^{A} (1 + s^2)^{1-\nu} \, ds \right\}^2. \tag{5.34}
\]

Hence there exists a number \( 0 < c < \infty \) such that

\[
\mathbf{D} \langle \hat{\theta}_A, f_\nu \rangle = \mathbf{E} \langle \hat{\theta}_A - \theta_A, f_\nu \rangle^2 \geq \frac{c}{n} A^{2-4\nu}, \tag{5.35}
\]

for \( n \) and \( A \) sufficiently large.

Let us write

\[
\langle \hat{\theta}_A - \theta_A, f_\nu \rangle = \sum_{k=1}^{n} \frac{1}{n} \langle \tilde{\chi}_k - \chi, 1_{[-A, A]}, \mathcal{F} f_\nu \rangle =: \sum_{k=1}^{n} T_{n,k}^* \tag{5.36}
\]

For any \( \delta > 0 \) we have

\[
\mathbf{E}|T_{n,k}^*|^{2+\delta} \leq \mathbf{E} \left\{ \frac{1}{n} \int_{-A}^{A} |\tilde{\chi}_k(t) - \chi(t)| (1 + t^2)^{1-\nu} \, dt \right\}^{2+\delta} \\
\leq (2C)^{2+\delta} n^{-2-\delta} A^{(3-2\nu)(2+\delta)}, \tag{5.37}
\]

where $0 < C < \infty$ does not depend on $n$. It follows that

$$
\sum_{k=1}^{n} \frac{E |T_{n,k}^*|^{2+\delta}}{(D \hat{\theta}_A)^{1+\delta/2}} \leq \left( \frac{2C}{\sqrt{c}} \right)^{2+\delta} n^{-\delta/2} A^{4+2\delta}. \quad (5.38)
$$

If we choose

$$A = A(n) = n^\xi \text{ for some } 0 < \xi < \frac{1}{4}, \quad (5.39)$$

there exists a $\delta > 0$ such that the upper bound on the right in (5.38) tends to 0, as $n \to \infty$, i.e., such that Lyapunov's condition is satisfied. Defining

$$\sigma_n^2 := nD(\hat{\theta}_A(n), f_\nu) \geq cn^{2(2-4\nu)\xi}, \quad (5.40)$$

we have proved that

$$\sqrt{n} \frac{\langle \hat{\theta}_A(n) - \theta, f_\nu \rangle}{\sigma_n} \xrightarrow{d} \mathcal{N}(0, 1) \text{ as } n \to \infty. \quad (5.41)$$

In order to center at the actual parameter $\theta$ we need to verify that $(\sqrt{n}/\sigma_n) \times \text{bias}$ tends to zero for large sample sizes. Since $\theta$ and $f_\nu$ are given by (5.28) and (5.29) respectively we can be more specific than in (5.17). In fact we have

$$\sqrt{n} \left| \frac{\langle \hat{\theta}_A(n) - \theta, f_\nu \rangle}{\sigma_n} \right| \leq 2 \frac{\sqrt{n}}{\sigma_n} \int_{A(n)}^\infty (1 + t^2)^{-1-\nu} dt \leq \frac{2}{c} n^{1/2 - (2-4\nu)\xi - (1+2\nu)\xi}$$

$$= \frac{2}{c} n^{1/2 - \xi(3-2\nu)} \to 0 \text{ as } n \to \infty \text{ for } \xi > \frac{1}{6 - 4\nu}. \quad (5.42)$$

When we choose $\xi = (5 - 2\nu)/[8(3 - 2\nu)]$, it satisfies both (5.39) and (5.42) so that we have the following result.

**Theorem 5.1.** There exist $\theta \in \Theta$, $f \in L^2(\mathbb{R}) \setminus \mathcal{L}$, and numbers $A(n) \to \infty$ and $\sigma(n) \to \infty$ as $n \to \infty$, such that

$$\sqrt{n} \frac{\langle \hat{\theta}_A(n) - \theta, f \rangle}{\sigma_n} \xrightarrow{d} \mathcal{N}(0, 1) \text{ as } n \to \infty. \quad (5.43)$$

In other words, weak convergence to a normal distribution is possible at a rate different from $\sqrt{n}$.

The authors are grateful to the editors and a referee for careful reading of the manuscript and useful comments.

**REFERENCES**


Поступила в редакцию
12.XI.1997
Переработанный вариант
28.IV.1998