

# A uniform central limit theorem and efficiency for deconvolution estimators

Jakob Söhl\* and Mathias Trabs\*

*Institut für Mathematik,*

*Humboldt-Universität zu Berlin,*

*Unter den Linden 6, 10099 Berlin, Germany*

*e-mail:* [soehl@math.hu-berlin.de](mailto:soehl@math.hu-berlin.de); [trabs@math.hu-berlin.de](mailto:trabs@math.hu-berlin.de)

**Abstract:** We estimate linear functionals in the classical deconvolution problem by kernel estimators. We obtain a uniform central limit theorem with  $\sqrt{n}$ -rate on the assumption that the smoothness of the functionals is larger than the ill-posedness of the problem, which is given by the polynomial decay rate of the characteristic function of the error. The limit distribution is a generalized Brownian bridge with a covariance structure that depends on the characteristic function of the error and on the functionals. The proposed estimators are optimal in the sense of semiparametric efficiency. The class of linear functionals is wide enough to incorporate the estimation of distribution functions. The proofs are based on smoothed empirical processes and mapping properties of the deconvolution operator.

**AMS 2000 subject classifications:** 62G05, 60F05.

**Keywords and phrases:** Deconvolution, Donsker theorem, efficiency, distribution function, smoothed empirical processes, Fourier multipliers.

Received September 2012.

## 1. Introduction

Our observations are given by  $n \in \mathbb{N}$  independent and identically distributed random variables

$$Y_j = X_j + \varepsilon_j, \quad j = 1, \dots, n, \quad (1)$$

where  $X_j$  and  $\varepsilon_j$  are independent of each other, the distribution of the errors  $\varepsilon_j$  is supposed to be known and the aim is statistical inference on the distribution of  $X_j$ . Let us denote the densities of  $X_j$  and  $\varepsilon_j$  by  $f_X$  and  $f_\varepsilon$ , respectively. We consider the case of ordinary smooth errors, which means that the characteristic function  $\varphi_\varepsilon$  of the errors  $\varepsilon_j$  decays with polynomial rate, determining the ill-posedness of the inverse problem. The contribution of this article to the well studied problem of deconvolution is twofold. First, we prove a uniform central limit theorem for kernel estimators of the distribution function of  $X_j$  in the setting of  $\sqrt{n}$  convergence rates. More precisely, the theorem does not only include the estimation of the distribution function, but covers translation classes

---

\*The authors thank Richard Nickl and Markus Reiß for helpful comments and discussions. The first author thanks the Statistical Laboratory in Cambridge for the hospitality during a stay in which part of the project was started. This research was supported by the Deutsche Forschungsgemeinschaft through the SFB 649 “Economic Risk”.

of linear functionals of the density  $f_X$  whenever the ill-posedness is smaller than the smoothness of the functionals. Second, we obtain more exact results than the minimax rates of convergence by showing that the used estimators are optimal in the sense of semiparametric efficiency.

The classical Donsker theorem plays a central role in statistics. For an independent, identically distributed sample  $X_1, \dots, X_n$  from a distribution function  $F$  on the real line, the empirical distribution function is defined by  $F_n(t) := \frac{1}{n} \sum_{j=1}^n \mathbb{1}_{\{X_j \leq t\}}$ . The classical Donsker theorem states that  $\sqrt{n}(F_n - F)$  converges in law to  $\mathbb{G}$  in the space of bounded functions  $\ell^\infty(\mathbb{R})$ , where  $\mathbb{G}$  is obtained by the composition of the distribution function  $F$  and a standard Brownian bridge. In the deconvolution model (1) the random variables  $X_1, \dots, X_n$  are not observed, but only  $Y_1, \dots, Y_n$ , which contain an additive error. In this model we consider estimators  $\hat{\vartheta}_t$  of linear functionals  $t \mapsto \vartheta_t := \int \zeta(x - t) f_X(x) dx$ , where the special case  $\zeta := \mathbb{1}_{(-\infty, 0]}$  leads to the estimation of the distribution function  $F$ . Our Donsker theorem states that  $\sqrt{n}(\hat{\vartheta}_t - \vartheta_t)_{t \in \mathbb{R}}$  converges in law to a centered Gaussian Borel random variable  $\mathbb{G}$  in  $\ell^\infty(\mathbb{R})$ . This generalization allows to consider functionals  $\vartheta_t$  as long as the smoothness of  $\zeta$  in an  $L^2$ -Sobolev sense compensates the ill-posedness of the problem. The limiting process  $\mathbb{G}$  in the uniform central limit theorem is a generalized Brownian bridge, whose covariance depends on the functional  $\zeta$  and through the deconvolution operator  $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$  also on the distribution of the errors. The used kernel estimators  $\hat{\vartheta}_t$  are minimax optimal since they converge with a  $\sqrt{n}$ -rate. So investigating optimality further leads naturally to the question whether the asymptotic variance of the estimators is minimal, as in the case of the empirical distribution function in the classical Donsker theorem. We prove that the estimator  $\hat{\vartheta}_\bullet$  is efficient in the sense of a Hájek–Le Cam convolution theorem. In particular, the asymptotic covariance matrices of the finite dimensional distributions achieve the Cramér–Rao information bound. By uniform convergence and efficiency the kernel estimator of  $f_X$  fulfills the ‘plug-in’ property of Bickel and Ritov (2003) in the deconvolution model (1).

The deconvolution problem has attracted much attention so we mention here only closely related works and refer the interested reader to the references therein. The classical works by Fan (1991a,b) contain asymptotic normality of kernel density estimators as well as minimax convergence rates for estimating the density and the distribution function. Butucea and Comte (2009) have treated the data-driven choice of the bandwidth for estimating functionals of  $f_X$  but assumed some minimal smoothness and integrability conditions on the functional  $\vartheta_t$ , which exclude, for example,  $\zeta := \mathbb{1}_{(-\infty, 0]}$  since it is not integrable. Dattner, Goldenshluger and Juditsky (2011) have studied minimax-optimal and adaptive estimation of the distribution function. Asymptotic normality of estimators for the distribution function has been shown for supersmooth errors by van Es and Uh (2005) and for ordinary smooth errors by Hall and Lahiri (2008). In contrast we consider the estimation of general linear functionals and are interested in uniform convergence. Uniform results have been studied for the density but not for the distribution function by Bissantz et al. (2007) and by Lounici and Nickl (2011). Nickl and Reiß (2012) have proved a Donsker theorem

for estimators of the distribution function of a Lévy measure. Their set-up is related to ours but concerns Lévy measures, while we are in a deconvolution context. We consider the more general problem of estimating linear functionals efficiently, which contains estimating the distribution function as a special case and provides clear insight in the interplay between smoothness of  $\zeta$  and the ill-posedness of the problem. Efficiency has been investigated in various semiparametric models, see Bickel et al. (1998) for an overview. In an abstract inverse problem van Rooij, Ruymgaart and van Zwet (1999) have derived a convolution theorem, which applies to the deconvolution set-up. However, their result only allows a rather restrictive class of linear functionals depending on the error density. In the related Lévy setting, Nickl and Reiß (2012) have shown heuristically that their estimator achieves the lower bound of the variance while a rigorous proof remained open.

In order to show the uniform central limit theorem in the deconvolution problem, we prove that the empirical process  $\sqrt{n}(\mathbb{P}_n - \mathbb{P})$  is tight in the space of bounded functions acting on the class

$$\mathcal{G} := \{\mathcal{F}^{-1}[1/\varphi_\varepsilon(-\bullet)] * \zeta_t \mid t \in \mathbb{R}\}, \quad \zeta_t := \zeta(\bullet - t),$$

where  $\mathbb{P}$  and  $\mathbb{P}_n = \frac{1}{n} \sum_{j=1}^n \delta_{Y_j}$  denote the true and the empirical probability measure of the observations  $Y_j$ , respectively. Since  $\mathcal{G}$  may consist of translates of an unbounded function, this is in general not a Donsker class. Nevertheless, Radulović and Wegkamp (2000) have observed that a smoothed empirical processes might converge even when the unsmoothed process does not. Giné and Nickl (2008) have further developed these ideas and have shown uniform central limit theorems for kernel density estimators. Nickl and Reiß (2012) used smoothed empirical processes in the inverse problem of estimating the distribution function of Lévy measures.

In order to show semiparametric efficiency in the deconvolution problem, the main problem is to show that the efficient influence function is indeed an element of the tangent space. If the regularity of  $\zeta$  is small, the standard methods given in the monograph of Bickel et al. (1998) do not apply in this ill-posed problem. Instead, we approximate  $\zeta$  by a sequence of smooth  $(\zeta_n)$  and show the convergence of the information bounds. Interestingly, this reveals a relation between the intrinsic metric of the limit  $\mathbb{G}$  and the metric which is induced by the inverse Fisher information. Additionally to techniques of smoothed empirical processes and the calculus of information bounds, our proofs rely on the Fourier multiplier property of the underlying deconvolution operator  $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$ , which is related to pseudo-differential operators as noted in the Lévy process setting by Nickl and Reiß (2012) and in the deconvolution context by Schmidt-Hieber, Munk and Dümbgen (2012). Important for our proofs are the mapping properties of  $\mathcal{F}^{-1}[1/\varphi_\varepsilon]$  on Besov spaces.

This paper is organized as follows: In Section 2 we formulate the Donsker theorem and discuss its consequences. Efficiency is then considered in Section 3. All proofs are deferred to Sections 4 and 5. In the Appendix we summarize definitions and properties of the function spaces used in the paper.

## 2. Uniform central limit theorem

### 2.1. The estimator

According to the observation scheme (1),  $Y_j$  are distributed with density  $f_Y = f_X * f_\varepsilon$  determining the probability measure  $\mathbb{P}$ . The characteristic function  $\varphi$  of  $\mathbb{P}$  can be estimated by its empirical version  $\varphi_n(u) = \frac{1}{n} \sum_{j=1}^n e^{iuY_j}$ ,  $u \in \mathbb{R}$ . For  $\zeta$  to be specified later and recalling  $\zeta_t = \zeta(\bullet - t)$ , our aim is to estimate functionals of the form

$$\vartheta_t := \langle \zeta_t, f_X \rangle = \int \zeta_t(x) f_X(x) dx. \quad (2)$$

Defining the Fourier transform by  $\mathcal{F}f(u) := \int e^{iux} f(x) dx$ ,  $u \in \mathbb{R}$ , the natural estimator of the functional  $\vartheta_t$  is given by

$$\hat{\vartheta}_t := \int \zeta_t(x) \mathcal{F}^{-1} \left[ \mathcal{F} K_h \frac{\varphi_n}{\varphi_\varepsilon} \right] (x) dx, \quad (3)$$

where  $K$  is a kernel,  $h > 0$  the bandwidth and we have written as usual  $K_h(x) = h^{-1}K(x/h)$ . Choosing  $\mathcal{F}K = \mathbb{1}_{[-\kappa, \kappa]}$  for some  $\kappa > 0$  leads to the estimator proposed by Butucea and Comte (2009). Throughout, we suppose that

- (i)  $K \in L^1(\mathbb{R}) \cap L^\infty(\mathbb{R})$  is symmetric and band-limited with  $\text{supp}(\mathcal{F}K) \subseteq [-1, 1]$ ,
- (ii) for  $l = 1, \dots, L$

$$\int K = 1, \quad \int x^l K(x) dx = 0, \quad \int |x^{L+1} K(x)| dx < \infty \quad \text{and} \quad (4)$$

- (iii)  $K \in C^1(\mathbb{R})$  satisfies, denoting  $\langle x \rangle := (1 + x^2)^{1/2}$ ,

$$|K(x)| + |K'(x)| \lesssim \langle x \rangle^{-2}. \quad (5)$$

Throughout, we write  $A_p \lesssim B_p$  if there exists a constant  $C > 0$  independent of the parameter  $p$  such that  $A_p \leq CB_p$ . If  $A_p \lesssim B_p$  and  $B_p \lesssim A_p$ , we write  $A_p \sim B_p$ . Examples of such kernels can be obtained by taking  $\mathcal{F}K$  to be a symmetric function in  $C^\infty(\mathbb{R})$  which is supported in  $[-1, 1]$  and constant to one in a neighborhood of zero. The resulting kernels are called flat top kernels and were used in deconvolution problems, for example, by Bissantz et al. (2007).

### 2.2. Statement of the theorem

Given a function  $\zeta$  specified later, our aim is to show a Donsker theorem for the estimator over the class of translations  $\zeta_t$ ,  $t \in \mathbb{R}$ . In view of the classical Donsker theorem in a model without additive errors, where no assumptions on the smoothness of the distribution are needed, we want to assume as less smoothness of  $f_X$  as possible still guaranteeing  $\sqrt{n}$ -rates. For some  $\delta > 0$  the following assumptions on the density  $f_X$  will be needed:

**Assumption 1.**

- (i) Let  $f_X$  be bounded and assume the moment condition  $\int |x|^{2+\delta} f_X(x) dx < \infty$ .
- (ii) Assume  $f_X \in H^\alpha(\mathbb{R})$  that is the density has Sobolev smoothness of order  $\alpha \geq 0$ .

We refer to the appendix for an exact definition of the Sobolev space  $H^\alpha(\mathbb{R})$ . Boundedness of the observation density  $f_Y$  follows immediately from (i) since  $\|f_Y\|_\infty \leq \|f_X\|_\infty \|f_\varepsilon\|_{L^1} < \infty$ . In addition to the smoothness of  $f_X$ , the smoothness of  $\zeta$  will be crucial. We assume for  $\gamma_s, \gamma_c > 0$

$$\zeta \in Z^{\gamma_s, \gamma_c} := \left\{ \zeta = \zeta^c + \zeta^s \mid \begin{array}{l} \zeta^s \in H^{\gamma_s}(\mathbb{R}) \text{ is compactly supported as well} \\ \text{as } \langle x \rangle^\tau (\zeta^c(x) - a(x)) \in H^{\gamma_c}(\mathbb{R}) \text{ for some } \tau > 0 \text{ and} \\ \text{some } a \in C^\infty(\mathbb{R}) \text{ such that } a' \text{ is compactly supported} \end{array} \right\} \quad (6)$$

and write for  $\zeta \in Z^{\gamma_s, \gamma_c}$  with a given decomposition  $\zeta = \zeta^s + \zeta^c$

$$\|\zeta\|_{Z^{\gamma_s, \gamma_c}} := \|\zeta^s\|_{H^{\gamma_s}} + \left\| \frac{1}{ix+1} \zeta^c(x) \right\|_{H^{\gamma_c}},$$

which is finite since the term  $\left\| \frac{1}{ix+1} \zeta^c(x) \right\|_{H^{\gamma_c}}$  can be bounded by  $\left\| \frac{a(x)}{ix+1} \right\|_{H^{\gamma_c}} + \left\| \frac{1}{(ix+1)\langle x \rangle^\tau} \|C^s\| \langle x \rangle^\tau (\zeta^c(x) - a(x)) \right\|_{H^{\gamma_c}} < \infty$  for any  $s > \gamma_c$ . Let us give two examples for  $\zeta$  and corresponding  $\gamma_s, \gamma_c$ .

*Example 1.* To estimate the distribution function of  $X_j$ , one has to consider translations of the indicator function  $\mathbb{1}_{(-\infty, 0]}(x)$ ,  $x \in \mathbb{R}$ . Let  $a$  be a monotone decreasing  $C^\infty(\mathbb{R})$  function, which is for some  $M > 0$  equal to zero for all  $x \geq M$  and equal to one for all  $x \leq -M$ . We define  $\zeta^s := \mathbb{1}_{(-\infty, 0]} - a$  and  $\zeta^c := a$ . From the bounded variation of  $\zeta^s$  follows  $\zeta^s \in B_{1,\infty}^1(\mathbb{R}) \subseteq H^{\gamma_s}(\mathbb{R})$  for any  $\gamma_s < 1/2$  by Besov smoothness of bounded variation functions (51) as well as by the Besov space embeddings (46) and (47). Since  $a \in C^\infty(\mathbb{R})$  and  $a'$  is compactly supported, the condition on  $\zeta^c$  is satisfied for any  $\gamma_c > 0$ . Hence,  $\mathbb{1}_{(-\infty, t]} \in Z^{\gamma_s, \gamma_c}$  if  $\gamma_s < 1/2$ . On the other hand, this cannot hold for  $\gamma_s > 1/2$  since  $H^{\gamma_s}(\mathbb{R}) \subseteq C^0(\mathbb{R})$  by Sobolev's embedding theorem or by (45), (46) and (47).

*Example 2.* In the context of M-estimation (or Z-estimation) the root of the equation

$$\langle \zeta(\cdot - t), f_X \rangle = 0$$

is used for inference, e.g., on the location of the distribution of  $X_j$ . A popular example in robust statistics is the Huber estimator where  $\zeta(x) = h_K(x) := ((-K) \vee x) \wedge K$  for some  $K > 0$ . In that case a similar decomposition as in Example 1 shows  $h_K \in Z^{\gamma_s, \gamma_c}$  for any  $\gamma_s < 3/2$ .

The ill-posedness of the problem is determined by the decay of the characteristic function of the errors. More precisely, we suppose

**Assumption 2.** Let the error distribution satisfy

- (i)  $\int |x|^{2+\delta} f_\varepsilon(x) dx < \infty$  thus  $\varphi_\varepsilon$  is twice continuously differentiable and
- (ii)  $|(\varphi_\varepsilon^{-1})'(u)| \lesssim \langle u \rangle^{\beta-1}$  for some  $\beta > 0$ , in particular  $|\varphi_\varepsilon^{-1}(u)| \lesssim \langle u \rangle^\beta, u \in \mathbb{R}$ .

Throughout, we write  $\varphi_\varepsilon^{-1} = 1/\varphi_\varepsilon$ . The Assumption (ii) on the distribution of the errors is similar to the classical decay assumption by Fan (1991a) and it is fulfilled for many ordinary smooth error laws such as gamma or Laplace distributions as discussed below. Assumption 2(ii) implies that  $\varphi_\varepsilon^{-1}$  is a Fourier multiplier on Besov spaces so that

$$B_{p,q}^s(\mathbb{R}) \ni f \mapsto \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} f] \in B_{p,q}^{s-\beta}(\mathbb{R})$$

for  $p, q \in [1, \infty], s \in \mathbb{R}$ , is a continuous linear map, which is essential in our proofs, compare Lemma 5. In the same spirit Schmidt–Hieber, Munk and Dümbgen (2012) discuss the behavior of the deconvolution operator as pseudo-differential operator. We define

$$g_t := \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta_t \quad \text{and} \quad \mathcal{G} = \{g_t | t \in \mathbb{R}\}. \quad (7)$$

Note that in general  $g_t$  may only exist in a distributional sense, but on Assumption 2 and for  $\zeta \in Z^{\gamma_s, \gamma_c}$  it can be rigorously interpreted by (see (19))

$$\begin{aligned} g_0(x) &= \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F} \zeta^s(u)](x) \\ &\quad + (1 + ix) \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F}[\frac{1}{iy+1} \zeta^c(y)](u)](x) \\ &\quad + \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u) \mathcal{F}[\frac{1}{iy+1} \zeta^c(y)](u)](x), \end{aligned}$$

which indicates why we have imposed an assumption on  $(\varphi_\varepsilon^{-1})'$  and have defined  $\|\bullet\|_{Z^{\gamma_s, \gamma_c}}$  as above.

It will turn out that  $\mathcal{G}$  is  $\mathbb{P}$ -pregaussian, but not Donsker in general. Denoting by  $\lfloor \alpha \rfloor$  the largest integer smaller or equal to  $\alpha$  and defining convergence in law on  $\ell^\infty(\mathbb{R})$  as Dudley (1999, p. 94), we state our main result

**Theorem 1.** Grant Assumptions 1 and 2 as well as  $\zeta \in Z^{\gamma_s, \gamma_c}$  with  $\gamma_s > \beta, \gamma_c > (1/2 \vee \alpha) + \gamma_s$  and  $\alpha + 3\gamma_s > 2\beta + 1$ . Furthermore, let the kernel  $K$  satisfy (4) with  $L = \lfloor \alpha + \gamma_s \rfloor$ . Let  $h_n^{2\alpha+2\gamma_s} n \rightarrow 0$  and if  $\gamma_s \leq \beta + 1/2$  let in addition  $h_n^\rho n \rightarrow \infty$  for some  $\rho > 4\beta - 4\gamma_s + 2$ , then

$$\sqrt{n}(\widehat{\vartheta}_t - \vartheta_t)_{t \in \mathbb{R}} \xrightarrow{\mathcal{L}} \mathbb{G} \quad \text{in } \ell^\infty(\mathbb{R})$$

as  $n \rightarrow \infty$ , where  $\mathbb{G}$  is a centered Gaussian Borel random variable in  $\ell^\infty(\mathbb{R})$  with covariance function given by

$$\Sigma_{s,t} := \int g_s(x) g_t(x) \mathbb{P}(dx) - \vartheta_s \vartheta_t$$

for  $g_s, g_t$  defined in (7) and  $s, t \in \mathbb{R}$ .

We illustrate the range of this theorem by the following examples.

*Example 3.* For estimating the distribution function Assumption 2 needs to be fulfilled for some  $\beta < 1/2$  owing to the condition  $\gamma_s > \beta$ . This is fulfilled by the gamma distribution  $\Gamma(\beta, \eta)$  with  $\beta \in (0, 1/2)$  and  $\eta \in (0, \infty)$ , that is

$$f_\varepsilon(x) := \gamma_{\beta, \eta}(x) := \frac{1}{\Gamma(\beta)\eta^\beta} x^{\beta-1} e^{-x/\eta} \mathbb{1}_{[0, \infty)}(x), \quad x \in \mathbb{R},$$

and  $\varphi_\varepsilon(u) = (1 - i\eta u)^{-\beta}$ ,  $u \in \mathbb{R}$ .

For the Huber estimator from Example 2 we required  $\beta < 3/2$ , which holds, for instance, for the chi-squared distribution with one or two degrees of freedom or for the exponential distribution.

*Example 4.* Butucea and Comte (2009) studied the case  $\beta > 1$  and derived  $\sqrt{n}$ -rates for  $\gamma_s > \beta$  in our notation. In particular, they considered supersmooth  $\zeta$ , that is  $\mathcal{F}\zeta$  decays exponentially. In this case  $\zeta \in H^s(\mathbb{R})$  for any  $s \in \mathbb{N}$ . Requiring the slightly stronger assumption that  $\langle x \rangle^\tau \zeta(x) \in H^s(\mathbb{R})$  for some arbitrary small  $\tau > 0$  and for all  $s \in \mathbb{N}$  we can choose  $\zeta^c := \zeta$  and  $\zeta^s := 0$ . Then  $\beta$  can be taken arbitrary large such that all gamma distributions, the Laplace distributions and convolutions of them can be chosen as error distributions.

### 2.3. Discussion

To have  $\sqrt{n}$ -rates we suppose  $\gamma_s > \beta$  in Theorem 1, which means that the smoothness of the functionals compensates the ill-posedness of the problem. This condition is natural in view of the abstract analysis in terms of Hilbert scales by Goldenshluger and Pereverzev (2003), who obtain the minimax rate  $n^{-(\alpha+\gamma_s)/(2\alpha+2\beta)} \vee n^{-1/2}$  in our notation. As a consequence of the condition on  $\gamma_s$  and  $\gamma_c$  we can bound the stochastic error term of the estimator  $\hat{\vartheta}_t$  uniformly in  $h \in (0, 1)$ . The bias term is of order  $h^{\alpha+\gamma_s}$ .

For  $\gamma_s > \beta + 1/2$  the class  $\mathcal{G}$  is a Donsker class. In this case the only condition on the bandwidth is that the bias tends faster than  $n^{-1/2}$  to zero. In the interesting but involved case  $\gamma_s \in (\beta, \beta + 1/2]$ , the class  $\mathcal{G}$  will in general not be a Donsker class. Estimating the distribution function as in Example 1 belongs to this case. In order to see that  $\mathcal{G}$  is in general not a Donsker class, let the error distribution be given by  $f_\varepsilon = \gamma_{\beta, \eta}(-\bullet)$  and  $\zeta = \gamma_{\sigma, \eta}$  with  $\sigma \in (\gamma_s + 1/2, \beta + 1)$ . Then  $g_t$  equals  $\gamma_{\sigma-\beta, \eta} * \delta_t$ . For the shape parameter holds  $\sigma - \beta \in (1/2, 1)$  and thus  $g_t$  is an  $L^2(\mathbb{R})$ -function unbounded at  $t$ . The Lebesgue density of  $\mathbb{P}$  is bounded by Assumption 1(i). Hence,  $\mathcal{G}$  consists of all translates of an unbounded function and thus cannot be Donsker, cf. Theorem 7 by Nickl (2006).

Therefore, for  $\gamma_s \in (\beta, \beta + 1/2]$  smoothed empirical processes are necessary, especially we need to ensure enough smoothing to be able to obtain a uniform central limit theorem. The bandwidth cannot tend too fast to zero, more precisely we require  $h_n^\rho n \rightarrow \infty$  as  $n \rightarrow \infty$  for some  $\rho$  with  $\rho > 4\beta - 4\gamma_s + 2$ . In combination with the bias condition  $h_n^{2\alpha+2\gamma_s} n \rightarrow 0$  as  $n \rightarrow \infty$  we obtain necessarily  $\alpha + \gamma_s > 2\beta - 2\gamma_s + 1$  leading to the assumption in the theorem. Since  $2\alpha + 2\gamma_s > \alpha + 2\beta - \gamma_s + 1 > 4\beta - 4\gamma_s + 2$  we can always choose  $h_n \sim n^{-1/(\alpha+2\beta-\gamma_s+1)}$ . In contrast to Fan (1991b); Butucea and Comte (2009);

Dattner, Goldenshluger and Juditsky (2011) our choice of the bandwidth  $h_n$  is not determined by the bias–variance trade–off, but rather by the amount of smoothing necessary to obtain a uniform central limit theorem. The classical bandwidth  $h_n \sim n^{-1/(2\alpha+2\beta)}$  is optimal for estimating the density in the sense that it achieves the minimax rate with respect to the mean integrated squared error (MISE), compare Fan (1991b) who assumes Hölder smoothness of  $f_X$  instead of  $L^2$ –Sobolev smoothness. For this choice the bias condition  $h_n^{2\alpha+2\gamma_s} n \rightarrow 0$  is satisfied. If  $\gamma_s \leq \beta + 1/2$  the classical bandwidth satisfies the additional minimal smoothness condition in the case of estimating the distribution function with mild conditions on  $f_X$ . It suffices for example that  $f_X$  is of bounded variation. Then  $\alpha$  and  $\gamma_s$  can be chosen large enough in  $(0, 1/2)$  such that  $2\alpha+2\beta > 4\beta-4\gamma_s+2$  and the classical bandwidth satisfies the conditions of the theorem. Whenever the classical bandwidth  $h_n \sim n^{-1/(2\alpha+2\beta)}$  satisfies the conditions of Theorem 1, then the corresponding density estimator is a ‘plug-in’ estimator in the sense of Bickel and Ritov (2003) meaning that the density is estimated rate optimal for the MISE, the functionals are estimated efficiently (see Section 3) and the estimators of the functionals converge uniformly over  $t \in \mathbb{R}$ .

The smoothness condition on the density  $f_X$  is then a consequence of the given choice of  $h_n$  together with the classical bias estimate for kernel estimators. As we have seen in Example 1 for estimating the distribution function, we have  $\zeta = \mathbb{1}_{(-\infty, 0]} \in Z^{\gamma_s, \gamma_c}$  with  $\gamma_s < 1/2$  arbitrary close to  $1/2$ . In the classical Donsker theorem which corresponds to the case  $\beta \rightarrow 0$  the condition  $\alpha + 3\gamma_s > 2\beta + 1$  would simplify to  $\alpha > -1/2$ . However, we suppose  $f_X$  to be bounded, which leads to much clearer proofs, and thus  $f_X \in H^0(\mathbb{R})$  is automatically satisfied. Assumption 1 allows to focus on the interplay between the functional  $\zeta$  and the deconvolution operator  $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}]$ . Nickl and Reiß (2012) have studied the case of unbounded densities, which is necessary in the Lévy process set–up, but considered  $\zeta_t = \mathbb{1}_{(-\infty, t]}$  only. The class  $Z^{\gamma_s, \gamma_c}$  is defined by  $L^2$ –Sobolev conditions so that bounded variation arguments for  $\zeta$  have to be avoided in the proofs.

An interesting aspect is the following: If we restrict the uniform convergence to  $(\zeta_t)_{t \in T}$  for some compact set  $T \subseteq \mathbb{R}$ , it is sufficient to assume  $\frac{1}{ix+1}\zeta^c \in H^{\gamma_c}(\mathbb{R})$  instead of requiring  $(1 \vee |x|^\tau)(\zeta^c(x) - a(x)) \in H^{\gamma_c}(\mathbb{R})$  for some  $\tau > 0$  and a function  $a \in C^\infty(\mathbb{R})$  such that  $a'$  is compactly supported as done in  $Z^{\gamma_s, \gamma_c}$ . In particular, slowly growing  $\zeta$  would be allowed. The stronger condition in the definition of  $Z^{\gamma_s, \gamma_c}$  is only needed to ensure polynomial covering numbers of  $\{g_t | t \in T\}$  for  $T \subseteq \mathbb{R}$  unbounded (cf. Theorem 7 below).

As a corollary of Theorem 1 we can weaken Assumption 2(ii). If the characteristic function of the errors  $\varepsilon$  is given by  $\tilde{\varphi}_\varepsilon = \varphi_\varepsilon \psi$  where  $\varphi_\varepsilon$  satisfies Assumption 2(ii) and there is a Schwartz distribution  $\nu \in \mathcal{S}'(\mathbb{R})$  such that  $\mathcal{F}\nu = \psi^{-1}$  and  $\nu * \zeta \in Z^{\gamma_s, \gamma_c}$  for  $\zeta \in Z^{\gamma_s, \gamma_c}$ , then for  $t \in \mathbb{R}$

$$\mathcal{F}^{-1}[\tilde{\varphi}_\varepsilon^{-1}] * \zeta(\bullet - t) = \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * (\nu * \zeta)(\bullet - t)$$

and thus we can proceed as before. For instance, for translated errors  $f_\varepsilon * \delta_\mu$  with  $\mu \neq 0$ , the distribution  $\nu$  would be given by  $\delta_{-\mu}$ . Thus even if  $\tilde{\varphi}_\varepsilon$  does not



satisfy Assumption 2(ii) the statement of Theorem 1 holds true if we can write  $\tilde{\varphi}_\varepsilon(u) = \varphi_\varepsilon(u)e^{i\mu u}$  with  $\varphi_\varepsilon$  satisfying Assumption 2(ii). The assumption by Fan and Liu (1997) for the asymptotic normality in the case of ordinary smooth deconvolution are weaker than the assumptions by Fan (1991a) since they allow exactly for this additional factor  $e^{i\mu u}$  in the assumptions on the characteristic function of the error.

As for the classical Donsker theorem the Donsker theorem for deconvolution estimators has many different applications, the most obvious being the construction of confidence bands. Further Donsker theorems may be obtained by applying the functional delta method to Hadamard differentiable maps. Let us illustrate the construction of confidence bands. By the continuous mapping theorem we infer

$$\sup_{t \in \mathbb{R}} \sqrt{n} |\hat{\vartheta}_t - \vartheta_t| \xrightarrow{\mathcal{L}} \sup_{t \in \mathbb{R}} |\mathbb{G}(t)|.$$

The construction of confidence bands reduces now to knowledge about the distribution of the supremum of  $\mathbb{G}$ . Suprema of Gaussian processes are well studied and information about their distribution can be either obtained from theoretical considerations as in van der Vaart and Wellner (1996, App. A.2) or from Monte Carlo simulations. Let  $q_{1-\alpha}$  be the  $(1-\alpha)$ -quantile of  $\sup_{t \in \mathbb{R}} |\mathbb{G}(t)|$  that is  $\mathbb{P}(\sup_{t \in \mathbb{R}} |\mathbb{G}(t)| \leq q_{1-\alpha}) = 1 - \alpha$ . Then

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \vartheta_t \in [\hat{\vartheta}_t - q_{1-\alpha} n^{-1/2}, \hat{\vartheta}_t + q_{1-\alpha} n^{-1/2}] \text{ for all } t \in \mathbb{R} \right) = 1 - \alpha$$

and thus the intervals  $[\hat{\vartheta}_t - q_{1-\alpha} n^{-1/2}, \hat{\vartheta}_t + q_{1-\alpha} n^{-1/2}]$  define a confidence band.

As often in the literature we assume the error distribution to be known. In a more realistic setting where an independent sample of pure noise observations is observable, Neumann (1997) constructs a density estimator for unknown errors. In this case the deconvolution operator  $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}]$  needs to be estimated leading to an inverse problem with an error in the operator. A Donsker theorem in this situation is related to the Lévy process setting of Nickl and Reiß (2012) since their estimator is based on the deconvolution with the unknown marginal distribution of the process. Another interesting aspect is the similarity of the deconvolution problem and the errors-in-variables regression, see for example Fan and Truong (1993) and the references therein. The corresponding statement to our result would be a uniform central limit theorem for linear functionals of the regression function, where similar methods may be applied.

### 3. Efficiency

Having established the asymptotic normality of our estimator, the natural question is whether it is optimal in the sense of the convolution Theorem 5.2.1 by Bickel et al. (1998). Typically, efficiency is investigated for estimators  $T_n$  which are (locally) regular, that is for any parametric submodel  $\eta \rightarrow f_{X,\eta}$  and  $n^{1/2}|\eta_n - \eta| \lesssim 1$  the law of  $n^{1/2}(T_n - \langle \zeta, f_{X,\eta} \rangle)$  under  $\eta_n$  converges for

$n \rightarrow \infty$  to a distribution independent of  $(\eta_n)$ . In Lemma 9 we show that the estimator  $\hat{\vartheta}_t$  from (3) is asymptotically linear with influence function  $x \mapsto \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta(y)(\delta_x - \mathbb{P})(dy)$  and thus  $\hat{\vartheta}_t$  is Gaussian regular.

In general, semiparametric lower bounds are constructed as the supremum of the information bounds over all regular parametric submodels. As it turns out, it suffices to apply the Cramér–Rao bound to the least favorable one-dimensional submodel  $\mathbb{P}_g$  of the form

$$f_{Y,\xi g} = f_{X,\xi g} * f_\varepsilon \quad \text{with} \quad f_{X,\xi g} := f_X + \xi g, \quad \text{for all } \xi \in (-\tau, \tau),$$

with some  $\tau > 0$  and a perturbation  $g$  satisfying

$$f_X \pm \tau g \geq 0 \quad \text{and} \quad \int g = 0. \quad (8)$$

Note that all laws  $\mathbb{P}_g$  are absolutely continuous with respect to  $\mathbb{P}$  assuming  $\text{supp}(f_X) = \mathbb{R}$ . Moreover, the submodels are regular with score function  $g * f_\varepsilon / f_Y$ , since for all  $\xi \in (-\tau, \tau) \setminus \{0\}$  we have the  $L^2$ -differentiability

$$\int \left( \frac{f_{Y,\xi g} - f_Y - \xi g * f_\varepsilon}{\xi f_Y} \right)^2 f_Y = 0.$$

Similarly to van der Vaart (1998, Chap. 25.5), we define the score operator  $Sg := (g * f_\varepsilon) f_Y^{-1/2}$  and thus the information operator of  $f_X$  is given by  $I := S^* S$ , where  $S^*$  denotes the adjoint of the linear operator  $S$ . This yields the Fisher information in direction  $g$

$$\langle I g, g \rangle = \langle S g, S g \rangle = \int \left( \frac{g * f_\varepsilon}{f_Y} \right)^2 f_Y \quad (9)$$

and we obtain the information bound

$$\mathcal{I}_\zeta := \sup_g \frac{\langle g, \zeta \rangle^2}{\langle S g, S g \rangle}, \quad (10)$$

where the supremum is taken over all  $g$  satisfying the conditions (8). In the notation of Bickel et al. (1998, Def. 3.3.2), we consider the tangent space  $\dot{Q} := \{(g * f_\varepsilon) / f_Y | g \text{ satisfies (8)}\}$ , representing the submodel  $\{\mathbb{P}_g\}$ , and the efficient influence function of the parameter  $\vartheta_\zeta : \dot{Q} \rightarrow \mathbb{R}, h \mapsto \langle h, \zeta \rangle$  needs to be determined.

Since we perturb the density additively with the restriction (8), the quotient  $|g / f_X|$  needs to be bounded and thus it is natural to assume a lower bound for the decay behavior of  $f_X$ . We state with some  $\delta > 0$  and  $M \in \mathbb{N}$

**Assumption 3.** *Let the following be satisfied*

- (i)  $f_X$  is bounded and fulfills the moment condition  $\int |x|^{2+\delta} f_X(x) dx < \infty$ ,
- (ii)  $f_X \in W_1^2(\mathbb{R})$  that is  $f_X$  has  $L^1$ -Sobolev regularity two,
- (iii)  $f_X(x) \gtrsim \langle x \rangle^{-M}$  for  $x \in \mathbb{R}$ .

A precise definition of the  $L^1$ -Sobolev space  $W_1^2(\mathbb{R})$  can be found in the appendix. Due to the Sobolev embedding  $W_1^2(\mathbb{R}) \subseteq H^\alpha(\mathbb{R})$  with  $\alpha < 3/2$  (cf. (44) and (46)), Assumption 3 implies the Assumption 1 in the previous section. The conditions on  $\varepsilon$  need to be strengthened, too.

**Assumption 4.** *We suppose*

- (i)  $\int |x|^{2+\delta} f_\varepsilon(x) dx < \infty$ ,
- (ii) *for some  $\beta \in (0, \infty) \setminus \mathbb{Z}$  and  $M$  from above let  $\varphi_\varepsilon \in C^{(\lfloor \beta \rfloor \vee M)+1}(\mathbb{R})$  satisfy for all  $k = 0, \dots, (\lfloor \beta \rfloor \vee M) + 1$*

$$\mathbb{1}_{\{k=0\}} \langle u \rangle^{-\beta-k} \lesssim |\varphi_\varepsilon^{(k)}(u)| \lesssim \langle u \rangle^{-\beta-k}.$$

Since  $M + 1 \geq 2$ , easy calculus shows that Assumption 2(ii) on  $\varphi_\varepsilon^{-1}$  follows from Assumption 4 on  $\varphi_\varepsilon$ . We supposed  $\beta \notin \mathbb{Z}$  mainly to simplify our proofs. Let us first show an information bound for smooth  $\zeta$ .

**Theorem 2.** *Grant Assumptions 3 and 4 and let  $\zeta \in \mathcal{S}(\mathbb{R})$  be a Schwartz function. For any regular estimator  $T$  of  $\vartheta_0 = \langle \zeta, f_X \rangle$  with asymptotic variance  $\sigma^2$  we obtain*

$$\sigma^2 \geq \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta)^2 f_Y - \vartheta_0^2. \quad (11)$$

*In particular, the supremum in (10) is attained at  $g^* := g^*(\zeta) := \mathbf{I}^{-1} \zeta - \langle \zeta, f_X \rangle f_X$ , where the inverses of  $S^*$  and  $\mathbf{I}$  are given by*

$$\begin{aligned} (S^*)^{-1} \zeta &= (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta) \sqrt{f_Y} \quad \text{and} \\ \mathbf{I}^{-1} \zeta &= S^{-1}(S^{-1})^* \zeta = \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * \{(\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta) f_Y\}. \end{aligned}$$

Therefore, the score function corresponding to  $g^*(\zeta)$  which is given by

$$\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta - \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta) f_Y$$

(compare (37) below) is the efficient influence function and, moreover, equals the influence function of  $\hat{\vartheta}_\zeta$ . This equality shows that the estimator is efficient for smooth functionals  $\vartheta_\zeta$ . Moreover, we found already the efficient influence function in the larger tangent set of all regular submodels.

Unfortunately, less smooth  $\zeta$  might be only in the domain of  $(S^*)^{-1}$  while  $\mathbf{I}^{-1} \zeta$  is not in  $L^2(\mathbb{R})$  and thus the formal maximizer  $g^*(\zeta)$  cannot be applied rigorously as the following example shows.

**Example 5.** Let  $\varepsilon_j$  be gamma distributed with density  $\gamma_{\beta,1}$  for  $\beta \in (1/4, 1/2)$  and consider  $\zeta(x) = e^x \mathbb{1}_{(-\infty, 0]}(x) = \gamma_{1,1}(-x)$  which is contained in  $Z^{\gamma_s, \gamma_c}$  for all  $\gamma_s < 1/2$  and  $\gamma_c$  arbitrary large. We obtain

$$(S^*)^{-1} \zeta = \gamma_{1-\beta,1}(-\bullet) \sqrt{f_Y} \quad \text{and} \quad \mathbf{I}^{-1} \zeta = \mathcal{F}^{-1} [(1 - iu)^\beta ((1 + iu)^{-1+\beta} * \varphi)].$$

While first term behaves nicely the Fourier transform of  $\mathbf{I}^{-1} \zeta$  is of the order  $|u|^{-1+2\beta} > |u|^{-1/2}$  for  $|u| \rightarrow \infty$  and thus  $\mathbf{I}^{-1} \zeta \notin L^2(\mathbb{R})$ .

Therefore, we choose an approximating sequence  $\zeta_n \rightarrow \zeta$  with  $(\zeta_n)_{n \in \mathbb{N}} \subseteq \mathcal{S}(\mathbb{R})$ . For  $n \in \mathbb{N}$  let  $g_n^* := g^*(\zeta_n) = \mathbf{I}^{-1} \zeta_n - \langle \zeta, f_X \rangle f_X$  be the least favorable direction in the estimation problem with respect to  $\langle f_X, \zeta_n \rangle$ . We obtain for every  $n \in \mathbb{N}$

$$\mathcal{I}_\zeta \geq \frac{\langle g_n^*, \zeta \rangle^2}{\langle Sg_n^*, Sg_n^* \rangle} = \frac{(\langle g_n^*, \zeta - \zeta_n \rangle + \langle g_n^*, \zeta_n \rangle)^2}{\langle Sg_n^*, Sg_n^* \rangle}.$$

This inequality suggests two possibilities to understand our strategy for obtaining the efficiency bound. First, the sequence  $(g_n^*)$  approximates the formal maximizer  $g^*(\zeta)$  and thus plugging  $g_n^*$  into the bound (10) might converge to the supremum. Second, any unbiased estimator of  $\vartheta_{\zeta_n} = \langle f_X, \zeta_n \rangle$  is at the same time a possibly biased estimator of  $\vartheta_\zeta$  with bias tending to zero. Therefore, the bound for the smooth problems should converge to the nonsmooth one. The following lemma provides a sufficient condition for the convergence of the Cramér–Rao bounds.

**Lemma 3.** *Let  $\zeta$  and  $(\zeta_n)$  satisfy  $(S^*)^{-1}\zeta \in L^2(\mathbb{R})$  and  $\zeta_n, \mathbf{I}^{-1}\zeta_n \in L^2(\mathbb{R})$  for all  $n \in \mathbb{N}$ . Then  $\vartheta_{\zeta_n} \rightarrow \vartheta_\zeta$  and  $\frac{\langle g_n^*, \zeta \rangle^2}{\langle Sg_n^*, Sg_n^* \rangle} \rightarrow \langle (S^*)^{-1}\zeta, (S^*)^{-1}\zeta \rangle - \langle \zeta, f_X \rangle^2$  hold as  $n \rightarrow \infty$  if*

$$\|(S^*)^{-1}(\zeta_n - \zeta)\|_{L^2} \rightarrow 0, \quad \text{as } n \rightarrow \infty.$$

Using mapping properties on Besov spaces, we will show that the underlying Fourier multiplier  $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}]$  and thus the inverse adjoint score operator  $(S^*)^{-1}$  are well-defined on the set  $Z^{\gamma_s, \gamma_c}$ . This allows the extension of Theorem 2 to all  $\zeta \in Z^{\gamma_s, \gamma_c}$  with  $\gamma_s > \beta$  and  $\gamma_c > \beta + 1/2$ .

Since  $\widehat{\vartheta}_t$  does not only estimate  $\vartheta_t$  pointwise but also as a process in  $\ell^\infty(\mathbb{R})$ , we want to generalize Theorem 2 in this direction, too. In view of Theorem 25.48 of van der Vaart (1998) the remaining ingredient is the tightness of the limiting object, which is already a necessary condition for the Donsker theorem. A regular estimator  $T_n$  of  $(\vartheta_t)_{t \in \mathbb{R}}$  in  $\ell^\infty(\mathbb{R})$  is efficient if the limiting distribution of  $\sqrt{n}(T_n - \vartheta)$  is a tight zero mean Gaussian process whose covariance structure is given by the information bound for the finite dimensional distributions (cf. the convolution Theorem 5.2.1 of (Bickel et al., 1998)). Interestingly, the class of efficient influence functions for  $t \in \mathbb{R}$  is not Donsker as discussed above and thus there exists no efficient estimator which is asymptotically linear in  $\ell^\infty(\mathbb{R})$  (cf. Kosorok, 2008, Thm. 18.8).

**Theorem 4.** *Let Assumptions 3 and 4 be satisfied as well as  $\zeta \in Z^{\gamma_s, \gamma_c}$  with  $\gamma_s > \beta$  and  $\gamma_c > \beta + 1/2$ . Then the estimator  $(\widehat{\vartheta}_t)_{t \in \mathbb{R}}$  defined in (3) is (uniformly) efficient.*

Additionally, the proof of Theorem 4 reveals the relation between the intrinsic metric  $d(s, t)^2 = \mathbb{E}[(\mathbb{G}_s - \mathbb{G}_t)^2]$  of the limit  $\mathbb{G}$ , which is essential to show tightness, and the metric  $d_{I^{-1}}(s, t)^2 = \langle (S^*)^{-1}(\zeta_t - \zeta_s), (S^*)^{-1}(\zeta_t - \zeta_s) \rangle$  which is induced by the inverse Fisher information, namely

$$d_{I^{-1}}(s, t)^2 = d(s, t)^2 + \langle \zeta_t - \zeta_s, f_X \rangle^2$$

(cf. equations (25) and (43) below) such that both metrics are equal up to some centering term which is another way of interpreting the efficiency of  $\hat{\vartheta}_\bullet$ .

#### 4. Proof of the Donsker theorem

First, we provide an auxiliary lemma, which describes the properties of the deconvolution operator  $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}]$ .

**Lemma 5.** *Grant Assumption 2.*

- (i) For all  $s \in \mathbb{R}, p, q \in [1, \infty]$  the deconvolution operator  $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]$  is a Fourier multiplier from  $B_{p,q}^s(\mathbb{R})$  to  $B_{p,q}^{s-\beta}(\mathbb{R})$ , that is the linear map

$$B_{p,q}^s(\mathbb{R}) \rightarrow B_{p,q}^{s-\beta}(\mathbb{R}), f \mapsto \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] \mathcal{F} f$$

is bounded.

- (ii) For any integer  $m$  strictly larger than  $\beta$  we have  $\mathcal{F}^{-1}[(1+iu)^{-m}\varphi_\varepsilon^{-1}] \in L^1(\mathbb{R})$  and if  $m > \beta + 1/2$  we also have  $\mathcal{F}^{-1}[(1+iu)^{-m}\varphi_\varepsilon^{-1}] \in L^2(\mathbb{R})$ .  
 (iii) Let  $\beta^+ > \beta$  and  $f, g \in H^{\beta^+}(\mathbb{R})$ . Then

$$\int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f)g = \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * g)f. \quad (12)$$

Using the kernel  $K$ , this equality extends to functions  $g \in L^2(\mathbb{R}) \cup L^\infty(\mathbb{R})$  and finite Borel measures  $\mu$ :

$$\int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] \mathcal{F} K_h * \mu)g = \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] \mathcal{F} K_h * g) d\mu. \quad (13)$$

*Proof.*

- (i) Analogously to Nickl and Reiß (2012), we deduce from Corollary 4.11 of Girardi and Weis (2003) that  $(1+iu)^{-\beta}\varphi_\varepsilon^{-1}(-u)$  is a Fourier multiplier on  $B_{p,q}^s$  by Assumption 2(ii). It remains to note that  $j : B_{p,q}^s(\mathbb{R}) \rightarrow B_{p,q}^{s-\beta}(\mathbb{R}), f \mapsto \mathcal{F}^{-1}[(1+iu)^\beta \mathcal{F} f]$  is a linear isomorphism (Triebel, 2010, Thm. 2.3.8).  
 (ii) Since the gamma density  $\gamma_{1,1}$  is of bounded variation, it is contained in  $B_{1,\infty}^1(\mathbb{R})$  by (51). Using the isomorphism  $j$  from (i), we deduce  $\gamma_{m,1} \in B_{1,\infty}^m(\mathbb{R})$  and thus by Besov embeddings (47) and (44)

$$\mathcal{F}^{-1}[(1+iu)^{-m}\varphi_\varepsilon^{-1}] \in B_{1,\infty}^{m-\beta}(\mathbb{R}) \subseteq B_{1,1}^0(\mathbb{R}) \subseteq L^1(\mathbb{R}).$$

If  $m-\beta > 1/2$ , we can apply the embedding  $B_{1,\infty}^{m-\beta}(\mathbb{R}) \subseteq B_{2,\infty}^{m-\beta-1/2}(\mathbb{R}) \subseteq L^2(\mathbb{R})$ .

- (iii) For  $f \in H^{\beta^+}(\mathbb{R})$ , (i) and the Besov embeddings (44), (46) and (47) yield

$$\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f\|_{L^2} \lesssim \|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f\|_{B_{2,1}^0} \lesssim \|f\|_{B_{2,1}^\beta} \lesssim \|f\|_{H^{\beta^+}} < \infty.$$

Therefore, it follows by Plancherel's equality

$$\begin{aligned} \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * f)(x)g(x) \, dx &= \frac{1}{2\pi} \int \varphi_\varepsilon^{-1}(-u) \mathcal{F} f(-u) \mathcal{F} g(u) \, du \\ &= \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * g)(x)f(x) \, dx. \end{aligned}$$

To prove the second part of the claim for  $g \in L^2(\mathbb{R})$ , we note that by Young's inequality

$$\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h]\|_{L^2} \leq \|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathbf{1}_{[-1/h, 1/h]}]\|_{L^2} \|K_h\|_{L^1} < \infty$$

due to the support of  $\mathcal{F} K$  and Assumption (5) on the decay of  $K$ . Since  $\mu$  is a finite measure and  $g$  is bounded, Fubini's theorem yields then

$$\begin{aligned} &\int g(x) (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h] * \mu)(x) \, dx \\ &= \int \int g(x) \mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h](x-y) \mu(dy) \, dx \\ &= \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} K_h] * g)(y) \mu(dy), \end{aligned}$$

where we have used the symmetry of the kernel. In order to apply Fubini's theorem for  $g \in L^\infty(\mathbb{R})$ , too, we have to show that  $\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h]\|_{L^1}$  is finite. We replace the indicator function by a function  $\chi \in C^\infty(\mathbb{R})$  which equals one on  $[-1/h, 1/h]$  and is compactly supported. We estimate

$$\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \mathcal{F} K_h]\|_{L^1} \leq \|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \chi]\|_{L^1} \|K_h\|_{L^1}. \quad (14)$$

Using that  $\varphi_\varepsilon^{-1} \chi$  is twice continuously differentiable and has got compact support, we obtain

$$\begin{aligned} \|(1+x^2) \mathcal{F}^{-1}[\varphi_\varepsilon^{-1} \chi](x)\|_\infty &\leq \|\mathcal{F}^{-1}[(\text{Id} - D^2) \varphi_\varepsilon^{-1} \chi](x)\|_\infty \\ &\leq \|(\text{Id} - D^2) \varphi_\varepsilon^{-1} \chi\|_{L^1} < \infty, \end{aligned}$$

where we denote the identity and the differential operator by  $\text{Id}$  and  $D$ , respectively. This shows that (14) is finite.  $\square$

#### 4.1. Convergence of the finite dimensional distributions

As usual, we decompose the error into a stochastic error term and a bias term:

$$\begin{aligned} \widehat{\vartheta}_t - \vartheta_t &= \widehat{\vartheta}_t - \mathbb{E}[\widehat{\vartheta}_t] + \mathbb{E}[\widehat{\vartheta}_t] - \vartheta_t \\ &= \int \zeta_t(x) \mathcal{F}^{-1} \left[ \mathcal{F} K_h \frac{\varphi_n - \varphi}{\varphi_\varepsilon} \right](x) \, dx + \int \zeta_t(x) (K_h * f_X - f_X)(x) \, dx. \end{aligned}$$

#### 4.1.1. The bias

The bias term can be estimated by the standard kernel estimator argument. Let us consider the singular and the continuous part of  $\zeta$  separately. Applying Plancherel's identity and Hölder's inequality, we obtain

$$\begin{aligned} & \int |\zeta_t^s(x)(K_h * f_X(x) - f_X(x))| dx \\ &= \frac{1}{2\pi} \int |\mathcal{F} \zeta_t^s(u)(\mathcal{F} K(hu) - 1) \mathcal{F} f_X(-u)| du \\ &\leq \|\langle u \rangle^{-(\alpha+\gamma_s)}(\mathcal{F} K(hu) - 1)\|_\infty \int \langle u \rangle^{\alpha+\gamma_s} |\mathcal{F} \zeta^s(u) \mathcal{F} f_X(u)| du \\ &\leq h^{\alpha+\gamma_s} \|u^{-(\alpha+\gamma_s)}(\mathcal{F} K(u) - 1)\|_\infty \|\zeta^s\|_{H^{\gamma_s}} \|f_X\|_{H^\alpha}. \end{aligned}$$

The term  $\|u^{-(\alpha+\gamma_s)}(\mathcal{F} K(u) - 1)\|_\infty$  is finite using the a Taylor expansion of  $\mathcal{F} K$  around 0 with  $(\mathcal{F} K)^{(l)} = 0$  for  $l = 1, \dots, \lfloor \alpha + \gamma_s \rfloor$  by the order of the kernel (4).

For the smooth part of  $\zeta_t$  Plancherel's identity yields

$$\begin{aligned} & \int |\zeta_t^c(x)(K_h * f_X - f_X)(x)| dx \\ &= \frac{1}{2\pi} \int |\mathcal{F}[\frac{1}{ix+1}\zeta_t^c(x)](\text{Id} + D)\{(\mathcal{F} K(hu) - 1) \mathcal{F} f_X(-u)\}| du \\ &\leq \int |\mathcal{F}[\frac{1}{ix+1}\zeta_t^c(x)](\mathcal{F} K(hu) - 1 + h \mathcal{F}[ixK](hu)) \mathcal{F} f_X(-u)| du \\ &\quad - \int |\mathcal{F}[\frac{1}{ix+1}\zeta_t^c(x)](\mathcal{F} K(hu) - 1) \mathcal{F}[ixf_X](-u)| du. \end{aligned}$$

The first term can be estimated as before and for the second term we note that  $xf_X(x) \in L^2(\mathbb{R}) = H^0(\mathbb{R})$  by Assumption 1(i) such that the additional smoothness of  $\frac{1}{ix+1}\zeta^c(x)$  yields the right order. Therefore, we have  $|\mathbb{E}[\hat{\vartheta}_t] - \vartheta_t| \lesssim h^{\alpha+\gamma_s}$  and thus by the choice of  $h$ , the bias term is of order  $o(n^{-1/2})$ .

#### 4.1.2. The stochastic error

We notice that  $\|\zeta^c - a\|_{H^{\gamma_c}} \lesssim \|\langle x \rangle^{-\tau}\|_{C^s} \|\langle x \rangle^\tau(\zeta^c(x) - a(x))\|_{H^{\gamma_c}} < \infty$  for any  $s > \gamma_c$ , where we used the pointwise multiplier property (48) as well as the Besov embeddings (47) and (45). We have  $\zeta^s \in L^2$  and by (44), (46) and (47)

$$\|\zeta^c\|_\infty \leq \|a\|_\infty + \|\zeta^c - a\|_\infty \leq \|a\|_\infty + \|\zeta^c - a\|_{H^{\gamma_c}} < \infty,$$

since  $\gamma_c > 1/2$ . Consequently, we can apply the smoothed adjoint equality (13) and obtain for the stochastic error term

$$\begin{aligned} & \int \zeta_t(x) \mathcal{F}^{-1} \left[ \mathcal{F} K_h \frac{\varphi_n - \varphi}{\varphi_\varepsilon} \right] (x) dx \\ &= \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} K_h] * \zeta_t(x) (\mathbb{P}_n - \mathbb{P})(dx). \end{aligned} \tag{15}$$

Therefore, it suffices for the convergence of the finite dimensional distributions to bound the term

$$\sup_{h \in (0,1)} \int |\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} K_h] * \zeta(x)|^{2+\delta} \mathbb{P}(\mathrm{d}x), \quad (16)$$

for any function  $\zeta \in Z^{\gamma_s, \gamma_c}$ . Then the stochastic error term converges in distribution to a normal random variable by the central limit theorem under the Lyapunov condition (i.e., Klenke, 2007, Thm. 15.43 together with Lem. 15.41). Finally, the Cramér-Wold device yields the convergence of the finite dimensional distributions in Theorem 1.

First, note that the moment conditions in Assumptions 1 and 2 and the estimate

$$|x|^p f_Y(x) \leq \int |x - y + y|^p f_X(x - y) f_\varepsilon(y) \mathrm{d}y \lesssim (|y|^p f_X) * f_\varepsilon + f_X * (|y|^p f_\varepsilon),$$

for  $x \in \mathbb{R}$ ,  $p \geq 1$ , yield finite  $(2 + \delta)$ th moments for  $\mathbb{P}$  since

$$\int |x|^{2+\delta} f_Y(x) \mathrm{d}x \lesssim \| |x|^{2+\delta} f_X \|_{L^1} \|f_\varepsilon\|_{L^1} + \|f_X\|_{L^1} \| |x|^{2+\delta} f_\varepsilon \|_{L^1} < \infty. \quad (17)$$

To estimate (16), we rewrite

$$\begin{aligned} \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta^c(x) &= \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u)(\mathrm{Id} + \mathrm{D}) \mathcal{F}[\frac{1}{iy+1} \zeta^c(y)](u)](x) \\ &= \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F}[\frac{1}{iy+1} \zeta^c(y)](u)](x) \\ &\quad + \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) (\mathcal{F}[\frac{1}{iy+1} \zeta^c(y)])'(u)](x) \\ &= (1 + ix) \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F}[\frac{1}{iy+1} \zeta^c(y)](u)](x) \\ &\quad + \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u) \mathcal{F}[\frac{1}{iy+1} \zeta^c(y)](u)](x), \end{aligned} \quad (18)$$

owing to the product rule for differentiation. Hence,

$$\begin{aligned} \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta(x) &= \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F} \zeta^s(u)](x) \\ &\quad + (1 + ix) \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F}[\frac{1}{iy+1} \zeta^c(y)](u)](x) \\ &\quad + \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u) \mathcal{F}[\frac{1}{iy+1} \zeta^c(y)](u)](x). \end{aligned} \quad (19)$$

While  $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta$  may exist only in distributional sense in general, it is defined rigorously through the right-hand side of the above display for  $\zeta \in Z^{\gamma_s, \gamma_c}$ . Considering  $\zeta * K_h$  instead of  $\zeta$ , we estimate separately all three terms in the following.

The continuity and linearity of the Fourier multiplier  $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)]$ , which was shown in Lemma 5(i), yield for the first term in (19)

$$\begin{aligned} \|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F} \zeta^s(u) \mathcal{F} K_h(u)]\|_{H^\delta} &= \|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F}[\zeta^s * K_h]]\|_{B_{2,2}^\delta} \\ &\lesssim \|\zeta^s * K_h\|_{B_{2,2}^{\beta+\delta}} \lesssim \|\zeta^s\|_{H^{\beta+\delta}}, \end{aligned}$$

where the last inequality holds by  $\|\mathcal{F} K_h\|_\infty \leq \|K\|_{L^1}$ . Using the boundedness of  $f_Y$  and the continuous Sobolev embedding  $H^{\delta/4}(\mathbb{R}) \subseteq L^{2+\delta}(\mathbb{R})$  by (44), (47)



and (46), we obtain

$$\begin{aligned}
& \| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F} \zeta^s(u) \mathcal{F} K_h(u)] \|_{L^{2+\delta}(\mathbb{P})} \\
& \lesssim \| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F} \zeta^s(u) \mathcal{F} K_h(u)] \|_{L^{2+\delta}} \\
& \lesssim \| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F} \zeta^s(u) \mathcal{F} K_h(u)] \|_{H^\delta} \\
& \lesssim \| \zeta^s \|_{H^{\beta+\delta}}
\end{aligned} \tag{20}$$

To estimate the second term in (19), we use the Cauchy–Schwarz inequality and Assumption 2(ii):

$$\begin{aligned}
& \| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F}[\frac{1}{ix+1} \zeta^c(x)](u) \mathcal{F} K_h(u)] \|_\infty \\
& \leq \| \varphi_\varepsilon^{-1}(-u) \mathcal{F}[\frac{1}{ix+1} \zeta^c] \mathcal{F} K_h(u) \|_{L^1} \\
& \lesssim \| \langle u \rangle^{-1/2-\beta-\delta} \varphi_\varepsilon^{-1}(-u) \|_{L^2} \| \langle u \rangle^{1/2+\beta+\delta} \mathcal{F}[\frac{1}{ix+1} \zeta^c(x)] \|_{L^2} \\
& \lesssim \| \frac{1}{ix+1} \zeta^c(x) \|_{H^{1/2+\beta+\delta}}.
\end{aligned}$$

Thus  $\int (1+x^2)^{(2+\delta)/2} f_Y(x) dx < \infty$  from (17) yields

$$\begin{aligned}
& \| (1+ix) \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F}[\frac{1}{iy+1} \zeta^c(y)](u) \mathcal{F} K_h(u)](x) \|_{L^{2+\delta}(\mathbb{P})} \\
& \lesssim \| \frac{1}{ix+1} \zeta^c(x) \|_{H^{1/2+\beta+\delta}}.
\end{aligned} \tag{21}$$

The last term in the decomposition (19) can be estimated similarly using the Cauchy–Schwarz inequality and Assumption 2(ii) for  $(\varphi^{-1})'$

$$\begin{aligned}
& \| \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u) \mathcal{F}[\frac{1}{ix+1} \zeta^c(x)](u) \mathcal{F} K_h(u)] \|_{L^{2+\delta}(\mathbb{P})} \\
& \lesssim \| (\varphi_\varepsilon^{-1})'(-u) \mathcal{F}[\frac{1}{ix+1} \zeta^c(x)](u) \|_{L^1} \\
& \leq \| \langle u \rangle^{1/2-\beta-\delta} (\varphi_\varepsilon^{-1})' \|_{L^2} \| \langle u \rangle^{-1/2+\beta+\delta} \mathcal{F}^{-1}[\frac{1}{ix+1} \zeta^c(x)](u) \|_{L^2} \\
& \lesssim \| \frac{1}{ix+1} \zeta^c(x) \|_{H^{-1/2+\beta+\delta}}.
\end{aligned} \tag{22}$$

Combining (20), (21) and (22), we obtain

$$\sup_{h \in (0,1)} \| \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} K_h] * \zeta(x) \|_{L^{2+\delta}(\mathbb{P})} \lesssim \| \zeta \|_{Z^{\beta+\delta, 1/2+\beta+\delta}}, \tag{23}$$

which is finite for  $\delta$  small enough satisfying  $\beta+\delta \leq \gamma_s$  and  $1/2+\beta+\delta \leq \gamma_c$ . Since  $\mathcal{F} K_h$  converges pointwise to one and  $|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} K_h] * \zeta(x)|^2$  is uniformly integrable by the bound of the  $(2+\delta)$ th moments, the variance converges to

$$\int |\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta(x)|^2 \mathbb{P}(dx).$$

#### 4.2. Tightness

Motivated by the representation (15) of the stochastic error, we introduce the empirical process

$$\nu_n(t) := \sqrt{n} \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} K_h] * \zeta_t(x) (\mathbb{P}_n - \mathbb{P})(dx), \quad t \in \mathbb{R}. \tag{24}$$

In order to show tightness of the empirical process, we first show some properties of the class of translations  $\mathcal{H} := \{\zeta_t | t \in \mathbb{R}\}$  for  $\zeta \in Z^{\gamma_s, \gamma_c}$ .

**Lemma 6.** For  $\zeta \in Z^{\gamma_s, \gamma_c}$  the following is satisfied:

- (i) The decomposition  $\zeta_t = \zeta_t^c + \zeta_t^s$  satisfies the conditions in the definition of  $Z^{\gamma_s, \gamma_c}$  with  $a_t$ . We have  $\sup_{t \in \mathbb{R}} \|\zeta_t\|_{Z^{\gamma_s, \gamma_c}} < \infty$ .
- (ii) For any  $\eta \in (0, \gamma_s)$ , there exists  $\tau > 0$  such that  $\|\zeta_t - \zeta_s\|_{Z^{\gamma_s - \eta, \gamma_c - \eta}} \lesssim |t - s|^\tau$  holds for all  $s, t \in \mathbb{R}$  with  $|t - s| \leq 1$ .

*Proof.*

- (i) Since  $\|\zeta_t^s\|_{H^{\gamma_s}}^2 = \int \langle u \rangle^{2\gamma_s} |e^{itu} \mathcal{F} \zeta^s(u)|^2 du = \|\zeta^s\|_{H^{\gamma_s}}^2$ , both claims hold for the singular part. Applying the pointwise multiplier property of Besov spaces (48) as well as the Besov embeddings (47) and (45), we obtain for some  $M > \gamma_c$  and  $a \in C^\infty(\mathbb{R})$  as in definition (6)

$$\begin{aligned} \|\langle x \rangle^\tau (\zeta_t^c(x) - a_t(x))\|_{H^{\gamma_c}} &\lesssim \|\frac{\langle x \rangle^\tau}{\langle x-t \rangle^\tau}\|_{C^M} \|\langle x-t \rangle^\tau (\zeta_t^c(x) - a_t(x))\|_{H^{\gamma_c}} \\ &= \|\frac{\langle x \rangle^\tau}{\langle x-t \rangle^\tau}\|_{C^M} \|\langle x \rangle^\tau (\zeta^c(x) - a(x))\|_{H^{\gamma_c}}, \end{aligned}$$

which is finite for all  $t \in \mathbb{R}$  since  $\langle x \rangle^\tau \langle x-t \rangle^{-\tau} \in C^M(\mathbb{R})$ . For the second claim we estimate similarly

$$\begin{aligned} \sup_{t \in \mathbb{R}} \|\frac{1}{ix+1} \zeta_t^c(x)\|_{H^{\gamma_c}} &\lesssim \sup_{t \in \mathbb{R}} \|\frac{a_t(x)}{ix+1}\|_{H^{\gamma_c}} + \|\frac{1}{ix+1}\|_{C^M} \sup_{t \in \mathbb{R}} \|\zeta_t^c - a_t\|_{H^{\gamma_c}} \\ &\lesssim \|\frac{1}{ix+1}\|_{H^{\gamma_c}} \|a\|_{C^M} + \|\frac{1}{ix+1}\|_{C^M} \|\zeta^c - a\|_{H^{\gamma_c}} < \infty. \end{aligned}$$

- (ii) For the singular part note that

$$\begin{aligned} &\|\zeta_t^s - \zeta_s^s\|_{H^{\gamma_s - \eta}} \\ &\leq \|\langle u \rangle^{\gamma_s} \mathcal{F} \zeta^s(u)\|_{L^2} \|\langle u \rangle^{-\eta} (1 - e^{i(t-s)u})\|_\infty \\ &\lesssim \|\langle u \rangle^{-\eta}\|_{L^\infty(\mathbb{R} \setminus (-|t-s|^{-1/2}, |t-s|^{-1/2}))} \\ &\quad \vee \|(1 - e^{i(t-s)u})\|_{L^\infty((-|t-s|^{-1/2}, |t-s|^{-1/2}))} \\ &\lesssim |t-s|^{\eta/2} \vee |t-s|^{1/2}. \end{aligned}$$

For  $\zeta^c$  we have

$$\begin{aligned} \left\| \frac{1}{ix+1} (\zeta_t^c(x) - \zeta_s^c(x)) \right\|_{H^{\gamma_c - \eta}} &\lesssim \left\| \frac{1}{ix+1} \zeta_t^c(x) - \left( \frac{1}{ix+1} \zeta_t^c(x) \right) * \delta_{s-t} \right\|_{H^{\gamma_c - \eta}} \\ &\quad + \left\| \frac{1}{i(x-s+t)+1} \zeta_s^c(x) - \frac{1}{ix+1} \zeta_s^c(x) \right\|_{H^{\gamma_c - \eta}}. \end{aligned}$$

The first term can be treated analogously to  $\zeta^s$ . Using some integer  $M \in \mathbb{N}$  strictly larger than  $\gamma_c$ , the second term can be estimated by

$$\begin{aligned} &\left\| \frac{1}{i(x-s+t)+1} \zeta_s^c(x) - \frac{1}{ix+1} \zeta_s^c(x) \right\|_{H^{\gamma_c - \eta}} \\ &\lesssim |t-s| \left\| \frac{1}{i(x-s+t)+1} \frac{1}{ix+1} \zeta_s^c(x) \right\|_{H^{\gamma_c - \eta}} \\ &\lesssim |t-s| \left\| \frac{1}{i(x-s+t)+1} \right\|_{C^M} \left\| \frac{1}{ix+1} \zeta_s^c(x) \right\|_{H^{\gamma_c - \eta}} \\ &\lesssim |t-s|, \end{aligned}$$

where we used again pointwise multiplier (48), embedding properties of Besov spaces (47) and (45) as well as (i).  $\square$

#### 4.2.1. Pregaussian limit process

Let  $\mathbb{G}$  be the stochastic process from Theorem 1. It induces the intrinsic covariance metric  $d(s, t) := \mathbb{E}[(\mathbb{G}_s - \mathbb{G}_t)^2]^{1/2}$ .

**Theorem 7.** *There exists a version of  $\mathbb{G}$  with uniformly  $d$ -continuous sample paths almost surely and with  $\sup_{t \in \mathbb{R}} |\mathbb{G}_t| < \infty$  almost surely.*

The proof of the theorem shows in addition that  $\mathbb{R}$  is totally bounded with respect to  $d$ . The boundedness of the sample paths follows from the totally bounded index set and the uniform continuity. Further we conclude that  $\mathcal{G}$  defined in (7) is  $\mathbb{P}$ -pregaussian by van der Vaart and Wellner (1996, p. 89). Thus  $\mathbb{G}$  is a tight Borel random variable in  $\ell^\infty(\mathbb{R})$  and the law of  $\mathbb{G}$  is uniquely defined through the covariance structure and the sample path properties in the theorem (van der Vaart and Wellner, 1996, Lem. 1.5.3).

*Proof.* To show that the class is pregaussian, it suffices to verify polynomial covering numbers. To that end we deduce that

$$d(s, t) = (\|g_t - g_s\|_{L^2(\mathbb{P})}^2 - \langle \zeta_t - \zeta_s, f_X \rangle^2)^{1/2} \leq \|g_t - g_s\|_{L^2(\mathbb{P})} \quad (25)$$

decreases polynomial for  $|t - s| \rightarrow 0$ , for  $\max(s, t) \rightarrow -\infty$  and for  $\min(s, t) \rightarrow \infty$ . Using the same estimates which show the moment bound (23) but replacing  $\mathcal{F}K_h = 1$ , we obtain

$$\|\mathcal{F}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta\|_{L^2(\mathbb{P})} \lesssim \|\zeta\|_{Z^{\beta+\delta, 1/2+\beta+\delta}} \quad (26)$$

and thus by choosing  $\delta$  and  $\eta$  small enough Lemma 6 yields  $d(s, t) \lesssim \|\zeta_t - \zeta_s\|_{Z^{\beta+\delta, 1/2+\beta+\delta}} \lesssim |t - s|^\tau$ . We now turn to the estimation of the tails. We will only consider the case  $s, t \geq N$  since the case  $s, t \leq N$  can be treated in the same way. Without loss of generality, let  $s < t$ .

For the smooth component of  $\zeta$  we have to show that  $\|\frac{1}{ix+1}(\zeta_t^c - \zeta_s^c)(x)\|_{H^{\gamma_c}}$  with  $t, s \geq N$  decays polynomially in  $N$ . It is sufficient to prove  $\|\frac{1}{ix+1}(\zeta_t^c - a_t)(x)\|_{H^{\gamma_c}}$  and  $\|\frac{1}{ix+1}(a_t - a_s)(x)\|_{H^{\gamma_c}}$  with  $a \in C^\infty(\mathbb{R})$  from definition (6) of  $Z^{\gamma_s, \gamma_c}$  both decay polynomially in  $N$ . Let  $M > \gamma_c$  and  $\psi \in C^M(\mathbb{R})$  with  $\psi(x) = 1$  for  $x \in \mathbb{R} \setminus [-\frac{1}{2}, \frac{1}{2}]$  and  $\psi(x) = 0$  for  $x \in [-\frac{1}{4}, \frac{1}{4}]$ . The pointwise multiplier property (48) yields

$$\begin{aligned} & \left\| \frac{1}{ix+1}(\zeta_t^c - a_t)(x) \right\|_{H^{\gamma_c}} \\ &= \left\| (\psi(x/N) + (1 - \psi(x/N))) \frac{1}{ix+it+1}(\zeta^c - a)(x) \right\|_{H^{\gamma_c}} \\ &\lesssim \left\| \frac{1}{ix+it+1} \right\|_{C^M} \|\psi(x/N)(\zeta^c - a)(x)\|_{H^{\gamma_c}} + \left\| \frac{1-\psi(x/N)}{ix+it+1} \right\|_{C^M} \|\zeta^c - a\|_{H^{\gamma_c}} \\ &\lesssim \|\langle x \rangle^{-\tau} \psi(x/N)\|_{C^M} \|\langle x \rangle^\tau (\zeta^c - a)(x)\|_{H^{\gamma_c}} + N^{-1} \|\zeta^c - a\|_{H^{\gamma_c}} \\ &\lesssim N^{-(\tau \wedge 1)} \end{aligned}$$

and for  $N$  large enough such that  $\text{supp}(a') \subseteq [-N/2, N/2]$  we obtain

$$\begin{aligned} & \left\| \frac{1}{ix+1} (a_t - a_s)(x) \right\|_{H^{\gamma_c}} \\ &= \left\| \frac{\psi(x/N)}{ix+1} (a_t - a_s)(x) \right\|_{H^{\gamma_c}} \lesssim \left\| \frac{\psi(x/N)}{ix+1} \right\|_{H^{\gamma_c}} \left\| (a_t - a_s)(x) \right\|_{C^M} \\ &\lesssim \left\| (ix+1)^{-3/4} \right\|_{H^{\gamma_c}} \left\| \psi(x/N)(ix+1)^{-1/4} \right\|_{C^M} \lesssim N^{-1/4}. \end{aligned}$$

To bound the singular part it suffices to show that

$$\left\| \mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet)] * \zeta_t^s \right\|_{L^2(\mathbb{P})}, \quad t \geq N,$$

decays polynomially in  $N$ . To this end, we split the integral domain into

$$\begin{aligned} \left\| \mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet)] * \zeta_t^s \right\|_{L^2(\mathbb{P})}^2 &= \int_{-\infty}^{-N/2} |\mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F} \zeta^s](x)|^2 f_Y(x+t) dx \\ &\quad + \int_{-N/2}^{\infty} |\mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F} \zeta^s](x)|^2 f_Y(x+t) dx. \end{aligned} \quad (27)$$

To estimate the first term, we use the following auxiliary calculations

$$\begin{aligned} & ix \mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F} \zeta^s](x) \\ &= -\mathcal{F}^{-1}[(\varphi_\epsilon^{-1})'(-\bullet) \mathcal{F} \zeta^s](x) + \mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F}[iy\zeta^s(y)]](x) \end{aligned}$$

and with an integer  $M \in \mathbb{N}$  strictly larger than  $\gamma_s$  and a function  $\chi \in C^M(\mathbb{R})$  which is equal to one on  $\text{supp}(\zeta^s)$  and has compact support

$$\begin{aligned} \|y\zeta^s(y)\|_{H^{\gamma_s}} &= \|y\chi(y)\zeta^s(y)\|_{H^{\gamma_s}} \lesssim \|y\chi(y)\|_{B_{\infty,2}^{\gamma_s}} \|\zeta^s(y)\|_{H^{\gamma_s}} \\ &\lesssim \|y\chi(y)\|_{C^M} < \infty, \end{aligned}$$

where we used the pointwise multiplier property (48) of Besov spaces as well as the Besov embeddings (47) and (45). Thus  $ix \mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F} \zeta^s](x) \in L^2(\mathbb{R})$ . Applying this and the boundedness of  $f_Y$  to the first term in (27) yields

$$\begin{aligned} & \int_{-\infty}^{-N/2} |\mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F} \zeta^s](x)|^2 f_Y(x+t) dx \\ &\lesssim \int_{-\infty}^{-N/2} |\mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F} \zeta^s](x)|^2 dx \\ &\leq 4N^{-2} \int_{-\infty}^{-N/2} |x \mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F} \zeta^s](x)|^2 dx \lesssim N^{-2}. \end{aligned}$$

Using Hölder's inequality and the boundedness of  $f_Y$ , we estimate the second term in (27) by

$$\begin{aligned} & \left\| \mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F} \zeta^s](x) \right\|_{L^{2+\delta}}^2 \left( \int_{-N/2}^{\infty} |f_Y(x+t)|^{(2+\delta)/\delta} dx \right)^{\delta/(2+\delta)} \\ &\lesssim \left\| \mathcal{F}^{-1}[\varphi_\epsilon^{-1}(-\bullet) \mathcal{F} \zeta^s](x) \right\|_{L^{2+\delta}}^2 \left( \int_{N/2}^{\infty} f_Y(x) dx \right)^{\delta/(2+\delta)}. \end{aligned}$$

While the first factor is finite according to our bound (20), which also holds when  $\mathcal{F}K_h$  is omitted, the second one is of order  $N^{-\delta}$  due to the finite  $(2+\delta)$ th moment of  $\mathbb{P}$ . Therefore, the second term in (27) decays polynomially.  $\square$

#### 4.2.2. Uniform central limit theorem

We recall the definition of the empirical process  $\nu_n$  in (24).

**Theorem 8.** *Grant Assumptions 1 and 2. Let*

$$(\nu_n(t_1), \dots, \nu_n(t_k)) \xrightarrow{\mathcal{L}} (\mathbb{G}_{t_1}, \dots, \mathbb{G}_{t_k})$$

for all  $t_1, \dots, t_k \in \mathbb{R}$  and for all  $k \in \mathbb{N}$ . If either  $\gamma_s \leq \beta + 1/2$  and  $h_n^\rho n^{1/4} \rightarrow \infty$  as  $n \rightarrow \infty$  for some  $\rho > \beta - \gamma_s + 1/2$  or if  $\gamma_s > \beta + 1/2$ , then

$$\nu_n \xrightarrow{\mathcal{L}} \mathbb{G} \quad \text{in } \ell^\infty(\mathbb{R}).$$

*Proof.* We split the empirical process  $\nu_n$  into three parts

$$\nu_n = \sqrt{n} \int (T_1(x) + T_2(x) + T_3(x))(\mathbb{P}_n - \mathbb{P})(dx),$$

where  $T_1$ ,  $T_2$  and  $T_3$  correspond to the three terms in decomposition (19) and are given by (28), (29) and (30) below. For the first term

$$T_1(x) = \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F}\zeta_t^s(u) \mathcal{F}K_h(u)](x) \quad (28)$$

we distinguish the two cases  $\gamma_s > \beta + 1/2$  and  $\gamma_s \leq \beta + 1/2$ . In the first case, we will show that  $T_1$  varies in a fixed Donsker class. In the second case, the process indexed by  $T_1$  is critical, this is where smoothed empirical processes and the condition on the bandwidth are needed. Tightness of  $T_1$  in this case will be shown in Section 4.2.3. We will further show that the second term  $T_2$  and the third term  $T_3$  are both varying in fixed Donsker classes for all  $\gamma_s > \beta$ . In particular the three processes indexed by  $T_1$ ,  $T_2$  and  $T_3$ , respectively, are tight. Applying the equicontinuity characterization of tightness (van der Vaart and Wellner, 1996, Thm. 1.5.7) with the maximum of the semimetrics yields that  $\nu_n$  is tight. Since we have assumed convergence of the finite dimensional distribution, the convergence of  $\nu_n$  in distribution follows (van der Vaart and Wellner, 1996, Thm. 1.5.4).

Here we consider only the first case  $\gamma_s > \beta + 1/2$ . We recall that  $\zeta_t^s$  is contained in  $H^{\gamma_s}(\mathbb{R})$ . By the Fourier multiplier property of the deconvolution operator in Lemma 5(i) and by  $\sup_{h>0, u} |\mathcal{F}K_h(u)| \leq \|K\|_{L^1} < \infty$  the functions  $T_1$  are contained in a bounded set of  $H^{1/2+\eta}(\mathbb{R})$  for some  $\eta > 0$  small enough. We apply (Nickl and Pötscher, 2007, Prop. 1) with  $p = q = 2$  and  $s = 1/2 + \eta$  and conclude that  $T_1$  varies in a universal Donsker class.

The second term is of the form

$$T_2(x) = (1 + ix) \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F}[\frac{1}{iy+1} \zeta_t^c(y)](u) \mathcal{F}K_h(u)](x). \quad (29)$$

By Assumption 2(ii) we have  $\varphi_\varepsilon^{-1}(u) \lesssim \langle u \rangle^\beta$ . For some  $\eta > 0$  sufficiently small, the functions  $\frac{1}{iy+1}\zeta_t^c(y)$ ,  $t \in \mathbb{R}$ , are contained in a bounded set of  $H^{\beta+\eta+1/2}(\mathbb{R})$  by Lemma 6. We obtain that the functions  $T_2(x)/(1+ix)$  are contained in a bounded subset of  $H^{1/2+\eta}(\mathbb{R})$ . Corollary 5 by Nickl and Pötscher (2007) yields with  $p = q = 2$ ,  $\beta = -1$ ,  $s = 1/2 + \eta$  and  $\gamma = \eta$  that  $T_2$  is contained in a fixed  $\mathbb{P}$ -Donsker class.

Similarly, we treat the third term

$$T_3(x) = \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1})'(-u) \mathcal{F}[\frac{1}{iy+1}\zeta_t^c(y)](u) \mathcal{F}K_h(u)](x). \quad (30)$$

By Assumption 2(ii) we have  $(\varphi_\varepsilon^{-1})' \lesssim \langle u \rangle^{\beta-1}$ . As above we conclude that the functions  $T_3$  are contained in a bounded set of  $H^{\eta+3/2}(\mathbb{R})$ . Applying Proposition 1 by Nickl and Pötscher (2007) with  $p = q = 2$  and  $s = \eta + 3/2$ , the term  $T_3$  varies in a universal Donsker class.  $\square$

#### 4.2.3. The critical term

In this section, we treat the first term  $T_1$  in the case  $\gamma_s \leq \beta + 1/2$ . We define

$$q_t := \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F}\zeta_t^s(u)]. \quad (31)$$

For simplicity in point (e) below it will be convenient to work with functions  $K_h$  of bounded support. Thus we fix  $\xi > 0$  and define the truncated kernel

$$K_h^{(0)} := K_h \mathbf{1}_{[-\xi, \xi]}.$$

By the Assumption (5) on the decay of  $K$  we have  $\sup_{h>0} \|K_h - K_h^{(0)}\|_{BV} < \infty$ . We conclude  $\mathcal{F}(K_h - K_h^{(0)})(u) \lesssim (1 + |u|)^{-1}$  with a constant independent of  $h > 0$ . By Assumption 2(ii) we have  $|\varphi_\varepsilon^{-1}(u)| \lesssim (1 + |u|)^\beta$ . The functions  $\zeta_t^s(u)$ ,  $t \in \mathbb{R}$ , are contained in a bounded set of  $H^{\gamma_s}(\mathbb{R})$ . Consequently,  $T_1$  with  $K_h - K_h^{(0)}$  instead of  $K_h$  is contained in a bounded set of  $H^{\gamma_s - \beta + 1}(\mathbb{R})$ . With the same argument as used for  $T_3$ , we see that this term is contained in a universal Donsker class because  $\gamma_s - \beta + 1 > 1$  by assumption. So it remains to consider  $T_1$  with the truncated kernel  $K_h^{(0)}$ .

In order to show tightness of the process indexed by  $T_1$  with the truncated kernel  $K_h^{(0)}$ , we check the assumptions of Theorem 3 by Giné and Nickl (2008) in the version of Nickl and Reiß (2012, Thm. 12) for the class  $Q = \{q_t | t \in \mathbb{R}\}$  and for  $\mu_n(dx) := K_{h_n}^{(0)}(x) dx$ , where  $q_t(x)$  was defined in (31). By Section 4.2.1 the class  $\mathcal{G}$  is  $\mathbb{P}$ -pregaussian. From the proof also follows that  $Q$  is  $\mathbb{P}$ -pregaussian since this is just the case  $\zeta^c = 0$ .

We write

$$Q'_\tau := \{r - q | r, q \in Q, \|r - q\|_{L^2(\mathbb{P})} \leq \tau\}.$$

Let  $\rho > \beta - \gamma_s + 1/2 \geq 0$  be such that  $h_n^\rho n^{1/4} \rightarrow \infty$ . We fix some  $\rho' \in (\beta - \gamma_s + 1/2, \rho \wedge 1)$  and obtain  $h_n^{\rho'} \log(n)^{-1/2} n^{1/4} \rightarrow \infty$ . We need to verify the following conditions.

- (a) We will show that the functions in  $\tilde{Q}_n := \{q_t * \mu_n | t \in \mathbb{R}\}$  are bounded by  $M_n := Ch_n^{-\rho'}$  for some constant  $C > 0$ . Since  $q_t$  is only a translation of  $q_0$  it suffices to consider  $q_0$ . By the definition of  $Z^{\gamma_s, \gamma_c}$  in (6), by Lemma 5(i) and by the Besov embedding (47) we have

$$q_0 = \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-u) \mathcal{F} \zeta^s(u)] \in B_{2,2}^{\gamma_s - \beta}(\mathbb{R}) \subseteq B_{2,\infty}^{1/2 - \rho'}(\mathbb{R}).$$

By our assumptions on the kernel (5) it follows that  $K'$  is integrable and thus that  $K$  is of bounded variation. Next, we apply continuous embeddings for Besov spaces (44) and (46), (49) as well as the estimate for  $\|K_{h_n}\|_{B_{1,1}^{\rho'}}$  in Giné and Nickl (2008, p. 384), which also applies to truncated kernels, and obtain

$$\|q_0 * K_{h_n}^{(0)}\|_\infty \lesssim \|q_0 * K_{h_n}^{(0)}\|_{B_{\infty,1}^0} \lesssim \|q_0 * K_{h_n}^{(0)}\|_{B_{2,1}^{1/2}} \lesssim \|K_{h_n}^{(0)}\|_{B_{1,1}^{\rho'}} \lesssim h_n^{-\rho'}. \quad (32)$$

- (b) For  $r \in Q'_\tau$  holds  $\|r * K_h^{(0)}\|_{L^2(\mathbb{P})} \leq \|r * K_h^{(0)} - r\|_{L^2(\mathbb{P})} + \tau$ . Thus it suffices to show that  $\|q * K_h^{(0)} - q\|_{L^2(\mathbb{P})} \rightarrow 0$  uniformly over  $q \in Q$ . We estimate

$$\|q_t * K_h^{(0)} - q_t\|_{L^2(\mathbb{P})} \lesssim \|\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} \zeta^s(\mathcal{F} K_h^{(0)} - 1)\|_{L^2}.$$

$\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} \zeta^s$  is an  $L^2$ -function and  $\mathcal{F} K_h^{(0)}$  is uniformly bounded and converges to one as  $h \rightarrow 0$ . By dominated convergence the integral converges to zero.

- (c) The estimates in (a) can be used to see that the classes  $\tilde{Q}_n$  have polynomial  $L^2(\mathbb{Q})$ -covering numbers, uniformly in all probability measures  $\mathbb{Q}$  and uniformly in  $n$ . The function  $q_0 * K_{h_n}^{(0)}$  is the convolution of two  $L^2$ -functions and thus continuous. The estimate (32) and embedding (50) yield that  $q_0 * K_{h_n}^{(0)}$  is of finite 2-variation. We argue as in Lemma 1 by Giné and Nickl (2009). As a function of bounded 2-variation  $q_0 * K_{h_n}^{(0)}$  can be written as a composition  $g_n \circ f_n$  of a nondecreasing function  $f_n$  and a function  $g_n$ , which satisfies a Hölder condition  $|g_n(u) - g_n(v)| \leq |u - v|^{1/2}$ , see, for example, Dudley (1992, p. 1971). More precisely, we can take  $f_n(x)$  to be the 2-variation of  $q_0 * K_{h_n}^{(0)}$  up to  $x$  and the envelopes of  $f_n$  to be multiples of  $M_n^2 = C^2 h_n^{-2\rho'}$ . The set  $F_n$  of all translates of the nondecreasing function  $f_n$  has VC-index 2 and thus polynomial  $L^1(\mathbb{Q})$ -covering numbers (de la Peña and Giné, 1999, Thm. 5.1.15). Since each  $\epsilon^2$ -covering of translates of  $f_n$  for  $L^1(\mathbb{Q})$  induces an  $\epsilon$ -covering of translates of  $g_n \circ f_n$  for  $L^2(\mathbb{Q})$ , we can estimate the covering numbers by

$$N(\tilde{Q}_n, L^2(\mathbb{Q}), \epsilon) \leq N(F_n, L^1(\mathbb{Q}), \epsilon^2) \lesssim (M_n/\epsilon)^4,$$

with constants independent of  $n$  and  $\mathbb{Q}$ . The conditions for inequality (22) by Giné and Nickl (2008) are fulfilled, where the envelopes are  $M_n = Ch_n^{-\rho'}$

and  $H_n(\eta) = H(\eta) = C_1 \log(\eta) + C_0$  with  $C_0, C_1 > 0$ . Consequently

$$\mathbb{E}^* \left\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \varepsilon_j f(X_j) \right\|_{(\tilde{Q}_n)'_{n^{-1/4}}} \lesssim \max \left( \frac{\sqrt{\log(n)}}{n^{1/4}}, \frac{h_n^{-\rho'}}{\sqrt{n}} \log(n) \right) \rightarrow 0$$

as  $n \rightarrow \infty$ .

(d) We apply Lemma 1 of Giné and Nickl (2008) to show that

$$\cup_{n \geq 1} \tilde{Q}_n = \bigcup_{n \geq 1} \left\{ x \mapsto \int_{\mathbb{R}} q_t(x-y) K_{h_n}^{(0)}(y) dy \mid t \in \mathbb{R} \right\}$$

is in the  $L^2(\mathbb{P})$ -closure of  $\|K\|_{L^1}$ -times the symmetric convex hull of the pregaussian class  $Q$ . The condition  $q_t(\bullet - y) \in L^2(\mathbb{P})$  is satisfied for all  $y \in \mathbb{R}$  since  $q_t \in L^2(\mathbb{R})$  and  $f_Y$  is bounded.  $q_t(x - \bullet) \in L^1(|\mu_n|)$  is fulfilled owing to  $K_{h_n}^{(0)}, q_t \in L^2(\mathbb{R})$ . The third condition that  $y \mapsto \|q_t(\bullet - y)\|_{L^2(\mathbb{P})}$  is in  $L^1(|\mu_n|)$  holds likewise since  $f_Y$  is bounded and  $K_{h_n}^{(0)} \in L^1(\mathbb{R})$ .

(e) The  $L^2(\mathbb{P})$ -distance of two functions in  $\tilde{Q}_n$  can be estimated by

$$\begin{aligned} & \mathbb{E} \left[ (q_t * K_h^{(0)}(X) - q_s * K_h^{(0)}(X))^2 \right]^{1/2} \\ &= \left\| \int q_t(\bullet - u) K_h^{(0)}(u) - q_s(\bullet - u) K_h^{(0)}(u) du \right\|_{L^2(\mathbb{P})} \\ &\leq \int |K_h^{(0)}(u)| \|q_t(\bullet - u) - q_s(\bullet - u)\|_{L^2(\mathbb{P})} du \\ &\leq \|K_h^{(0)}\|_{L^1} \sup_{|u| \leq \xi} \|q_t(\bullet - u) - q_s(\bullet - u)\|_{L^2(\mathbb{P})} \\ &= \|K_h^{(0)}\|_{L^1} \sup_{|u| \leq \xi} \|q_{t+u} - q_{s+u}\|_{L^2(\mathbb{P})}. \end{aligned}$$

As seen in the proof that  $Q$  is pregaussian, the covering numbers grow at most polynomially. We take  $N$  large enough such that  $N \geq 2\xi$ . Then  $s, t > N$  implies  $s+u, t+u > N/2$  and  $s, t < -N$  implies  $s+u, t+u < -N/2$ . Since this is only a polynomial change in  $N$ , the growth of the covering numbers remains at most polynomial. This leads to the entropy bound  $H(\tilde{Q}_n, L^2(\mathbb{P}), \eta) \lesssim \log(\eta^{-1})$  for  $\eta$  small enough and independent of  $n$ . We define  $\lambda_n(\eta) := \log(\eta^{-1})\eta^2$ . The bound in the condition is of the order  $\log(n)^{-1/2}n^{1/4}$ . As seen before (a) this growth faster than  $M_n = Ch_n^{-\rho'}$ .

## 5. Proof of the lower bound

First, we show asymptotic linearity of  $\hat{\vartheta}_\zeta$ .

**Lemma 9.** *Supposing Assumptions 1 and 2 and  $\zeta \in Z^{\gamma_s, \gamma_c}$  with  $\gamma_s > \beta$  and  $\gamma_c > (1/2 \vee \alpha) + \gamma_s$ , the estimator  $\hat{\vartheta}_\zeta$  with  $h_n = o(n^{-1/(2\alpha+2\gamma_s)})$  is asymptotically linear with influence function  $x \mapsto \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta(y)(\delta_x - \mathbb{P})(dy)$  and thus  $\hat{\vartheta}_\zeta$  is Gaussian regular.*



*Proof.* The analysis of the bias of  $\widehat{\vartheta}$  in Section 4.1.1 yields

$$\begin{aligned}\widehat{\vartheta} &= \vartheta + \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet) \mathcal{F} K_h] * \zeta(y)(\mathbb{P}_n - \mathbb{P})(dy) + o_P(n^{-1/2}) \\ &= \vartheta + \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta(y)(\mathbb{P}_n - \mathbb{P})(dy) \\ &\quad + \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)(\mathcal{F} K_h - 1)] * \zeta(y)(\mathbb{P}_n - \mathbb{P})(dy) + o_P(n^{-1/2}).\end{aligned}$$

Since

$$\mathbb{E} \left[ \left| \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta(d\delta_x - d\mathbb{P}) \right|^2 \right] \leq 4 \mathbb{E} \left[ \int |\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta|^2 d\mathbb{P} \right]$$

is finite and  $\mathbb{E}[\int \mathcal{F}^{-1}[(\varphi_\varepsilon^{-1}(-\bullet)] * \zeta)(d\delta_x - d\mathbb{P})] = 0$  by (23) it suffices to show

$$\int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)(\mathcal{F} K_h - 1)] * \zeta(y)(\mathbb{P}_n - \mathbb{P})(dy) = o_P(n^{-1/2}). \quad (33)$$

For convenience we write  $\psi_h := \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)(\mathcal{F} K_h - 1)] * \zeta$  and let  $\tau > 0$ . Since  $(Y_j)$  are independent and identically distributed, we obtain

$$\begin{aligned}\mathbb{P} \left( \left| n^{1/2} \int \psi_h(y)(\mathbb{P}_n - \mathbb{P})(dy) \right| > \tau \right) &\leq \tau^{-2} n \mathbb{E} \left[ \left| \int \psi_h(y)(\mathbb{P}_n - \mathbb{P})(dy) \right|^2 \right] \\ &= \tau^{-2} n \mathbb{E} \left[ \int \int \psi_h(y) \overline{\psi_h}(z) (\mathbb{P}_n - \mathbb{P})(dy) (\mathbb{P}_n - \mathbb{P})(dz) \right] \\ &= \tau^{-2} n^{-1} \sum_{j,k=1}^n \mathbb{E} \left[ \int \int \psi_h(y) \overline{\psi_h}(z) (\delta_{Y_j} - \mathbb{P})(dy) (\delta_{Y_k} - \mathbb{P})(dz) \right] \\ &= \tau^{-1} \mathbb{E} \left[ \left| \int \psi_h(y) (\delta_{Y_j} - \mathbb{P})(dy) \right|^2 \right] \\ &\leq 4\tau^{-1} \int |\psi_h(y)|^2 \mathbb{P}(dy).\end{aligned}$$

By uniform integrability of  $\psi_h^2$  with respect to  $\mathbb{P}$  by (23) and pointwise convergence  $\psi_h \rightarrow 0$  as  $h \rightarrow 0$  we conclude  $\int |\psi_h(y)|^2 \mathbb{P}(dy) \rightarrow 0$  and thus (33). From asymptotic linearity follows Gaussian regularity by Proposition 2.2.1 by Bickel et al. (1998).  $\square$

Let us now briefly discuss the consequence of Assumption 4 in terms of Fourier multipliers. Standard calculus yields  $|(\varphi_\varepsilon^{-1})^{(k)}(u)| \lesssim \langle u \rangle^{\beta-k}$  for  $k = 0, \dots, (\lfloor \beta \rfloor \vee M) + 1$ . With the same arguments as in the proof of Lemma 5(i) we deduce that

$$(1 + iu)^{\beta+k} \varphi_\varepsilon^{(k)}(u) \text{ and } (1 + iu)^{-\beta+k} (\varphi_\varepsilon^{-1})^{(k)}(u) \quad (34)$$

are Fourier multipliers on  $B_{p,q}^s(\mathbb{R})$  for all  $s \in \mathbb{R}, p, q \in [1, \infty]$  and  $k = 0, \dots, \lfloor \beta \rfloor \vee M$ .

### 5.1. Information bound for smooth functionals

In this subsection we prove Theorem 2.

**Step 1:** To determine the solution of the maximization problem (10), we define  $h := Sg = (g * f_\varepsilon) f_Y^{-1/2}$  with score operator  $S$  such that the Fisher information (9) satisfies  $\langle Ig, g \rangle = \|h\|_{L^2}^2$ . Therefore, we obtain  $g = S^{-1}h = \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * (\sqrt{f_Y}h)$ . Owing to the adjoint equation (12),

$$\langle g, \zeta \rangle = \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta) \sqrt{f_Y}h = \langle h, (S^{-1})^* \zeta \rangle$$

holds. Ignoring all restrictions on  $g$ , the supremum is thus attained at

$$h^* := (S^{-1})^* \zeta = (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta) \sqrt{f_Y}. \quad (35)$$

Let us define  $\bar{\beta} := \lfloor \beta + 1/2 \rfloor + 1$  and  $r := \mathcal{F}^{-1}[(1 + iu)^{-\bar{\beta}} \varphi_\varepsilon^{-1}(u)]$ . Because of Lemma 5(ii) we obtain  $r \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  and  $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(u)] = r * (\text{Id} - D)^{\bar{\beta}}$ . Therefore, the condition  $\int g = 0$ , Fubini's theorem and the fundamental theorem of calculus, provided  $(\sqrt{f_Y}h)^{(k)} \in L^1(\mathbb{R})$ ,  $k = 0, \dots, \bar{\beta}$ , imply

$$\begin{aligned} 0 &= \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * (\sqrt{f_Y}h) = \int r * \left( \sqrt{f_Y}h + \sum_{k=1}^{\bar{\beta}} \binom{\bar{\beta}}{k} (-1)^k (\sqrt{f_Y}h)^{(k)} \right) \\ &= \int r \left( \int \sqrt{f_Y}h + \sum_{k=1}^{\bar{\beta}} \binom{\bar{\beta}}{k} (-1)^k \int (\sqrt{f_Y}h)^{(k)} \right). \end{aligned}$$

For each  $k = 1, \dots, \bar{\beta}$  the integrability of  $(\sqrt{f_Y}h)^{(l)}$ ,  $l = k-1, k$ , yields then  $\int (\sqrt{f_Y}h)^{(k)} = \lim_{x \rightarrow \infty} (\sqrt{f_Y}h)^{(k-1)}(x) - (\sqrt{f_Y}h)^{(k-1)}(-x) = 0$  and thus

$$0 = \int \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * (\sqrt{f_Y}h) = \int \sqrt{f_Y}h, \quad (36)$$

since  $\int r = \mathcal{F}r(0) = 1$ . Hence, we should project the solution  $h^*$  onto the  $L^2$ -orthogonal space  $\text{span}\{\sqrt{f_Y}\}^\perp$ :

$$\begin{aligned} h^{**} &:= h^* - \frac{\langle h^*, \sqrt{f_Y} \rangle}{\|\sqrt{f_Y}\|_{L^2}^2} \sqrt{f_Y} \\ &= \left( \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta - \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta) f_Y \right) \sqrt{f_Y} \\ &= \left( \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta - \int \zeta f_X \right) \sqrt{f_Y}, \end{aligned} \quad (37)$$

where we used  $\int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta) f_Y = \int \zeta f_X$  by (12). This leads to the candidate for the maximization of (10) given by

$$\begin{aligned} g^* &= S^{-1}h^{**} = S^{-1}(S^{-1})^* \zeta - \langle \zeta, f_X \rangle S^{-1} \sqrt{f_Y} = \text{I}^{-1} \zeta - \langle \zeta, f_X \rangle f_X \\ &= \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * \left\{ \left( \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta \right) f_Y \right\} - \left( \int \zeta f_X \right) f_X \end{aligned}$$

and (12) yields  $\langle g^*, \zeta \rangle = \langle \mathbb{I}g^*, g^* \rangle$  and the bound

$$\mathcal{I}_\zeta = \frac{\langle g^*, \zeta \rangle^2}{\langle \mathbb{I}g^*, g^* \rangle} = \int (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta)^2 f_Y - \left( \int \zeta f_X \right)^2. \quad (38)$$

Inequality (11) holds then by the local version of the Hájek–Le Cam convolution theorem (Bickel et al., 1998, Thm. 2.3.1). It remains to check the conditions in (8),  $(\sqrt{f_Y}h^{**})^{(k)} \in L^1(\mathbb{R})$  for  $k = 0, \dots, \bar{\beta}$  and that the three-fold application of the adjoint equality is allowed. The latter will follow from  $\sqrt{f_Y}h^{**}, f_Y \in H^{\beta^+}(\mathbb{R})$  for some  $\beta^+ > \beta$ .

**Step 2:** We prove now the integrability of  $\sqrt{f_Y}h^{**} = (\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta - \int \zeta f_X) f_Y$  and its derivatives up to order  $\bar{\beta}$  which makes the calculation (36) rigorous.

For convenience we denote

$$\kappa := \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}(-\bullet)] * \zeta = r * \left( \sum_{k=0}^{\bar{\beta}} \binom{\bar{\beta}}{k} (-1)^k \zeta^{(k)} \right).$$

Owing to Young's inequality together with  $r \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$  and  $\zeta^{(k)} \in L^2(\mathbb{R})$  for any  $k \geq 0$ , we obtain  $\kappa \in C^s(\mathbb{R}) \cap H^s(\mathbb{R})$  for any  $s \geq 0$ . It suffices to show  $f_Y^{(k)} \in L^1(\mathbb{R})$  for  $k = 0, \dots, \bar{\beta}$ . Note that by (34)

$$\|(\text{Id} + \text{D})^k f_\varepsilon\|_{L^1} \lesssim \|\mathcal{F}^{-1}[(1 - iu)^k \varphi_\varepsilon]\|_{B_{1,1}^0} \lesssim \|\mathcal{F}^{-1}[(1 - iu)^{k-\beta}]\|_{B_{1,1}^0}$$

is finite for  $\beta > k$  since then  $\mathcal{F}^{-1}[(1 - iu)^{k-\beta}] = \gamma_{\beta-k,1} \in B_{1,\infty}^{\beta-k}(\mathbb{R}) \subseteq B_{1,1}^0(\mathbb{R})$  by the proof of Lemma 5(ii). Recalling that  $\beta \notin \mathbb{Z}$ , we conclude iteratively  $f_\varepsilon^{(k)} \in L^1(\mathbb{R})$  for  $k = 0, \dots, \lfloor \beta \rfloor$ . Therefore,

$$\|f_Y^{(\bar{\beta})}\|_{L^1} \leq \|f_X^{(\bar{\beta}-\lfloor \beta \rfloor)}\|_{L^1} \|f_\varepsilon^{(\lfloor \beta \rfloor)}\|_{L^1} < \infty$$

by Assumption 3 and similarly for derivatives of lower order.

Moreover, we conclude for  $\bar{\beta}^- \in (\beta + 1/2, \bar{\beta})$  that

$$f_Y \in B_{1,1}^{\bar{\beta}^-}(\mathbb{R}) \subseteq B_{2,1}^{\bar{\beta}^- - 1/2}(\mathbb{R}) \subseteq H^{\beta^+}(\mathbb{R})$$

for some  $\beta^+ > \beta$  by the embeddings (46) and (46). Since also  $\kappa f_Y \in H^{\beta^+}(\mathbb{R})$ , using  $\kappa \in C^s(\mathbb{R})$  for  $s > \beta$ , we can apply the adjoint equality (12) in Step 1.

**Step 3:** We will show now  $\|g^*/f_X\|_\infty < \infty$  which justifies  $f_X \pm \tau g^* \geq 0$  for some choice of  $\tau > 0$  small enough.

By Step 1  $g^* = \mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * (\kappa f_Y) - \langle \zeta, f_X \rangle f_X$ . For the second term Assumption 3 implies  $\|\langle \zeta, f_X \rangle f_X / f_X\|_\infty \leq \|\zeta\|_{L^2} \|f_X\|_\infty \|f_X\|_{L^1} < \infty$ . Hence, we only need to show  $\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * (\kappa f_Y) \lesssim f_X$ . Using the Besov embedding (45), the Fourier multiplier property of (34) and the pointwise multiplier property of Besov spaces (48), we obtain for some  $\beta^+ \in (\beta, \lfloor \beta \rfloor + 1)$

$$\|\mathcal{F}^{-1}[\varphi_\varepsilon^{-1}] * (\kappa f_Y)\|_\infty \lesssim \|\kappa f_Y\|_{B_{\infty,1}^\beta} \lesssim \|\kappa\|_{B_{\infty,1}^\beta} \|f_Y\|_{B_{\infty,1}^\beta} \lesssim \|\kappa\|_{C^s} \|f_Y\|_{C^{\beta^+}}.$$

for any  $s > \beta$ . In Step 2 we have seen that  $\kappa \in C^s(\mathbb{R})$ . Moreover,

$$\begin{aligned} \|f_Y\|_{C^{\beta+}} &= \sum_{k=0}^{\lfloor \beta \rfloor} \|f_Y^{(k)}\|_{\infty} + \sup_{x \neq y} \left| \frac{f_Y^{(\lfloor \beta \rfloor)}(x) - f_Y^{(\lfloor \beta \rfloor)}(y)}{(x-y)^{\beta+ - \lfloor \beta \rfloor}} \right| \\ &\leq \sum_{k=0}^{\lfloor \beta \rfloor} \|f_X\|_{\infty} \|f_{\varepsilon}^{(k)}\|_{L^1} + \sup_{x \neq y} \int \frac{|f_X(x-z) - f_X(y-z)|}{|x-y|^{\beta+ - \lfloor \beta \rfloor}} f_{\varepsilon}^{(\lfloor \beta \rfloor)}(z) dz \\ &\leq \|f_X\|_{\infty} \sum_{k=0}^{\lfloor \beta \rfloor} \|f_{\varepsilon}^{(k)}\|_{L^1} + \|f_X\|_{C^{\beta+ - \lfloor \beta \rfloor}} \|f_{\varepsilon}^{(\lfloor \beta \rfloor)}\|_{L^1} < \infty, \end{aligned} \quad (39)$$

using the Besov embedding  $f_X \in W_1^2(\mathbb{R}) \subseteq B_{1,1}^{\beta+ - \lfloor \beta \rfloor + 1} \subseteq C^{\beta+ - \lfloor \beta \rfloor}$ . Hence,  $g^* \in L^{\infty}(\mathbb{R})$ . Since  $f_X$  is a continuous, strictly positive function, we conclude that the quotient  $g^*/f_X$  is bounded on every compact subset of  $\mathbb{R}$ . Therefore, it suffices to estimate the tails. For  $|x|$  large enough Assumption 3 implies, using again (34),

$$\begin{aligned} \frac{|\mathcal{F}^{-1}[\varphi_{\varepsilon}^{-1}] * (\kappa f_Y)(x)|}{f_X(x)} &\lesssim |x^M (\mathcal{F}^{-1}[\varphi_{\varepsilon}^{-1}] * (\kappa f_Y))(x)| \\ &\leq \sum_{k=0}^M \binom{M}{k} \left| \mathcal{F}^{-1} [(\varphi_{\varepsilon}^{-1})^{(k)} \mathcal{F}[y^{M-k} \kappa f_Y]](x) \right| \\ &\lesssim \sum_{k=0}^M \|y^{M-k} \kappa f_Y\|_{B_{\infty,1}^{\beta+ - k}}. \end{aligned}$$

Note that the above calculation shows that  $\varphi_{\varepsilon}^{-1}$  is a Fourier multiplier on the weighted Besov space with weight function  $\langle x \rangle^M$  (cf. Edmunds and Triebel, 1996, Def. 4.2.1/2 and Thm. 5.4.2). Each term in the above sum can be estimated by

$$\begin{aligned} &\|y^{M-k} \kappa\|_{C^s} \|f_Y\|_{C^{\beta+}} \\ &= \left\| \sum_{l=0}^{M-k} \binom{M-k}{l} (-1)^l \mathcal{F}^{-1} [(\varphi_{\varepsilon}^{-1})^{(l)}(-u) \mathcal{F}[(ix)^{M-k-l} \zeta]] \right\|_{C^s} \|f_Y\|_{C^{\beta+}}, \end{aligned}$$

where with abuse of notation  $\beta^+ < \lfloor \beta \rfloor + 1$  is slightly larger in the last line and  $s > \beta^+$ . By (39) we have  $f_Y \in C^{\beta^+}(\mathbb{R})$ . Now,  $(ix)^{M-k-l} \zeta \in \mathcal{S}(\mathbb{R})$  is again a Schwartz function and thus it suffices to show  $\mathcal{F}^{-1}[(\varphi_{\varepsilon}^{-1})^{(k)}(-u) \mathcal{F} \chi] \in C^s(\mathbb{R})$  for  $s > \beta$ ,  $\chi \in \mathcal{S}(\mathbb{R})$  and  $k = 0, \dots, M$ . For  $k = 0$  this is already done in Step 2. We proceed analogously: for any integer  $s \geq 0$  we have

$$\begin{aligned} &\|\mathcal{F}^{-1} [(\varphi_{\varepsilon}^{-1})^{(k)}(-u) \mathcal{F} \chi]^{(s)}\|_{\infty} \\ &= \left\| \left( \mathcal{F}^{-1} [(1+iu)^{(-\bar{\beta}+k) \wedge 0} (\varphi_{\varepsilon}^{-1})^{(k)}(-u)] * ((\text{Id} - D)^{(\bar{\beta}-k) \vee 0} \chi) \right)^{(s)} \right\|_{\infty} \\ &\leq \|(1+iu)^{(-\bar{\beta}+k) \wedge 0} (\varphi_{\varepsilon}^{-1})^{(k)}(-u)\|_{L^2} \|D^s (\text{Id} - D)^{(\bar{\beta}-k) \vee 0} \chi\|_{L^2} \\ &\lesssim \|\langle u \rangle^{((- \bar{\beta}+k) \wedge 0) + \beta - k}\|_{L^2} \|D^s (\text{Id} - D)^{(\bar{\beta}-k) \vee 0} \chi\|_{L^2}. \end{aligned}$$

Owing to  $\bar{\beta} > \beta + 1/2$ , the first factor is finite since

$$((-\bar{\beta} + k) \wedge 0) + \beta - k \leq \begin{cases} \beta - \bar{\beta} < -1/2, & \text{for } \bar{\beta} \geq k, \\ \bar{\beta} - 1/2 - k < -1/2, & \text{for } \bar{\beta} < k \end{cases}$$

and the second factor is the  $L^2$ -norm of a Schwartz function and thus finite, too.

### 5.2. Approximation lemma

To prove convergence of the information bounds it suffices to show that

$$\begin{aligned} \langle g_n^*, \zeta \rangle &\rightarrow \langle (S^*)^{-1} \zeta, (S^*)^{-1} \zeta \rangle - \langle \zeta, f_X \rangle^2 \quad \text{and} \quad (40) \\ \langle Sg_n^*, Sg_n^* \rangle &= \langle g_n^*, \zeta_n \rangle \rightarrow \langle (S^*)^{-1} \zeta, (S^*)^{-1} \zeta \rangle - \langle \zeta, f_X \rangle^2 \quad (41) \end{aligned}$$

where we used the equality  $\langle g_n^*, \zeta_n \rangle = \langle I g_n^*, g_n^* \rangle = \langle Sg_n^*, Sg_n^* \rangle$ , which holds naturally for the maximizer of the information bound  $\mathcal{I}_{\zeta_n}$ . For (40) we note

$$\langle g_n^*, \zeta \rangle = \langle I^{-1} \zeta_n, \zeta \rangle - \langle \zeta_n, f_X \rangle \langle \zeta, f_X \rangle = \langle (S^*)^{-1} \zeta_n, (S^*)^{-1} \zeta \rangle - \langle \zeta_n, f_X \rangle \langle \zeta, f_X \rangle$$

where the Cauchy–Schwarz inequality yields

$$|\langle f_X, \zeta_n - \zeta \rangle| = |\langle (S^*)^{-1}(\zeta_n - \zeta), \sqrt{f_Y} \rangle| \leq \| (S^*)^{-1}(\zeta_n - \zeta) \|_{L^2} \rightarrow 0 \quad (42)$$

and

$$\begin{aligned} |\langle (S^*)^{-1}(\zeta_n - \zeta), (S^*)^{-1} \zeta \rangle| &= |\langle (S^*)^{-1}(\zeta_n - \zeta), (S^*)^{-1} \zeta \rangle| \\ &\leq \| (S^*)^{-1}(\zeta_n - \zeta) \|_{L^2} \| (S^*)^{-1} \zeta \|_{L^2} \rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ . Analogously follows (41), where we use that the assumption of the lemma implies  $\langle (S^*)^{-1} \zeta_n, (S^*)^{-1} \zeta_n \rangle \rightarrow \langle (S^*)^{-1} \zeta, (S^*)^{-1} \zeta \rangle$  as  $n \rightarrow \infty$ . The second part of the claim  $\vartheta_{\zeta_n} \rightarrow \vartheta_{\zeta}$  has already been shown in the estimate (42).

### 5.3. Information bound for non-regular functionals

To prove the efficiency of  $\hat{\vartheta}_t$  for  $t \in \mathbb{R}$  in Theorem 4, it suffices by Lemma 3 and (35) to show

$$\langle (S^*)^{-1}(\zeta_n - \zeta), (S^*)^{-1}(\zeta_n - \zeta) \rangle^{1/2} = \| \mathcal{F}[\varphi_\varepsilon^{-1}(-\bullet)] * (\zeta_n - \zeta) \|_{L^2(\mathbb{P})} \rightarrow 0 \quad (43)$$

as  $n \rightarrow \infty$ . Using the moment bound (23) replacing  $\mathcal{F}K_h$  by 1, we obtain

$$\| \mathcal{F}[\varphi_\varepsilon^{-1}(-\bullet)] * (\zeta_n - \zeta) \|_{L^2(\mathbb{P})} \lesssim \| \zeta_n - \zeta \|_{Z^{\beta+\delta, 1/2+\beta+\delta}}.$$

By assumption we have  $Z^{\beta+\delta, 1/2+\beta+\delta} \subseteq Z^{\gamma_s, \gamma_c}$  for  $\delta$  small enough. Because the space of Schwartz-functions is dense in every Sobolev space  $H^s(\mathbb{R})$ ,  $s \geq 0$ ,  $\mathcal{S}(\mathbb{R})$  is also dense in  $Z^{\gamma_s, \gamma_c}$  and thus the information bound (11) holds for all  $\zeta \in Z^{\gamma_s, \gamma_c}$ . Finally, applying Theorem 25.48 by van der Vaart (1998) and Theorem 7 from above completes the proof of Theorem 4.

## Appendix: Function spaces

Let us define the  $L^p$ -Sobolev space for  $p \in (0, \infty)$  and  $m \in \mathbb{N}$

$$W_p^m(\mathbb{R}) := \left\{ f \in L^p(\mathbb{R}) \mid \sum_{k=0}^m \|f_X^{(k)}\|_{L^p} < \infty \right\}$$

In particular,  $W_p^0(\mathbb{R}) = L^p(\mathbb{R})$ . Due to the Hilbert space structure, the case  $p = 2$  is crucial. It can be described equivalently with the notation  $\langle u \rangle = (1 + u^2)^{1/2}$  by,  $\alpha \geq 0$ ,

$$H^\alpha(\mathbb{R}) := \left\{ f \in L^2(\mathbb{R}) \mid \|f\|_{H^\alpha}^2 := \int \langle u \rangle^{2\alpha} |\mathcal{F}f(u)|^2 du < \infty \right\},$$

which we call Sobolev space, too. Obviously,  $W_2^m(\mathbb{R}) = H^m(\mathbb{R})$ . Also frequently used are the Hölder spaces. Denoting the space of all bounded, continuous functions with values in  $\mathbb{R}$  as  $C(\mathbb{R})$  we define,  $\alpha \geq 0$ ,

$$C^\alpha(\mathbb{R}) := \left\{ f \in C(\mathbb{R}) \mid \|f\|_{C^\alpha} := \sum_{k=0}^{[\alpha]} \|f^{(k)}\|_\infty + \sup_{x \neq y} \frac{|f^{([\alpha])}(x) - f^{([\alpha])}(y)|}{|x - y|^{\alpha - [\alpha]}} < \infty \right\}$$

where  $[\alpha]$  denotes the largest integer smaller or equal to  $\alpha$ . A unifying approach which contains all function spaces defined so far, is given by Besov spaces (Triebel, 2010, Sect. 2.3.1) which we will discuss in the sequel. Let  $\mathcal{S}(\mathbb{R})$  be the Schwartz space of all rapidly decreasing infinitely differentiable functions with values in  $\mathbb{C}$  and  $\mathcal{S}'(\mathbb{R})$  its dual space, that is the space of all tempered distributions. Let  $0 < \psi \in \mathcal{S}(\mathbb{R})$  with  $\text{supp } \psi \subseteq \{x | 1/2 \leq |x| \leq 2\}$  and  $\psi(x) > 0$  if  $\{x | 1/2 < |x| < 2\}$ . Then define  $\varphi_j(x) := \psi(2^{-j}x)(\sum_{k=-\infty}^{\infty} \psi(2^{-k}x))^{-1}$ ,  $j = 1, 2, \dots$ , and  $\varphi_0(x) := 1 - \sum_{j=1}^{\infty} \varphi_j(x)$  such that the sequence  $\{\varphi_j\}_{j=0}^{\infty}$  is a smooth resolution of unity. In particular,  $\mathcal{F}^{-1}[\varphi_j \mathcal{F}f]$  is an entire function for all  $f \in \mathcal{S}'(\mathbb{R})$ . For  $s \in \mathbb{R}$  and  $p, q \in (0, \infty]$  the Besov spaces are defined by

$$B_{p,q}^s := \left\{ f \in \mathcal{S}'(\mathbb{R}) \mid \|f\|_{B_{p,q}^s} := \left( \sum_{j=0}^{\infty} 2^{sjq} \|\mathcal{F}^{-1}[\varphi_j \mathcal{F}f]\|_{L^p}^q \right)^{1/q} < \infty \right\}.$$

We omit the dependence of  $\|\bullet\|_{B_{p,q}^s}$  on  $\psi$  since any function with the above properties defines an equivalent norm. Setting the Besov spaces in relation to the more elementary function spaces, we first note that the Schwartz functions  $\mathcal{S}(\mathbb{R})$  are dense in every Besov space  $B_{p,q}^s$  with  $p, q < \infty$  and  $H^\alpha(\mathbb{R}) = B_{2,2}^\alpha(\mathbb{R})$  as well as  $C^\alpha(\mathbb{R}) = B_{\infty,\infty}^\alpha(\mathbb{R})$ , where the latter holds only if  $\alpha$  is not an integer (Triebel, 2010, Thms. 2.3.3 and 2.5.7). Frequently used are the following continuous embeddings, which can be found in (Triebel, 2010, Sect. 2.5.7, Thms. 2.3.2(1), 2.7.1): For  $p \geq 1, m \in \mathbb{Z}$ ,

$$B_{p,1}^m(\mathbb{R}) \subseteq W_p^m(\mathbb{R}) \subseteq B_{p,\infty}^m(\mathbb{R}) \quad \text{and} \quad B_{\infty,1}^0(\mathbb{R}) \subseteq L^\infty(\mathbb{R}) \subseteq B_{\infty,\infty}^0(\mathbb{R}) \quad (44)$$

and for  $s \geq 0$

$$B_{\infty,1}^s(\mathbb{R}) \subseteq C^s(\mathbb{R}) \subseteq B_{\infty,\infty}^s(\mathbb{R}). \quad (45)$$

Furthermore, for  $0 < p_0 \leq p_1 \leq \infty, q \geq 0$  and  $-\infty < s_1 \leq s_0 < \infty$

$$B_{p_0,q}^{s_0}(\mathbb{R}) \subseteq B_{p_1,q}^{s_1}(\mathbb{R}) \quad \text{if} \quad s_0 - \frac{1}{p_0} \geq s_1 - \frac{1}{p_1} \quad (46)$$

and for  $0 < p, q_0, q_1 \leq \infty$  and  $-\infty < s_1 < s_0 < \infty$

$$B_{p,q_0}^{s_0}(\mathbb{R}) \subseteq B_{p,q_1}^{s_1}(\mathbb{R}). \quad (47)$$

Another important relation is the pointwise multiplier property of Besov spaces (Triebel, 2010, (24) on p. 143) that is

$$\|fg\|_{B_{p,q}^s} \lesssim \|f\|_{B_{\infty,q}^s} \|g\|_{B_{p,q}^s} \quad (48)$$

for  $s > 0, 1 \leq p \leq \infty$  and  $0 < q \leq \infty$ .

The Besov norm of a convolution can be bounded by Lemma 7 (i) by Qui (1981). Let  $1 \leq p, q, r, s \leq \infty, -\infty < \alpha, \beta < \infty, 0 \leq 1/u = 1/p + 1/r - 1 \leq 1, 0 \leq 1/v = 1/q + 1/s \leq 1$ . For  $f \in B_{p,q}^\alpha(\mathbb{R})$  and  $g \in B_{r,s}^\beta(\mathbb{R})$

$$\|f * g\|_{B_{u,v}^{\alpha+\beta}} \lesssim \|f\|_{B_{p,q}^\alpha} \|g\|_{B_{r,s}^\beta}. \quad (49)$$

Using for any function  $f : \mathbb{R} \rightarrow \mathbb{R}$  and  $h \in \mathbb{R}$  the difference operators  $\Delta_h^1 f(x) := f(x+h) - f(x)$  and  $(\Delta_h^l f)(x) := \Delta_h^1(\Delta_h^{l-1} f)(x), l \in \mathbb{N}$ , the Besov can be equivalently described by

$$\|f\|_{B_{pq}^s} \sim \|f\|_{L^p} + \|f\|_{\dot{B}_{pq}^s} \quad \text{with} \quad \|f\|_{\dot{B}_{pq}^s} := \left( \int |h|^{-sq-1} \|\Delta_h^M f\|_{L^p}^q dh \right)^{1/q}$$

for  $s > 0, p, q \geq 1$  and any integer  $M > s$  (Triebel, 2010, Thm. 2.5.12). The space of all  $f \in \mathcal{S}'(\mathbb{R})$  for which  $\|f\|_{\dot{B}_{pq}^s}$  is finite is called homogeneous Besov space  $\dot{B}_{pq}^s(\mathbb{R})$  (Triebel, 2010, Def. 5.1.3/2, Thm. 2.2.3/2) and thus  $B_{pq}^s = L^p(\mathbb{R}) \cap \dot{B}_{pq}^s(\mathbb{R})$  for  $s > 0, p, q \geq 1$ . Of interest is the relation of homogeneous Besov spaces to functions of bounded  $p$ -variation. Let  $\mathcal{BV}_p(\mathbb{R})$  denote the space of measurable functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  such that there is a function  $g$  which coincides with  $f$  almost everywhere and satisfies

$$\sup \left\{ \sum_{i=1}^n |g(x_i) - g(x_{i-1})|^p \mid -\infty < x_1 < \dots < x_n < \infty, n \in \mathbb{N} \right\} < \infty$$

and we define  $BV_p(\mathbb{R})$  as the quotient set  $\mathcal{BV}_p(\mathbb{R})$  modulo equality almost everywhere. Then,

$$\dot{B}_{p1}^{1/p}(\mathbb{R}) \subseteq BV_p(\mathbb{R}) \subseteq \dot{B}_{p,\infty}^{1/p}(\mathbb{R}), \quad \text{for } p > 1 \quad (50)$$

by (Bourdaud, Lanza de Cristoforis and Sickel, 2006, Thm. 5). For  $p = 1$  holds by (Giné and Nickl, 2008, Lem. 8)

$$BV_1(\mathbb{R}) \cap L^1(\mathbb{R}) \subseteq B_{1,\infty}^1(\mathbb{R}). \quad (51)$$

## References

- BICKEL, P. J. and RITOV, Y. (2003). Nonparametric estimators which can be “plugged-in”. *Ann. Statist.* **31** 1033–1053. [MR2001641](#)
- BICKEL, P. J., KLAASSEN, C. A. J., RITOV, Y. and WELLNER, J. A. (1998). *Efficient and Adaptive Estimation for Semiparametric Models*. Springer, New York. [MR1623559](#)
- BISSANTZ, N., DÜMBGEN, L., HOLZMANN, H. and MUNK, A. (2007). Nonparametric confidence bands in deconvolution density estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **69** 483–506. [MR2323764](#)
- BOURDAUD, G., LANZA DE CRISTOFORIS, M. and SICKEL, W. (2006). Superposition operators and functions of bounded  $p$ -variation. *Rev. Mat. Iberoam.* **22** 455–487. [MR2294787](#)
- BUTUCEA, C. and COMTE, F. (2009). Adaptive estimation of linear functionals in the convolution model and applications. *Bernoulli* **15** 69–98. [MR2546799](#)
- DATTNER, I., GOLDENSHLUGER, A. and JUDITSKY, A. (2011). On deconvolution of distribution functions. *The Annals of Statistics* **39** 2477–2501. [MR2906875](#)
- DE LA PEÑA, V. H. and GINÉ, E. (1999). *Decoupling: From Dependence to Independence*. Springer, New York. [MR1666908](#)
- DUDLEY, R. M. (1992). Fréchet differentiability,  $p$ -variation and uniform Donsker classes. *Ann. Probab.* **20** 1968–1982. [MR1188050](#)
- DUDLEY, R. M. (1999). *Uniform Central Limit Theorems*. Cambridge University Press, Cambridge. [MR1720712](#)
- EDMUNDS, D. E. and TRIEBEL, H. (1996). *Function Spaces, Entropy Numbers, Differential Operators*. Cambridge University Press, Cambridge.
- FAN, J. (1991a). Asymptotic Normality for Deconvolution Kernel Density Estimators. *Sankhyā: The Indian Journal of Statistics* **53** 97–110. [MR1177770](#)
- FAN, J. (1991b). On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *The Annals of Statistics* **19** 1257–1272. [MR1126324](#)
- FAN, Y. and LIU, Y. (1997). A note on asymptotic normality for deconvolution kernel density estimators. *Sankhyā Ser. A* **59** 138–141. [MR1665140](#)
- FAN, J. and TRUONG, Y. K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900–1925. [MR1245773](#)
- GINÉ, E. and NICKL, R. (2008). Uniform central limit theorems for kernel density estimators. *Probab. Theory Related Fields* **141** 333–387. [MR2391158](#)
- GINÉ, E. and NICKL, R. (2009). Uniform limit theorems for wavelet density estimators. *Ann. Probab.* **37** 1605–1646. [MR2546757](#)
- GIRARDI, M. and WEIS, L. (2003). Operator-valued Fourier multiplier theorems on Besov spaces. *Mathematische Nachrichten* **251** 34–51. [MR1960803](#)
- GOLDENSHLUGER, A. and PEREVERZEV, S. V. (2003). On adaptive inverse estimation of linear functionals in Hilbert scales. *Bernoulli* **9** 783–807. [MR2047686](#)
- HALL, P. and LAHIRI, S. N. (2008). Estimation of distributions, moments and quantiles in deconvolution problems. *Ann. Statist.* **36** 2110–2134. [MR2458181](#)
- KLENKE, A. (2007). *Probability Theory: A Comprehensive Course (University text)*, 1st ed. Springer, London. [MR2372119](#)



- KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*, 1st ed. *Springer Series in Statistics*. Springer, New York. [MR2724368](#)
- LOUNICI, K. and NICKL, R. (2011). Global uniform risk bounds for wavelet deconvolution estimators. *Ann. Statist.* **39** 201–231. [MR2797844](#)
- NEUMANN, M. H. (1997). On the effect of estimating the error density in non-parametric deconvolution. *J. Nonparametr. Statist.* **7** 307–330. [MR1460203](#)
- NICKL, R. (2006). Empirical and Gaussian processes on Besov classes. In *High dimensional probability. IMS Lecture Notes Monogr. Ser.* **51** 185–195. [MR2387769](#)
- NICKL, R. and PÖTSCHER, B. M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type. *J. Theoret. Probab.* **20** 177–199. [MR2324525](#)
- NICKL, R. and REISS, M. (2012). A Donsker theorem for Lévy measures. *J. Funct. Anal.* **263** 3306–3332. [MR2973342](#)
- QUI, B. H. (1981). Bernstein’s theorem and translation invariant operators. *Hiroshima Math. J.* **11** 81–96. [MR0606835](#)
- RADULOVIĆ, D. and WEGKAMP, M. (2000). Weak convergence of smoothed empirical processes: beyond Donsker classes. In *High dimensional probability, II. Progr. Probab.* **47** 89–105. Birkhäuser, Boston. [MR1857317](#)
- SCHMIDT-HIEBER, J., MUNK, A. and DÜMBGEN, L. (2012). Multiscale Methods for Shape Constraints in Deconvolution: Confidence Statements for Qualitative Features. arXiv 1107.1404.
- TRIEBEL, H. (2010). *Theory of Function Spaces*. Birkhäuser Verlag, Basel. Reprint of the 1983 Edition. [MR0781540](#)
- VAN DER VAART, A. W. (1998). *Asymptotic statistics. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.
- VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak convergence and empirical processes. Springer Series in Statistics*. Springer, New York. [MR1385671](#)
- VAN ES, B. and UH, H.-W. (2005). Asymptotic normality of kernel-type deconvolution estimators. *Scand. J. Statist.* **32** 467–483. [MR2204630](#)
- VAN ROOIJ, A. C. M., RUYMGAART, F. H. and VAN ZWET, W. R. (1999). Asymptotic efficiency of inverse estimators. *Teor. Veroyatnost. i Primenen.* **44** 826–844. Translation in *Theory Probab. Appl.* 44(4), 722–738. [MR1811134](#)