Confidence bands for Horvitz-Thompson estimators using sampled noisy functional data

Hervé Cardot[a], David Degras[b], Etienne Josserand[a]
(a) Institut de Mathématiques de Bourgogne, UMR 5584, Université de Bourgogne,
9 Avenue Alain Savary, 21078 Dijon, France.
email: {herve.cardot,etienne.josserand}@u-bourgogne.fr
(b) Statistical and Applied Mathematical Sciences Institute,
19 T.W. Alexander Drive, P.O. Box 14006, Research Triangle Park, NC 27709, USA.
email: ddegras@samsi.info

May 12, 2011

Abstract

When collections of functional data are too large to be exhaustively observed, survey sampling techniques provide an effective way to estimate global quantities such as the population mean function. Assuming functional data are collected from a finite population according to a probabilistic sampling scheme, with the measurements being discrete in time and noisy, we propose to first smooth the sampled trajectories with local polynomials and then estimate the mean function with a Horvitz-Thompson estimator. Under mild conditions on the population size, observation times, regularity of the trajectories, sampling scheme, and smoothing bandwidth, we prove a Central Limit Theorem in the space of continuous functions. We also establish the uniform consistency of a covariance function estimator and apply the former results to build global confidence bands for the mean function. The bands attain nominal coverage and are obtained through Gaussian process simulations conditional on the estimated covariance function. To select the bandwidth, we propose a cross-validation method that accounts for the sampling weights. A simulation study assesses the performance of our approach and highlights the influence of the sampling scheme and bandwidth choice.

Keywords : CLT, functional data, local polynomial smoothing, maximal inequalities, space of continuous functions, suprema of Gaussian processes, survey sampling, weighted cross-validation.

1 Introduction

The recent development of automated sensors has given access to very large collections of signals sampled at fine time scales. However, exhaustive transmission, storage, and analysis of such massive functional data may incur very large investments. In this context, when the goal is to assess a global indicator like the mean temporal signal, survey sampling techniques are appealing solutions as they offer a good trade-off between statistical accuracy and global cost of the analysis. In particular they are competitive with signal compression techniques (Chiky and Hébrail, 2008). The previous facts provide some explanation why, although survey sampling and functional data analysis have been long-established statistical fields, motivation for studying them jointly only recently emerged in the literature. In this regard Cardot *et al.* (2010a) examine the theoretical properties of functional principal components analysis (FPCA) in the survey sampling framework. Cardot *et al.* (2010b) harness FPCA for model-assisted estimation by relating the unobserved principal component scores to available auxiliary information. Focusing on sampling schemes, Cardot and Josserand (2011) estimate the mean electricity consumption curve in a population of about 19,000 customers whose electricity meters were read every 30 minutes during one week. Assuming exact measurements, they first perform a linear interpolation of the discretized signals and then consider a functional version of the Horvitz-Thompson estimator. For a fixed sample size, they show that estimation can be greatly improved by utilizing stratified sampling over simple random sampling and they extend the Neyman optimal allocation rule (see *e.g.* Fuller (2009)) to the functional setup. Note however that the finite-sample and asymptotic properties of their estimator rely heavily on the assumption of error-free measurements, which is not always realistic in practice.

The first contribution of the present work is to generalize the framework of Cardot and Josserand (2011) to noisy functional data. Assuming data are observed with errors that may be correlated over time, we replace the interpolation step in their procedure by a data smoothing step based on local polynomials. We extend the previous asymptotic theory by establishing a functional CLT for the resulting mean function estimator and proving the uniform consistency of a related covariance estimator.

In relation to mean function estimation, a key statistical task is to build confidence regions. There exists a vast and still active literature on confidence bands in nonparametric regression. See *e.g.* Sun and Loader (1994), Eubank and Speckman (1993), Claeskens and van Keilegom (2003), Krivobokova *et al.* (2010), and the references therein. When data are functional the literature is much less abundant. One possible approach is to obtain confidence balls for the mean function in a L^2-space. Mas (2007) exploits this idea in a goodness-of-fit test based on the functional sample mean and regularized inverse covariance operator. Using adaptive projection estimators, Bunea *et al.* (2011) build conservative confidence regions for the mean of a Gaussian process. Another approach consists in deriving results in a space C of continuous functions equipped with the supremum norm. This allows to build confidence bands which can be visualized and interpreted as opposed to L^2-confidence balls. It is adopted for example by Faraway (1997) to build bootstrap bands in a varying-coefficients model, by Cuevas *et al.* (2006) to derive various bootstrap bands for functional location parameters, by Degras (2009, 2010) to obtain normal and bootstrap bands using noisy functional data, and by Cardot and Josserand (2011) in the context of a finite population. In the latter work, the strategy was to first establish a CLT in the space C and then derive confidence bands based on a simple but rough approximation to the supremum of a Gaussian process (Landau and Shepp (1970)). Unfortunately, the associated bands depend on the data-generating process only through its variance structure and not its correlation structure, which may cause the empirical coverage to differ from the nominal level. The second innovation of our paper is to propose confidence bands that are easy to implement and attain nominal coverage in the survey sampling/finite population setting. To do so we use Gaussian process simulations as in Cuevas *et al.* (2006) or Degras (2010). Our contribution is to provide the theoretical underpinning of the construction method, thereby guaranteeing that nominal coverage is attained. The theory we derive involves random entropy numbers, maximal inequalities, and large covariance matrix theory.

Finally, the implementation of the mean function estimator developed in this paper requires to select a bandwidth in the data smoothing step. Objective, data-driven bandwidth selection methods are desirable for this purpose. As explained by Opsomer and Miller (2005), bandwidth selection in the survey estimation context poses specific problems (in particular, the necessity to take the sampling design into account) that make usual cross-validation or mean square error optimization methods inadequate. In view of the model-assisted survey estimation of a population total, these authors propose a cross-validation method that aims at minimizing the variance of the estimator, the bias component being negligible in their setting. In our functional and design-based framework, the bias is however no longer negligible. We therefore devise a novel cross-validation criterion based on weighted least squares, with weights proportional to the sampling weights. For the particular case of simple random sampling without replacement, this criterion reduces to the cross validation technique of Rice and Silverman (1991).

The paper is organized as follows. We fix notations and define our estimators in section 2. In section 3, we introduce our asymptotic framework based on superpopulation models (see Isaki and Fuller, 1982), establish a CLT for the estimator of the mean trajectory in the space of continuous functions, and show the uniform consistency of a covariance function estimator. After that, we prove that by simulating the limiting Gaussian process conditional on its estimated covariance, one can build confidence bands that have asymptotically correct coverage. Simulations are performed in section 4, where different sampling schemes and bandwidth choices are compared to assess the numerical performance of our methodology. The paper ends with a short discussion on topics for future research. Proofs are gathered in an Appendix.

2 Notations and estimators

Consider a finite population $U_N = \{1, \ldots, N\}$ of size N and suppose that to each unit $k \in U_N$ corresponds a real function X_k on $[0, T]$, with $T < \infty$. We assume that each trajectory X_k belongs to the space of continuous functions $C([0, T])$. Our target is the mean trajectory $\mu_N(t),\ t \in [0, T]$, defined as follows:

$$\mu_N(t) \quad = \quad \frac{1}{N} \sum_{k \in U} X_k(t). \tag{1}$$

We consider a random sample s drawn from U_N without replacement according to a fixed-size sampling design $p_N(s)$, where $p_N(s)$ is the probability of drawing the sample s. The size n_N of s is nonrandom and we suppose that the first and second order inclusion probabilities satisfy

- $\pi_k := \mathbb{P}(k \in s) > 0$ for all $k \in U_N$
- $\pi_{kl} := \mathbb{P}(k \& l \in s) > 0$ for all $k, l \in U_N$

so that each unit and each pair of units can be drawn with a non null probability from the population. Note that for simplicity of notation the subscript N has been omitted. Also, by convention, we write $\pi_{kk} = \pi_k$ for all $k \in U_N$.

Assume that noisy measurements of the sampled curves are available at $d = d_N$ fixed discretization points $0 = t_1 < t_2 < \ldots < t_d = T$. For all unit $k \in s$, we observe

$$Y_{jk} = X_k(t_j) + \epsilon_{jk} \tag{2}$$

where the measurement errors ϵ_{jk} are centered random variables that are independent across the index k (units) but not necessarily across j (possible temporal dependence). It is also assumed that the random sample s is independent of the noise ϵ_{jk} and the trajectories $X_k(t), t \in [0, T]$ are deterministic.

Our goal is to estimate μ_N as accurately as possible and to build asymptotic confidence bands, as in Degras (2010) and Cardot and Josserand (2011). For this, we must have a uniformly consistent estimator of its covariance function.

2.1 Linear smoothers and the Horvitz-Thompson estimator

For each (potentially observed) unit $k \in U_N$, we aim at recovering the curve X_k by smoothing the corresponding discretized trajectory $(Y_{1k}, \ldots, Y_{dk})$ with a linear smoother (e.g. spline, kernel, or local polynomial):

$$\widehat{X}_k(t) = \sum_{j=1}^{d} W_j(t) Y_{jk}. \tag{3}$$

Note that the reconstruction can only be performed for the observed units $k \in s$.

Here we use local linear smoothers (see *e.g.* Fan and Gijbels (1997)) because of their wide popularity, good statistical properties, and mathematical convenience. The weight functions $W_j(t)$ can be expressed as

$$W_j(t) = \frac{\frac{1}{dh}\left\{s_2(t) - (t_j - t)s_1(t)\right\} K\left(\frac{t_j - t}{h}\right)}{s_2(t)s_0(t) - s_1^2(t)}, \quad j = 1, \ldots, d, \tag{4}$$

where K is a kernel function, $h > 0$ is a bandwidth, and

$$s_l(x) = \frac{1}{dh}\sum_{j=1}^{d}(t_j - t)^l K\left(\frac{t_j - t}{h}\right), \quad l = 0, 1, 2. \tag{5}$$

We suppose that the kernel K is nonnegative, has compact support, satisfies $K(0) > 0$ and $|K(s) - K(t)| \leq C|s - t|$ for some finite constant C and for all $s, t \in [0, T]$.

The classical Horvitz-Thompson estimator (see *e.g.* Fuller (2009)) of the mean curve is

$$\begin{aligned} \widehat{\mu}_N(t) &= \frac{1}{N}\sum_{k \in s}\frac{\widehat{X}_k(t)}{\pi_k} \\ &= \frac{1}{N}\sum_{k \in U}\frac{\widehat{X}_k(t)}{\pi_k} I_k, \end{aligned} \tag{6}$$

where I_k is the sample membership indicator ($I_k = 1$ if $k \in s$ and $I_k = 0$ otherwise). It holds that $\mathbb{E}(I_k) = \pi_k$ and $\mathbb{E}(I_k I_l) = \pi_{kl}$.

2.2 Covariance estimation

The covariance function of $\widehat{\mu}_N$ can be written as

$$\mathrm{Cov}\left(\widehat{\mu}_N(s), \widehat{\mu}_N(t)\right) = \frac{1}{N}\gamma_N(s, t) \tag{7}$$

for all $s, t \in [0, T]$, where

$$\gamma_N(s, t) = \frac{1}{N}\sum_{k,l \in U}\Delta_{kl}\frac{\tilde{X}_k(s)}{\pi_k}\frac{\tilde{X}_l(t)}{\pi_l} + \frac{1}{N}\sum_{k \in U}\frac{1}{\pi_k}\mathbb{E}\left(\tilde{\epsilon}_k(s)\tilde{\epsilon}_k(t)\right) \tag{8}$$

with

$$\begin{cases} \tilde{X}_k(t) & = \sum_{j=1}^{d} W_j(t) X_k(t_j), \\ \tilde{\epsilon}_k(t) & = \sum_{j=1}^{d} W_j(t) \epsilon_{kj}, \\ \Delta_{kl} & = \mathrm{Cov}(I_k, I_l) = \pi_{kl} - \pi_k \pi_l. \end{cases} \tag{9}$$

A natural estimator of $\gamma_N(s,t)$ (see *e.g.* Fuller (2009)) is given by

$$\widehat{\gamma}_N(s,t) = \frac{1}{N} \sum_{k,l \in U} \frac{\Delta_{kl}}{\pi_{kl}} \left(\frac{I_k}{\pi_k} \frac{I_l}{\pi_l} \right) \widehat{X}_k(s) \widehat{X}_l(t). \tag{10}$$

It is unbiased and its uniform mean square consistency is established in Section 3.2.

3 Asymptotic theory

We consider the superpopulation framework introduced by Isaki and Fuller (1982) and discussed in detail by Fuller (2009). Specifically, we study the behaviour of the estimators $\widehat{\mu}_N$ and $\widehat{\gamma}_N$ as population $U_N = \{1, \ldots, N\}$ increases to infinity with N. Recall that the sample size n, inclusion probabilities π_k and π_{kl}, and grid size d all depend on N. In what follows we use the notations c and C for finite, positive constants whose value may vary from place to place. The following assumptions are needed for our asymptotic study.

(A1) (*Sampling design*) $\dfrac{n}{N} \geq c$, $\pi_k \geq c$, $\pi_{kl} \geq c$, and $n|\pi_{kl} - \pi_k \pi_l| \leq C$ for all $k, l \in U_N$ $(k \neq l)$ and $N \geq 1$.

(A2) (*Trajectories*) $|X_k(s) - X_k(t)| \leq C|s-t|^{\beta}$ and $|X_k(0)| \leq C$ for all $k \in U_N$, $N \geq 1$, and $s, t \in [0, T]$, where $\beta > \frac{1}{2}$ is a finite constant.

(A3) (*Growth rates*) $c \leq d(t_{j+1} - t_j) \leq C$ for all $1 \leq j \leq d$, $N \geq 1$, and $\dfrac{d(\log \log N)}{N} \to 0$ as $N \to \infty$.

(A4) (*Measurement errors*) The random vectors $(\epsilon_{k1}, \ldots, \epsilon_{kd})'$, $k \in U_N$, are i.i.d. and follow the multivariate normal distribution with mean zero and covariance matrix $\mathbf{V}_N$. The largest eigenvalue of the covariance matrix satisfies $\|\mathbf{V}_N\| \leq C$ for all $N \geq 1$.

Assumption (A1) deals with the properties of the sampling design. It states that the sample size must be at least a positive fraction of the population size, that the one- and two-fold inclusion probabilities must be larger than a positive number, and that the two-fold inclusion probabilities should not be to far from independence. The latter is fulfilled for example for stratified sampling with sampling without replacement within each stratum (Robinson and Särndal (1983)). Assumption (A2) imposes Hölder continuity on the trajectories, a mild regularity condition. Assumption (A3) states that the design points have a quasi-uniform repartition (this holds in particular for equidistant designs and designs generated by a regular density function) and that the grid size is essentially negligible compared to the population size (for example if $d_N \propto N^\alpha$ for some $\alpha \in (0,1)$). In fact the results of this paper also hold if d_N/N stays bounded away from zero and infinity as $N \to \infty$ (see Section 5). Finally (A4) imposes joint normality, short range temporal dependence, and bounded variance for the measurement errors $\epsilon_{kj}, 1 \leq j \leq d$. It is trivially satisfied if the $\epsilon_{kj} \sim N(0, \sigma_j^2)$ are independent with variances $\mathrm{Var}(\epsilon_{kj}) \leq C$. It is also verified if the ϵ_{kj} arise from a discrete time Gaussian process with short term temporal correlation such as ARMA or stationary mixing processes. Note that the Gaussian assumption is not central to our derivations: it can be weakened and replaced by moment conditions on the error distributions at the expense of much more complicated proofs.

3.1 Limit distribution of the Horvitz-Thompson estimator

Proceeding further, we would now like to derive the asymptotic distribution of our estimator $\widehat{\mu}_N$ in order to build asymptotic confidence intervals and bands. Obtaining the asymptotic normality of estimators in survey sampling is a technical and difficult issue even for simple quantities such as means or totals of real numbers. Although confidence intervals are commonly used in the survey sampling community, the Central Limit Theorem (CLT) has only been checked rigorously, as far as we know, for a few sampling designs. Erdös and Rényi (1959) and Hàjek (1960) proved that the Horvitz-Thompson estimator is asymptotically Gaussian for simple random sampling without replacement. These results were extended more recently to stratified sampling (Bickel and Freedman (1994)) and some particular cases of two-phase sampling designs (Chen and Rao (2007)). Let us assume that the Horvitz-Thompson estimator satisfies a CLT for real valued quantities.

(A5) (*Univariate CLT*) For any fixed $t \in [0, T]$, it holds that

$$\frac{\widehat{\mu}_N(t) - \mu_N(t)}{\sqrt{\mathrm{Var}\big(\widehat{\mu}_N(t)\big)}} \rightsquigarrow N(0,1)$$

as $N \to \infty$, where $\rightsquigarrow$ stands for convergence in distribution.

We recall here the definition of the weak convergence in $C([0,T])$ equipped with the supremum norm $\|\cdot\|_\infty$ (e.g. van der Vaart and Wellner (2000)). A sequence (ξ_N) of random elements of $C([0,T])$ is said to converge weakly to a limit ξ in $C([0,T])$ if $\mathbb{E}(\phi(\xi_N)) \to \mathbb{E}(\phi(\xi))$ as $N \to \infty$ for all bounded, uniformly continuous functional ϕ on $(C([0,T]), \|\cdot\|_\infty)$.

To establish the limit distribution of $\widehat{\mu}_N$ in $C([0,T])$, we need to assume the existence of a limit covariance function

$$\gamma(s,t) = \lim_{N\to\infty} \frac{1}{N} \sum_{k,l \in U_N} \Delta_{kl} \frac{X_k(s)}{\pi_k} \frac{X_l(t)}{\pi_l} \, .$$

In the following theorem we state the asymptotic normality of the estimator $\widehat{\mu}_N$ in the space $C([0,T])$ equipped with the sup norm.

Theorem 1. *Assume (A1)–(A5) and that* $\sqrt{N}h^\beta \to 0$ *and* $dh/\log d \to \infty$ *as* $N \to \infty$. *Then*

$$\sqrt{N}\,(\widehat{\mu}_N - \mu_N) \rightsquigarrow G$$

in $C([0,T])$*, where* G *is a Gaussian process with mean zero and covariance function* γ.

Theorem 1 provides a convenient way to infer the local features of μ_N. It is applied in Section 3.3 to the construction of simultaneous confidence bands, but it can also be used for a variety of statistical tests based on supremum norms (see Degras (2010)).

Observe that the conditions on the bandwidth h and design size d are not very constraining. Suppose for example that $d \propto N^\eta$ and $h \propto N^{-\nu}$ for some $\eta, \nu > 0$. Then d and h satisfy the conditions of Theorem 1 as soon as $(2\beta)^{-1} < \nu < \eta < 1$. Thus, for more regular trajectories, *i.e.* larger β, the bandwidth h can be chosen with more flexibility.

The proof of Theorem 1 is similar in spirit to that of Theorem 1 in Degras (2010) and Proposition 3 in Cardot and Josserand (2011). Essentially, it breaks down into: (i) controlling uniformly on $[0,T]$ the bias of $\widehat{\mu}_N$, (ii) establishing the functional asymptotic normality of the local linear smoother applied to the sampled curves X_k, and (iii) controlling uniformly on $[0,T]$ (in probability) the local linear smoother applied to the errors ϵ_{jk}. Part (i) is easily handled with standard results on approximation properties of local polynomial estimators (see *e.g.* Tsybakov (2009)). Part (ii) mainly consists in proving an asymptotic tightness property, which entails the computation of entropy numbers and the use of maximal inequalities (see van der Vaart and Wellner (2000)). Part (iii) requires first to show the finite-dimensional convergence of the smoothed error process to zero and then to establish its tightness with similar arguments as in part (ii).

3.2 Uniform consistency of the covariance estimator

We first note that under (A1)–(A4), by the approximation properties of local linear smoothers, γ_N converges uniformly to γ on $[0,T]^2$ as $h \to 0$ and $N \to \infty$. Hence the consistency of $\widehat{\gamma}_N$ can be stated with respect to γ instead of γ_N. In alignment with the related Proposition 2 in Cardot and Josserand (2010) and Theorem 3 in Breidt and Opsomer (2000), we need to make some assumption on the two-fold inclusion probabilities of the sampling design p_N:

(A6)

$$\lim_{N\to\infty} \max_{(k_1,k_2,k_3,k_4)\in D_{4,N}} \left|\mathbb{E}\{(I_{k_1}I_{k_2} - \pi_{k_1k_2})(I_{k_3}I_{k_4} - \pi_{k_3k_4})\}\right| = 0$$

where $D_{4,N}$ is the set of all quadruples (k_1,k_2,k_3,k_4) in U_N with distinct elements.

This assumption is discussed in detail in Breidt and Opsomer (2000) and is fulfilled for example for stratified sampling.

Theorem 2. *Assume (A1)–(A4), (A6), and that $h \to 0$ and $dh^{1+\alpha} \to \infty$ for some $\alpha > 0$ as $N \to \infty$. Then*

$$\lim_{N\to\infty} \mathbb{E}\left(\sup_{s,t\in[0,T]^2} \left|\widehat{\gamma}_N(s,t) - \gamma(s,t)\right|^2 \right) = 0.$$

Note the additional condition on the bandwidth h in Theorem 2. If we suppose, as in the remark in Section 3.1, that $d \propto N^{\eta}$ and $h \propto N^{-\nu}$ for some $(2\beta)^{-1} < \nu < \eta < 1$, then condition $dh^{1+\alpha} \to \infty$ as $N \to \infty$ is fulfilled with *e.g.* $\alpha = 1 - \eta/2\nu$.

3.3 Global confidence bands

In this section we build global confidence bands for μ_N of the form

$$\left\{ \left[\widehat{\mu}_N(t) \pm c\, \frac{\widehat{\sigma}_N(t)}{N^{1/2}} \right],\ t \in [0,T] \right\}, \tag{11}$$

where c is a suitable number and $\widehat{\sigma}_N(t) = \widehat{\gamma}_N(t,t)^{1/2}$. More precisely, given a confidence level $1-\alpha \in (0,1)$, we seek $c = c_\alpha$ that approximately satisfies

$$\mathbb{P}\left(|G(t)| \le c\,\sigma(t),\ \forall t \in [0,T]\right) = 1 - \alpha, \tag{12}$$

where G is a Gaussian process with mean zero and covariance function γ, and where $\sigma(t) = \gamma(t,t)^{1/2}$. Exact bounds for the supremum of Gaussian processes have only been derived for only a few particular cases (Adler and Taylor, 2007, Chapter 4). Computing accurate and as explicit as possible bounds in a general setting is a difficult issue and would require additional strong conditions such as stationarity which have no reason to be fulfilled in our setting.

In view of Theorems 1-2 and Slutski's Theorem, the bands defined in (11) with c chosen as in (12) will have approximate coverage level $1-\alpha$. The following result provides a simulation-based method to compute c.

Theorem 3. *Assume (A1)–(A6) and $dh^{1+\alpha} \to \infty$ for some $\alpha > 0$ as $N \to \infty$. Let G be a Gaussian process with mean zero and covariance function γ. Let $(\widehat{G}_N)$ be a sequence of processes such that for each N, conditionally on $\widehat{\gamma}_N$, $\widehat{G}_N$ is Gaussian with mean zero and covariance $\widehat{\gamma}_N$ defined in* (10)*. Then for all $c > 0$, as $N \to \infty$, the following convergence holds in probability:*

$$\mathbb{P}\left(|\widehat{G}_N(t)| \le c\,\widehat{\sigma}_N(t),\ \forall t \in [0,T] \,\middle|\, \widehat{\gamma}_N\right) \to \mathbb{P}\left(|G(t)| \le c\,\sigma(t),\ \forall t \in [0,T]\right).$$

Theorem 3 is derived by showing the weak convergence of $(\widehat{G}_N)$ to G in $C([0,T])$, which stems from Theorem 2 and the Gaussian nature of the processes $\widehat{G}_N$. As in the first two theorems, maximal inequalities are used to obtain the above weak convergence. The practical importance of Theorem 3 is that it allows to estimate the number c in (12) via simulation: (with the previous notations), conditionally on $\widehat{\gamma}_N$, one can simulate a large number of sample paths of the Gaussian process $(\widehat{G}_N/\widehat{\sigma}_N)$ and compute their supremum norms. One then obtains a precise approximation to the distribution of $\|\widehat{G}_N/\widehat{\sigma}_N\|_\infty$, and it suffices to set c as the quantile of order $(1-\alpha)$ of this distribution:

$$\mathbb{P}\left(|\widehat{G}_N(t)| \leq c\,\widehat{\sigma}_N(t),\ \forall t \in [0,T] \,\middle|\, \widehat{\gamma}_N\right) = 1-\alpha. \tag{13}$$

Corollary 1. *Assume (A1)–(A6). Under the conditions of Theorems 1-2-3, the bands defined in* (11) *with the real* $c = c(\widehat{\gamma}_N)$ *chosen as in* (13) *have asymptotic coverage level* $1-\alpha$, *i.e.*

$$\lim_{N\to\infty} \mathbb{P}\left(\mu_N(t) \in \left[\widehat{\mu}_N(t) \pm c\,\frac{\widehat{\sigma}_N(t)}{N^{1/2}}\right], \quad \forall t \in [0,T]\right) = 1-\alpha.$$

4 A simulation study

In this section, we evaluate the performances of the mean curve estimator as well as the coverage and the width of the confidence bands for different bandwidth selection criteria and different levels of noise. The simulations are conducted in the R environment.

4.1 Simulated data and sampling designs

We have generated a population of $N = 20000$ curves discretized at $d = 200$ and $d = 400$ equidistant instants of time in $[0,1]$. The curves of the population are generated so that they have approximately the same distribution as the electricity consumption curves analyzed in Cardot & Josserand (2011) and each individual curve X_k, for $k \in U$, is simulated as follows

$$X_k(t) = \mu(t) + \sum_{\ell=1}^{3} Z_\ell\ v_\ell(t), \quad t \in [0,1], \tag{14}$$

where the mean function μ is drawn in Figure 2 and the random variables Z_ℓ are independent realizations of a centered Gaussian random variable with variance σ_ℓ^2. The three basis function v_1, v_2 and v_3 are orthonormal functions which represent the main mode of variation of the signals, they are represented in Figure 1. Thus, the covariance function of the population $\gamma(s,t)$ is simply

$$\gamma(s,t) \quad = \quad \sum_{\ell=1}^{3} \sigma_\ell^2\ v_\ell(s) v_\ell(t). \tag{15}$$

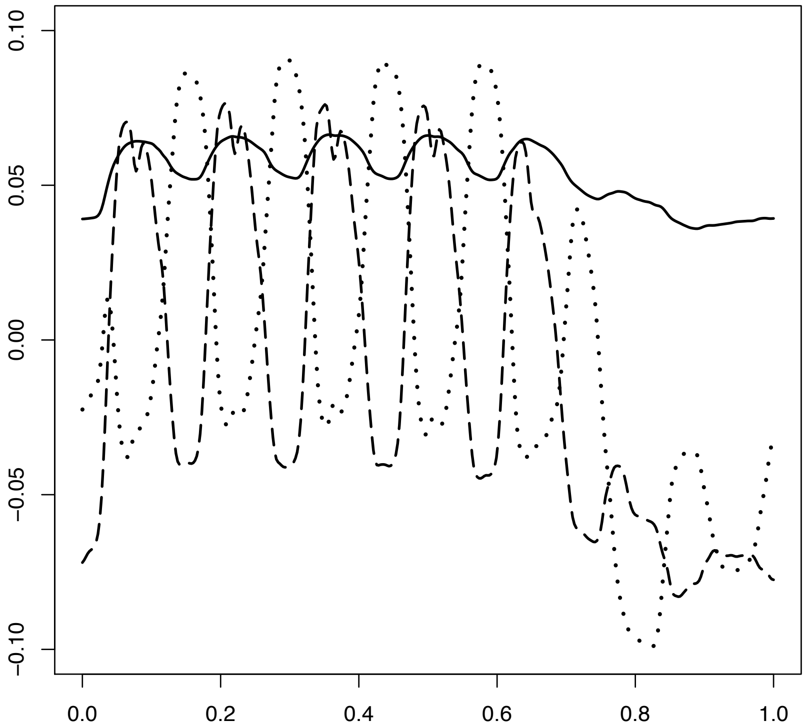

Figure 1: Basis functions v_1 (solid line), v_2 (dashed line) and v_3 (dotted line).

To select the samples, we have considered two probabilistic selection procedures, with fixed sample size, $n = 1000$,

- Simple random sampling without replacement (SRSWR).
- Stratified sampling with SRSWR in all strata. The population U is divided into a fixed number of $G = 5$ strata built by considering the quantiles $q_{0.5}, q_{0.7}, q_{0.85}$ and $q_{0.95}$ of the total consumption $\int_0^1 X_k(t)dt$ for all units $k \in U$. For example, the first strata contains all the units k such that $\int_0^1 X_k(t)dt \leq q_{0.5}$, and thus its size is half of the population size N. The sample size n_g in stratum g is determined by a Neyman-like allocation, as suggested in Cardot and Josserand (2011), in order to get a Horvitz-Thompson estimator of the mean trajectory whose variance is as small as possible. The sizes of the different strata, which are optimal according to this mean variance criterion, are reported in Table 1.

We suppose we observe, for each unit k in the sample s, the discretized trajectories, at

Stratum number	1	2	3	4	5
Stratum size	10000	4000	3000	2000	1000
Allocation	655	132	98	68	47

Table 1: Strata sizes and optimal allocations.

d equispaced points, $0 = t_1 < \ldots < t_d = 1$,

$$Y_{jk} = X_k(t_j) + \delta\epsilon_{jk} \tag{16}$$

where the $\epsilon_{jk} \sim N(0, \gamma(t_j, t_j))$ are independent random variables and the parameter δ allows to control the noise level. As an illustrative example, a sample of $n = 10$ noisy discretized curves are plotted in Figure 2.

4.2 Weighted cross-validation for bandwidth selection

Assuming we can access the exact trajectories X_k, $k \in s$, (which is the case in simulations) we consider the oracle-type estimator

$$\widehat{\mu}_s = \sum_{k \in s} \frac{X_k}{\pi_k}\,, \tag{17}$$

which will be a benchmark in our numerical study. We compare different interpolation and smoothing strategies for estimating the X_k, $k \in s$:

- Linear interpolation of the Y_{jk} as in Cardot and Josserand (2011).
- Local linear smoothing of the Y_{jk} with bandwidth h as in (3).

The crucial element here is h. To evaluate the interest of smoothing and the performances of data-driven bandwidth selection criteria, we consider an error measure that compares the oracle $\widehat{\mu}_s$ to any estimator $\widehat{\mu}$ based on the noisy data Y_{jk}, $k \in s$, $j = 1, \ldots d$:

$$L(\widehat{\mu}) = \int_0^T \left(\widehat{\mu}_s(t) - \widehat{\mu}(t)\right)^2 dt\,. \tag{18}$$

Considering the estimator defined in (6), we denote by h_{oracle} the bandwidth h that minimizes (18) and call smooth oracle the corresponding estimator.

When $\sum_{k \in s} \pi_k^{-1} = N$, as in SRSWR and stratified sampling, it can be easily checked that $\widehat{\mu}_s$ is the minimum argument of the weighted least squares functional

$$\sum_{k \in s} w_k \int_0^T \left(X_k(t) - \mu(t)\right)^2 dt \tag{19}$$

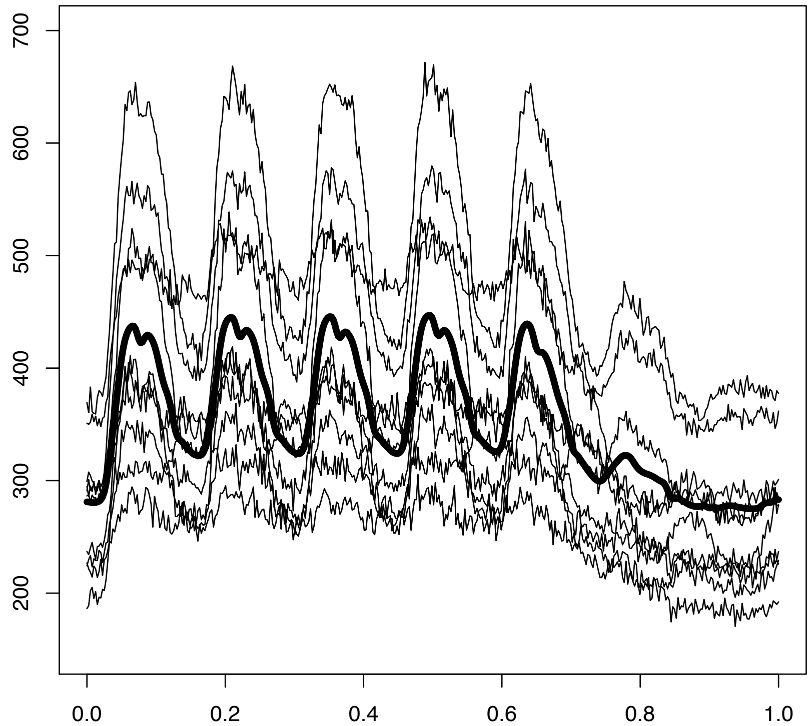

Figure 2: A sample of 10 curves for $\delta = 0.05$. The mean profile is plotted in bold line.

with respect to $\mu \in L^2([0,T])$, where the weights are $w_k = (N\pi_k)^{-1}$. Then, a simple and natural way to select bandwidth h is to consider the following design-based cross validation

$$\mathrm{WCV}(h) = \sum_{k\in s} w_k \sum_{j=1}^{d} \left(Y_{jk} - \widehat{\mu}_N^{-k}(t_j)\right)^2. \tag{20}$$

where

$$\widehat{\mu}_N^{-k}(t) = \sum_{\ell\in s, \ell\neq k} \widetilde{w}_\ell \ \widehat{X}_\ell(t),$$

with new weights $\widetilde{w}_\ell$. A heuristic justification for this approach is that, given s, we have $\mathbb{E}\left[\epsilon_{jk}(X_k(t_j) - \widehat{\mu}_N^{-k}(t_j))|s\right] = 0$ for $j = 1, \dots, d$ and $k \in s$. Thus,

$$\begin{aligned}
\mathbb{E}\left[\mathrm{WCV}(h)|s\right] &= \sum_{k\in s} w_k \sum_{j=1}^{d} \left\{ \mathbb{E}\left[\left(X_k(t_j) - \widehat{\mu}_N^{-k}(t_j)\right)^2 |s\right] + 2\mathbb{E}\left[\epsilon_{jk}(X_k(t_j) - \widehat{\mu}_N^{-k}(t_j))|s\right] + \mathbb{E}\left[\epsilon_{jk}^2\right] \right\} \\
&= \sum_{k\in s} w_k \sum_{j=1}^{d} \mathbb{E}\left[\left(X_k(t_j) - \widehat{\mu}_N^{-k}(t_j)\right)^2 |s\right] + \mathrm{tr}(\mathbf{V}_N)
\end{aligned}$$

and, up to $\text{tr}(\mathbf{V}_N)$ which does not depend on h, the minimum value of the expected cross validation criterion should be attained for estimators which are not too far from $\widehat{\mu}_s$.

This weighted cross validation criterion is simpler than the cross validation criteria based on the estimated variance proposed in Opsomer and Miller (2005). Indeed, in our case, the bias may be non negligible and focusing only on the variance part of the error leads to too large selected values for the bandwidth. Furthermore, Opsomer and Miller (2005) suggested to consider weights defined as follows $\widetilde{w}_\ell = w_\ell/(1-w_k)$. For SRSWR, since $w_k = n^{-1}$ one has $\widetilde{w}_k = (n-1)^{-1}$, so that the weighted cross validation criterion defined in (20) is exactly the cross validation criterion introduced by Rice and Silverman (1991) in the independent case. We denote in the following by h_{cv} the bandwidth value minimizing this criterion.

For stratified sampling, a better approximation which keeps the design-based properties of the estimator $\widehat{\mu}_N^{-k}$ can be obtained by taking into account the sampling rates in the different strata. We have G strata with sizes N_g, $g = 1, \dots, G$ and we sample n_g observations, with SRSWR, in each stratum g. If unit k comes from strata g, we have $w_k = N_g(Nn_g)^{-1}$. Thus, we take $\widetilde{w}_\ell = (N_g - 1)\{(N-1)(n_g-1)\}^{-1}$ for all the units $\ell \neq k$ in stratum g and just scale the weights for all the units ℓ' of the sample that do not belong to stratum g, $\widetilde{w}_{\ell'} = N(N-1)^{-1}w_{\ell'}$. We denote by h_{wcv} the bandwidth value minimizing (20).

4.3 Estimation errors and confidence bands

We draw 1000 samples in the population of curves and compare the different estimators of Section 4.2 with the L^2 loss criterion

$$R(\widehat{\mu}) = \int_0^T (\widehat{\mu}(t) - \mu(t))^2 \, dt \tag{21}$$

for different values of δ and d in (16).

The empirical mean as well as the first, second and third quartiles of the estimation error $R(\widehat{\mu})$ are given in Table 2 for $d = 200$ and in Table 3 for $d = 400$. We can first note that stratified sampling allows to improve much the estimation of the mean curve. We also remark that, for such large samples, linear interpolation performs nearly as well as the smooth oracle estimator, especially when the noise level is low ($\delta \leq 15\%$). As far as bandwidth selection is concerned, we can note that the usual cross validation criterion h_{cv} is not adapted to unequal probability sampling and does not perform as well as linear interpolation for stratified sampling by selecting too large values for the bandwidth. On the other hand, the weighted cross-validation criterion seems to be effective to select good bandwidth values and produce estimators whose estimation errors are very close to the oracle and perform better than the other estimators when the noise level is moderate or high ($\delta \geq 20\%$).

This is clearer when we look at criterion $L(\widehat{\mu})$, defined in (18), which only focus on the part of the estimation error which is due to the noise. Results are presented in Table 4 for $d = 200$ and in Table 5 for $d = 400$. We can also note that there is a significant effect of the number of discretization points on the accuracy of the smoothed estimators. Our individual trajectories, which have roughly the same shape as load curves, are actually not very smooth so that smoothing approaches are only really interesting, compared to linear interpolation when the number of discretization points d is large enough.

We now examine in Table 6 and Table 7 the empirical coverage and the width of the confidence bands, which are built as described in Section 3.3. For each sample, we estimate the covariance function $\widehat{\gamma}_N$ and draw 10000 realizations of a centered Gaussian process with variance function $\widehat{\gamma}_N$ in order to obtain a suitable coefficient c with a confidence level of $1 - \alpha = 0.95$ as explained in equation (13). The area of the confidence band is then $\int_0^T 2c\sqrt{\widehat{\gamma}(t,t)}\ \mathrm{d}t$. The results highlight now the interest of considering smoothing strategies combined with the weighted cross validation bandwidth selection criterion (20). It appears that linear interpolation, which does not intend to get rid of the noise, always gives larger confidence bands than the smoothed estimators based on h_{wcv}. As before, smoothing approaches become more interesting as the number of discretization points and the noise level increase. The empirical coverage of the smoothed estimator is lower than the linear interpolation estimator but remains slightly higher than the nominal one.

As a conclusion of this simulation study, it appears that smoothing is not a crucial aspect when the only target is the estimation of the mean, and that bandwidth values should be chosen by a cross validation criterion that takes the sampling weights into account. When the goal is also to build confidence bands, smoothing with weighted cross validation criteria lead to narrower bands compared to interpolation techniques, without deteriorating the empirical coverage.

5 Concluding remarks

We have studied in this paper the use of survey sampling methods for estimating a population mean temporal signal. This type of approach is extremely effective when data transmission or storage costs are important, in particular for large networks of distributed sensors. In view of noisy functional data, we have built a functional estimator by first smoothing the sampled curves and then setting up a Horvitz-Thompson estimator based on the smoothed curves. It has been shown that the estimator satisfies a CLT in the space of continuous functions and that its covariance can be estimated uniformly and consistently. These results have been exploited to show that by simulating Gaussian processes conditional on the estimated covariance, one obtains global confidence bands with asymptotic correct coverage. The problem of bandwidth selection, which is particularly difficult in the survey sampling context, has been addressed. We have devised a weighted cross-validation method that aims at mimicking an oracle estimator. This method has displayed very good performances in our numerical study; however, its rigorous theoretical study of remains to be done. Our numerical study has also revealed that in comparison to SRSWR, unequal probability sampling (e.g. stratified sampling) yields far superior performances and that when the noise level in the data is moderate to high, incorporating a smoothing step in the estimation procedure greatly enhances the accuracy in comparison to interpolation. Furthermore, we have seen that even when the noise level is low, smoothing can be highly beneficial for build global confidence bands. Indeed, smoothing the data leads to estimators that have higher temporal correlation, which in turn makes the confidence bands narrower and more stable. Our method for confidence bands is simple and quick to implement. It gives satisfactory coverage (a little conservative) when the bandwidth is chosen correctly, e.g. with our weighted cross-validation method. Such confidence bands can find a variety of applications in statistical testing. They can be used to compare mean functions in different sub-populations, or to test for a parametric shape or for periodicity, among others. Examples of applications can be found in Degras (2010).

This work also raises some questions which deserve further investigation. A straightforward extension could be to relax the normality assumption made on the measurement errors. It is possible to consider more general error distributions under additional assumptions on the moments and much longer proofs. In another direction, it would be worthwhile to see whether our methodology can be extended to build confidence bands for other functional parameters such as population quantile or covariance functions. Also, as mentioned earlier, the weighted cross-validation proposed in this work seems a promising candidate for automatic bandwidth selection. However it is for now only based on heuristic arguments and its theoretical underpinning should be investigated.

Finally, it is well known that taking account of auxiliary information, which can be made available for all the units of the population at a low cost, can lead to substantial improvements with model assisted estimators (Särndal *et al.* 1992). In a functional context, an interesting strategy consists in first reducing the dimension through a functional principal components analysis shaped for the sampling framework (Cardot *et. al* 2010a) and then consider semi parametric models relating the principal components scores to the auxiliary variables (Cardot *et al.* 2010b). It is still possible to get consistent estimators of the covariance function of the limit process but further investigations are needed to prove the functional asymptotic normality and deduce that Gaussian simulations based approaches still lead to accurate confidence bands.

Acknowledgement. Etienne Josserand thanks the *Conseil Régional de Bourgogne, France* for its financial support (FABER PhD grant).

Appendix

Throughout the proofs we use the letter C to denote a generic constant whose value may vary from place to place. This constant does not depend on N nor on the arguments $s,t \in [0,T]$.

Proof of Theorem 1. We first decompose the difference between the estimator $\widehat{\mu}_N(t)$ and its target $\mu_N(t)$ as the sum of two stochastic components, one pertaining to the sampling variability and the other to the measurement errors, and of a deterministic bias component:

$$\widehat{\mu}_N(t) - \mu_N(t) = \frac{1}{N}\sum_{k\in U}\left(\frac{I_k}{\pi_k} - 1\right)\tilde{X}_k(t) + \frac{1}{N}\sum_{k\in U}\frac{I_k}{\pi_k}\tilde{\epsilon}_k(t) + \frac{1}{N}\sum_{k}\left(\tilde{X}_k(t) - X_k(t)\right) \quad (22)$$

where $\tilde{X}_k(t)$ and $\tilde{\epsilon}_k(t)$ are defined in (9).

Bias term.
To study the bias term $N^{-1}\sum_k \left(\tilde{X}_k(t) - X_k(t)\right) = \mathbb{E}(\widehat{\mu}_N(t)) - \mu_N(t)$ in (22), it suffices to use classical results on local linear smoothing (e.g. Tsybakov (2009), Proposition 1.13) together with the Hölder continuity (A2) of the X_k to see that

$$\sup_{t\in[0,T]}\left|\frac{1}{N}\sum_{k}\left(\tilde{X}_k(t) - X_k(t)\right)\right| \leq \frac{1}{N}\sum_{k}\sup_{t\in[0,T]}\left|\tilde{X}_k(t) - X_k(t)\right| \leq Ch^{\beta}. \quad (23)$$

Hence, for the bias to be negligible in the normalized estimator, it is necessary that the bandwidth satisfy $\sqrt{N}h^{\beta} \to 0$ as $N \to \infty$.

Error term.
We now turn to the measurement error term in (22), which can be seen as a sequence of random functions. We first show that this sequence goes pointwise to zero in mean square (a fortiori in probability) at a rate $(Ndh)^{-1}$. We then establish its tightness in $C([0,T])$, when premultiplied by $\sqrt{N}$, to prove the uniformity of the convergence over $[0,T]$.

Writing the vector of local linear weights at point t as follows

$$W(t) = (W_1(t), \dots, W_d(t))'$$

and using the i.i.d assumption (A4) on the $(\epsilon_{k1},\dots,\epsilon_{kd})'$, $k\in U_N$, we first obtain that

$$\begin{aligned}\mathbb{E}\left(\frac{1}{N}\sum_{k\in U}\frac{I_k}{\pi_k}\tilde{\epsilon}_k(t)\right)^2 &= \frac{1}{N^2}\sum_{k\in U}\frac{1}{\pi_k}\mathbb{E}\left(\tilde{\epsilon}_k(t)\right)^2\\ &= \frac{1}{N^2}\sum_{k\in U}\frac{1}{\pi_k}W(t)'\mathbf{V}_N W(t).\end{aligned}$$

Then, considering the facts that $\min_k \pi_k > c$ by (A2), $\|\mathbf{V}_N\|$ is uniformly bounded in N by (A4), and exploiting a classical bound on the weights of the local linear smoother (e.g. Tsybakov (2009), Lemma 1.3), we deduce that

$$\begin{aligned}\mathbb{E}\left(\frac{1}{N}\sum_{k\in U}\frac{I_k}{\pi_k}\tilde{\epsilon}_k(t)\right)^2 &\le \frac{N}{(\min\pi_k)N^2}\|W(t)\|^2\|\mathbf{V}_N\|\\ &\le \frac{C}{Ndh}\,. \end{aligned}\tag{24}$$

We can now prove the tightness of the sequence of processes $(N^{-1/2}\sum_k (I_k/\pi_k)\,\tilde{\epsilon}_k)$. Let us define the associated pseudo-metric

$$d_\epsilon^2(s,t)=\mathbb{E}\left(\frac{1}{\sqrt{N}}\sum_{k\in U}\frac{I_k}{\pi_k}\left(\tilde{\epsilon}_k(s)-\tilde{\epsilon}_k(t)\right)\right)^2.$$

We use the following maximal inequality holding for sub-Gaussian processes (van der Vaart and Wellner (2000), Corollary 2.2.8):

$$\mathbb{E}\left(\sup_{t\in[0,T]}\left|\frac{1}{\sqrt{N}}\sum_{k\in U}\frac{I_k}{\pi_k}\tilde{\epsilon}_k(t)\right|\right)\le \mathbb{E}\left(\left|\frac{1}{\sqrt{N}}\sum_{k\in U}\frac{I_k}{\pi_k}\tilde{\epsilon}_k(t_0)\right|\right)+K\int_0^\infty\sqrt{\log N(x,d_\epsilon)}dx\,,\tag{25}$$

where t_0 is an arbitrary point in $[0,T]$ and the covering number $N(x,d_\epsilon)$ is the minimal number of d_ϵ-balls of radius $x>0$ needed to cover $[0,T]$. Note the equivalence of working with packing or covering numbers in maximal inequalities, see *ibid* p. 98. Also note that the sub-Gaussian nature of the smoothed error process $N^{-1/2}\sum_{k\in U}(I_k/\pi_k)\,\tilde{\epsilon}_k$ stems from the i.i.d. multivariate normality of the random vectors $(\epsilon_{k1},\dots,\epsilon_{kd})'$ and the boundedness of the I_k for $k\in U_N$.

By the arguments used in (24) and an elementary bound on the increments of the weight function vector W (see e.g. Lemma 1 in Degras (2010)), one obtains that

$$\begin{aligned}d_\epsilon^2(s,t)&=\frac{1}{N}\sum_{k\in U}\frac{1}{\pi_k}\mathbb{E}\left(\tilde{\epsilon}_k(s)-\tilde{\epsilon}_k(t)\right)^2\\ &\le\frac{1}{\min\pi_k}\|W(s)-W(t)\|^2\|\mathbf{V}_N\|\\ &\le\frac{C}{dh}\left(\frac{|s-t|^2}{h^2}\wedge 1\right).\end{aligned}\tag{26}$$

It follows that the covering numbers satisfy

$$\begin{cases} N(x, d_\epsilon) = 1, & \text{if } \frac{C}{dh} \leq x^2, \\ N(x, d_\epsilon) \leq \frac{\sqrt{C}}{h\sqrt{dh}x}, & \text{if } \frac{C}{dh} > x^2. \end{cases}$$

Plugging this bound and the pointwise convergence (24) in the maximal inequality (25), we get after a simple integral calculation (see Eq. (17) in Degras (2010) for details) that

$$\mathbb{E}\left(\sup_{t\in[0,T]} \left| \frac{1}{\sqrt{N}} \sum_{k\in U} \frac{I_k}{\pi_k} \tilde{\epsilon}_k(t) \right| \right) \leq \frac{C}{dh} + C\sqrt{\frac{|\log(h)|}{dh}}\,. \tag{27}$$

Thanks to Markov's inequality, the previous bound guarantees the uniform convergence in probability of $N^{-1/2}\sum_{k\in U}(I_k/\pi_k)\tilde{\epsilon}_k$ to zero, provided that $|\log(h)|/(dh) \to 0$ as $N \to \infty$. The last condition is equivalent to $\log(d)/(dh) \to 0$ by the fact that $dh \to \infty$ and by the properties of the logarithm.

Main term: sampling variability.
Finally, we look at the process $N^{-1}\sum_{k\in U}(I_k/\pi_k - 1)\tilde{X}_k$ in (22), which is asymptotically normal in $C([0,T])$ as we shall see. We first establish the finite-dimensional asymptotic normality of this process normalized by $\sqrt{N}$, after which we will prove its tightness thanks to a maximal inequality.

Let us start by verifying that the limit covariance function of the process is indeed the function γ defined in Section 3.1. The finite-sample covariance function expresses as

$$\begin{aligned} \mathbb{E}\left\{ \left(\frac{1}{\sqrt{N}} \sum_{k\in U} \left(\frac{I_k}{\pi_k} - 1\right) \tilde{X}_k(s) \right) \left(\frac{1}{\sqrt{N}} \sum_{l\in U} \left(\frac{I_l}{\pi_l} - 1\right) \tilde{X}_l(t) \right) \right\} \\ = \frac{1}{N} \sum_{k,l\in U} \frac{\Delta_{kl}}{\pi_k\pi_l} \tilde{X}_k(s)\tilde{X}_l(t) \\ = \frac{1}{N} \sum_{k,l\in U} \frac{\Delta_{kl}}{\pi_k\pi_l} X_k(s)X_l(t) + \mathcal{O}\big(h^\beta\big) \\ = \gamma(s,t) + o(1) + \mathcal{O}\big(h^\beta\big). \end{aligned} \tag{28}$$

To derive the previous relation we have used the facts that

$$\max_{k,l\in U} \sup_{s,t\in[0,T]} \left| \tilde{X}_k(s)\tilde{X}_l(t) - X_k(s)X_l(t) \right| \leq Ch^\beta$$

by (23) and the uniform boundedness of the X_k arising from (A2) and that, by (A1),

$$\begin{aligned} \frac{1}{N} \sum_{k,l\in U} \frac{|\Delta_{kl}|}{\pi_k\pi_l} &= \frac{1}{N} \sum_{k\neq l} \frac{|\Delta_{kl}|}{\pi_k\pi_l} + \frac{1}{N} \sum_{k} \frac{\Delta_{kk}}{\pi_k^2} \\ &\leq \frac{1}{N} \frac{N(N-1)}{2} \frac{\max_{k,l}(n|\Delta_{kl}|)}{n} + \frac{1}{N} \sum_{k} \frac{1-\pi_k}{\pi_k} \leq C. \end{aligned} \tag{29}$$

We now check the finite-dimensional convergence of $N^{-1/2}\sum_{k\in U}(I_k/\pi_k-1)\tilde{X}_k$ to a centered Gaussian process with covariance γ. In light of the Cramer-Wold theorem, this convergence is easily shown with characteristic functions and appears as a straightforward consequence of(A5). It suffices for us to check that the uniform boundedness of the trajectories X_k derived from (A2) is preserved by local linear smoothing, so that the $\tilde{X}_k$ are uniformly bounded as well.

It remains to establish the tightness of the previous sequence of processes so as to obtain its asymptotic normality in $C([0,T])$. To that intent we use the maximal inequality of the Corollary 2.2.5 in van der Vaart and Wellner (2000). With the notations of this result, we consider the pseudo-metric $d^2_{\tilde{X}}(s,t)=\mathbb{E}\{N^{-1/2}\sum_{k\in U}(I_k/\pi_k-1)(\tilde{X}_k(s)-\tilde{X}_k(t))\}^2$ and the function $\psi(t)=t^2$ for the Orlicz norm. We get the following bound for the second moment of the maximal increment:

$$\mathbb{E}\left\{\sup_{d_{\tilde{X}}(s,t)\leq\delta}\left|\frac{1}{\sqrt{N}}\sum_{k\in U}\left(\frac{I_k}{\pi_k}-1\right)\left(\tilde{X}_k(s)-\tilde{X}_k(t)\right)\right|\right\}^2 \leq C\left(\int_0^\eta\psi^{-1}(N(x,d_{\tilde{X}}))dx+\delta\psi^{-1}(N^2(\eta,d_{\tilde{X}}))\right)^2 \quad (30)$$

for any arbitrary constants $\eta,\delta>0$. Observe that the maximal inequality (30) is weaker than (25) where an additional assumption of sub-Gaussianity is made (no log factor in the integral above). Employing again the arguments of (28), we see that

$$\begin{aligned} d^2_{\tilde{X}}(s,t) &= \frac{1}{N}\sum_{k,l}\frac{\Delta_{kl}}{\pi_k\pi_l}\left(\tilde{X}_k(s)-\tilde{X}_k(t)\right)\left(\tilde{X}_l(s)-\tilde{X}_l(t)\right) \\ &\leq \frac{C}{N}\,\frac{N(N-1)}{2n}|s-t|^{2\beta}+\frac{C}{N}\,N|s-t|^{2\beta} \\ &\leq C|s-t|^{2\beta}. \end{aligned} \quad (31)$$

It follows that the covering number satisfies $N(x,d_{\tilde{X}})\leq Cx^{-1/\beta}$ and that the integral in (30) is smaller than $C\int_0^\eta x^{-0.5/\beta}dx=C\eta^{1-0.5/\beta}$, which can be made arbitrarily small since $\beta>0.5$. Once η is fixed, δ can be adjusted to make the other term in the right-handside of (30) arbitrarily small as well. With Markov's inequality, we deduce that the sequence $(N^{-1/2}\sum_{k\in U}(I_k/\pi_k-1)\tilde{X}_k)_{N\geq 1}$ is asymptotically $d_{\tilde{X}}$-equicontinuous in probability (with the terminology of van der Vaart and Wellner (2000)), which guarantees its tightness in $C([0,T])$. □

Proof of Theorem 2.

Mean square convergence.

We first decompose the distance between $\widehat{\gamma}_N(s,t)$ and its target $\gamma_N(s,t)$ as follows:

$$\begin{aligned}\widehat{\gamma}_N(s,t)-\gamma_N(s,t) &= \frac{1}{N}\sum_{k,l\in U}\frac{\Delta_{kl}}{\pi_k\pi_l}\left(\frac{I_kI_l}{\pi_{kl}}-1\right)\tilde{X}_k(s)\tilde{X}_l(t)\\ &+\frac{2}{N}\sum_{k,l\in U}\frac{\Delta_{kl}}{\pi_k\pi_l}\frac{I_kI_l}{\pi_{kl}}\tilde{X}_k(s)\tilde{\epsilon}_l(t)\\ &+\frac{1}{N}\sum_{k,l\in U}\frac{\Delta_{kl}}{\pi_k\pi_l}\frac{I_kI_l}{\pi_{kl}}\tilde{\epsilon}_k(s)\tilde{\epsilon}_l(t)\\ &-\frac{1}{N}\sum_{k\in U}\frac{1}{\pi_k}\mathbb{E}\big(\tilde{\epsilon}_k(s)\tilde{\epsilon}_k(t)\big)\\ &:= A_{1,N}+A_{2,N}+A_{3,N}-A_{4,N}. \end{aligned}\tag{32}$$

To establish the mean square convergence of $(\widehat{\gamma}_N(s,t)-\gamma_N(s,t))$ to zero as $N\to\infty$, it is enough to show that $\mathbb{E}(A_{i,N}^2)\to 0$ for $i=1,\dots,4$, by the Cauchy-Schwarz inequality.

Let us start with

$$\mathbb{E}(A_{1,N}^2)=\frac{1}{N^2}\sum_{k,l}\sum_{k',l'}\frac{\Delta_{kl}\Delta_{k'l'}}{\pi_k\pi_l\pi_{k'}\pi_{l'}}\frac{\mathbb{E}\{(I_kI_l-\pi_{kl})(I_{k'}I_{l'}-\pi_{k'l'})\}}{\pi_{kl}\pi_{k'l'}}\widetilde{X}_k(s)\widetilde{X}_l(t)\widetilde{X}_{k'}(s)\widetilde{X}_{l'}(t). \tag{33}$$

It can be shown that this sum converges to zero by strictly following the proof of the Theorem 3 in Breidt and Opsomer (2000). The idea of the proof is to partition the set of indexes in (33) into (i) $k=l$ and $k'=l'$, (ii) $k=l$ and $k'\neq l'$ or vice-versa, (iii) $k\neq l$ and $k'\neq l'$, and study the related subsums. The convergence to zero is then handled with assumption (A1) (mostly) in case (i), with (A1)-(A6) in case (iii), and thanks to the previous results and Cauchy-Schwarz inequality in case (ii). More precisely, it holds that

$$\begin{aligned}\mathbb{E}(A_{1,N}^2) &\leq \frac{C\max_{k\neq l} n|\Delta_{kl}|}{(\min\pi_k)^4 n}+\frac{C}{(\min\pi_k)^3N}\\ &+\left(\frac{C(\max_{k\neq l} n|\Delta_{kl}|)N}{(\min\pi_k)^2(\min_{k\neq l}\pi_{kl})n}\right)^2\max_{(k,l,k',l')\in D_{4,N}}\big|\mathbb{E}\{(I_kI_l-\pi_{kl})(I_{k'}I_{l'}-\pi_{k'l'})\}\big|.\end{aligned}\tag{34}$$

For the (slightly simpler) study of $\mathbb{E}(A_{2,N}^2)$, we provide an explicit decomposition:

$$\begin{aligned}\mathbb{E}(A_{2,N}^2) &= \frac{4}{N^2}\sum_{k,l}\sum_{k'}\frac{\Delta_{kl}\Delta_{k'l}}{\pi_k\pi_{k'}\pi_l^2}\tilde{X}_k(s)\tilde{X}_{k'}(t)\,\mathbb{E}\big(\tilde{\epsilon}_l(s)\tilde{\epsilon}_l(t)\big)\\ &=\frac{4}{N^2}\sum_{k\in U}\frac{\Delta_{kk}^2}{\pi_k^5}\tilde{X}_k(s)\tilde{X}_k(t)\mathbb{E}\big(\tilde{\epsilon}_k(s)\tilde{\epsilon}_k(t)\big)\\ &+\frac{4}{N^2}\sum_{k\neq k'}\frac{\Delta_{kk}\Delta_{kk'}}{\pi_k^4\pi_{k'}}\tilde{X}_k(s)\tilde{X}_{k'}(t)\,\mathbb{E}(\tilde{\epsilon}_k(s)\tilde{\epsilon}_k(t))\\ &+\frac{4}{N^2}\sum_{k\neq l}\sum_{k':k'\neq l}\frac{\Delta_{kl}\Delta_{k'l}}{\pi_k\pi_{k'}\pi_l^2}\tilde{X}_k(s)\tilde{X}_{k'}(t)\,\mathbb{E}\big(\tilde{\epsilon}_l(s)\tilde{\epsilon}_l(t)\big).\end{aligned}\tag{35}$$

Note that the expression of $\mathbb{E}(A_{2,N}^2)$ as a quadruple sum over $k, l, k', l' \in U_N$ reduces to a triple sum since $\mathbb{E}(\tilde{\epsilon}_l(s)\tilde{\epsilon}_{l'}(t)) = 0$ if $l \neq l'$ by (A4). With the bound $|\mathbb{E}(\tilde{\epsilon}_k(s)\tilde{\epsilon}_k(t))| = |W(s)'\mathbf{V}_N W(t)| \leq \|W(s)\|\|\mathbf{V}_N\|\|W(t)\| \leq C/(dh)$, it follows that

$$\begin{aligned}\mathbb{E}(A_{2,N}^2) \leq \frac{CN}{N^2}\frac{\|\mathbf{V}_N\|}{dh} + \frac{CN^2}{N^2}\frac{\max_{k\neq k'} n|\Delta_{kk'}|}{n}\frac{\|\mathbf{V}_N\|}{dh} \\ + \frac{CN^3}{N^2}\frac{(\max_{k\neq l} n|\Delta_{kl}|)^2}{n^2}\frac{\|\mathbf{V}_N\|}{dh} = \frac{C}{Ndh}.\end{aligned} \tag{36}$$

To study the term $\mathbb{E}(A_{3,N}^2)$, we start with the same partition of the quadruple sum as the one used with $\mathbb{E}(A_{1,N}^2)$. Here, due to the independence assumption (A4) on the error vectors, the partition simplifies further into (i) $k = l$, $k' = l'$, $k \neq k'$, and (ii) $k = l = k' = l'$:

$$\begin{aligned}\mathbb{E}(A_{3,N}^2) = \frac{1}{N^2}\sum_{k\neq k'}\frac{\Delta_{kk'}}{\pi_k\pi_{k'}}\frac{I_k I_{k'}}{\pi_{kk'}}\mathbb{E}(\tilde{\epsilon}_k(s)\tilde{\epsilon}_k(t))\,\mathbb{E}(\tilde{\epsilon}_{k'}(s)\tilde{\epsilon}_{k'}(t)) \\ + \frac{1}{N^2}\sum_{k}\frac{\Delta_{kk}}{\pi_k^2}\frac{I_k}{\pi_{kk}}\mathbb{E}(\tilde{\epsilon}_k^2(s)\tilde{\epsilon}_k^2(t)).\end{aligned} \tag{37}$$

Forgoing the calculations already done before, we focus on the main task which for this term is to bound the quantity $\mathbb{E}(\tilde{\epsilon}_k^2(s)\tilde{\epsilon}_k^2(t))$ (recall that $\mathbb{E}(\tilde{\epsilon}_k(s)\tilde{\epsilon}_k(t)) \leq C/(dh)$ as seen before). We first note that $\mathbb{E}(\tilde{\epsilon}_k^2(s)\tilde{\epsilon}_k^2(t)) \leq \left\{\mathbb{E}(\tilde{\epsilon}_k^4(s))\right\}^{1/2}\left\{\mathbb{E}(\tilde{\epsilon}_k^4(t))\right\}^{1/2}$. Writing $\boldsymbol{\epsilon} \sim N(0, \mathbf{V}_N)$, it holds that $\mathbb{E}(\tilde{\epsilon}_k^4(t)) = \mathbb{E}((W(t)'\boldsymbol{\epsilon})^4) = 3(W(t)'\mathbf{V}_N W(t))^2$ by the moment properties of the normal distribution. Plugging this expression in (37), we find that

$$\mathbb{E}(A_{3,N}^2) \leq \frac{C}{(dh)^2} + \frac{C}{N(dh)^2}. \tag{38}$$

Finally, note that an expression very similar to the deterministic term $A_{4,N}(s,t)$ has been studied in (24). One easily concludes that $A_{4,N}(s,t)$ is dominated by $(dh)^{-1}$ uniformly in $s, t \in [0, T]$.

Tightness.
To prove the tightness of the sequence $(\widehat{\gamma}_N - \gamma_N)_{N\geq 1}$ in $C([0,T]^2)$, we study separately each term in the decomposition (32) and we call again to the maximal inequalities of van der Vaart and Wellner (2000).

For the first term $A_{1,N} = A_{1,N}(s,t)$, we consider the pseudo-metric d defined as the L^4-norm of the increments: $d_1^4((s,t),(s',t')) = \mathbb{E}|A_{1,N}(s,t) - A_{1,N}(s',t')|^4$. (The need to use here the L^4-norm and not the usual L^2-norm is justified hereafter by a dimension argument.) With (A1)-(A2) and the approximation properties of local linear smoothers, one sees that

$$\left|\frac{1}{N}\sum_{k,l\in U}\frac{\Delta_{kl}}{\pi_k\pi_l}\left(\frac{I_k I_l}{\pi_{kl}} - 1\right)\left(\tilde{X}_k(s)\tilde{X}_l(t) - \tilde{X}_k(s')\tilde{X}_l(t')\right)\right| \leq C\left(|s-s'|^\beta + |t-t'|^\beta\right).$$

Hence $d_1(s,t) \leq C\left(|s-s'|^\beta + |t-t'|^\beta\right)$ and for all $x>0$, the covering number $N(x,d_1)$ is no larger than the size of a two-dimensional square grid of mesh $x^{1/\beta}$, i.e. $N(x,d_1) \leq Cx^{-2/\beta}$. (Compare to the proof of Theorem 1 where, for the main term $N^{-1/2}\sum_k (I_k/\pi_k)\tilde{X}_k$, we have $N(x, d_{\tilde{X}}) \leq Cx^{-1/\beta}$ because the index set $[0,T]$ is of dimension 1.) Using Theorem 2.2.4 of van der Vaart and Wellner (2000) with $\psi(t)=t^4$, it follows that for all $\eta, \delta > 0$,

$$\mathbb{E}\left\{\sup_{d_1((s,t),(s',t')\leq\delta} |A_{1,N}(s,t) - A_{1,N}(s',t')|^4\right\} \leq C\left(\int_0^\eta \psi^{-1}(N(x,d_1))dx + \delta\psi^{-1}(N^2(\eta,d_1))\right)^4$$
$$\leq C\left(\eta^{1-0.5/\beta} + \delta\eta^{-1/\beta}\right)^4.$$

The upper bound above can be made arbitrarily small by varying η first and δ next since $\beta > 0.5$. Hence, with Markov's inequality, we deduce that the processes $A_{1,N}$ are tight in $C([0,T]^2)$.

The bivariate processes $(A_{2,N})_{N\geq 1}$ are sub-Gaussian for the same reasons as the univariate processes $N^{-1/2}\sum_{k\in U}(I_k/\pi_k)\,\tilde{\epsilon}_k$ are in the proof of Theorem 1, namely the independence and multivariate normality of the error vectors $(\epsilon_{k1},\ldots,\epsilon_{kd})'$ and the boundedness of the sample membership indicators I_k for $k\in U_N$. Therefore, although the covering number $N(x,d_2)$ grows to $O(x^{-2/\beta})$ in dimension 2, with d_2 being the L^2-norm on $[0,T]^2$, this does not affect significantly the integral upper bound $\int_0^\infty \sqrt{\log(N(x,d_2)}dx$ in a maximal inequality like (25). As a consequence, one obtains the tightness of $(A_{2,N})$ in $C([0,T]^2)$.

To study the term $A_{3,N}(s,t)$ in (32), we start with the following bound:

$$\begin{aligned}\left|A_{3,N}(s,t)\right| &\leq \frac{1}{N}\sum_{k,l}\frac{|\Delta_{kl}|}{\pi_k\pi_l}\frac{I_kI_l}{\pi_{kl}}\frac{\tilde{\epsilon}_k^2(s)+\tilde{\epsilon}_l^2(t)}{2}\\ &= \frac{1}{N}\sum_k\left(\sum_l\frac{|\Delta_{kl}|}{2\pi_l}\frac{I_l}{\pi_{kl}}\right)\frac{I_k}{\pi_k}\tilde{\epsilon}_k^2(s) + \frac{1}{N}\sum_l\left(\sum_k\frac{|\Delta_{kl}|}{2\pi_k}\frac{I_k}{\pi_{kl}}\right)\frac{I_l}{\pi_l}\tilde{\epsilon}_l^2(t)\\ &\leq \frac{C}{N}\sum_k\tilde{\epsilon}_k^2(s) + \frac{C}{N}\sum_l\tilde{\epsilon}_l^2(t).\end{aligned}$$

The two-dimensional study is thus reduced to an easier one-dimensional problem.

To apply the Corollary 2.2.5 of van der Vaart and Wellner (2000), we consider the function $\psi(t)=t^m$ and the pseudo-metric $d_3^m(s,t) = \mathbb{E}|N^{-1}\sum_k(\tilde{\epsilon}_k^2(s)-\tilde{\epsilon}_k^2(t))|^m$, where $m\geq 1$ is an arbitrary integer. We have that

$$\mathbb{E}\left\{\sup_{s,t\in[0,T]}\left|\frac{1}{N}\sum_k\left(\tilde{\epsilon}_k^2(s)-\tilde{\epsilon}_k^2(t)\right)\right|^m\right\} \leq C\left(\int_0^{D_T}(N(x,d_3))^{1/m}dx\right)^m \qquad (39)$$

where $D_T = \sup_{s,t\in[0,T]} d_3(s,t)$ is the diameter of $[0,T]$ for d_3. Using the classical inequality, $|\sum_{k=1}^n a_k|^m \leq n^{m-1}\sum_{k=1}^n |a_k|^m$, for $m>1$ and arbitrary real number $a_1,\ldots,a_n$, we get, with the Cauchy-Schwarz inequality and the moment properties of Gaussian random vectors, that

$$
\begin{aligned}
d_3^m(s,t) &\le \frac{1}{N}\sum_k \mathbb{E}\left|\tilde{\epsilon}_k^2(s)-\tilde{\epsilon}_k^2(t)\right|^m \\
&\le \frac{1}{N}\sum_k \left\{\mathbb{E}\left|\tilde{\epsilon}_k(s)-\tilde{\epsilon}_k(t)\right|^{2m}\right\}^{1/2}\left\{\mathbb{E}\left|\tilde{\epsilon}_k(s)+\tilde{\epsilon}_k(t)\right|^{2m}\right\}^{1/2} \\
&\le \frac{C_m}{N}\sum_k \|W(s)-W(t)\|_{\mathbf{V}_N}^m \|W(s)+W(t)\|_{\mathbf{V}_N}^m \\
&\le \frac{C_m'}{(dh)^m}\left(\frac{|s-t|}{h}\wedge 1\right)^m, \qquad (40)
\end{aligned}
$$

where $\|\mathbf{x}\|_{\mathbf{V}_N}=(\mathbf{x}'\mathbf{V}_N\mathbf{x})^{1/2}$ and C_m and C_m' are constants that only depends on m.

We deduce from (40) that the diameter D_T is at most of order $1/(dh)$ and that for all $0<x\le 1/(dh)$, the covering number $N(x,d_3)$ is of order $1/(xdh^2)$. Hence the integral bound in (39) is of order $\int_0^{1/(dh)}(dh^2x)^{-1/m}dx \le C(dh^2)^{-1/m}(dh)^{(1-1/m)}) = C/(dh)^{1+1/m}$. Therefore, if $dh^{1+\alpha}\to\infty$ for some $\alpha>0$, the sequence $(N^{-1}\sum_k(\tilde{\epsilon}_k^2))_{N\ge 1}$ tends uniformly to zero in probability which concludes the study of the term $(A_{3,N})_{N\ge 1}$ and the proof. □

Proof of Theorem 3.

We show here the weak convergence of $(\widehat{G}_N)$ to G in $C([0,T])$ conditionally on $\widehat{\gamma}_N$. This convergence, together with the uniform convergence of $\widehat{\gamma}_N$ to γ presented in Theorem 2, is stronger than the result of Theorem 3 required to build simultaneous confidence bands.

First, the finite-dimensional convergence of $(\widehat{G}_N)$ to G conditionally on $\widehat{\gamma}_N$ is a trivial consequence of Theorem 2.

Second, we show the tightness of $(\widehat{G}_N)$ in $C([0,T])$ (conditionally on $\widehat{\gamma}_N$) similarly to the study of $(A_{3,N})$ in the proof of Theorem 2. We start by considering the random pseudo-metric $\hat{d}_\gamma^m(s,t)=\mathbb{E}[(\widehat{G}_N(s)-\widehat{G}_N(t))^m \,|\, \widehat{\gamma}_N]$, where $m\ge 1$ is an arbitrary integer. By the moment properties of Gaussian random variables, it holds that

$$
\begin{aligned}
\hat{d}_\gamma^m(s,t) &= C_m\left[\frac{1}{N}\sum_{k,l\in U}\frac{\Delta_{kl}}{\pi_{kl}}\frac{I_kI_l}{\pi_k\pi_l}\left(\widehat{X}_k(s)-\widehat{X}_k(t)\right)\left(\widehat{X}_l(s)-\widehat{X}_l(t)\right)\right]^{m/2} \\
&\le C_m\left[\frac{1}{N}\sum_{k,l\in U}\frac{|\Delta_{kl}|}{\pi_{kl}}\frac{I_kI_l}{\pi_k\pi_l}\left(\widehat{X}_k(s)-\widehat{X}_k(t)\right)^2\right]^{m/2} \\
&\le C_m\left[\frac{2}{N}\sum_{k,l\in U}\frac{|\Delta_{kl}|}{\pi_{kl}}\frac{I_kI_l}{\pi_k\pi_l}\left(\tilde{X}_k(s)-\tilde{X}_k(t)\right)^2+\frac{2}{N}\sum_{k,l\in U}\frac{|\Delta_{kl}|}{\pi_{kl}}\frac{I_kI_l}{\pi_k\pi_l}\left(\tilde{\epsilon}_k(s)-\tilde{\epsilon}_k(t)\right)^2\right]^{m/2} \\
&\le \frac{C_m}{2}\left[\frac{1}{N}\sum_k\left(\tilde{X}_k(s)-\tilde{X}_k(t)\right)^2\right]^{m/2}+\frac{C_m}{2}\left[\frac{1}{N}\sum_k\left(\tilde{\epsilon}_k(s)-\tilde{\epsilon}_k(t)\right)^2\right]^{m/2}. \qquad (41)
\end{aligned}
$$

Clearly, the first sum in the right-handside of (41) is dominated by $|s-t|^{m\beta}$ thanks to (A2) and the approximation properties of local linear smoothers. The second sum can be viewed as a random quadratic form. Introducing the square root $\mathbf{V}_N^{1/2}$ of $\mathbf{V}_N$, we note that $\boldsymbol{\epsilon}_k = \mathbf{V}_N^{1/2}\boldsymbol{Z}_k$, with equality in distribution, for $k=1,\dots,N$, where the $\boldsymbol{Z}_k$ are i.i.d centered d-dimensional Gaussian vectors with identity covariance matrix.

Thus,

$$\begin{aligned}\frac{1}{N}\sum_k (\tilde{\epsilon}_k(s)-\tilde{\epsilon}_k(t))^2 &= (W(s)-W(t))'\left(\frac{1}{N}\sum_k \boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k'\right)(W(s)-W(t)) \\ &\leq \|W(s)-W(t)\|^2 \left\|\frac{1}{N}\sum_k \boldsymbol{\epsilon}_k\boldsymbol{\epsilon}_k'\right\| \\ &\leq \|W(s)-W(t)\|^2 \|\mathbf{V}_N\| \left\|\frac{1}{N}\sum_k \mathbf{Z}_k\mathbf{Z}_k'\right\| \qquad (42)\end{aligned}$$

Now, the vector norm $\|W(s)-W(t)\|^2$ has already been studied in (26) and the sequence $(\|\mathbf{V}_N\|)$ is bounded by (A4). The remaining matrix norm in (42) is smaller than the largest eigenvalue, up to a factor N^{-1}, of a d-variate Wishart matrix with N degrees of freedom. By (A3) it holds that $d=o(N/\log\log N)$ and one can apply Theorem 3.1 in Fey *et al.* (2008), which states that for any fixed $\alpha\geq 1$,

$$\lim_{N\to\infty} -\frac{1}{N}\log\mathbb{P}\left(\left\|\frac{1}{N}\sum_k \mathbf{Z}_k\mathbf{Z}_k'\right\|\geq\alpha\right) = \frac{1}{2}\left(\alpha-1-\log\alpha\right). \qquad (43)$$

A immediate consequence of (43) is that $\left\|\frac{1}{N}\sum_k \mathbf{Z}_k\mathbf{Z}_k'\right\|$ remains almost surely bounded as $N\to\infty$. Note that the same result holds if instead of (A3), (d/N) remains bounded away from zero and infinity, thanks to the pioneer work of Geman (1980) on the norm of random matrices. Thus, there exists a deterministic constant $C\in(0,\infty)$ such that

$$\hat{d}_\gamma^m(s,t)\leq C|s-t|^{m\beta}+\frac{C}{(dh)^{m/2}}\left(\frac{|s-t|}{h}\wedge 1\right)^m \qquad (44)$$

for all $s,t\in[0,T]$, with probability tending to 1 as $N\to\infty$. Similarly to the previous entropy calculations, one can show that there exists a constant $C\in(0,\infty)$ such that $N(x,\hat{d}_\gamma)\leq C(x^{-1/\beta}+(dh^3)^{-1/2}x^{-1})$ for all $x\leq(dh)^{-1}$ with probability tending to 1 as $N\to\infty$. Applying the maximal inequality of van der Vaart and Wellner (2000) (Th. 2.2.4) to the conditional increments of $\widehat{G}_N$, with $\phi(t)=t^m$ (usual L^m-norm), one finds a covering integral $\int_0^{1/(ph)}(N(x,\hat{d}_\gamma))^{1/2}dx$ of the order of $(dh)^{1/(m\beta)-1}+(dh^3)^{-1/(2m)}(dh)^{1/m-1}$. Hence the covering integral tends to zero in probability, provided that $h\to 0$ and $dh^{\frac{1+1/(2m)}{1-1/(2m)}}\to\infty$ as $N\to\infty$. Obvisouly, the latter condition on h holds for some integer $m\geq 1$ if $dh^{1+\alpha}\to\infty$ for some real $\alpha>0$. Under this condition, the sequence $(\widehat{G}_N)$ is tight in $C([0,T])$ and therefore converges to G. □

Acknowledgements. Etienne Josserand thanks the Conseil Régional de Bourgogne for its financial support (Faber PhD grant).

Bibliography

Adler, R. J. and Taylor, J. (2007). *Random Fields and Geometry.* Springer, New York.

Bickel, P. J. and Freedman, D. A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *Annals of Statistics*, **12**, 470-482.

Breidt, F. J. and Opsomer, J. D. (2000). Local polynomial regression estimators in survey sampling. *Annals of Statistics*, **28**, 1026-1053.

Bunea, F., Ivanescu, A. and M. Wegkamp (2011). Adaptive inference for the mean of a Gaussian process in functional data. *J. Roy. Statist. Soc. Ser. B*, to appear.

Cardot, H., Chaouch, M., Goga, C. and C. Labruère (2010a). Properties of Design-Based Functional Principal Components Analysis, *J. Statist. Planning and Inference*, **140**, 75-91.

Cardot, H., Dessertaine, A., Josserand, E. (2010b). Semiparametric models with functional responses in survey sampling setting : model assisted estimation of electricity consumption curves. *Compstat 2010*, Eds Lechevallier, Y. and Saporta, G. Physica-Verlag, Springer, 411-420.

Cardot, H. and Josserand, E. (2011). Horvitz-Thompson estimators for functional data: asymptotic confidence bands and optimal allocation for stratified sampling. *Biometrika*, **98**, 107-118.

Chiky, R. and Hébrail, G. (2008). Summarizing distributed data streams for storage in data warehouses. In DaWaK 2008, I-Y. Song, J. Eder and T. M. Nguyen, Eds. *Lecture Notes in Computer Science*, Springer, 65-74.

Claeskens, G. and van Keilegom, I. (2003). Bootstrap confidence bands for regression curves and their derivatives. *Ann. Statist.*, **31**, 18521884.

Cuevas, A., Febrero, M. and Fraiman, R. (2006). On the use of the bootstrap for estimating functions with functional data. *Computational Statistics and Data Analysis*, **51**, 1063-1074.

Degras, D. (2009). Nonparametric estimation of a trend based upon sampled continuous processes. *C. R. Math. Acad. Sci. Paris*, **347**, 191-194.

Degras, D. (2010). Simultaneous confidence bands for nonparametric regression with functional data. Accepted for publication at *Statistica Sinica.* http://arxiv.org/abs/0908.1980

Erdös, P. and Rényi, A. (1959). On the central limit theorem for samples from a finite population. *Publ. Math. Inst. Hungar. Acad. Sci.*, **4**, 49-61.

Eubank, R.L. and Speckman P.L. (1993). Confidence Bands in Nonparametric Regression. *J. Amer. Statist. Assoc.*, **88**, 1287-1301.

Faraway, J.T. (1997). Regression analysis for a functional response. *Technometrics* **39**, 254-261.

Fey, A., van der Hofstad, R. and Klok, M. (2008). Large deviations for eigenvalues of sample covariance matrices, with applications to mobile communication systems. *Adv. in Appl. Probab.* **40**, 1048-1071.

Fuller, W.A. (2009). *Sampling Statistics.* John Wiley and Sons.

Geman, S. (1980). A Limit Theorem for the Norm of Random Matrices. *Ann. Probab.*, **8**, 252-261

Hart, J. D. and Wehrly, T. E. (1993). Consistency of cross-validation when the data are curves. *Stoch. Proces. Applic.*, **45**, 351361.

Isaki, C.T. and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *J. Am. Statist. Ass.*, **77**, 89-96.

Krivobokova, T., Kneib, T. and G. Claeskens (2010). Simultaneous confidence bands for penalized spline estimators. *J. Am. Statist. Ass.*, **105**, 852-863.

Landau, H. and Shepp, L.A. (1970), On the supremum of a Gaussian process. *Sankhyã*, **32**, 369-378

Mas A. (2007). Testing for the mean of random curves: a penalization approach. *Statistical Inference for Stochastic Processes,* **10**, 147-163

Opsomer, J. D. and Miller, C. P. (2005). Selecting the amount of smoothing in nonparametric regression estimation for complex surveys. *J. Nonparametric Statistics,* **17**, 593-611.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B*, **53**, 233-243.

Robinson, P. M. and Särndal, C. E. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhya : The Indian Journal of Statistics*, **45**, 240-248.

Sun, J. and Loader, C.R. (1994). Simultaneous confidence bands for linear regression and smoothing. *Annals of Statistics*, **22**, 1328-1345.

Särndal, C. E., Swensson, B. and J. Wretman, J. (1992). *Model Assisted Survey Sampling.* Springer-Verlag.

Tsybakov, A.B. (2009). *Introduction to Nonparametric Estimation.* Springer, New York.

van der Vaart, A. D. and Wellner, J. A. (2000). *Weak Convergence and Empirical Processes. With Applications to Statistics.* Springer-Verlag, New York.

Zhang, J. T. and Chen, J. (2007). Statistical inferences for functional data. *Annals of Statistics*, **35**, 1052-1079.

		SRSWR				Stratified sampling			
δ	h	Mean	1Q	Median	3Q	Mean	1Q	Median	3Q
5%	lin	17.65	3.078	8.729	23.50	4.221	1.444	2.791	5.594
	h_{cv}	17.65	3.066	8.710	23.51	6.493	3.605	5.356	8.028
	h_{wcv}	17.65	3.066	8.710	23.51	4.221	1.448	2.781	5.555
	h_{oracle}	17.65	3.068	8.725	23.50	4.220	1.446	2.778	5.571
	$\widehat{\mu}_s$	17.60	3.011	8.698	23.36	4.174	1.378	2.758	5.548
10%	lin	17.18	3.226	9.019	22.20	4.335	1.535	3.040	5.675
	h_{cv}	17.17	3.201	8.975	22.26	6.688	3.699	5.354	8.091
	h_{wcv}	17.17	3.201	8.975	22.26	4.342	1.613	3.068	5.637
	h_{oracle}	17.17	3.209	8.969	22.23	4.330	1.573	3.053	5.627
	$\widehat{\mu}_s$	17.00	3.092	8.780	22.11	4.136	1.359	2.835	5.486
15%	lin	18.08	3.616	9.589	23.58	4.263	1.755	3.050	5.614
	h_{cv}	18.06	3.633	9.557	23.44	6.473	3.641	5.336	8.064
	h_{wcv}	18.06	3.633	9.557	23.44	4.238	1.703	3.041	5.682
	h_{oracle}	18.06	3.634	9.568	23.43	4.225	1.702	3.012	5.631
	$\widehat{\mu}_s$	17.69	3.282	9.263	22.77	3.812	1.307	2.625	5.131
20%	lin	16.98	3.722	9.222	21.31	4.870	2.187	3.755	6.047
	h_{cv}	16.91	3.657	9.226	21.28	7.025	4.022	5.878	8.838
	h_{wcv}	16.91	3.657	9.226	21.28	4.791	2.110	3.683	6.014
	h_{oracle}	16.90	3.668	9.221	21.29	4.779	2.110	3.681	6.000
	$\widehat{\mu}_s$	16.27	3.040	8.606	20.71	4.086	1.373	2.964	5.283
25%	lin	17.69	3.940	8.989	21.52	5.257	2.625	4.148	6.535
	h_{cv}	17.53	3.826	8.755	21.53	6.982	4.287	5.829	8.469
	h_{wcv}	17.53	3.826	8.755	21.53	5.017	2.388	3.893	6.331
	h_{oracle}	17.52	3.806	8.778	21.52	5.007	2.369	3.883	6.269
	$\widehat{\mu}_s$	16.58	2.847	7.870	20.01	4.069	1.457	2.944	5.278

Table 2: Estimation errors according to $R(\widehat{\mu})$ for different noise levels and bandwidth values, with $d = 200$ time instants. Units are selected by simple random sampling without replacements (SRSWR) or stratified sampling.

		SRSWR				Stratified sampling			
δ	h	Mean	1Q	Median	3Q	Mean	1Q	Median	3Q
5%	lin	18.03	3.388	9.243	23.27	4.049	1.446	2.858	5.353
	h_{cv}	18.02	3.384	9.257	23.34	6.092	3.244	4.868	7.555
	h_{wcv}	18.02	3.384	9.257	23.34	4.047	1.449	2.821	5.398
	h_{oracle}	18.02	3.387	9.269	23.32	4.043	1.433	2.828	5.388
	$\widehat{\mu}_s$	17.98	3.353	9.199	23.17	4.000	1.388	2.809	5.294
10%	lin	16.97	3.084	8.058	21.46	4.294	1.683	3.208	5.797
	h_{cv}	16.93	2.979	7.916	21.45	6.207	3.308	5.137	7.800
	h_{wcv}	16.93	2.979	7.916	21.45	4.233	1.577	3.123	5.741
	h_{oracle}	16.93	2.970	7.914	21.44	4.229	1.579	3.118	5.703
	$\widehat{\mu}_s$	16.81	2.876	7.811	21.45	4.099	1.512	3.006	5.608
15%	lin	19.03	3.761	10.09	24.80	4.528	1.772	3.367	5.994
	h_{cv}	18.87	3.642	9.899	24.54	6.259	3.446	5.327	7.829
	h_{wcv}	18.87	3.642	9.899	24.54	4.335	1.630	3.188	5.828
	h_{oracle}	18.87	3.642	9.888	24.52	4.330	1.612	3.178	5.826
	$\widehat{\mu}_s$	18.61	3.414	9.665	24.24	4.080	1.340	2.918	5.538
20%	lin	17.06	3.635	8.545	22.95	4.749	2.128	3.643	6.144
	h_{cv}	16.69	3.288	8.060	22.82	6.362	3.489	5.205	8.009
	h_{wcv}	16.69	3.288	8.060	22.82	4.353	1.755	3.231	5.675
	h_{oracle}	16.69	3.267	8.044	22.77	4.347	1.734	3.216	5.694
	$\widehat{\mu}_s$	16.35	2.993	7.860	22.13	3.960	1.333	2.867	5.414
25%	lin	18.16	3.885	9.427	22.86	5.254	2.845	4.236	6.572
	h_{cv}	17.55	3.303	8.889	22.09	6.452	3.770	5.372	8.114
	h_{wcv}	17.55	3.303	8.889	22.09	4.566	2.121	3.486	5.809
	h_{oracle}	17.55	3.282	8.898	22.09	4.561	2.107	3.477	5.812
	$\widehat{\mu}_s$	17.04	2.750	8.383	21.87	4.043	1.602	3.018	5.306

Table 3: Estimation errors according to $R(\widehat{\mu})$ for different noise levels and bandwidth values, with $d = 400$ time instants. Units are selected by simple random sampling without replacements (SRSWR) or stratified sampling.

		SRSWR				Stratified sampling			
δ	h	Mean	1Q	Median	3Q	Mean	1Q	Median	3Q
5%	lin	0.044	0.041	0.044	0.047	0.049	0.046	0.049	0.053
	h_{cv}	0.044	0.041	0.044	0.048	2.520	2.083	2.852	3.032
	h_{wcv}	0.044	0.041	0.044	0.048	0.058	0.054	0.058	0.062
	h_{oracle}	0.044	0.041	0.044	0.047	0.049	0.045	0.049	0.052
10%	lin	0.175	0.163	0.175	0.186	0.196	0.181	0.194	0.208
	h_{cv}	0.170	0.158	0.170	0.182	2.626	2.162	2.902	3.110
	h_{wcv}	0.170	0.158	0.170	0.182	0.202	0.188	0.201	0.216
	h_{oracle}	0.169	0.156	0.168	0.180	0.188	0.174	0.187	0.200
15%	lin	0.396	0.366	0.394	0.424	0.443	0.411	0.440	0.474
	h_{cv}	0.368	0.342	0.366	0.394	2.743	2.264	2.972	3.206
	h_{wcv}	0.368	0.342	0.366	0.394	0.417	0.388	0.413	0.446
	h_{oracle}	0.365	0.339	0.364	0.391	0.404	0.378	0.403	0.432
20%	lin	0.706	0.654	0.702	0.754	0.784	0.724	0.779	0.837
	h_{cv}	0.628	0.58	0.626	0.672	3.002	2.441	3.122	3.417
	h_{wcv}	0.628	0.580	0.626	0.672	0.699	0.646	0.698	0.748
	h_{oracle}	0.622	0.575	0.620	0.667	0.682	0.630	0.679	0.731
25%	lin	1.087	1.011	1.080	1.156	1.214	1.134	1.210	1.287
	h_{bcv}	0.905	0.837	0.901	0.970	3.155	2.638	3.260	3.602
	h_{wcv}	0.905	0.837	0.901	0.970	1.009	0.936	1.004	1.076
	h_{oracle}	0.898	0.830	0.894	0.962	0.990	0.919	0.988	1.055

Table 4: Estimation errors according to $L(\widehat{\mu})$ for different noise levels and bandwidth values, with $d = 200$ time instants. Units are selected by simple random sampling without replacements (SRSWR) or stratified sampling.

		SRSWR				Stratified sampling			
δ	h	Mean	1Q	Median	3Q	Mean	1Q	Median	3Q
5%	lin	0.044	0.042	0.044	0.047	0.049	0.047	0.049	0.051
	h_{cv}	0.040	0.038	0.040	0.042	2.231	1.612	1.917	2.806
	h_{wcv}	0.040	0.038	0.040	0.042	0.052	0.049	0.052	0.055
	h_{oracle}	0.040	0.038	0.040	0.042	0.044	0.041	0.044	0.046
10%	lin	0.175	0.166	0.175	0.184	0.196	0.186	0.196	0.205
	h_{cv}	0.128	0.120	0.127	0.135	2.258	1.648	1.969	2.844
	h_{wcv}	0.128	0.120	0.127	0.135	0.145	0.135	0.144	0.154
	h_{oracle}	0.127	0.119	0.127	0.134	0.138	0.13	0.137	0.146
15%	lin	0.397	0.377	0.397	0.416	0.444	0.420	0.445	0.466
	h_{cv}	0.233	0.217	0.232	0.247	2.293	1.682	1.991	2.901
	h_{wcv}	0.233	0.217	0.232	0.247	0.257	0.238	0.255	0.272
	h_{oracle}	0.231	0.215	0.230	0.246	0.250	0.234	0.250	0.266
20%	lin	0.708	0.672	0.706	0.744	0.79	0.749	0.791	0.829
	h_{cv}	0.351	0.327	0.350	0.374	2.442	1.763	2.152	3.107
	h_{wcv}	0.351	0.327	0.350	0.374	0.388	0.359	0.384	0.413
	h_{oracle}	0.349	0.326	0.348	0.373	0.381	0.355	0.380	0.406
25%	lin	1.089	1.030	1.087	1.142	1.219	1.155	1.212	1.280
	h_{cv}	0.498	0.462	0.495	0.535	2.591	1.932	2.344	3.254
	h_{wcv}	0.498	0.462	0.495	0.535	0.552	0.509	0.549	0.594
	h_{oracle}	0.497	0.460	0.494	0.533	0.547	0.505	0.545	0.586

Table 5: Estimation errors according to $L(\widehat{\mu})$ for different noise levels and bandwidth values, with $d = 400$ time instants. Units are selected by simple random sampling without replacements (SRSWR) or stratified sampling.

		SRSWR					Stratified sampling				
δ	h	$1-\widehat{\alpha}$	Mean	1Q	Median	3Q	$1-\widehat{\alpha}$	Mean	1Q	Median	3Q
5%	lin	97.2	10.91	10.74	10.90	11.07	98.1	5.946	5.868	5.946	6.019
	h_{cv}	97.3	10.89	10.73	10.89	11.06	47.5	5.681	5.600	5.680	5.760
	h_{wcv}	97.3	10.89	10.73	10.89	11.06	97.5	5.918	5.840	5.913	6.000
	h_{oracle}	97.2	10.9	10.72	10.90	11.07	98.0	5.941	5.862	5.942	6.018
	$\widehat{\mu}_s$	97.3	10.54	10.36	10.54	10.70	98.2	5.593	5.513	5.596	5.671
10%	lin	98.1	11.43	11.25	11.42	11.60	97.1	6.455	6.374	6.458	6.531
	h_{cv}	98.0	11.38	11.20	11.37	11.55	50.2	5.903	5.819	5.902	5.980
	h_{wcv}	98.0	11.38	11.20	11.37	11.55	96.4	6.358	6.277	6.355	6.433
	h_{oracle}	98.1	11.39	11.22	11.39	11.56	96.7	6.414	6.335	6.416	6.496
	$\widehat{\mu}_s$	97.7	10.54	10.36	10.53	10.72	97.1	5.597	5.515	5.598	5.671
15%	lin	98.0	12.03	11.84	12.03	12.19	98.4	7.024	6.942	7.023	7.104
	h_{cv}	97.8	11.88	11.71	11.89	12.04	51.4	6.161	6.066	6.159	6.252
	h_{wcv}	97.8	11.88	11.71	11.89	12.04	98.2	6.804	6.720	6.799	6.891
	h_{oracle}	97.8	11.89	11.71	11.89	12.06	98.4	6.876	6.782	6.878	6.964
	$\widehat{\mu}_s$	97.3	10.56	10.38	10.56	10.72	98.2	5.598	5.519	5.594	5.679
20%	lin	98.5	12.61	12.44	12.62	12.80	97.5	7.631	7.537	7.628	7.724
	h_{cv}	98.4	12.29	12.12	12.29	12.45	54.0	6.418	6.316	6.410	6.509
	h_{wcv}	98.4	12.29	12.12	12.29	12.45	97.0	7.198	7.105	7.195	7.286
	h_{oracle}	98.3	12.32	12.14	12.31	12.50	97.3	7.306	7.212	7.305	7.393
	$\widehat{\mu}_s$	98.2	10.54	10.38	10.55	10.70	97.0	5.595	5.512	5.597	5.676
25%	lin	97.7	13.23	13.06	13.22	13.41	98.3	8.270	8.185	8.269	8.357
	h_{cv}	97.2	12.66	12.49	12.65	12.83	64.7	6.704	6.603	6.691	6.790
	h_{wcv}	97.2	12.66	12.49	12.65	12.83	97.3	7.563	7.479	7.564	7.645
	h_{oracle}	97.3	12.70	12.50	12.70	12.87	97.5	7.683	7.575	7.678	7.788
	$\widehat{\mu}_s$	97.0	10.53	10.37	10.52	10.70	97.7	5.589	5.514	5.586	5.663

Table 6: Empirical covering levels $1-\widehat{\alpha}$ and confidence band areas for different noise levels and bandwidth values, with $d = 200$ time instants. Units are selected by simple random sampling without replacements (SRSWR) or stratified sampling.

		SRSWR					Stratified sampling				
δ	h	$1-\widehat{\alpha}$	Mean	1Q	Median	3Q	$1-\widehat{\alpha}$	Mean	1Q	Median	3Q
5%	lin	97.4	10.97	10.81	10.97	11.15	97.9	6.027	5.948	6.024	6.106
	h_{cv}	97.5	10.90	10.73	10.91	11.06	48.4	5.640	5.566	5.634	5.717
	h_{wcv}	97.5	10.90	10.73	10.91	11.06	97.6	5.894	5.816	5.889	5.971
	h_{oracle}	97.4	10.92	10.75	10.92	11.09	97.6	5.964	5.883	5.959	6.040
	$\widehat{\mu}_s$	97.3	10.54	10.38	10.54	10.70	97.8	5.597	5.519	5.591	5.675
10%	lin	97.8	11.58	11.41	11.57	11.76	97.8	6.589	6.504	6.590	6.669
	h_{cv}	97.6	11.23	11.06	11.22	11.40	49.5	5.788	5.705	5.786	5.864
	h_{wcv}	97.6	11.23	11.06	11.22	11.40	97.1	6.173	6.090	6.173	6.254
	h_{oracle}	97.7	11.25	11.08	11.24	11.41	97.3	6.257	6.180	6.256	6.333
	$\widehat{\mu}_s$	97.5	10.55	10.39	10.54	10.71	97.5	5.592	5.517	5.592	5.668
15%	lin	97.4	12.23	12.05	12.23	12.40	98.0	7.218	7.130	7.212	7.307
	h_{cv}	96.8	11.50	11.33	11.50	11.67	52.7	5.965	5.880	5.962	6.048
	h_{wcv}	96.8	11.50	11.33	11.50	11.67	97.3	6.453	6.366	6.447	6.533
	h_{oracle}	96.8	11.51	11.34	11.51	11.67	97.7	6.503	6.417	6.497	6.589
	$\widehat{\mu}_s$	96.7	10.55	10.38	10.56	10.72	97.7	5.590	5.510	5.588	5.668
20%	lin	98.2	12.90	12.72	12.91	13.08	98.2	7.891	7.805	7.892	7.972
	h_{cv}	97.7	11.80	11.63	11.80	11.96	55.1	6.153	6.071	6.154	6.235
	h_{wcv}	97.7	11.80	11.63	11.80	11.96	97.4	6.764	6.673	6.759	6.841
	h_{oracle}	97.7	11.79	11.62	11.79	11.96	97.6	6.785	6.700	6.783	6.868
	$\widehat{\mu}_s$	97.9	10.56	10.39	10.56	10.72	97.5	5.598	5.518	5.598	5.672
25	lin	98.0	13.58	13.40	13.58	13.75	98.3	8.587	8.491	8.588	8.676
	h_{cv}	97.5	12.11	11.95	12.10	12.28	58.1	6.344	6.244	6.343	6.437
	h_{wcv}	97.5	12.11	11.95	12.10	12.28	97.6	7.088	6.998	7.081	7.172
	h_{oracle}	97.5	12.12	11.94	12.12	12.29	97.8	7.101	7.011	7.099	7.188
	$\widehat{\mu}_s$	97.4	10.56	10.39	10.55	10.73	97.6	5.592	5.509	5.590	5.668

Table 7: Empirical covering levels $1-\widehat{\alpha}$ and confidence band areas for different noise levels and bandwidth values, with $d = 400$ time instants. Units are selected by simple random sampling without replacements (SRSWR) or stratified sampling.