



HAL
open science

Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data

Hervé Cardot, Camelia Goga, Pauline Lardin

► **To cite this version:**

Hervé Cardot, Camelia Goga, Pauline Lardin. Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data. *Electronic Journal of Statistics*, 2013, 7, pp.562-596. 10.1214/13-EJS779 . hal-00691943v2

HAL Id: hal-00691943

<https://hal.science/hal-00691943v2>

Submitted on 25 Nov 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Uniform convergence and asymptotic confidence bands for model-assisted estimators of the mean of sampled functional data

Hervé CARDOT^(a), Camelia GOGA^(a) and Pauline LARDIN^(a,b)

(a) Université de Bourgogne, Institut de Mathématiques de Bourgogne,
9 av. Alain Savary, 21078 DIJON, FRANCE

(b) EDF, R&D, ICAME-SOAD, 1 av. du Général de Gaulle,
92141 CLAMART, FRANCE

November 19, 2012

Abstract

When the study variable is functional and storage capacities are limited or transmission costs are high, selecting with survey sampling techniques a small fraction of the observations is an interesting alternative to signal compression techniques, particularly when the goal is the estimation of simple quantities such as means or totals. We extend, in this functional framework, model-assisted estimators with linear regression models that can take account of auxiliary variables whose totals over the population are known. We first show, under weak hypotheses on the sampling design and the regularity of the trajectories, that the estimator of the mean function as well as its variance estimator are uniformly consistent. Then, under additional assumptions, we prove a functional central limit theorem and we assess rigorously a fast technique based on simulations of Gaussian processes which is employed to build asymptotic confidence bands. The accuracy of the variance function estimator is evaluated on a real dataset of sampled electricity consumption curves measured every half an hour over a period of one week.

Keywords : calibration, covariance function, functional linear model, GREG, Hájek estimator, Horvitz-Thompson estimator, linear interpolation, survey sampling.

1 Introduction

Survey sampling techniques, which consist in randomly selecting only a part of the elements of a population, are interesting alternatives to signal compression when one has to deal with very large populations of quantities that evolve along time. With the development of automatic sensors such very large datasets of temporal data are not unusual anymore and survey sampling techniques offer a good trade-off between accuracy of the estimators and size of the analyzed data. Examples can be found in different domains such as internet traffic monitoring (see Callado et al. (2009)) or estimation of energy consumption measured by individual smart

meters. Motivated by the estimation of mean consumption electricity profiles measured every half an hour over one week, Cardot and Josserand (2011) have introduced Horvitz-Thompson estimators of the mean function and have shown, under weak hypotheses on the regularity of the functional trajectories and the sampling design, that one gets uniformly convergent estimators. They also prove a functional central limit theorem, in the space of continuous functions, that can, in part, justify the construction of asymptotic confidence bands. More recently, Cardot et al. (2012b) made a comparison, in terms of precision of the mean estimators of electricity load curves and width of the confidence bands, of different sampling approaches that can take auxiliary information into account. One of the conclusions of this empirical study was that very simple strategies based on simple sampling designs (such as simple random sampling without replacement) could be improved much if some well chosen auxiliary information, whose total is known for the whole population, is also taken into account at the estimation stage, with model-assisted estimators. Important variables for the electricity consumption such as temperature or geographical location were not available for these datasets so that only one auxiliary information, the mean past consumption over the previous period, was taken into account. Its correlation with the current consumption is always very high (see Figure 1) so that linear regression models are natural candidates for assisting the Horvitz-Thompson estimator. More generally, one advantage of linear approaches is that they only require the knowledge of the auxiliary variable totals in the population. More sophisticated nonlinear or nonparametric approaches would have required to know the values of the auxiliary variables for all the elements of the population.

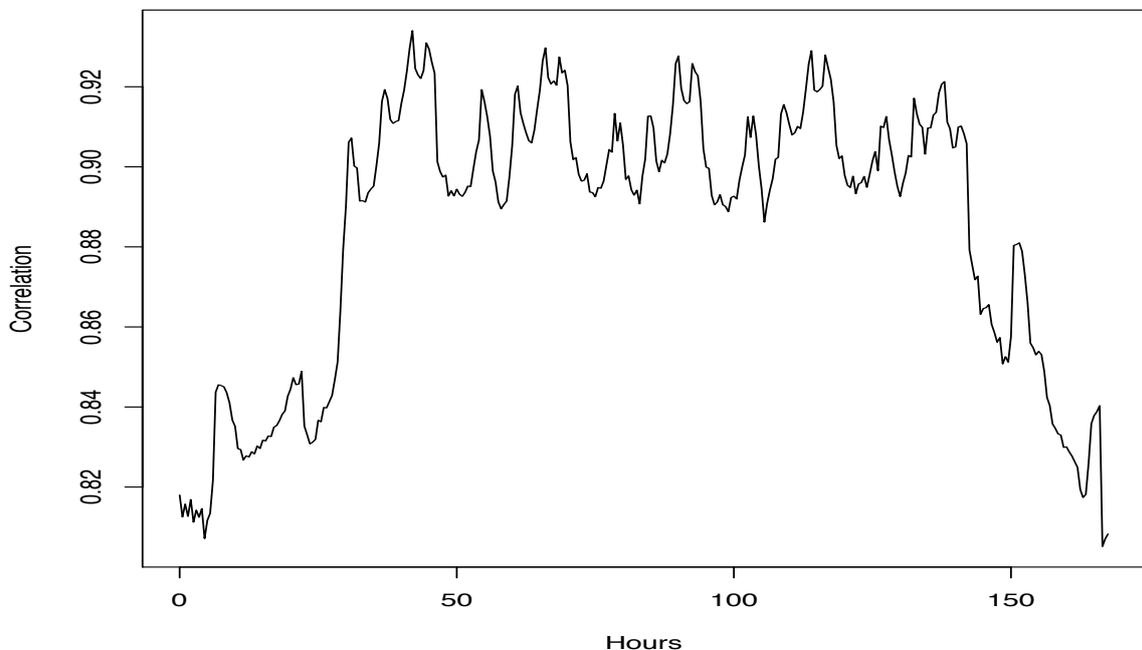


Figure 1: Correlation between the current consumption at each instant t of the week under study and the total past consumption of the week before.

Thus, we focus in this paper on linear relationships between the set of auxiliary variables and the response at each instant t of the current period. The regression coefficients vary in time (see Faraway (1997) or Ramsay and Silverman (2005)) so that the model-assisted estimator can be seen as a direct extension, to a functional or varying-time context, of the generalized regression (GREG) estimators studied in Robinson and Särndal (1983) and Särndal et al. (1992). Note also that from another point of view, the model-assisted estimator can be obtained using a calibration technique (Deville and Särndal (1992)).

Confidence bands are then built using a simulation technique developed in Faraway (1997), Cuevas et al. (2006) and Degras (2011). We first estimate the covariance function of the mean estimator and then, assuming asymptotic normality, perform simulations of a centered Gaussian process whose covariance function is the covariance function estimated at the previous step. We can, this way, obtain an approximation to the law of the "sup" and deduce confidence bands for the mean trajectory. In a recent work, Cardot et al. (2012a) have given a rigorous mathematical justification of this technique for sampled functional data and Horvitz-Thompson estimators for the mean. The required theoretical ingredients that can justify such a procedure are the functional central limit theorem for the mean estimator, in the space of continuous functions equipped with the sup-norm, as well as a uniformly consistent estimator of the variance function.

The aim of this paper is to study the asymptotic properties of model-assisted estimators and to show that we obtain, under classical assumptions, a uniformly consistent estimator of the mean as well as of its variance function. One additional difficulty is that, for model-assisted estimators, the variance function cannot be derived exactly and we can only have asymptotic approximations. Then, we deduce that the confidence bands built via simulations have asymptotically the desired coverage. In Section 2, we introduce notations and we suggest a slight modification of the model-assisted estimators which permits control of the variance of the regression coefficient estimator. Under classical assumptions on the sampling design and on the regularity of the trajectories, we state, in Section 3, the uniform convergence of the model assisted-estimators to the mean function. Under additional assumptions on the design we also prove that we can get a consistent estimator of the covariance function and a functional central limit theorem that can justify rigorously that the confidence bands built with the procedure based on Gaussian process simulations attain asymptotically the desired level of confidence. In Section 4, we assess the precision of the variance estimator on the real dataset consisting of electricity consumption curves studied in Cardot et al. (2012b) and observe that, in our context, the approximation error is negligible compared to the sampling error. A brief discussion about possible extensions and future investigation is proposed in Section 5. All the proofs are gathered in an Appendix.

2 Notations and estimators

2.1 The Horvitz Thompson estimator for functional data

Let us consider a finite population $U_N = \{1, \dots, N\}$ of size N supposed to be known, and suppose that, for each unit k of the population U_N , we can observe a deterministic curve $Y_k = (Y_k(t))_{t \in [0, T]}$. The target is the mean trajectory $\mu_N(t)$, $t \in [0, T]$, defined as follows:

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

We consider a sample s , with size n , drawn from U_N according to a fixed-size sampling design $p_N(s)$, where $p_N(s)$ is the probability of drawing the sample s . For simplicity of notations, the subscript N is omitted when there is no ambiguity. We suppose that the first and second order inclusion probabilities satisfy $\pi_k = \mathbb{P}(k \in s) > 0$, for all $k \in U$, and $\pi_{kl} = \mathbb{P}(k \& l \in s) > 0$ for all $k, l \in U_N$, $k \neq l$. Without auxiliary information, the population mean curve $\mu(t)$ is often estimated by the Horvitz-Thompson estimator, defined as follows for $t \in [0, T]$,

$$\hat{\mu}(t) = \frac{1}{N} \sum_{k \in s} \frac{Y_k(t)}{\pi_k} = \frac{1}{N} \sum_{k \in U} \frac{Y_k(t)}{\pi_k} \mathbb{1}_k, \quad (2)$$

where $\mathbb{1}_k$ is the sample membership indicator, $\mathbb{1}_k = 1$ if $k \in s$ and $\mathbb{1}_k = 0$ otherwise. For each $t \in [0, T]$, the estimator $\hat{\mu}(t)$ is design-unbiased for $\mu(t)$, *i.e.* $\mathbb{E}_p(\hat{\mu}(t)) = \mu(t)$, where $\mathbb{E}_p[\cdot]$ denotes expectation with respect to the sampling design.

The Horvitz-Thompson covariance function of $\hat{\mu}$ between two instants r and t , computed with respect to the sampling design, is defined as follows

$$\text{Cov}_p(\hat{\mu}(r), \hat{\mu}(t)) = \frac{1}{N^2} \sum_{k \in U} \sum_{l \in U} (\pi_{kl} - \pi_k \pi_l) \frac{Y_k(r)}{\pi_k} \cdot \frac{Y_l(t)}{\pi_l} \quad r, t \in [0, T]. \quad (3)$$

Note that for $r = t$, we obtain the Horvitz-Thompson variance function.

2.2 The mean curve estimator assisted by a functional linear model

Let us suppose now that for each unit $k \in U$ we can also observe p real variables, X_1, \dots, X_p , and let us denote by $\mathbf{x}_k = (x_{k1}, \dots, x_{kp})'$, the value of the auxiliary variable vector for each unit k in the population. We introduce an estimator based on a linear regression model that can use these variables in order to improve the accuracy of $\hat{\mu}$. By analogy to the real case (see *e.g.* Särndal et al. (1992)) we suppose that the relationship between the functional variable of interest and the auxiliary variables is modeled by the superpopulation model ξ defined as follows:

$$\xi : \quad Y_k(t) = \mathbf{x}'_k \boldsymbol{\beta}(t) + \epsilon_{kt}, \quad t \in [0, T] \quad (4)$$

where $\boldsymbol{\beta}(t) = (\beta_1(t), \dots, \beta_p(t))'$ is the vector of functional regression coefficients, ϵ_{kt} are independent (across units) and centered continuous time processes, $\mathbb{E}_\xi(\epsilon_{kt}) = 0$, with covariance

function $\text{Cov}_\xi(\epsilon_{kt}, \epsilon_{kr}) = \Gamma(t, r)$, for $(t, r) \in [0, T] \times [0, T]$. This model is a direct extension to several variables of the functional linear model proposed by Faraway (1997).

If \mathbf{x}_k and Y_k are known for all units $k \in U$ and if the matrix $\mathbf{G} = \frac{1}{N} \sum_{k \in U} \mathbf{x}_k \mathbf{x}'_k$ is invertible, it is possible, under the model ξ , to estimate $\boldsymbol{\beta}(t)$ by $\tilde{\boldsymbol{\beta}}(t) = \mathbf{G}^{-1} \frac{1}{N} \sum_{k \in U} \mathbf{x}_k Y_k(t)$, the ordinary least squares estimator. Then, the mean curve $\mu(t)$ can be estimated by the generalized difference estimator (see Särndal et al. (1992), Chapter 6) defined as follows for all $t \in [0, T]$,

$$\begin{aligned} \tilde{\mu}(t) &= \frac{1}{N} \sum_{k \in U} \mathbf{x}'_k \tilde{\boldsymbol{\beta}}(t) - \frac{1}{N} \sum_{k \in s} \frac{\mathbf{x}'_k \tilde{\boldsymbol{\beta}}(t) - Y_k(t)}{\pi_k} \\ &= \frac{1}{N} \sum_{k \in U} \tilde{Y}_k(t) - \frac{1}{N} \sum_{k \in s} \frac{\tilde{Y}_k(t) - Y_k(t)}{\pi_k}, \end{aligned} \quad (5)$$

where $\tilde{Y}_k(t) = \mathbf{x}'_k \tilde{\boldsymbol{\beta}}(t)$.

In practice, we do not know Y_k except for $k \in s$, and it is not possible to compute $\tilde{\boldsymbol{\beta}}(t)$. An estimator of $\mu(t)$ is obtained by substituting each total in $\tilde{\boldsymbol{\beta}}(t)$ by its Horvitz-Thompson estimator. Thus, if the matrix $\hat{\mathbf{G}} = \frac{1}{N} \sum_{k \in s} \frac{\mathbf{x}_k \mathbf{x}'_k}{\pi_k}$ is invertible, $\tilde{\boldsymbol{\beta}}(t)$ is estimated by:

$$\hat{\boldsymbol{\beta}}(t) = \hat{\mathbf{G}}^{-1} \frac{1}{N} \sum_{k \in s} \frac{\mathbf{x}_k Y_k(t)}{\pi_k}, \quad t \in [0, T].$$

Remark that the denominator N is used in the expression of $\tilde{\boldsymbol{\beta}}(t)$ for asymptotic purposes and need not be estimated. The model-assisted estimator $\hat{\mu}_{MA}(t)$ is then defined by replacing $\tilde{\boldsymbol{\beta}}(t)$ by $\hat{\boldsymbol{\beta}}(t)$ in (5),

$$\hat{\mu}_{MA}(t) = \frac{1}{N} \sum_{k \in U} \hat{Y}_k(t) - \frac{1}{N} \sum_{k \in s} \frac{\hat{Y}_k(t) - Y_k(t)}{\pi_k}, \quad t \in [0, T], \quad (6)$$

where $\hat{Y}_k(t) = \mathbf{x}'_k \hat{\boldsymbol{\beta}}(t)$. Since $\sum_{k \in U} \hat{Y}_k(t) = (\sum_{k \in U} \mathbf{x}_k)' \hat{\boldsymbol{\beta}}(t)$, the only required information to build $\hat{\mu}_{MA}(t)$ is \mathbf{x}_k and $Y_k(t)$ for all the units $k \in s$ as well as the population totals of the auxiliary variables, $\sum_{k \in U} \mathbf{x}_k$.

Remark 1. *If the vector of auxiliary information contains the intercept (constant term), then it can be shown (see Särndal (1980)) that the Horvitz-Thompson estimator of the estimated residuals $\hat{Y}_k(t) - Y_k(t)$ is equal to zero for each $t \in [0, T]$. This means that the model-assisted estimator $\hat{\mu}_{MA}$ reduces in this case to the mean in the population of the predicted values \hat{Y}_k ,*

$$\hat{\mu}_{MA}(t) = \frac{1}{N} \sum_{k \in U} \hat{Y}_k(t), \quad t \in [0, T].$$

Moreover, if only the intercept term is used, namely $Y_k(t) = \beta(t) + \epsilon_{kt}$ for all $k \in U$, then the estimator $\hat{\mu}_{MA}$ is simply the well known Hájek estimator,

$$\hat{\mu}_{MA}(t) = \frac{\sum_{k \in s} \pi_k^{-1} Y_k(t)}{\sum_{k \in s} \pi_k^{-1}}, \quad t \in [0, T],$$

which is sometimes preferred to the Horvitz-Thompson estimator (see e.g. Särndal et al. (1992), Chapter 5.7).

Remark 2. Estimator $\hat{\mu}_{MA}(t)$ may also be obtained by using a calibration approach (Deville and Särndal (1992)) which consists in looking for weights $w_{ks}, k \in s$, that are as close as possible, according to some distance, to the sampling weights $1/\pi_k$ while estimating exactly the population totals of the auxiliary information,

$$\sum_{k \in s} w_{ks} \mathbf{x}_k = \sum_{k \in U} \mathbf{x}_k.$$

Considering the chi-square distance leads to the following choice of weights

$$w_{ks} = \frac{1}{\pi_k} - \left(\sum_{l \in s} \frac{\mathbf{x}_l}{\pi_l} - \sum_{l \in U} \mathbf{x}_l \right)' \left(\sum_{l \in s} \frac{\mathbf{x}_l \mathbf{x}_l'}{\pi_l} \right)^{-1} \frac{\mathbf{x}_k}{\pi_k}$$

and the calibration estimator $\sum_s w_{ks} Y_k(t)/N$ for the mean $\mu(t)$ is equal to $\hat{\mu}_{MA}(t)$ defined in (6).

2.3 A regularized estimator for asymptotics

The construction of the estimator $\hat{\mu}_{MA}(t)$ is based on the assumption that the matrix $\hat{\mathbf{G}}$ is invertible. To show the uniform convergence, we consider a modification of $\hat{\mathbf{G}}$ which will permit control of the expected norm of its inverse. Such a trick has already been used for example in Bosq (2000) and Guillas (2001). Since $\hat{\mathbf{G}}$ is a $p \times p$ symmetric and non negative matrix it is possible to write it as follows

$$\hat{\mathbf{G}} = \sum_{j=1}^p \eta_{j,n} \mathbf{v}_{jn} \mathbf{v}_{jn}',$$

where $\eta_{j,n}$ is the j th eigenvalue, $\eta_{1,n} \geq \dots \geq \eta_{p,n} \geq 0$, and \mathbf{v}_{jn} is the corresponding orthonormal eigenvector. Let us consider a real number $a > 0$ and define the following regularized estimator of \mathbf{G} ,

$$\hat{\mathbf{G}}_a = \sum_{j=1}^p \max(\eta_{j,n}, a) \mathbf{v}_{jn} \mathbf{v}_{jn}'.$$

It is clear that $\hat{\mathbf{G}}_a$ is always invertible and

$$\|\hat{\mathbf{G}}_a^{-1}\| \leq a^{-1}, \quad (7)$$

where $\|\cdot\|$ is the spectral norm for matrices. Furthermore, if $\eta_{p,n} \geq a$ then $\hat{\mathbf{G}} = \hat{\mathbf{G}}_a$. If $a > 0$ is small enough, we show under standard conditions on the moments of the variables X_1, \dots, X_p and on the first and second order inclusion probabilities that $\mathbb{P}(\hat{\mathbf{G}} \neq \hat{\mathbf{G}}_a) = \mathbb{P}(\eta_{p,n} < a) = O(n^{-1})$ (see Lemma A.1 in the Appendix).

Consequently, it is possible to estimate the mean function $\mu_N(t)$ by the following estimator

$$\hat{\mu}_{MA,a}(t) = \frac{1}{N} \sum_{k \in U} \hat{Y}_{k,a}(t) - \frac{1}{N} \sum_{k \in s} \frac{\hat{Y}_{k,a}(t) - Y_k(t)}{\pi_k}, \quad t \in [0, T], \quad (8)$$

where $\hat{Y}_{k,a}(t) = \mathbf{x}_k' \hat{\boldsymbol{\beta}}_a(t)$ and $\hat{\boldsymbol{\beta}}_a(t) = \hat{\mathbf{G}}_a^{-1} \frac{1}{N} \sum_{k \in s} \frac{\mathbf{x}_k Y_k(t)}{\pi_k}$.

2.4 Discretized observations

Note finally that with real data, we do not observe $Y_k(t)$ at all instants t in $[0, T]$ but only for a finite set of D measurement times, $0 = t_1 < \dots < t_D = T$. In functional data analysis, when the noise level is low and the grid of discretization points is fine, it is usual to perform a linear interpolation or to smooth the discretized trajectories in order to obtain approximations of the trajectories at every instant $t \in [0, T]$ (see Ramsay and Silverman (2005)).

If there are no measurement errors, if the trajectories are regular enough (but not necessarily differentiable) and if the grid of discretization points is dense enough, Cardot and Josserand (2011) showed that linear interpolation can provide sufficiently accurate approximations to the trajectories so that the approximation error can be neglected compared to the sampling error for the Horvitz-Thompson estimator. Note also that even if the observations are corrupted by noise, it has been shown by simulations in Cardot et al. (2012a) that smoothing does really improve the accuracy of the Horvitz-Thompson estimator only when the noise level is high.

Thus, for each unit k in the sample s , we build the interpolated trajectory

$$Y_{k,d}(t) = Y_k(t_i) + \frac{Y_k(t_{i+1}) - Y_k(t_i)}{t_{i+1} - t_i}(t - t_i) \quad t \in [t_i, t_{i+1}]$$

and we define $\widehat{\beta}_{a,d}(t)$ as the estimator of $\beta(t)$ based on the discretized observations as follows

$$\begin{aligned} \widehat{\beta}_{a,d}(t) &= \widehat{\mathbf{G}}_a^{-1} \frac{1}{N} \sum_{k \in s} \mathbf{x}_k Y_{k,d}(t) \\ &= \widehat{\beta}_a(t_i) + \frac{\widehat{\beta}_a(t_{i+1}) - \widehat{\beta}_a(t_i)}{t_{i+1} - t_i}(t - t_i). \end{aligned}$$

Therefore, the estimator of the mean population curve $\mu(t)$ based on the discretized observations is obtained by linear interpolation between $\widehat{\mu}_{MA,a}(t_i)$ and $\widehat{\mu}_{MA,a}(t_{i+1})$. For $t \in [t_i, t_{i+1}]$,

$$\begin{aligned} \widehat{\mu}_{MA,d}(t) &= \frac{1}{N} \sum_{k \in U} \widehat{Y}_{k,d}(t) - \frac{1}{N} \sum_{k \in s} \frac{(\widehat{Y}_{k,d}(t) - Y_{k,d}(t))}{\pi_k} \\ &= \widehat{\mu}_{MA,a}(t_i) + \frac{\widehat{\mu}_{MA,a}(t_{i+1}) - \widehat{\mu}_{MA,a}(t_i)}{t_{i+1} - t_i}(t - t_i) \end{aligned} \quad (9)$$

where $\widehat{Y}_{k,d}(t) = \mathbf{x}'_k \widehat{\beta}_{a,d}(t)$.

3 Asymptotic properties under the sampling design

All the proofs are postponed in an Appendix.

3.1 Assumptions

To derive the asymptotic properties under the sampling design $p(\cdot)$ of $\widehat{\mu}_{MA,d}$ we must suppose that both the sample size and the population size become large. More precisely, we consider the superpopulation framework introduced by Isaki and Fuller (1982) with a sequence of

growing and nested populations U_N with size N tending to infinity and a sequence of samples s_N of size n_N drawn from U_N according to the sampling design $p_N(s_N)$. The first and second order inclusion probabilities are respectively denoted by π_{kN} and π_{klN} . For simplicity of notations and when there is no ambiguity, we drop the subscript N . To prove our asymptotic results we need the following assumptions.

A1. We assume that $\lim_{N \rightarrow \infty} \frac{n}{N} = \pi \in (0, 1)$.

A2. We assume that $\min_{k \in U} \pi_k \geq \lambda > 0$, $\min_{k \neq l} \pi_{kl} \geq \lambda^* > 0$ and

$$\limsup_{N \rightarrow \infty} n \max_{k \neq l \in U} |\pi_{kl} - \pi_k \pi_l| < C_1 < \infty$$

A3. There are two positive constants C_2 and C_3 and $1 \geq \beta > 1/2$ such that, for all N and for all $(r, t) \in [0, T] \times [0, T]$,

$$\frac{1}{N} \sum_{k \in U} Y_k(0)^2 < C_2 \quad \text{and} \quad \frac{1}{N} \sum_{k \in U} \{Y_k(t) - Y_k(r)\}^2 < C_3 |t - r|^{2\beta}.$$

A4. We assume that there is a positive constant C_4 such that for all $k \in U$, $\|\mathbf{x}_k\|^2 < C_4$.

A5. We assume that, for $N > N_0$, the matrix \mathbf{G} is invertible and that the number a chosen before satisfies $\|\mathbf{G}^{-1}\| < a^{-1}$.

Assumptions **A1** and **A2** are classical hypotheses in survey sampling and deal with the first and second order inclusion probabilities. They are satisfied for many usual sampling designs with fixed size (see for example Hájek (1981), Robinson and Särndal (1983) and Breidt and Opsomer (2000)).

Assumption **A3** is a minimal regularity condition already required in Cardot and Josserand (2011). Even if pointwise consistency, for each fixed value of t , can be proved without any condition on the Hölder coefficient β , this regularity condition is necessary to get a uniform convergence result. A counterexample is given in Hahn (1977) when $\beta \leq 1/2$. More precisely it is shown that the sample mean i.i.d copies of a uniformly bounded continuous random function defined on a compact interval may not satisfy the Central Limit Theorem in the space of continuous functions. The hypothesis $\beta > 1/2$ also implies that the trajectories of the residual processes ϵ_{kt} , see (4), are regular enough (but not necessarily differentiable). Assumption **A4** could certainly be weakened at the expense of longer proofs. Assumption **A5** means that for all $\mathbf{u} \in \mathbb{R}$, with $\mathbf{u} \neq 0$, we have $\mathbf{u}'\mathbf{G}\mathbf{u} \geq a\mathbf{u}'\mathbf{u}$. The same kind of assumption is required in Isaki and Fuller (1982) to get the pointwise convergence in probability whereas Robinson and Särndal (1983) introduce a much stronger condition (condition A7 in their article) which directly deals with the mean square convergence of the estimator of the vector β of regression coefficients.

3.2 Uniform consistency of $\hat{\mu}_{MA,d}$

We aim at showing that $\hat{\mu}_{MA,d}$ is uniformly consistent for μ , namely that, for all $\varepsilon > 0$,

$$\mathbb{P} \left(\sup_{t \in [0, T]} |\hat{\mu}_{MA,d}(t) - \mu(t)| > \varepsilon \right) \rightarrow 0,$$

when N tends to infinity. The suitable space for proving the uniform convergence is the space of continuous functions on $[0, T]$, denoted by $C[0, T]$, equipped with its natural distance ρ ; for two elements $f, g \in C[0, T]$, the distance between f and g is $\rho(f, g) = \sup_{t \in [0, T]} |f(t) - g(t)|$. It results that the uniform consistency of $\hat{\mu}_{MA,d}$ is simply the convergence in probability of $\hat{\mu}_{MA,d}$ to μ in the space $C[0, T]$. Remark that with assumption **A3** the trajectories Y_k are continuous for all $k \in U$, and thus the mean curve μ belongs to $C[0, T]$ as well as its estimator $\hat{\mu}_{MA,d}$, by construction.

We first state the uniform consistency of the estimator $\hat{\beta}_{a,d}(t)$ towards its population counterpart $\tilde{\beta}(t)$ under conditions on the number and the repartition of discretization points.

Proposition 3.1. *Let assumptions (A1)-(A5) hold. If the discretization scheme satisfies $\lim_{N \rightarrow \infty} \max_{i=\{1, \dots, D_N-1\}} |t_{i+1} - t_i|^{2\beta} = o(n^{-1})$ then there is a constant $C > 0$ such that, for all n ,*

$$\sqrt{n} \mathbb{E}_p \left\{ \sup_{t \in [0, T]} \left\| \hat{\beta}_{a,d}(t) - \tilde{\beta}(t) \right\| \right\} \leq C.$$

We can now state a similar type of result for the estimator of the mean function.

Proposition 3.2. *Let assumptions (A1)-(A5) hold. If the discretization scheme satisfies $\lim_{N \rightarrow \infty} \max_{i=\{1, \dots, D_N-1\}} |t_{i+1} - t_i|^{2\beta} = o(n^{-1})$ then there is a constant $C > 0$ such that, for all n ,*

$$\sqrt{n} \mathbb{E}_p \left\{ \sup_{t \in [0, T]} |\hat{\mu}_{MA,d}(t) - \mu(t)| \right\} \leq C.$$

We deduce from Proposition 3.2 that estimator $\hat{\mu}_{MA,d}(t)$ is asymptotically unbiased as well as design consistent. Note that the approximation error (with linear interpolation) is negligible, compared to the sampling variability, under the additional assumption on the repartition of the discretization points. This assumption also tells us that less discretization points are required for smoother trajectories.

Let us also remark that, for each t ,

$$\hat{\mu}_{MA,d}(t) - \tilde{\mu}(t) = \frac{1}{N} \sum_{k \in U} \left(1 - \frac{\mathbb{1}_k}{\pi_k} \right) \mathbf{x}'_k \left(\hat{\beta}_a(t) - \tilde{\beta}(t) \right), \quad (10)$$

where $\mathbb{1}_k$ is the sample membership, so that it is not difficult to prove, under previous assumptions and by using lemma A.4 in the Appendix, that

$$\sqrt{n} (\hat{\mu}_{MA,d}(t) - \tilde{\mu}(t)) = o_p(1), \quad t \in [0, T]. \quad (11)$$

3.3 Covariance function estimation under the sampling design

We undertake in this section a detailed study of the covariance function of estimator $\hat{\mu}_{MA,d}$. The covariance function is computed with respect to the sampling design $p(\cdot)$ and from relation (9), we can deduce that $\hat{\mu}_{MA,d}$ is a nonlinear function of Horvitz-Thompson estimators, so the usual Horvitz-Thompson covariance formula given by (3) can not be used anymore. Nevertheless, in light of relation (11), the covariance function of $\hat{\mu}_{MA,d}$ between two instants r and t may be approximated by the covariance $\text{Cov}_p(\tilde{\mu}(r), \tilde{\mu}(t))$, which in turn is equal to the Horvitz-Thompson covariance applied to the residuals $Y_k - \tilde{Y}_k$. Let us denote by γ_{MA} the approximative covariance function of $\hat{\mu}_{MA,d}$ defined as follows

$$\begin{aligned} \gamma_{MA}(r, t) &= \frac{1}{N^2} \text{Cov}_p \left(\sum_{k \in s} \frac{Y_k(r) - \tilde{Y}_k(r)}{\pi_k}, \sum_{k \in s} \frac{Y_k(t) - \tilde{Y}_k(t)}{\pi_k} \right) \\ &= \frac{1}{N^2} \sum_{k \in U} \sum_{l \in U} (\pi_{kl} - \pi_k \pi_l) \frac{Y_k(r) - \tilde{Y}_k(r)}{\pi_k} \frac{Y_l(t) - \tilde{Y}_l(t)}{\pi_l}, \quad r, t \in [0, T]. \end{aligned} \quad (12)$$

This approximation explains that model-assisted estimators will perform much better than Horvitz-Thompson estimators if the residuals $Y_k(t) - \tilde{Y}_k(t)$ are small compared to $Y_k(t)$. The covariance function $\gamma_{MA}(r, t)$ can be estimated by the Horvitz-Thompson variance estimator for the estimated residuals $Y_{k,d}(t) - \hat{Y}_{k,d}(t)$,

$$\hat{\gamma}_{MA,d}(r, t) = \frac{1}{N^2} \sum_{k, l \in s} \frac{\pi_{kl} - \pi_k \pi_l}{\pi_{kl}} \cdot \frac{Y_{k,d}(r) - \hat{Y}_{k,d}(r)}{\pi_k} \cdot \frac{Y_{l,d}(t) - \hat{Y}_{l,d}(t)}{\pi_l}, \quad r, t \in [0, T], \quad (13)$$

where $\hat{Y}_{k,d}(t) = \mathbf{x}'_k \hat{\beta}_{a,d}(t)$.

To prove the consistency of the covariance estimator $\hat{\gamma}_{MA,d}(r, t)$, let us introduce additional assumptions that involve higher-order inclusion probabilities as well as conditions on the fourth order moments of the trajectories.

A6. We assume that $\lim_{N \rightarrow \infty} \max_{(k, l, k', l') \in D_{4,n}} |\mathbb{E}_p\{(\mathbb{1}_{kl} - \pi_{kl})(\mathbb{1}_{k'l'} - \pi_{k'l'})\}| = 0$ where $D_{t,N}$ is the set of all distinct t -tuples (i_1, \dots, i_t) from U_N and $\mathbb{1}_{kl} = \mathbb{1}_k \mathbb{1}_l$.

A7. There are two positive constants C_5 and C_6 such that $N^{-1} \sum_U Y_k(0)^4 < C_5$ and $N^{-1} \sum \{Y_k(t) - Y_k(r)\}^4 < C_6 |t - r|^{4\beta}$, for all $(r, t) \in [0, T]^2$

Condition **A6** has already been assumed by Breidt and Opsomer (2000) in a nonparametric model-assisted context and in Cardot and Josserand (2011) for Horvitz-Thompson estimators. It can be checked that it is fulfilled for simple random sampling without replacement (SRSWOR) or stratified sampling with SRSWOR within each strata. More generally, it is fulfilled for high entropy sampling designs. Boistard et al. (2012) prove that it is fulfilled for the rejective sampling whereas Cardot et al. (2012c) check that it is true for sampling designs, such as Sampford sampling, whose Kullback-Leibler divergence with respect to rejective sampling, tends to zero when the population size increases.

Proposition 3.3. *Assume (A1)-(A7) hold and the sequence of discretization schemes satisfy $\lim_{N \rightarrow \infty} \max_{i=\{1, \dots, D_N-1\}} |t_{i+1} - t_i| = o(1)$. Then, as N tends to infinity, we have for all $(r, t) \in [0, T]^2$,*

$$n \mathbb{E}_p \{ |\hat{\gamma}_{MA,d}(r, t) - \gamma_{MA}(r, t)| \} \rightarrow 0$$

and

$$n \mathbb{E}_p \left\{ \sup_{t \in [0, T]} |\hat{\gamma}_{MA,d}(t, t) - \gamma_{MA}(t, t)| \right\} \rightarrow 0.$$

Since $n\gamma_{MA}(r, t)$ remains bounded, the previous proposition tells us that $\hat{\gamma}_{MA,d}$ is consistent pointwise and the variance function estimator is uniformly convergent. Note also that the condition on the number of discretization points is much weaker than in Proposition 3.2 because we do not give here rates of convergence. To obtain such rates, we would also need to impose additional assumptions on the sampling design.

3.4 Asymptotic normality and confidence bands

We assume a supplementary assumption in order to get the asymptotic normality of the functional estimator $\hat{\mu}_{MA,d}$ in the space of continuous functions.

A8. We assume that for each fixed value of $t \in [0, 1]$,

$$\{\gamma_{MA}(t, t)\}^{-1/2} (\tilde{\mu}(t) - \mu(t)) \rightarrow \mathcal{N}(0, 1)$$

in distribution when N tends to infinity.

This assumption is satisfied for usual sampling designs (see *e.g.* Fuller (2009), Chapter 2.2). Note that using relation (11), we can write for any fixed value $t \in [0, T]$,

$$\hat{\mu}_{MA,d}(t) - \mu(t) = \tilde{\mu}(t) - \mu(t) + o_p(n^{-1/2}),$$

and deduce that $\sqrt{n}(\hat{\mu}_{MA,d}(t) - \mu(t))$ is also pointwise asymptotically Gaussian when conditions of Proposition 3.1 hold. Let us state now a much stronger result which indicates that the convergence to a Gaussian distribution also occurs for the trajectories, in the space of continuous functions (see Billingsley (1968), Chapter 2).

Proposition 3.4. *Assume (A1)-(A5) and (A8) hold. If the discretization scheme satisfies $\lim_{N \rightarrow \infty} \max_{i=\{1, \dots, D_N-1\}} |t_{i+1} - t_i|^{2\beta} = o(n^{-1})$, we have when n tends to infinity*

$$\sqrt{n} \{ \hat{\mu}_{MA,d} - \mu \} \rightsquigarrow Z$$

where \rightsquigarrow indicates the convergence in distribution in $C[0, T]$ with the uniform topology and Z is a Gaussian process taking values in $C[0, T]$ with mean 0 and covariance function $\gamma_Z(r, t) = \lim_{n \rightarrow +\infty} n\gamma_{MA}(r, t)$.

The "sup" functional defined on the space of continuous functions being continuous, the Proposition 3.4 implies that the real random variable $\sup_t |\sqrt{n} \{\widehat{\mu}_{MA,d}(t) - \mu(t)\}|$ converges in distribution to $\sup_t |Z(t)|$. We thus consider confidence bands for μ of the form

$$\left\{ \left[\widehat{\mu}_{MA,d}(t) \pm c \frac{\widehat{\sigma}(t)}{\sqrt{n}} \right], t \in [0, T] \right\} \quad (14)$$

where c is a suitable number and $\widehat{\sigma}(t) = \sqrt{n \widehat{\gamma}_{MA,d}(t, t)}$. Note that the fact that μ belongs to the confidence band defined in (14) is equivalent to

$$\sup_{t \in [0, T]} \frac{\sqrt{n}}{\widehat{\sigma}(t)} |\widehat{\mu}_{MA,d}(t) - \mu(t)| \leq c.$$

Given a confidence level $1 - \alpha \in (0, 1)$, one way to build such confidence band, that is to say one way to find an adequate value for c_α , is to perform simulations of a centered Gaussian functions \widehat{Z} defined on $[0, T]$ with mean 0 and covariance function $n \widehat{\gamma}_{MA,d}(r, t)$ and then compute the quantile of order $1 - \alpha$ of $\sup_{t \in [0, T]} |\widehat{Z}(t)/\widehat{\sigma}(t)|$. In other words, we look for a constant c_α , which is in fact a random variable since it depends on the estimated covariance function $\widehat{\gamma}_{MA,d}$, such that

$$\mathbb{P} \left(|\widehat{Z}(t)| \leq c_\alpha \frac{\widehat{\sigma}(t)}{\sqrt{n}}, \forall t \in [0, T] \mid \widehat{\gamma}_{MA,d} \right) = 1 - \alpha$$

The asymptotic coverage of this simulation based procedure has been rigorously studied for the Horvitz-Thompson estimators of the mean of sampled and noisy trajectories in Cardot et al. (2012a) whereas Cardot et al. (2012b) have successfully employed this approach on real load curves with model-assisted estimators. The next proposition, which can be seen as a functional version of Slutsky's Lemma, provides a rigorous justification of this latter technique.

Proposition 3.5. *Assume (A1)-(A8) hold and the discretization scheme satisfies*

$$\lim_{N \rightarrow \infty} \max_{i=\{1, \dots, D_N-1\}} |t_{i+1} - t_i|^{2\beta} = o(n^{-1}).$$

Let Z be a Gaussian process with mean zero and covariance function γ_Z (as in Proposition 3.4). Let (\widehat{Z}_N) be a sequence of processes such that for each N , conditionally on the estimator $\widehat{\gamma}_{MA,d}$ defined in (13), \widehat{Z}_N is Gaussian with mean zero and covariance $n \widehat{\gamma}_{MA,d}$. Suppose that $\gamma_Z(t, t)$ is a continuous function and $\inf_t \gamma_Z(t, t) > 0$. Then, as $N \rightarrow \infty$, the following convergence holds uniformly in c ,

$$\mathbb{P} \left(|\widehat{Z}_N(t)| \leq c \widehat{\sigma}(t), \forall t \in [0, T] \mid \widehat{\gamma}_{MA,d} \right) \rightarrow \mathbb{P} (|Z(t)| \leq c \sigma(t), \forall t \in [0, T]),$$

where $\widehat{\sigma}(t) = \sqrt{n \widehat{\gamma}_{MA,d}(t, t)}$ and $\sigma(t) = \sqrt{\gamma_Z(t, t)}$.

As in Cardot et al. (2012a), it is possible to deduce from the previous proposition that the chosen value $\widehat{c}_\alpha = c_\alpha(\widehat{\gamma}_{MA,d})$ provides asymptotically the desired coverage since it satisfies

$$\lim_{N \rightarrow \infty} \mathbb{P} \left(\mu(t) \in \left[\widehat{\mu}_{MA,d}(t) \pm \widehat{c}_\alpha \frac{\widehat{\sigma}(t)}{\sqrt{n}} \right], \forall t \in [0, T] \right) = 1 - \alpha.$$

4 An illustration on electricity consumption curves

We consider, as in Cardot et al. (2012b), a population of $N = 15069$ electricity consumption curves, measured every 30 minutes over a period of one week. Each element k of the population is thus a vector with size 336, denoted by $(Y_k(t), t \in \{1, \dots, 336\})$. The auxiliary information X of values $x_k, k \in U$ is simply the mean consumption of each meter $k \in U$ recorded during the week before the sample is drawn. As shown in Figure 1, the real variable X is strongly correlated with the consumption at each instant t of the current period of estimation so that a linear model with a functional response is well adapted for model-assisted estimation.

We draw samples s_i of size n , for $i = 1, \dots, I = 10000$ with simple random sampling without replacement (SRSWOR) so that $\pi_k = n/N$ for $k \in \{1, \dots, N\}$. We define, for each sample s_i , the model-assisted estimator of the mean curve,

$$\hat{\mu}_{MA,d}^{(i)}(t) = \frac{1}{N} \sum_{k \in U} \hat{Y}_k^{(i)}(t) - \frac{1}{N} \sum_{k \in s_i} \frac{\hat{Y}_k^{(i)}(t) - Y_k(t)}{n/N} \quad (15)$$

where $\mathbf{x}'_k = (1, x_k)$, $\hat{Y}_k^{(i)}(t) = \mathbf{x}'_k \hat{\boldsymbol{\beta}}^{(i)}(t)$, and $\hat{\boldsymbol{\beta}}^{(i)}(t) = \hat{\mathbf{G}}^{-1} \frac{1}{N} \sum_{k \in s_i} \frac{\mathbf{x}_k Y_k(t)}{n/N}$ for $t \in \{1, \dots, 336\}$. Cardot et al. (2012b) noted that, for the same sample size, the mean square error of estimation of the mean curve is divided by four compared to the Horvitz-Thompson estimator with SRSWOR when considering the model-assisted estimator defined in (15). There is only one covariate in this study and we did not encounter any problem with the invertibility of matrix $\hat{\mathbf{G}}$, the value of parameter a is thus $a = 0$.

We also define $\hat{\mu}(t) = \frac{1}{I} \sum_{i=1}^I \hat{\mu}_{MA,d}^{(i)}(t)$, $t \in \{1, \dots, 336\}$. The true variance function of the model-assisted estimator being unknown, we approximate it with a Monte Carlo approach based on the $I = 10000$ samples drawn with simple random sampling without replacement. The approximation to the true variance function is thus given by

$$\gamma_{emp}(r, t) = \frac{1}{I} \sum_{i=1}^I (\hat{\mu}_{MA,d}^{(i)}(t) - \hat{\mu}(t)) (\hat{\mu}_{MA,d}^{(i)}(r) - \hat{\mu}(r)) \quad (16)$$

for $(r, t) \in \{1, \dots, 336\}$.

The following quadratic loss criterion which measures a relative error is used to evaluate, for each sample, the accuracy of the variance estimator defined in (13),

$$E_r(\hat{\gamma}_{MA,d}) = \frac{1}{336} \sum_{t=1}^{336} \frac{|\hat{\gamma}_{MA,d}(t, t) - \gamma_{emp}(t, t)|^2}{\gamma_{emp}(t, t)^2} dt \quad (17)$$

We also decompose, over the $I = 10000$ estimations, the relative mean square error (RMSE) of the estimator into an approximation error ($RB(\hat{\gamma}_{MA,d})^2$) and a variance term ($VR(\hat{\gamma}_{MA,d})$) that can be related to the sampling error,

$$\begin{aligned} RMSE(\hat{\gamma}_{MA,d}) &= \frac{1}{I} \sum_{i=1}^I E_r^{(i)}(\hat{\gamma}_{MA,d}) \\ &= RB(\hat{\gamma}_{MA,d})^2 + VR(\hat{\gamma}_{MA,d}) \end{aligned}$$

where $E_r^{(i)}(\hat{\gamma}_{MA,d})$ is the value of $E_r(\hat{\gamma}_{MA,d})$ for the i th sample. The relative bias of the estimator $\hat{\gamma}_{MA,d}$ may be written as

$$RB(\hat{\gamma}_{MA,d})^2 = \frac{1}{336} \sum_{t=1}^{336} \left(\frac{\bar{\gamma}_{MA,d}(t,t) - \gamma_{emp}(t,t)}{\gamma_{emp}(t,t)} \right)^2$$

where $\bar{\gamma}_{MA,d}(t,t) = \frac{1}{I} \sum_{i=1}^I \hat{\gamma}_{MA,d}^{(i)}(t,t)$.

Sample size	$RMSE(\hat{\gamma}_{MA,d})$	$RB(\hat{\gamma}_{MA,d})^2$	$E_r(\hat{\gamma}_{MA,d})$				
			q_5	q_{25}	Median	q_{75}	q_{95}
250	0.1315	0.0027	0.0264	0.0455	0.0707	0.117	0.4945
500	0.0697	0.0016	0.0166	0.029	0.0459	0.0794	0.1945
1500	0.0238	0.0003	0.0076	0.0125	0.0186	0.028	0.0569

Table 1: Summary statistics for $E_r(\hat{\gamma}_{MA,d}, \gamma_{emp})$, with I=10000 samples.

The RMSE as well as the approximation error and statistics (quantiles) for E_r are given in Table 1. We can note that logically the RMSE decreases as the sample size increases and that even for moderate sample sizes, the estimations are rather precise. A closer look on how the RMSE is decomposed reveals that estimation error is mainly due to the sampling error, via the variance term whereas the approximation error term $RB(\hat{\gamma}_{MA,d})^2$ is negligible. This fact can be observed in Figure 2 where we plot the true variance function γ_{emp} over the considered period, its approximation γ_{MA} as well as an estimation $\hat{\gamma}_{MA,d}$ with a sample with size $n = 1500$, whose error according to criterion (17) is close to the mean error ($E_r \approx 0.02$).

We have also plotted in Figure 3 the difference between the empirical covariance function γ_{emp} and its approximation γ_{MA} and in Figure 4 the difference between γ_{MA} and its estimation $\hat{\gamma}_{MA,d}$ for a sample with size $n = 1500$ whose error, $E_r \approx 0.02$, is close to the mean value. Once again, it is clearly seen that the approximation error to the true covariance function (see Figure 3) is much smaller than the sampling error (see Figure 4). We can also remark some strong periodic pattern which reflects the natural daily periodicity in the electricity consumption behavior and that is related to the temporal correlation of the unknown residual process ϵ_{kt} defined in (4).

5 Concluding remarks

We have made in this paper an asymptotic study of model-assisted estimators, with linear regression models with functional response, when the target is the mean of functional data with discrete observations in time. This work can be extended in many directions. For example, one could consider more sophisticated regression models than model (4) such as non linear or nonparametric models with functional response by adapting, in a survey sampling context, models studied in the functional data analysis literature (see Chiou et al. (2004), Cardot (2007), or Ferraty et al. (2011)). However, one important drawback of such more

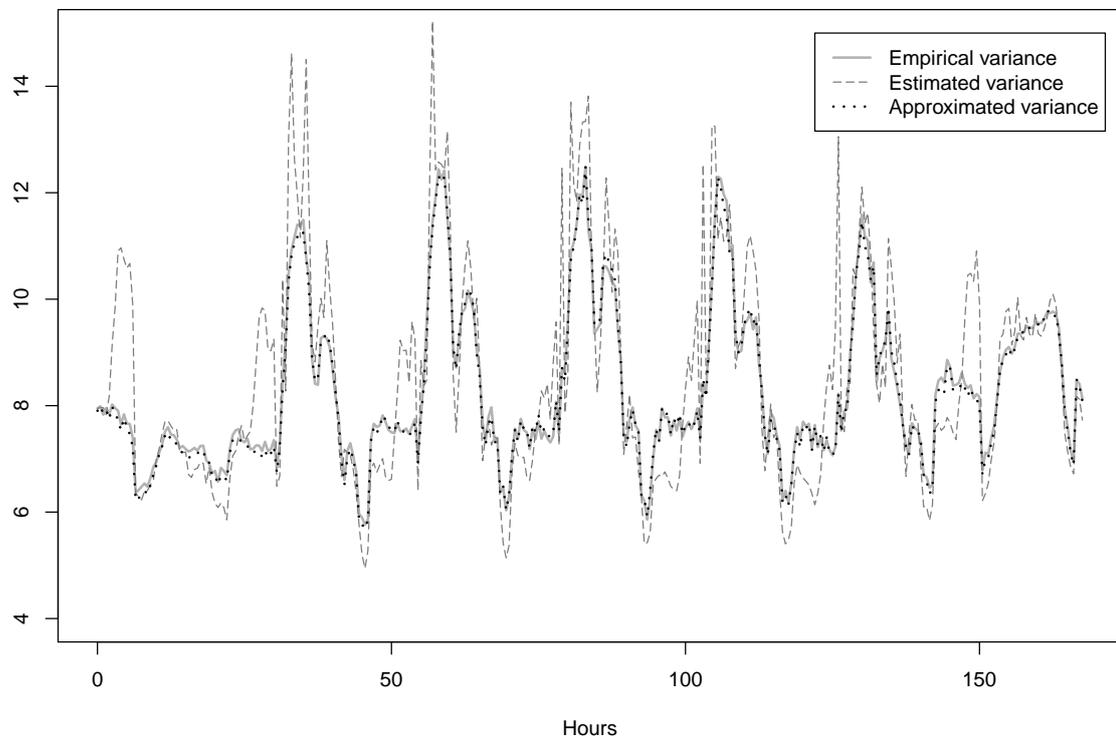


Figure 2: Empirical variance function γ_{emp} , approximated variance γ_{MA} and estimated variance $\hat{\gamma}_{MA,d}$ obtained with a sample of size $n = 1500$.

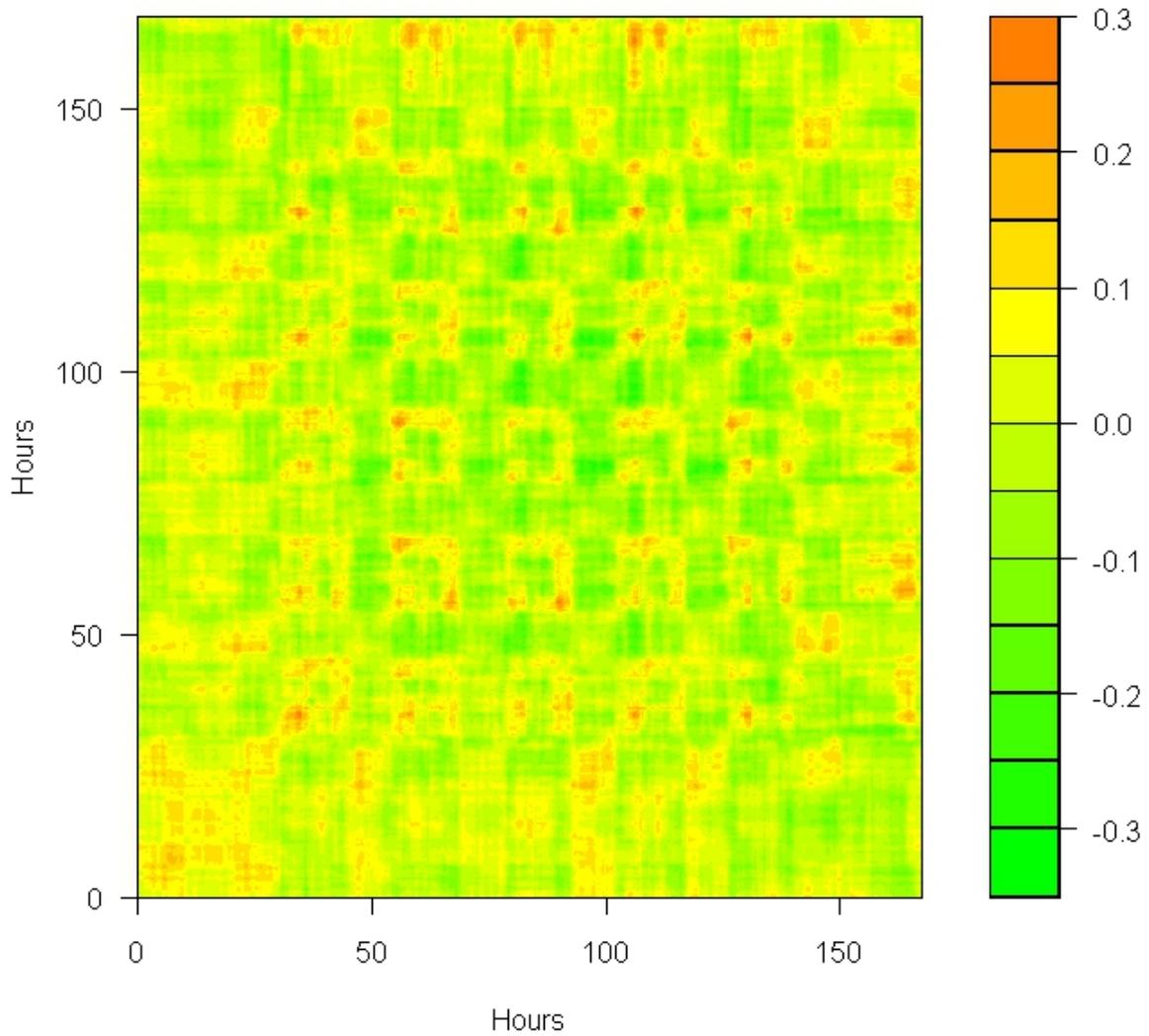


Figure 3: (Approximation error) difference between the covariance function and its approximation, $\gamma_{emp}(t, r) - \gamma_{MA}(t, r)$, for a sample with size $n = 1500$.

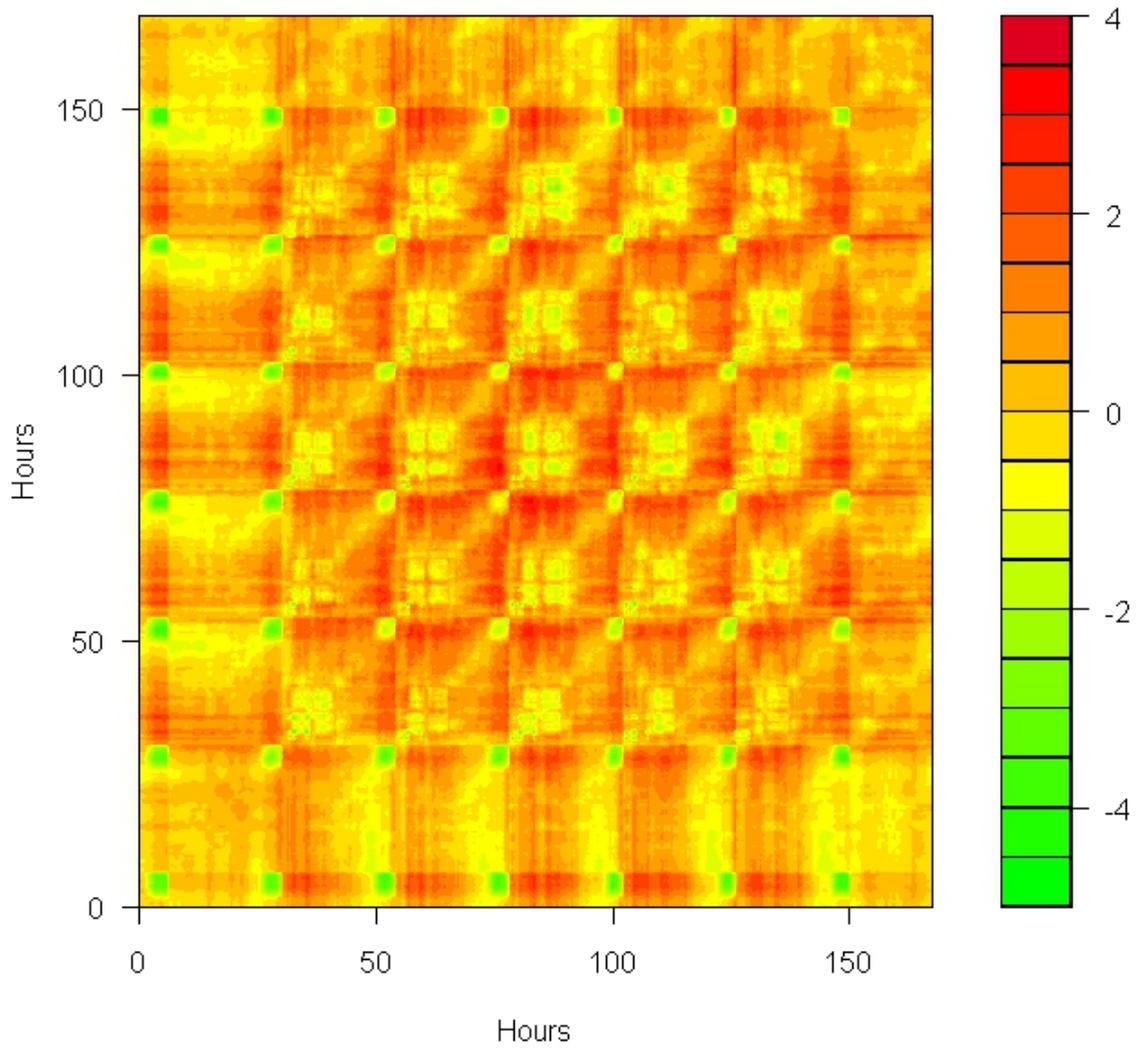


Figure 4: (Sampling error) difference between the approximated covariance function and its estimation, $\gamma_{MA}(t, r) - \hat{\gamma}_{MA,d}(t, r)$, for a sample with size $n = 1500$.

sophisticated approaches is that they would require to know \mathbf{x}_k for all the units k in the population as opposed to only their population totals.

An interesting direction for future investigation would be to consider noisy and possibly sparse measurements in time. For the Horvitz-Thompson estimator, local polynomials are employed in Cardot et al. (2012a) in order to first smooth the trajectories and it would certainly be possible to adapt the techniques developed in this work to the model-assisted estimation procedure.

Another promising direction for future research would be to adapt model-assisted estimators for time-varying samples. When one works with large networks of sensors it can be possible to consider a sequence of samples $s(t)$ that evolve along time. A preliminary work (see Degras (2012)), which focuses on Horvitz-Thompson estimators and stratified sampling clearly shows that such time-varying samples can outperform sampling designs that are fixed in time.

Acknowledgements. We thank the two anonymous referees for a careful reading and interesting suggestions that have permitted a great improvement of the original manuscript.

A Proofs

Throughout the proofs we use the letter C to denote a generic constant whose value may vary from place to place. We also denote by $\alpha_k = \frac{1}{\pi_k} - 1$, $k \in U$ and by $\Delta_{kl} = \pi_{kl} - \pi_k \pi_l$, $k, l \in U$.

A.1 Some useful Lemmas

Note that the result showed in the first Lemma is sometimes stated as an assumption (see *e.g* Robinson and Särndal (1983)). It is used to prove the convergence of the estimator of the mean in terms of mean square error.

Lemma A.1. *Let assumptions (A1), (A2) and (A4), (A5) hold. Then, there is a constant C such that*

$$n \mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a^{-1} - \mathbf{G}^{-1}\|^2 \right) \leq C.$$

Proof. The proof follows the lines of (Bosq (2000), Theorem 8.4) and (Cardot et al. (2010), Proposition 3.1). Using assumption (A5) and inequality (7), we have

$$\begin{aligned} \|\widehat{\mathbf{G}}_a^{-1} - \mathbf{G}^{-1}\| &\leq \|\widehat{\mathbf{G}}_a^{-1}\| \cdot \|\widehat{\mathbf{G}}_a - \mathbf{G}\| \cdot \|\mathbf{G}^{-1}\| \\ &\leq a^{-2} \|\widehat{\mathbf{G}}_a - \mathbf{G}\|, \end{aligned}$$

which implies

$$\mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a^{-1} - \mathbf{G}^{-1}\|^2 \right) \leq a^{-4} \mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a - \mathbf{G}\|^2 \right). \quad (18)$$

To bound $\mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a - \mathbf{G}\|^2 \right)$, we use the following decomposition.

$$\begin{aligned} \mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a - \mathbf{G}\|^2 \right) &= \mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a - \mathbf{G}\|^2 \mathbb{1}_{\{\widehat{\mathbf{G}}_a = \widehat{\mathbf{G}}\}} \right) + \mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a - \mathbf{G}\|^2 \mathbb{1}_{\{\widehat{\mathbf{G}}_a \neq \widehat{\mathbf{G}}\}} \right) \\ &\leq \mathbb{E}_p \left(\|\widehat{\mathbf{G}} - \mathbf{G}\|^2 \right) + 2\mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a - \widehat{\mathbf{G}}\|^2 \mathbb{1}_{\{\widehat{\mathbf{G}}_a \neq \widehat{\mathbf{G}}\}} \right) + 2\mathbb{E}_p \left(\|\widehat{\mathbf{G}} - \mathbf{G}\|^2 \mathbb{1}_{\{\widehat{\mathbf{G}}_a \neq \widehat{\mathbf{G}}\}} \right) \\ &\leq 3\mathbb{E}_p \left(\|\widehat{\mathbf{G}} - \mathbf{G}\|^2 \right) + 2\mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a - \widehat{\mathbf{G}}\|^2 \mathbb{1}_{\{\widehat{\mathbf{G}}_a \neq \widehat{\mathbf{G}}\}} \right). \end{aligned} \quad (19)$$

To bound the first term from the right-side of (19), we use the fact that the spectral norm is majored by the trace norm $\|\cdot\|_2$ defined by $\|\mathbf{A}\|_2^2 = \text{tr}(\mathbf{A}'\mathbf{A})$. Next, we show (see also Cardot et al. (2010), proof of Proposition 3.1) that,

$$\mathbb{E}_p \|\widehat{\mathbf{G}} - \mathbf{G}\|_2^2 = O(n^{-1}). \quad (20)$$

We have, with assumptions (A1), (A2) and (A4) that,

$$\begin{aligned} \mathbb{E}_p \|\widehat{\mathbf{G}} - \mathbf{G}\|_2^2 &= \frac{1}{N^2} \mathbb{E}_p \left(\sum_{k \in U} \sum_{l \in U} \alpha_k \alpha_l \text{tr}[\mathbf{x}_k \mathbf{x}'_k \mathbf{x}_l \mathbf{x}'_l] \right) \\ &\leq \frac{1}{N^2} \frac{1}{\lambda} \sum_{k \in U} \|\mathbf{x}_k \mathbf{x}'_k\|_2^2 + \max_{k \neq l \in U} |\Delta_{kl}| \frac{1}{N^2 \lambda^2} \sum_{k \in U} \sum_{l \in U} \|\mathbf{x}_k\|^2 \|\mathbf{x}_l\|^2 \\ &\leq \frac{1}{n} \left(\frac{n}{N} \frac{1}{\lambda} + n \max_{k \neq l \in U} |\Delta_{kl}| \frac{1}{\lambda^2} \right) C_2^2 \\ &\leq \frac{C}{n}. \end{aligned}$$

On the other hand,

$$\mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a - \widehat{\mathbf{G}}\|^2 \mathbb{1}_{\{\widehat{\mathbf{G}}_a \neq \widehat{\mathbf{G}}\}} \right) \leq a^2 \mathbb{P}(\widehat{\mathbf{G}}_a \neq \widehat{\mathbf{G}})$$

since

$$\begin{aligned} \|\widehat{\mathbf{G}}_a - \widehat{\mathbf{G}}\|^2 &= \left\| \sum_{j=1}^p [\max(\eta_{j,n}, a) - \eta_{j,n}] \mathbf{v}_{jn} \mathbf{v}'_{jn} \right\|^2 \\ &\leq \sup_{j=1, \dots, p} |\max(\eta_{j,n}, a) - \eta_{j,n}|^2 \\ &\leq a^2. \end{aligned}$$

Moreover, since $a < \eta_p = \|\mathbf{G}^{-1}\|^{-1}$ and by Chebychev inequality, we can bound

$$\begin{aligned} \mathbb{P}(\widehat{\mathbf{G}}_a \neq \widehat{\mathbf{G}}) &= \mathbb{P}(\eta_{p,n} < a) \\ &\leq \mathbb{P} \left(|\eta_{p,n} - \eta_p| \geq \frac{|\eta_p - a|}{2} \right), \\ &\leq \frac{4}{(\eta_p - a)^2} \mathbb{E}_p (|\eta_{p,n} - \eta_p|^2) \\ &\leq \frac{4}{(\eta_p - a)^2} \mathbb{E}_p \left(\|\widehat{\mathbf{G}} - \mathbf{G}\|^2 \right), \end{aligned}$$

because it is known that the eigenvalue map is Lipschitzian for symmetric matrices (see Bhatia (1997), Chapter 3). This means that for two $p \times p$ symmetric matrices \mathbf{A} and \mathbf{B} , with eigenvalues $\eta_1(\mathbf{A}) \geq \eta_2(\mathbf{A}) \geq \dots \geq \eta_p(\mathbf{A})$ (resp. $\eta_1(\mathbf{B}) \geq \dots \geq \eta_p(\mathbf{B})$), we have

$$\max_{j \in \{1, \dots, p\}} |\eta_j(\mathbf{A}) - \eta_j(\mathbf{B})| \leq \|\mathbf{A} - \mathbf{B}\|.$$

Hence, for some constant C

$$\begin{aligned}\mathbb{E}_p \left(\|\widehat{\mathbf{G}}_a - \mathbf{G}\|^2 \right) &\leq 3\mathbb{E}_p \left(\|\widehat{\mathbf{G}} - \mathbf{G}\|^2 \right) + 2a^2\mathbb{P}(\widehat{\mathbf{G}}_a \neq \widehat{\mathbf{G}}) \\ &\leq \frac{C}{n}.\end{aligned}\tag{21}$$

Combining (18), (19), (20) and (21), the proof is complete. \square

Lemma A.2. *Under assumptions (A1), (A2) and (A4), there is a constant C such that, for all n ,*

$$n \mathbb{E}_p \left\| \frac{1}{N} \sum_{k \in U} \left(\frac{\mathbb{1}_k}{\pi_k} - 1 \right) \mathbf{x}_k \right\|^2 \leq C.$$

Proof. Expanding the square norm, we have

$$\begin{aligned}n\mathbb{E}_p \left\| \frac{1}{N} \sum_{k \in U} \alpha_k \mathbf{x}_k \right\|^2 &= n\mathbb{E}_p \left(\frac{1}{N^2} \sum_{k \in U} \sum_{l \in U} \alpha_k \alpha_l \mathbf{x}'_k \mathbf{x}_l \right) \\ &\leq \frac{n}{N^2} \sum_{k \in U} \sum_{l \in U} \left| \frac{\Delta_{kl}}{\pi_k \pi_l} \right| \mathbf{x}'_k \mathbf{x}_l \\ &\leq \left[\frac{n}{N} \frac{1}{\lambda} + \frac{1}{\lambda^2} n \max_{k \neq l \in U} |\Delta_{kl}| \right] \frac{1}{N} \sum_{k \in U} \|\mathbf{x}_k\|^2\end{aligned}$$

and the result follows with hypotheses (A1), (A2) and (A4). \square

Lemma A.3. *Under assumptions (A2)-(A5), we have*

- i) $\|\tilde{\boldsymbol{\beta}}(t) - \tilde{\boldsymbol{\beta}}(r)\|^2 \leq a^{-2} C_3 C_4 |t - r|^{2\beta}$.
- ii) $\|\widehat{\boldsymbol{\beta}}_a(t) - \widehat{\boldsymbol{\beta}}_a(r)\|^2 \leq \frac{a^{-2}}{\lambda^2} C_3 C_4 |t - r|^{2\beta}$.

Proof For i), we just need to remark that, under hypotheses (A3), (A4) and (A5),

$$\begin{aligned}\|\tilde{\boldsymbol{\beta}}(t) - \tilde{\boldsymbol{\beta}}(r)\|^2 &= \left\| \mathbf{G}^{-1} \frac{1}{N} \sum_{k \in U} \mathbf{x}_k (Y_k(t) - Y_k(r)) \right\|^2 \\ &\leq \|\mathbf{G}^{-1}\|^2 \left(\frac{1}{N} \sum_{k \in U} \|\mathbf{x}_k\|^2 \right) \left(\frac{1}{N} \sum_{k \in U} (Y_k(t) - Y_k(r))^2 \right) \\ &\leq a^{-2} C_4 C_3 |t - r|^{2\beta}.\end{aligned}$$

The proof of point ii) is similar, but also requires the use of lower bounds on the first order inclusion probabilities (assumption (A2)),

$$\begin{aligned}
\|\widehat{\boldsymbol{\beta}}_a(t) - \widehat{\boldsymbol{\beta}}_a(r)\|^2 &= \left\| \widehat{\mathbf{G}}_a^{-1} \frac{1}{N} \sum_{k \in U} \frac{\mathbb{1}_k}{\pi_k} \mathbf{x}_k (Y_k(t) - Y_k(r)) \right\|^2 \\
&\leq \frac{1}{\lambda^2} \|\widehat{\mathbf{G}}_a^{-1}\|^2 \left(\frac{1}{N} \sum_{k \in U} \|\mathbf{x}_k\|^2 \right) \left(\frac{1}{N} \sum_{k \in U} (Y_k(t) - Y_k(r))^2 \right) \\
&\leq a^{-2} \frac{1}{\lambda^2} C_4 C_3 |t - r|^{2\beta}.
\end{aligned}$$

□

The following Lemma states the pointwise mean square convergence for any fixed value of $t \in [0, T]$.

Lemma A.4. *Suppose that assumptions (A1)-(A5) hold. Then, there is a positive constant ζ_1 such that, for all $t \in [0, T]$,*

$$n\mathbb{E}_p \left(\|\widehat{\boldsymbol{\beta}}_a(t) - \tilde{\boldsymbol{\beta}}(t)\|^2 \right) \leq \zeta_1.$$

Proof. The demonstration is similar to the proof of Lemma A.5 and is thus omitted. □

Lemma A.5. *Suppose that assumptions (A1)-(A5) hold. Then, there is a positive constant ζ_2 such that*

$$n\mathbb{E}_p \left(\|\widehat{\boldsymbol{\beta}}_a(t) - \tilde{\boldsymbol{\beta}}(t) - \widehat{\boldsymbol{\beta}}_a(r) + \tilde{\boldsymbol{\beta}}(r)\|^2 \right) \leq \zeta_2 |t - r|^{2\beta}.$$

Proof. A direct decomposition leads to

$$\begin{aligned}
&n\|\widehat{\boldsymbol{\beta}}_a(t) - \tilde{\boldsymbol{\beta}}(t) - \widehat{\boldsymbol{\beta}}_a(r) + \tilde{\boldsymbol{\beta}}(r)\|^2 \\
&\leq \left\| (\widehat{\mathbf{G}}_a^{-1} - \mathbf{G}^{-1}) \frac{1}{N} \sum_{k \in U} \frac{\mathbb{1}_k}{\pi_k} \mathbf{x}_k (Y_k(t) - Y_k(r)) + \mathbf{G}^{-1} \frac{1}{N} \sum_{k \in U} \left(\frac{\mathbb{1}_k}{\pi_k} - 1 \right) \mathbf{x}_k (Y_k(t) - Y_k(r)) \right\|^2 \\
&\leq 2A_{1N}^2 + 2A_{2N}^2, \tag{22}
\end{aligned}$$

where $A_{1N}^2 = n\|\widehat{\mathbf{G}}_a^{-1} - \mathbf{G}^{-1}\|^2 \left\| \frac{1}{N} \sum_{k \in U} \frac{\mathbb{1}_k}{\pi_k} \mathbf{x}_k (Y_k(t) - Y_k(r)) \right\|^2$ and

$A_{2N}^2 = n\|\mathbf{G}^{-1}\|^2 \left\| \frac{1}{N} \sum_{k \in U} \alpha_k \mathbf{x}_k (Y_k(t) - Y_k(r)) \right\|^2$. Using now assumptions (A2)-(A4) and the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
A_{1N}^2 &\leq n\|\widehat{\mathbf{G}}_a^{-1} - \mathbf{G}^{-1}\|^2 \left(\frac{1}{\lambda^2} \frac{1}{N} \sum_{k \in U} \|\mathbf{x}_k\|^2 \right) \left(\frac{1}{N} \sum_{k \in U} (Y_k(t) - Y_k(r))^2 \right) \\
&\leq n\|\widehat{\mathbf{G}}_a^{-1} - \mathbf{G}^{-1}\|^2 \frac{1}{\lambda^2} C_3 C_4 |t - r|^{2\beta}.
\end{aligned}$$

Using now Lemma A.1, we can bound

$$\mathbb{E}_p(A_{1N}^2) \leq C|t - r|^{2\beta}, \quad (23)$$

for some constant C . Now, with assumptions (A1)-(A5) and following the same arguments as in the proof of Lemma A.2, we also have

$$\begin{aligned} \mathbb{E}_p(A_{2N}^2) &\leq n\|\mathbf{G}^{-1}\|^2\mathbb{E}_p\left(\left\|\frac{1}{N}\sum_{k\in U}\alpha_k\mathbf{x}_k(Y_k(t) - Y_k(r))\right\|^2\right) \\ &\leq\left(\frac{n}{N}\frac{1}{\lambda} + \frac{n\max_{k\neq l\in U}|\Delta_{kl}|}{\lambda^2}\right)C_3C_4a^{-2}|t - r|^{2\beta} \leq C|t - r|^{2\beta}. \end{aligned} \quad (24)$$

for some positive constant C . Combining (22), (23) and (24), the result is proved. \square

A.2 Proof of Proposition 3.1 and Proposition 3.2

The proof of Proposition 3.1 is omitted. It is analogous to the proof of Proposition 3.2, which is given below. The different steps are similar to the proof of Proposition 1 in Cardot and Josserand (2011).

Let us decompose, for $t \in [0, T]$,

$$\sup_{t\in[0,T]}|\widehat{\mu}_{\mathbf{MA},d}(t) - \mu(t)| \leq \sup_{t\in[0,T]}|\widehat{\mu}_{\mathbf{MA},d}(t) - \widehat{\mu}_{\mathbf{MA},a}(t)| + \sup_{t\in[0,T]}|\widehat{\mu}_{\mathbf{MA},a}(t) - \mu(t)| \quad (25)$$

and study each term at the right-hand side of the inequality separately.

Step 1. The interpolation error $\sup_{t\in[0,T]}|\widehat{\mu}_{\mathbf{MA},d}(t) - \widehat{\mu}_{\mathbf{MA},a}(t)|$.

Consider $t \in [t_i, t_{i+1})$ and write

$$|\widehat{\mu}_{\mathbf{MA},d}(t) - \widehat{\mu}_{\mathbf{MA},a}(t)| \leq |\widehat{\mu}_{\mathbf{MA},a}(t_i) - \widehat{\mu}_{\mathbf{MA},a}(t)| + |\widehat{\mu}_{\mathbf{MA},a}(t_{i+1}) - \widehat{\mu}_{\mathbf{MA},a}(t_i)|.$$

Under assumptions (A2)-(A5) and using Lemma A.3, ii), we get

$$\begin{aligned} |\widehat{\mu}_{\mathbf{MA},a}(t) - \widehat{\mu}_{\mathbf{MA},a}(r)| &\leq \left|\frac{1}{N}\sum_{k\in U}\alpha_k\mathbf{x}'_k(\widehat{\boldsymbol{\beta}}_a(t) - \widehat{\boldsymbol{\beta}}_a(r))\right| + \frac{1}{N}\sum_{k\in s}\frac{|Y_k(t) - Y_k(r)|}{\pi_k} \\ &\leq\left(1 + \frac{1}{\lambda}\right)\sqrt{C_4}\|\widehat{\boldsymbol{\beta}}_a(t) - \widehat{\boldsymbol{\beta}}_a(r)\| + \frac{1}{\lambda}\left(\frac{1}{N}\sum_{k\in U}(Y_k(t) - Y_k(r))^2\right)^{1/2} \\ &\leq((1 + \lambda^{-1})C_4a^{-1} + 1)\lambda^{-1}\sqrt{C_3}|t - r|^\beta. \end{aligned}$$

So, there is a positive constant C such that

$$|\widehat{\mu}_{\mathbf{MA},a}(t) - \widehat{\mu}_{\mathbf{MA},a}(r)| \leq C|t - r|^\beta$$

and consequently,

$$\begin{aligned} |\widehat{\mu}_{\text{MA},d}(t) - \widehat{\mu}_{\text{MA},a}(t)| &\leq C[|t_i - t|^\beta + |t_{i+1} - t_i|^\beta] \\ &\leq 2C|t_{i+1} - t_i|^\beta. \end{aligned}$$

Hence, since by hypothesis, $\lim_{N \rightarrow \infty} \max_{i=\{1, \dots, d_{N-1}\}} |t_{i+1} - t_i|^\beta = o(n^{-1/2})$, we have

$$\sup_{t \in [0, T]} \sqrt{n} |\widehat{\mu}_{\text{MA},d}(t) - \widehat{\mu}_{\text{MA},a}(t)| = o(1). \quad (26)$$

Step 2. The estimation error $\sup_{t \in [0, T]} |\widehat{\mu}_{\text{MA},a}(t) - \mu(t)|$.

We use the following decomposition:

$$\sup_{t \in [0, T]} |\widehat{\mu}_{\text{MA},a}(t) - \mu(t)| \leq |\widehat{\mu}_{\text{MA},a}(0) - \mu(0)| + \sup_{r, t \in [0, T]} |\widehat{\mu}_{\text{MA},a}(t) - \mu(t) - \widehat{\mu}_{\text{MA},a}(r) + \mu(r)|. \quad (27)$$

Writing,

$$\begin{aligned} \widehat{\mu}_{\text{MA},a}(0) - \mu(0) &= \frac{1}{N} \sum_{k \in U} \alpha_k Y_k(0) - \frac{1}{N} \sum_{k \in U} \alpha_k \widehat{Y}_k(0) \\ &= \frac{1}{N} \sum_{k \in U} \alpha_k Y_k(0) - \frac{1}{N^2} \sum_{k \in U} \alpha_k \mathbf{x}'_k \widehat{\mathbf{G}}_a^{-1} \sum_{l \in s} \frac{\mathbf{x}_l Y_l(0)}{\pi_l} \end{aligned}$$

we directly get, with hypotheses A1-A3 and with similar arguments as in the proof of Lemma A.2, that for some constant C ,

$$\mathbb{E}_p (\widehat{\mu}_{\text{MA},a}(0) - \mu(0))^2 \leq \frac{C}{n}. \quad (28)$$

The second term at the right-hand side in (27) is dealt with using maximal inequalities. More exactly, we use Corollary 2.2.5 in van der Vaart and Wellner (2000). Consider for this, the Orlicz norm of some random variable X which is defined as follows

$$\|X\|_\psi = \sqrt{\mathbb{E}_p(\psi(X))}.$$

For the particular case $\psi(u) = u^2$, the Orlicz norm is simply the well-known L^2 norm, $\|X\|_\psi = \sqrt{\mathbb{E}_p(X^2)}$. Let us introduce for $(r, t) \in [0, T]^2$, the semimetric $d(r, t)$ defined by

$$\begin{aligned} d^2(r, t) &= \left\| \sqrt{n} |\widehat{\mu}_{\text{MA},a}(t) - \mu(t) - \widehat{\mu}_{\text{MA},a}(r) + \mu(r)| \right\|_\psi^2 \\ &= n \mathbb{E}_p (|\widehat{\mu}_{\text{MA},a}(t) - \mu(t) - \widehat{\mu}_{\text{MA},a}(r) + \mu(r)|^2) \end{aligned}$$

and consider $D(\epsilon, d)$, the packing number, which is defined as the maximum number of points in $[0, T]$ whose distance between each pair is strictly larger than ϵ . Then, Corollary 2.2.5 in van der Vaart and Wellner (2000) states that there is a constant $K > 0$ such that

$$\left\| \sup_{(r, t) \in [0, T]^2} \sqrt{n} |\widehat{\mu}_{\text{MA},a}(t) - \mu(t) - \widehat{\mu}_{\text{MA},a}(r) + \mu(r)| \right\|_\psi \leq K \int_0^T \psi^{-1}(D(\epsilon, d)) d\epsilon. \quad (29)$$

We show below that there is a constant C such that $d^2(r, t) \leq C|t - r|^{2\beta}$ and thus, since $\beta > 1/2$, the integral at the right-hand side of (29) is finite.

Let us first decompose

$$d^2(r, t) \leq 2d_1^2(r, t) + 2d_2^2(r, t) \quad (30)$$

where

$$d_1^2(r, t) = n\mathbb{E}_p(|\widehat{\mu}_{\text{MA},a}(t) - \tilde{\mu}(t) - \widehat{\mu}_{\text{MA},a}(r) + \tilde{\mu}(r)|^2)$$

and

$$d_2^2(r, t) = n\mathbb{E}_p(|\tilde{\mu}(t) - \mu(t) - \tilde{\mu}(r) + \mu(r)|^2).$$

By assumptions (A2)-(A4) and Lemma A.5, we can bound, for some constant C ,

$$\begin{aligned} d_1^2(r, t) &\leq \mathbb{E}_p \left\{ n \left\| \frac{1}{N} \sum_{k \in U} \alpha_k \mathbf{x}_k \right\|^2 \|\widehat{\boldsymbol{\beta}}_a(t) - \tilde{\boldsymbol{\beta}}(t) - \widehat{\boldsymbol{\beta}}_a(r) + \tilde{\boldsymbol{\beta}}(r)\|^2 \right\} \\ &\leq \left(1 + \frac{1}{\lambda}\right)^2 C_4 \zeta_2 |t - r|^{2\beta} := C|t - r|^{2\beta}. \end{aligned} \quad (31)$$

Considering now $d_2(r, t)$, we have

$$\begin{aligned} d_2^2(r, t) &= n\mathbb{E}_p \left[\frac{1}{N} \sum_{k \in U} \alpha_k [Y_k(t) - Y_k(r) - \mathbf{x}'_k(\tilde{\boldsymbol{\beta}}(t) - \tilde{\boldsymbol{\beta}}(r))] \right]^2 \\ &\leq 2\mathbb{E}_p(A_N^2) + 2\mathbb{E}_p(B_N^2) \end{aligned} \quad (32)$$

where $A_N^2 = n \left(\frac{1}{N} \sum_{k \in U} \alpha_k [Y_k(t) - Y_k(r)] \right)^2$ and $B_N^2 = n \left(\frac{1}{N} \sum_{k \in U} \alpha_k \mathbf{x}'_k(\tilde{\boldsymbol{\beta}}(t) - \tilde{\boldsymbol{\beta}}(r)) \right)^2$. With hypotheses (A1)-(A3), one can easily obtain that there is a positive constant C such that

$$\mathbb{E}_p(A_N^2) \leq C|t - r|^{2\beta} \quad (33)$$

and thanks to Lemma A.2 and to Lemma A.3, we can bound

$$\begin{aligned} \mathbb{E}_p(B_N^2) &\leq \mathbb{E}_p \left[n \left\| \frac{1}{N} \sum_{k \in U} \alpha_k \mathbf{x}_k \right\|^2 \right] \|\tilde{\boldsymbol{\beta}}(t) - \tilde{\boldsymbol{\beta}}(r)\|^2 \\ &\leq C|t - r|^{2\beta}. \end{aligned} \quad (34)$$

Combining now (33) and (34) with (30) and (31), we get that

$$d^2(r, t) \leq C|t - r|^{2\beta}, \quad (35)$$

for some constant C .

Using now (35), it is clear that the packing number is bounded as follows: $D(\epsilon, d) = O(\epsilon^{-1/\beta})$. Consequently, the integral at the right-hand side of (29) is finite when $\beta > 1/2$. Inserting (28) and (29) in (27), the proof of step 2 is complete.

A.3 Proof of the consistency of the covariance function

We first prove that for any $(r, t) \in [0, T]^2$, the estimator $\widehat{\gamma}_{\text{MA},d}(r, t)$ of the covariance function converges to $\gamma_{\text{MA}}(r, t)$.

Then we prove the uniform convergence of the variance estimator $\widehat{\gamma}_{\text{MA},d}(t, t)$ by showing its convergence in distribution to zero in the space of continuous functions. The proof is decomposed into two classical steps (see for example Theorem 8.1 in Billingsley (1968)). We first show the pointwise convergence, by considering the convergence of all finite linear combinations, and then we check that the sequence is tight by bounding the increments.

Step 1. Pointwise convergence

We want to show, that for each $(t, r) \in [0, T]^2$, we have when N tends to infinity,

$$n\mathbb{E}_p \{ | \widehat{\gamma}_{\text{MA},d}(r, t) - \gamma_{\text{MA}}(r, t) | \} \rightarrow 0.$$

Let us decompose

$$n(\widehat{\gamma}_{\text{MA},d}(r, t) - \gamma_{\text{MA}}(r, t)) = n(\widehat{\gamma}_{\text{MA},d}(r, t) - \widehat{\gamma}_{\text{MA},a}(r, t)) + n(\widehat{\gamma}_{\text{MA},a}(r, t) - \gamma_{\text{MA}}(r, t))$$

where $\widehat{\gamma}_{\text{MA},a}(r, t)$ is defined by

$$\widehat{\gamma}_{\text{MA},a}(r, t) = \frac{1}{N^2} \sum_{k,l \in s} \frac{\Delta_{kl}}{\pi_{kl}} \frac{Y_k(r) - \widehat{Y}_{k,a}(r)}{\pi_k} \cdot \frac{Y_l(t) - \widehat{Y}_{l,a}(t)}{\pi_l}$$

We study separately the interpolation and the estimation errors.

Interpolation error

Let us suppose that $t \in [t_i, t_{i+1})$, $r \in [t_{i'}, t_{i'+1})$. We have $n(\widehat{\gamma}_{\text{MA},d}(r, t) - \widehat{\gamma}_{\text{MA},a}(r, t)) \leq A + B$, with

$$\begin{aligned} A &= \frac{n}{N^2} \sum_{k,l \in U} \frac{|\Delta_{kl}|}{\pi_{kl}\pi_k\pi_l} |(Y_{k,d}(r) - Y_k(r))(Y_{l,d}(t) - Y_l(t)) \\ &\quad + (Y_{k,d}(r) - Y_k(r))(Y_l(t) - \widehat{Y}_{l,d}(t)) + (Y_k(r) - \widehat{Y}_{k,d}(r))(Y_{l,d}(t) - Y_l(t))| \end{aligned}$$

and

$$\begin{aligned} B &= \frac{n}{N^2} \sum_{k,l \in U} \frac{|\Delta_{kl}|}{\pi_{kl}\pi_k\pi_l} \left| (Y_k(r) - \widehat{Y}_{k,d}(r)) (Y_l(t) - \widehat{Y}_{l,d}(t)) - (Y_k(r) - \widehat{Y}_{k,a}(r)) (Y_l(t) - \widehat{Y}_{l,a}(t)) \right| \\ &= \frac{n}{N^2} \sum_{k,l \in U} \frac{|\Delta_{kl}|}{\pi_{kl}\pi_k\pi_l} \left| Y_k(r)(\widehat{Y}_{l,a}(t) - \widehat{Y}_{l,d}(t)) + Y_l(t)(\widehat{Y}_{k,a}(r) - \widehat{Y}_{k,d}(r)) + \widehat{Y}_{k,d}(r)\widehat{Y}_{l,d}(t) - \widehat{Y}_{k,a}(r)\widehat{Y}_{l,a}(t) \right|. \end{aligned}$$

For $t \in [t_i, t_{i+1}]$, we can write

$$|Y_{l,d}(t) - Y_l(t)| \leq |Y_l(t_i) - Y_l(t)| + |Y_l(t_{i+1}) - Y_l(t_i)|$$

and

$$|\widehat{Y}_{l,a}(t) - \widehat{Y}_{l,d}(t)| \leq |\widehat{Y}_{l,a}(t) - \widehat{Y}_{l,a}(t_i)| + |\widehat{Y}_{l,a}(t_{i+1}) - \widehat{Y}_{l,d}(t_i)|$$

We have that $\frac{1}{N} \sum_{l \in U} (Y_{l,d}(t) - Y_l(t))^2 \leq C[|t_i - t|^{2\beta} + |t_{i+1} - t_i|^{2\beta}]$ and $\frac{1}{N} \sum_{l \in U} (Y_l(t) - \hat{Y}_{l,d}(t))^2 = O(1)$. Thanks to Lemma A.3, we can bound

$$|\hat{Y}_{l,a}(t_i) - \hat{Y}_{l,a}(t)| \leq C_4 a^{-1} \frac{1}{\lambda} C_3^{1/2} |t_i - t|^\beta \leq C_4 a^{-1} \frac{1}{\lambda} C_3^{1/2} |t_{i+1} - t_i|^\beta.$$

Under the assumption on the grid of discretization points, one can get after some algebra that

$$n|\hat{\gamma}_{\text{MA},d}(r, t) - \hat{\gamma}_{\text{MA},a}(r, t)| = o(1).$$

Estimation error

Consider now,

$$\begin{aligned} n(\hat{\gamma}_{\text{MA},a}(r, t) - \gamma_{\text{MA}}(r, t)) &= \frac{n}{N^2} \sum_U \sum_U \frac{\Delta_{kl}}{\pi_k \pi_l} \left(\frac{\mathbb{1}_{kl}}{\pi_{kl}} - 1 \right) [Y_k(t) - \tilde{Y}_k(t)][Y_l(r) - \tilde{Y}_l(r)] \\ &\quad + \frac{n}{N^2} \sum_{k \in U} \sum_{l \in U} \frac{\Delta_{kl}}{\pi_k \pi_l} \frac{\mathbb{1}_{kl}}{\pi_{kl}} [Y_k(t) - \tilde{Y}_k(t)][\tilde{Y}_l(r) - \hat{Y}_{l,a}(r)] \\ &\quad + \frac{n}{N^2} \sum_{k \in U} \sum_{l \in U} \frac{\Delta_{kl}}{\pi_k \pi_l} \frac{\mathbb{1}_{kl}}{\pi_{kl}} [\tilde{Y}_k(t) - \hat{Y}_{k,a}(t)][Y_l(r) - \tilde{Y}_l(r)] \\ &\quad + \frac{n}{N^2} \sum_{k \in U} \sum_{l \in U} \frac{\Delta_{kl}}{\pi_k \pi_l} \frac{\mathbb{1}_{kl}}{\pi_{kl}} [\tilde{Y}_k(t) - \hat{Y}_{k,a}(t)][\tilde{Y}_l(r) - \hat{Y}_{l,a}(r)] \\ &:= A_1(r, t) + A_2(r, t) + A_3(r, t) + A_4(r, t). \end{aligned} \quad (36)$$

Let us define $\tilde{e}_k(t) = Y_k(t) - \tilde{Y}_k(t)$ and first show that $\mathbb{E}_p(A_1(r, t)^2) \rightarrow 0$ when $N \rightarrow \infty$.

$$\begin{aligned} \mathbb{E}_p(A_1(r, t)^2) &= \mathbb{E}_p \left[\frac{n^2}{N^4} \sum_{k, l \in U} \sum_{k', l' \in U} \frac{\Delta_{kl}}{\pi_k \pi_l} \left(\frac{\mathbb{1}_{kl}}{\pi_{kl}} - 1 \right) \frac{\Delta_{k'l'}}{\pi_{k'} \pi_{l'}} \left(\frac{\mathbb{1}_{k'l'}}{\pi_{k'l'}} - 1 \right) \tilde{e}_k(t) \tilde{e}_l(r) \tilde{e}_{k'}(t) \tilde{e}_{l'}(r) \right] \\ &= \mathbb{E}_p \left[\frac{n^2}{N^4} \sum_{k \in U} \sum_{k' \in U} \frac{1 - \pi_k}{\pi_k} \left(\frac{\mathbb{1}_k}{\pi_k} - 1 \right) \frac{1 - \pi_{k'}}{\pi_{k'}} \left(\frac{\mathbb{1}_{k'}}{\pi_{k'}} - 1 \right) \tilde{e}_k(t) \tilde{e}_k(r) \tilde{e}_{k'}(t) \tilde{e}_{k'}(r) \right] \\ &\quad + 2\mathbb{E}_p \left[\frac{n^2}{N^4} \sum_{k \in U} \sum_{k' \neq l' \in U} \frac{1 - \pi_k}{\pi_k} \left(\frac{\mathbb{1}_k}{\pi_k} - 1 \right) \frac{\Delta_{k'l'}}{\pi_{k'} \pi_{l'}} \left(\frac{\mathbb{1}_{k'l'}}{\pi_{k'l'}} - 1 \right) \tilde{e}_k(t) \tilde{e}_k(r) \tilde{e}_{k'}(t) \tilde{e}_{l'}(r) \right] \\ &\quad + \mathbb{E}_p \left[\frac{n^2}{N^4} \sum_{k \neq l \in U} \sum_{k' \neq l' \in U} \frac{\Delta_{kl}}{\pi_k \pi_l} \left(\frac{\mathbb{1}_{kl}}{\pi_{kl}} - 1 \right) \frac{\Delta_{k'l'}}{\pi_{k'} \pi_{l'}} \left(\frac{\mathbb{1}_{k'l'}}{\pi_{k'l'}} - 1 \right) \tilde{e}_k(t) \tilde{e}_l(r) \tilde{e}_{k'}(t) \tilde{e}_{l'}(r) \right] \\ &:= a_1 + a_2 + a_3. \end{aligned} \quad (37)$$

The hypotheses on the moments of the inclusion probabilities and Lemma A.6 give us

$$a_1 \leq \left(\frac{n^2}{N^3} \frac{1}{\lambda^3} + \frac{n^2}{N^2} \frac{\max_{k \neq k' \in U} |\Delta_{kk'}|}{\lambda^4} \right) \zeta_4$$

as well as

$$a_3 \leq \frac{C}{N} + \frac{(n \max_{k \neq l \in U} |\Delta_{kl}|)^2}{\lambda^4 \lambda^{*2}} \max_{(k, l, k', l') \in D_{4, n}} |\mathbb{E}_p\{(\mathbb{1}_{kl} - \pi_{kl})(\mathbb{1}_{k'l'} - \pi_{k'l'})\}| \zeta_5$$

so that $a_1 \rightarrow 0$ and $a_3 \rightarrow 0$ when $N \rightarrow \infty$. Then, the Cauchy-Schwarz inequality allows us to get that $a_2 \rightarrow 0$ when $N \rightarrow \infty$ and $\mathbb{E}_p(A_1(r, t)^2) \rightarrow 0$ when $N \rightarrow \infty$.

Let us show now that $\mathbb{E}_p(|A_4(r, t)|) \rightarrow 0$ when $N \rightarrow \infty$. With Lemma A.4, and assumptions (A1)-(A5), we have

$$\begin{aligned} \mathbb{E}_p(|A_4(r, t)|) &\leq n\mathbb{E}_p\left(\frac{1}{N^2} \sum_{k \in U} \sum_{l \in U} \frac{|\Delta_{kl}|}{\pi_k \pi_l} \frac{1}{\pi_{kl}} \|\mathbf{x}_k\| \|\mathbf{x}_l\| \|\tilde{\boldsymbol{\beta}}(t) - \hat{\boldsymbol{\beta}}_a(t)\| \|\tilde{\boldsymbol{\beta}}(r) - \hat{\boldsymbol{\beta}}_a(r)\|\right) \\ &\leq \frac{1}{n} \left[\frac{n}{\lambda^2 N} + \frac{n \max_{k \neq l \in U} |\Delta_{kl}|}{\lambda^2 \lambda^*} \right] C_4 \zeta_1 \end{aligned}$$

so that $\mathbb{E}_p(|A_4(r, t)|) \rightarrow 0$ when $N \rightarrow \infty$.

In a similar way, we can bound $\mathbb{E}_p(|A_2(r, t)|)$ as follows,

$$\begin{aligned} \mathbb{E}_p(|A_2(r, t)|) &\leq \frac{n}{N^2} \sum_{k \in U} \sum_{l \in U} \frac{|\Delta_{kl}|}{\pi_k \pi_l} \frac{1}{\pi_{kl}} \mathbb{E}_p|\tilde{e}_k(t) \hat{e}_l(r)| \\ &\leq \frac{n}{N^2} \sum_{k \in U} \sum_{l \in U} \frac{|\Delta_{kl}|}{\pi_k \pi_l} \frac{\|\mathbf{x}_l\|}{\pi_{kl}} |Y_k(t) - \tilde{Y}_k(t)| \cdot \mathbb{E}_p(\|\tilde{\boldsymbol{\beta}}(r) - \hat{\boldsymbol{\beta}}_a(r)\|) \\ &\leq \left(\frac{\sqrt{n}}{\lambda^2 N} + \frac{\sqrt{n} \max_{k \neq l \in U} |\Delta_{kl}|}{\lambda^2 \lambda^*} \right) C_4^{1/2} \zeta_1^{1/2} \frac{1}{N} \sum_{k \in U} |Y_k(t) - \tilde{Y}_k(t)|, \end{aligned}$$

where $\hat{e}_k(t) = \tilde{Y}_k(t) - \hat{Y}_{k,a}(t) = \mathbf{x}'_k(\tilde{\boldsymbol{\beta}}(t) - \hat{\boldsymbol{\beta}}_a(t))$. Thus, there is constant C such that,

$$\mathbb{E}_p(|A_2(r, t)|) \leq \frac{C}{\sqrt{n}}$$

and $\mathbb{E}_p(|A_2(r, t)|) \rightarrow 0$ when $N \rightarrow \infty$. We can show in a similar way that $\mathbb{E}_p(|A_3(r, t)|) \rightarrow 0$ when $N \rightarrow \infty$.

Finally, we have that for all $(r, t) \in [0, T]^2$,

$$n\mathbb{E}_p\{|\hat{\gamma}_{\text{MA},a}(r, t) - \gamma_{\text{MA}}(r, t)|\} \rightarrow 0, \quad \text{when } N \rightarrow \infty. \quad (38)$$

Step 2. Uniform convergence of the variance estimator

The pointwise convergence of the variance function proved in the previous step clearly implies the convergence of all finite linear combinations : for all $p \in \{1, 2, \dots\}$, for all $(c_1, \dots, c_p) \in \mathbb{R}^p$ and for all $(t_1, \dots, t_p) \in [0, T]^p$, we have

$$\sum_{\ell=1}^p c_\ell n (\hat{\gamma}_{\text{MA},a}(t_\ell, t_\ell) - \gamma_{\text{MA}}(t_\ell, t_\ell)) \rightarrow 0 \quad (39)$$

in probability as N tends to infinity. Thus, we deduce with the Cramer-Wold device that the vector $n(\hat{\gamma}_{\text{MA},a}(t_1, t_1) - \gamma_{\text{MA}}(t_1, t_1), \dots, \hat{\gamma}_{\text{MA},a}(t_p, t_p) - \gamma_{\text{MA}}(t_p, t_p))$ converges in distribution to 0 (in \mathbb{R}^p).

We need now to prove that the sequence of random functions $\hat{\gamma}_{\text{MA},a}(t, t)$ is tight in $C[0, T]$ by using a bound on its increments. Let us introduce the following criterion,

$$d_\gamma^2(t, r) = n^2 \mathbb{E}_p(|\hat{\gamma}_{\text{MA},a}(t, t) - \gamma_{\text{MA}}(t, t) - \hat{\gamma}_{\text{MA},a}(r, r) + \gamma_{\text{MA}}(r, r)|^2).$$

To conclude we show in the following that $d_\gamma^2(t, r) \leq C|t - r|^{2\beta}$ for a constant C and all $(r, t) \in [0, T]^2$. Using (36), the distance is decomposed into four parts.

Let us define $\phi_{kl}(t, r) = \tilde{e}_k(t)\tilde{e}_l(t) - \tilde{e}_k(r)\tilde{e}_l(r)$ and first consider $d_{A_1}^2 = \mathbb{E}_p(|A_1(t, t) - A_1(r, r)|^2)$. We have

$$\begin{aligned} d_{A_1}^2 &= \mathbb{E}_p \left[\frac{n^2}{N^4} \sum_{k \in U} \sum_{k' \in U} \frac{1 - \pi_k}{\pi_k} \left(\frac{\mathbb{1}_k}{\pi_k} - 1 \right) \frac{1 - \pi_{k'}}{\pi_{k'}} \left(\frac{\mathbb{1}_{k'}}{\pi_{k'}} - 1 \right) \phi_{kk}(t, r) \phi_{k'k'}(t, r) \right] \\ &+ 2\mathbb{E}_p \left[\frac{n^2}{N^4} \sum_{k \in U} \sum_{k' \neq l' \in U} \frac{1 - \pi_k}{\pi_k} \left(\frac{\mathbb{1}_k}{\pi_k} - 1 \right) \frac{\Delta_{k'l'}}{\pi_{k'}\pi_{l'}} \left(\frac{\mathbb{1}_{k'l'}}{\pi_{k'l'}} - 1 \right) \phi_{kk}(t, r) \phi_{k'l'}(t, r) \right] \\ &+ \mathbb{E}_p \left[\frac{n^2}{N^4} \sum_{k \neq l \in U} \sum_{k' \neq l' \in U} \frac{\Delta_{kl}}{\pi_k\pi_l} \left(\frac{\mathbb{1}_{kl}}{\pi_{kl}} - 1 \right) \frac{\Delta_{k'l'}}{\pi_{k'}\pi_{l'}} \left(\frac{\mathbb{1}_{k'l'}}{\pi_{k'l'}} - 1 \right) \phi_{kl}(t, r) \phi_{k'l'}(t, r) \right] \\ &:= b_1 + b_2 + b_3 \end{aligned} \quad (40)$$

Thanks to Lemma A.8, we get

$$\begin{aligned} b_1 &\leq \left(\frac{n^2}{N^3} \frac{1}{\lambda^3} + \frac{n^2 \max_{k \neq k' \in U} |\Delta_{kk'}|}{N^2 \lambda^4} \right) \frac{1}{N} \sum_{k \in U} |\phi_{kk}(t, r)|^2 \\ &\leq C|t - r|^{2\beta} \end{aligned} \quad (41)$$

and

$$\begin{aligned} b_3 &\leq \frac{C}{N} |t - r|^{2\beta} + \frac{(n \max_{k \neq l \in U} |\Delta_{kl}|)^2}{\lambda^4 \lambda^{*2}} \max_{(k, l, k', l') \in D_{4, n}} |\mathbb{E}_p\{(\mathbb{1}_{kl} - \pi_{kl})(\mathbb{1}_{k'l'} - \pi_{k'l'})\}| \left(\frac{1}{N^2} \sum_{k, l \in U} |\phi_{kl}(t, r)| \right)^2 \\ &\leq C|t - r|^{2\beta}. \end{aligned} \quad (42)$$

The Cauchy-Schwarz inequality together with bounds (41) and (42) allows us to get $b_2 \leq C|t - r|^{2\beta}$ so that

$$d_{A_1}^2 \leq C|t - r|^{2\beta}. \quad (43)$$

Let us bound now $d_{A_2}^2 = \mathbb{E}_p(|A_2(t, t) - A_2(r, r)|^2)$ and define $\tilde{\phi}_{kl}(t, r) = \tilde{e}_k(t)\widehat{e}_l(t) - \tilde{e}_k(r)\widehat{e}_l(r)$. Thanks to Lemma A.9, we get

$$\begin{aligned} d_{A_2}^2 &\leq \frac{2n^2}{N^2 \lambda^4} \mathbb{E}_p \left(\frac{1}{N} \sum_{k \in U} \tilde{\phi}_{kk}(t, r) \right)^2 + \frac{2n^2 \max_{k \neq l \in U} |\Delta_{kl}|^2}{\lambda^4 \lambda^{*2}} \mathbb{E}_p \left(\frac{1}{N^2} \sum_{k, l \in U} |\tilde{\phi}_{k,l}(t, r)| \right)^2 \\ &\leq C|t - r|^{2\beta}. \end{aligned} \quad (44)$$

Let us study now the last term, $d_{A_4}^2 = \mathbb{E}_p(|A_4(t, t) - A_4(r, r)|^2)$ and define $\widehat{\phi}_{kl}(t, r) = \widehat{e}_k(t)\widehat{e}_l(t) - \widehat{e}_k(r)\widehat{e}_l(r)$. Thanks to Lemma A.7, we have

$$\begin{aligned} d_{A_4}^2 &\leq \frac{2n^2}{N^2 \lambda^4} \mathbb{E}_p \left(\frac{1}{N} \sum_{k \in U} \widehat{\phi}_{kk}(t, r) \right)^2 + \frac{2n^2 \max_{k \neq l \in U} |\Delta_{kl}|^2}{\lambda^4 \lambda^{*2}} \mathbb{E}_p \left(\frac{1}{N^2} \sum_{k, l \in U} |\widehat{\phi}_{k,l}(t, r)| \right)^2 \\ &\leq C|t - r|^{2\beta}. \end{aligned} \quad (45)$$

Finally, we can deduce, with inequalities (36), (43), (44) and (45), that

$$\begin{aligned} d_\gamma^2(t, r) &= n^2 \mathbb{E}_p(|\widehat{\gamma}_{\text{MA},a}(t, t) - \gamma_{\text{MA}}(t, t) - \widehat{\gamma}_{\text{MA},a}(r, r) + \gamma_{\text{MA}}(r, r)|^2) \\ &\leq C|t - r|^{2\beta}. \end{aligned} \tag{46}$$

The end of the proof is a direct application of Theorem 12.3 of Billingsley (1968). Since $\beta > 1/2$, the sequence $n(\widehat{\gamma}_{\text{MA},a}(t, t) - \gamma_{\text{MA}}(t, t))$ is tight in $C([0, T])$ and converges in distribution to 0. The proof is complete with a direct application of the definition of weak convergence in $C([0, T])$ considering the bounded and continuous "sup" functional. \square

A.4 Proofs related to the asymptotic normality and the confidence bands

The steps of the proof of Proposition 3.4 are similar to the steps of the proof of Proposition 3.3. We first examine the finite combinations and invoke the Cramer-Wold device. Then we prove the tightness thanks to inequalities on the increments.

Let us first deal with the interpolation error, which is negligible under the assumption on the grid of discretization points, as shown in (26).

Then, in light of (10), Lemma A.2 and Lemma A.4, we clearly have that, for each value of t ,

$$\sqrt{n}(\widehat{\mu}_{\text{MA},a}(t) - \widetilde{\mu}(t)) = o_p(1),$$

and consequently, as n tends to infinity,

$$\sqrt{n}(\widehat{\mu}_{\text{MA},a}(t) - \mu(t)) \rightarrow \mathcal{N}(0, \gamma_Z(t, t)) \quad \text{in distribution,}$$

where the covariance-function of $\widetilde{\mu}$, which defined in (12), satisfies $\lim_{N \rightarrow \infty} n\gamma_{\text{MA}} = \gamma_Z$.

If we now consider p distinct discretization instants $0 \leq t_1 < t_2 \dots < t_p \leq 1$, it is immediate to check that for any vector $\mathbf{c} \in \mathbb{R}^p$, $\sqrt{n} \left(\sum_{j=1}^p c_j (\widetilde{\mu}(t_j) - \mu(t_j)) \right) \rightarrow \mathcal{N}(0, \sigma_c^2)$ where

$$\sigma_c^2 = \sum_{j=1}^p \sum_{\ell=1}^p c_j c_\ell \gamma_Z(t_j, t_\ell).$$

Indeed, by linearity, there exists a vector of random weights (w_1, \dots, w_N) which does not depend on time t such that

$$\widetilde{\mu}(t) = \sum_{k \in U} w_k Y_k(t),$$

and $\sum_{j=1}^p c_j \widetilde{\mu}(t_j) = \sum_{k \in U} w_k \left(\sum_{j=1}^p c_j Y_k(t_j) \right)$ also satisfies a CLT, with asymptotic variance σ_c^2 , under the moment conditions (A7). Thus, any finite linear combination is asymptotically Gaussian and we can conclude that the vector $\sqrt{n}(\widetilde{\mu}(t_1) - \mu(t_1), \dots, \widetilde{\mu}(t_p) - \mu(t_p))$ is asymptotically Gaussian with the Cramer-Wold device.

It remains to check the tightness of the functional process and this is a direct consequence of (30) and (35). Indeed, denoting by $Z_n(t) = \sqrt{n}(\widehat{\mu}_{\text{MA},a}(t) - \mu(t))$, there is a constant C such that, for all $(r, t) \in [0, T]^2$,

$$\mathbb{E}_p \left([Z_n(t) - Z_n(r)]^2 \right) \leq C|t - r|^{2\beta},$$

and, since $\beta > 1/2$, the sequence Z_n is tight in $C[0, T]$, in view of Theorem 12.3 of Billingsley (1968). \square

We prove now Proposition 3.5, the last result of the paper. The proof consists in showing the weak convergence of the sequence of distributions (\widehat{Z}_N) to the law of Z in $C([0, T])$.

For any vector of p points $0 \leq t_1 < \dots < t_p \leq T$, the finite dimensional convergence of the distribution of the Gaussian vector $(\widehat{Z}_N(t_1), \dots, \widehat{Z}_N(t_p))$ to the distribution of $(Z(t_1), \dots, Z(t_p))$ is an immediate consequence of the uniform convergence of the covariance function stated in Proposition 3.3. We can conclude with Slutsky's Lemma noting that for any $(c_1, \dots, c_p) \in \mathbb{R}^p$,

$$\sum_{j=1}^p \sum_{\ell=1}^p c_j c_\ell \widehat{\gamma}_{\text{MA},d}(t_j, t_\ell) \rightarrow \sum_{j=1}^p \sum_{\ell=1}^p c_j c_\ell \gamma_{\text{MA}}(t_j, t_\ell) \quad \text{in probability.} \quad (47)$$

Now, we need to check the tightness of (\widehat{Z}_N) in $C([0, T])$. Given $\widehat{\gamma}_{\text{MA},d}$, we have for $(r, t) \in [0, T]^2$,

$$\mathbb{E}_p \left[\left(\widehat{Z}_N(t) - \widehat{Z}_N(r) \right)^2 \mid \widehat{\gamma}_{\text{MA},d} \right] = n \left(\widehat{\gamma}_{\text{MA},d}(t, t) - 2\widehat{\gamma}_{\text{MA},d}(r, t) + \widehat{\gamma}_{\text{MA},d}(r, r) \right)$$

and after some algebra, we obtain thanks to Assumption (A2) that

$$\mathbb{E}_p \left[\left(\widehat{Z}_N(t) - \widehat{Z}_N(r) \right)^2 \mid \widehat{\gamma}_{\text{MA},d} \right] \leq \frac{C}{N} \sum_{k \in U} \left[(Y_{k,d}(t) - Y_{k,d}(r))^2 + \left(\widehat{Y}_{k,d}(t) - \widehat{Y}_{k,d}(r) \right)^2 \right]. \quad (48)$$

Let us first study the term $\sum_{k \in U} (Y_{k,d}(t) - Y_{k,d}(r))^2$ in the previous inequality and without loss of generality suppose that $t > r$. To check the continuity of the trajectories, we only need to consider points r and t that are close to each other. If t and r belong to the same interval, say $[t_i, t_{i+1}]$, then it is easy to check, with Assumption (A4) that

$$\begin{aligned} \frac{1}{N} \sum_{k \in U} (Y_{k,d}(t) - Y_{k,d}(r))^2 &= \frac{(t-r)^2}{(t_{i+1} - t_i)^2} \frac{1}{N} \sum_{k \in U} (Y_k(t_{i+1}) - Y_k(t_i))^2 \\ &\leq C(t-r)^{2\beta}. \end{aligned} \quad (49)$$

If we suppose now that $r \in [t_{i-1}, t_i]$ and $t \in [t_i, t_{i+1}]$, then we have

$$\begin{aligned} \frac{|Y_{k,d}(t) - Y_{k,d}(r)|}{t-r} &\leq \max \left(\frac{|Y_k(t_{i+1}) - Y_k(t_i)|}{t_{i+1} - t_i}, \frac{|Y_k(t_i) - Y_k(t_{i-1})|}{t_i - t_{i-1}} \right) \\ &\leq \frac{|Y_k(t_{i+1}) - Y_k(t_i)|}{t_{i+1} - t_i} + \frac{|Y_k(t_i) - Y_k(t_{i-1})|}{t_i - t_{i-1}} \end{aligned}$$

and using the same decomposition as in (49), we directly get that $\sum_{k \in U} (Y_{k,d}(t) - Y_{k,d}(r))^2 \leq C(t-r)^{2\beta}$. The second term at the right-hand side of inequality (48) is dealt with similar arguments and the decomposition used in the proof of Lemma A.3, so that $\frac{1}{N} \sum_{k \in U} \left(\widehat{Y}_{k,d}(t) - \widehat{Y}_{k,d}(r) \right)^2 \leq C|t-r|^{2\beta}$.

Thus, the trajectories of the Gaussian process are continuous on $[0, T]$ whenever $\beta > 0$ (see *e.g.* Theorem 1.4.1 in Adler and Taylor (2007)) and the sequence (\widehat{Z}_N) converges weakly to Z in $C([0, T])$ equipped with the supremum norm.

Using again Proposition 3.3, we have, uniformly in t , $\widehat{\sigma}_Z(t) = \sigma_Z(t) + o_p(1)$, where $\widehat{\sigma}_Z^2(t) = n\widehat{\gamma}_{MA,d}(t, t)$. Since, by hypothesis $\sigma_Z^2(t) = \gamma_Z(t, t)$ is a continuous function and $\inf_t \gamma_Z(t, t) > 0$, we get with Slutsky's lemma that $(\widehat{Z}_N/\widehat{\sigma}_Z)$ converges weakly to Z/σ_Z in $C([0, T])$. By definition of the weak convergence in $C([0, T])$ and the continuous mapping theorem, we also deduce that the real random variable $\widehat{M}_N = \sup_{t \in [0, T]} |\widehat{Z}_N(t)|/\widehat{\sigma}_Z(t)$ converges in distribution to $M = \sup_{t \in [0, T]} |Z(t)|/\sigma_Z(t)$, so that for each $c \geq 0$,

$$\mathbb{P} \left(\sup_{t \in [0, T]} |\widehat{Z}_N(t)|/\widehat{\sigma}_Z(t) \leq c \right) \rightarrow \mathbb{P} \left(\sup_{t \in [0, T]} |Z(t)|/\sigma_Z(t) \leq c \right).$$

Note finally, that under the previous hypotheses on γ_Z (see *e.g.* Pitt and Tran (1979)), the real random variable $M = \sup_{t \in [0, T]} (|Z(t)|/\sigma_Z(t))$ has an absolutely continuous and bounded density function so that the convergence holds uniformly in c (see *e.g.* Lemma 2.11 in van der Vaart (1998)). \square

A.5 Some useful lemmas

We state here without any proof some results that are needed for the study of the convergence of the covariance function. They rely on applications of the Cauchy-Schwarz inequality and on the assumptions on the moments of the trajectories and the inclusion probabilities.

Lemma A.6. *Assume (A2)-(A5) and (A7) hold. There are two constants ζ_4 and ζ_5 such that*

$$\frac{1}{N} \sum_{k \in U} \tilde{e}_k(t)^2 \tilde{e}_k(r)^2 \leq \zeta_4$$

and

$$\frac{1}{N^2} \sum_{k \in U} \sum_{l \in U} \tilde{e}_k(t)^2 \tilde{e}_l(r)^2 \leq \zeta_5,$$

where $\tilde{e}_k(t) = Y_k(t) - \tilde{Y}_k(t)$.

Lemma A.7. *Assume (A2)-(A5) and (A7) hold. There are two constants ζ_6 and ζ_7 such that*

$$\mathbb{E}_p \left(\frac{1}{N} \sum_{k \in U} \widehat{\phi}_{kk}(t, r)^2 \right) \leq \zeta_6 |t - r|^{2\beta}$$

and

$$\mathbb{E}_p \left(\frac{1}{N^2} \sum_{k, l \in U} \widehat{\phi}_{kl}(t, r) \right)^2 \leq \zeta_7 [|t - r|^{2\beta}]^2$$

where $\widehat{\phi}_{kl}(t, r) = \widehat{e}_k(t)\widehat{e}_l(t) - \widehat{e}_k(r)\widehat{e}_l(r)$ and $\widehat{e}_k(t) = \tilde{Y}_k(t) - \widehat{Y}_{k,a}(t)$.

Lemma A.8. Assume (A2)-(A5) and (A7) hold. There are two constant constants ζ_8 and ζ_9 such that

$$\frac{1}{N} \sum_{k \in U} \phi_{kk}^2(t, r) \leq \zeta_8 |t - r|^{2\beta}$$

and

$$\left(\frac{1}{N^2} \sum_{k, l \in U} \phi_{kl}(t, r) \right)^2 \leq \zeta_9 |t - r|^{2\beta}$$

where $\phi_{kl}(t, r) = \tilde{e}_k(t)\tilde{e}_l(t) - \tilde{e}_k(r)\tilde{e}_l(r)$ and $\tilde{e}_k(t) = Y_k(t) - \tilde{Y}_k(t)$.

Lemma A.9. Assume (A2)-(A5) and (A7) hold. There are two constants ζ_{10} and ζ_{11} such that

$$\mathbb{E}_p \left(\frac{1}{N} \sum_{k \in U} \tilde{\phi}_{kk}(t, r)^2 \right) \leq \zeta_{10} |t - r|^{2\beta}$$

and

$$\mathbb{E}_p \left(\frac{1}{N^2} \sum_{k, l \in U} \tilde{\phi}_{kl}(t, r) \right)^2 \leq \zeta_{11} |t - r|^{2\beta}$$

where $\tilde{\phi}_{kl}(t, r) = \tilde{e}_k(t)\hat{e}_l(t) - \tilde{e}_k(r)\hat{e}_l(r)$, $\tilde{e}_k(t) = Y_k(t) - \tilde{Y}_k(t)$ and $\hat{e}_k(t) = \tilde{Y}_k(t) - \hat{Y}_{k,a}(t)$.

References

- Adler, R. J. and Taylor, J. E. (2007). *Random Fields and Geometry*. Springer-Verlag, New York.
- Bhatia, R. (1997). *Matrix Analysis*. Springer-Verlag, New York.
- Billingsley, P. (1968). *Convergence of Probability Measures*. John Wiley and Sons.
- Boistard, H., Lopuhaä, H. P., and Ruiz-Gazen, A. (2012). Approximation of rejective sampling inclusion probabilities and application to higher order correlation. *Electronic J. of Statistics*, 6:1967–1983.
- Bosq, D. (2000). *Linear Processes in Function Spaces: Theory and Applications*, volume 149 of *Lecture notes in Statistics*. Springer-Verlag, New York.
- Breidt, F. J. and Opsomer, J. D. (2000). Local polynomial regression estimators in survey sampling. *Ann. Statist.*, 28(4):1023–1053.
- Callado, A., Kamiński, C., Szabó, G., Gerö, B., Kelner, J., Fernandes, S. and Sadok, D. (2009). A Survey on Internet Traffic Identification and Classification. *IEEE Communications Surveys and Tutorials*, 11:37–52.
- Cardot, H. (2007). Conditional functional principal components analysis. *Scandinavian J. of Statistics*, 34:317–335.

- Cardot, H., Chaouch, M., Goga, C., and Labruère, C. (2010). Properties of design-based functional principal components analysis. *J. of Statistical Planning and Inference*, 140:75–91.
- Cardot, H., Degras, D., and Josserand, E. (2012a). Confidence bands for Horvitz-Thompson estimators using sampled noisy functional data. To appear in *Bernoulli*.
- Cardot, H., Dessertaine, A., Goga, C., Josserand, E., and Lardin, P. (2012b). Comparaison de différents plans de sondage et construction de bandes de confiance pour l’estimation de la moyenne de données fonctionnelles : une illustration sur la consommation électrique. To appear in *Techniques d’Enquêtes/Survey Methodology*.
- Cardot, H., Goga, C., and Lardin, P. (2012c). Variance estimation and asymptotic confidence bands for the mean estimator of sampled functional data with high entropy unequal probability sampling designs. *Arxiv:1209.6503*
- 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.
- Chiou, J., Müller, H., and Wang, J. (2004). Functional response models. *Statistica Sinica*, 14:675–693.
- 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. (2011). Simultaneous confidence bands for parametric regression with functional data. *Statistica Sinica*, 21(4):1735–1765.
- Degras, D. (2012). Rotation sampling for functional data. <http://arxiv.org/abs/1204.4494>.
- Deville, J. C. and Särndal, C. E. (1992). Calibration estimators in survey sampling. *J. Amer. Statist. Assoc.*, 87:376–382.
- Faraway, J. (1997). Regression analysis for a functional response. *Technometrics*, 39(3):254–261.
- Ferraty, F., Laksaci, A., Tadj, A., and Vieu, P. (2011). Kernel regression with functional response. *Electronic J. of Statist.*, 5:159–171.
- Fuller, W. A. (2009). *Sampling Statistics*. John Wiley and Sons, Hoboken, New Jersey.
- Guillas, S. (2001). Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes. *Statist. and Probability Letters*, 55:281–291.
- Hahn, M. (1977). Conditions for sample-continuity and the central limit theorem. *Annals of Probability*, 5:351–360.

- Hájek, J. (1981). *Sampling From a Finite Population*. Statistics: Textbooks and Monographs. Marcel Dekker, New York.
- Isaki, C. and Fuller, W. (1982). Survey design under the regression superpopulation model. *J. Amer. Statist. Assoc.*, 77:49–61.
- Pitt, L. D. and Tran, L. T. (1979). Local sample path properties of Gaussian fields. *Annals of Probability*, 7:477–493.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer-Verlag, New York, second edition.
- Robinson, P. and Särndal, C. (1983). Asymptotic properties of the generalized regression estimator in probability sampling. *Sankhya : The Indian Journal of Statistics*, 45:240–248.
- Särndal, C. (1980). On π inverse weighting versus best linear unbiased weighting in probability sampling. *Biometrika*, 67:639–50.
- Särndal, C. E., Swensson, B., and Wretman, J. (1992). *Model Assisted Survey Sampling*. Springer series in statistics. Springer-Verlag, New York.
- van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.
- van der Vaart, A. W. and Wellner, J. A. (2000). *Weak Convergence and Empirical Processes. With Applications to Statistics*. Springer-Verlag, New York.