HAL
archives-ouvertes.fr

Estimation of the Sobol indices in a linear functional multidimensional model

Jean-Claude Fort, Thierry Klein, Agnès Lagnoux, Béatrice Laurent

▶ **To cite this version:**

Jean-Claude Fort, Thierry Klein, Agnès Lagnoux, Béatrice Laurent. Estimation of the Sobol indices in a linear functional multidimensional model. Journal of Statistical Planning and Inference, Elsevier, 2013, 143 (9), pp.1590-1605. <hal-00685998>

HAL Id: hal-00685998
https://hal.archives-ouvertes.fr/hal-00685998

Submitted on 6 Apr 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.

Estimation of the Sobol indices in a linear functional multidimensional model

Jean-Claude Fort [*], Thierry Klein[†], Agnès Lagnoux[‡], Béatrice Laurent [§]

April 6, 2012

Abstract

We consider a functional linear model where the explicative variables are stochastic processes taking values in a Hilbert space, the main example is given by Gaussian processes in $\mathbb{L}^2([0,1])$. We propose estimators of the Sobol indices in this functional linear model. Our estimators are based on $U-$statistics. We prove the asymptotic normality and the efficiency of our estimators and we compare them from a theoretical and practical point of view with classical estimators of Sobol indices.

Mathematics Subject Classification:
Keywords Karhunen-Loève expansion, fractional Gaussian process, semi-parametric efficient estimation, sensitivity analysis, quadratic functionals.

Introduction

Many mathematical models encountered in applied sciences involve a large number of poorly-known parameters as inputs. It is important for the practitioner to assess the impact of this uncertainty on the model output. An aspect of this assessment is sensitivity analysis, which aims to identify the most sensitive parameters, that is, parameters having the largest influence on the output. In global stochastic sensitivity analysis (see for example [17] and [19] and references therein) the input variables are assumed to be independent random variables. Their probability distributions account for the practitioner's belief about the input uncertainty. This turns the model output into a random variable, whose total variance can be split down into different partial variances (this is the so-called Hoeffding decomposition see [24]). Each of these partial variances measures the incertitude on the output induced by each input variable uncertainty. By considering the ratio of each partial variance to the total variance, we obtain a measure of importance for each input variable that is called the *Sobol index*

[*]Université Paris Descartes, 45 rue des Saints Pères, 75006 Paris, France
[†]Institut de Mathématiques de Toulouse, Université Toulouse 3, 31062 Toulouse Cédex 9, France
[‡]Institut de Mathématiques de Toulouse, Université Toulouse 2, 31058 Toulouse Cédex 9, France
[§]Institut de Mathématiques de Toulouse, INSA, 135 av. de Rangueil, 31077 Toulouse Cédex 4, France

or *sensitivity index* of the variable [20]; the most sensitive parameters can then be identified and ranked as the parameters with the largest Sobol indices.

Once the Sobol indices have been defined, the question of their effective computation or estimation remains open. In practice, one has to estimate (in a statistical sense) those indices using a finite sample of evaluations of model outputs [6]. Many Monte Carlo or quasi Monte Carlo approaches have been developed by the experimental sciences and engineering communities. This includes the FAST methods (see for example [3], [23] and references therein) and the Sobol pick-freeze (SPF) scheme (see [20, 22]). Nevertheless, those methods require many evaluations of model outputs which can be a strong limitation when thoses evaluations are expensive. Many approaches have been developped to overcome this issue. The most popular are Bayesian approach (see for example [17]) or the construction of metamodels. As mentioned in Kleijnen [11] (see equation (1) page 121) one can use functional linear regression as metamodel. In this paper, we study the particular context of the functional linear regression and propose a different way of estimation. We consider nonparametric estimators of quadratic functionals by projection methods, which are related to the procedures developed by Laurent (see [12, 13]) in a density model and by Da Veiga and Gamboa in [4] in a regression model. This method allows us to estimate simultaneously all the Sobol indices with a single sample of reasonable size.

More precisely we consider a separable Hilbert space $\mathbb{H}$ endowed with the scalar product $<,>$ and $X^1, \ldots, X^p$, p independent centered, $\mathbb{H}$-valued, stochastic processes. The model that we consider is a linear regression model :

$$Y = \mu + \sum_{k=1}^{p} < \beta^k, X^k > +\varepsilon.$$

where $\beta_i, 1 \leq i \leq p$ are elements of $\mathbb{H}$, μ is in $\mathbb{R}$ and ε is a centered noise independent of the processes $X_1, \ldots, X_p$.

Our approach is based on the so-called Karhunen-Loève decomposition of the processes X^k ([16, 1]). Thanks to this decomposition we construct natural estimators of the Sobol indices for whom we prove asymptotic normality and efficiency. (Asymptotic efficiency is a natural property which generalizes the notion of minimum variance unbiased estimator, see [24] chapters 8 and 25 or [8] for more details.)

This paper is organized as follows: in the first section, we set up the notations for the model, review the definition of Sobol indices and present our estimators. In the second section, we prove asymptotic normality and efficiency when we consider a simple functional linear regression model. These two properties are generalized in the third section in the general setting of the multiple functional linear regression. In Section 3, we also compare this method with the classical SPF. The fourth section gives numerical illustrations on a benchmark model.

1 Setting and notations

Let $\mathbb{H}$ be a separable Hilbert space endowed with the scalar product $<,>$ and let $X^1, \ldots, X^p$ be p independent centered, $\mathbb{H}$-valued, stochastic processes.

The generic model that we consider is the following

$$Y = \mu + \sum_{k=1}^{p} < \beta^k, X^k > +\varepsilon, \tag{1}$$

where $\beta^k, 1 \leq k \leq p$ are elements of $\mathbb{H}$, μ is in $\mathbb{R}$ and ε is a centered noise with variance η^2 independent of the processes $X_1, \ldots, X_p$, and with finite fourth order moment.
We define, as usual in a finite dimensional setting, the Sobol index with respect to the entry (explicative variable) number k :

$$S^k = \frac{\mathrm{Var}(\mathbb{E}(Y|X^k))}{\mathrm{Var}(Y)}.$$

From the observation of $\left(X_i^1, \ldots, X_i^p, Y_i\right)_{1 \leq i \leq n}$ i.i.d. obeying to Model (1), our aim is to estimate the vector $\left(S^1, \ldots, S^p\right)$. Since $\mathrm{Var}(Y)$ can be easily estimated by the empirical variance based on $(Y_1, \ldots, Y_n)$, the main purpose of the paper is to estimate for all k the quantity

$$V^k = \mathrm{Var}(\mathbb{E}(Y|X^k)).$$

We will assume that we know the distributions of the input processes $\left(X^k\right)_{1 \leq k \leq p}$.
In the next section we consider the simple case where $p = 1$. In this setting the Sobol index is of less interest, but the computations then easily extend to the generic model.

2 Simple functional linear regression model

Using the same notations as in Section 1, we consider, in this section, the case $p = 1$, which leads to the model

$$Y = \mu + < \beta, X > +\varepsilon. \tag{2}$$

We observe a n-sample of (X, Y), that we denote $(X_i, Y_i), 1 \leq i \leq n$.
We have $\mathbb{E}(Y|X) = \mu + < \beta, X >$ and $V^X = \mathrm{Var}\left(\mathbb{E}(Y|X)\right) = \mathrm{Var}(< \beta, X >)$.
The Sobol index S^X is defined by

$$S^X = \frac{V^X}{\mathrm{Var}(Y)}.$$

We assume that $\mathbb{E}\left(\|X\|^2\right) < \infty$, so that the covariance operator defined for all $f \in \mathbb{H}$ by $\Gamma(f) = \mathbb{E}\left[< X, f > X\right]$ is Hilbert-Shmidt and is diagonalizable via the Karhunen-Loève expansion in an orthonormal basis of eigenfunctions $(\varphi_l, 1 \leq l)$, with decreasing eigenvalues $(\lambda_l, 1 \leq l)$ such that $\sum_{l=1}^{\infty} \lambda_l < \infty$. See e.g. [14, 15, 10]. We set $< X, \varphi_l >= \sqrt{\lambda_l}\xi_l$. The variables $(\xi_l)_{l \geq 1}$ are centered, uncorrelated, with variance 1. When X is a Gaussian process, the variables $(\xi_l)_{l \geq 1}$ are i.i.d. standard Gaussian variables.

Example 2.1 (Examples of Karhunen-Loève expansions ([16, 1]))**.**

1. *Let $B = (B_t)_{t\in[0,1]}$ a Brownian motion and $\mathbb{H} = \mathbb{L}^2([0,1], dt)$. Then*

$$\lambda_l = \frac{1}{(\pi(l-\frac{1}{2}))^2} \quad \textit{and} \quad \varphi_l(t) = \sqrt{2}\sin(\frac{t}{\sqrt{\lambda_l}}), \quad l \geq 1.$$

2. *Let $W = (W_t)_{t\in[0,1]}$ a Brownian bridge and $\mathbb{H} = \mathbb{L}^2([0,1], dt)$.*

$$\lambda_l = \frac{1}{(\pi l)^2} \quad \textit{and} \quad \varphi_l(t) = \sqrt{2}\sin(\pi l t), \quad l \geq 1.$$

3. *Let $X = (X_t)_{t\in[0,1]}$ a fractional Brownian motion with Hurst exponent H. Then the asymptotic of the eigenvalues is given by*

$$\lambda_l = \frac{\sin(\pi H)\Gamma(2H+1)}{(l\pi)^{2H+1}} + o\left(l^{\frac{(2H+2)(4H+3)}{4H+5}+\delta}\right).$$

In the case $H = 1/2$, the Brownian case, the result agrees with the exact result $\lambda_l = \frac{1}{(\pi(l-1/2))^2}$.

In the basis $(\varphi_l)_{l\geq 1}$, β has the expansion $\beta = \sum_{l=1}^{\infty} \gamma_l \varphi_l$, and we have :

$$\mathbb{E}(YX) = \mathbb{E}(< X, \beta > X) = \Gamma(\beta) = \sum_{l=1}^{\infty} \lambda_l \gamma_l \varphi_l.$$

Setting $g = \sum_{l=1}^{\infty} \lambda_l \gamma_l \varphi_l$, we have : $\gamma_l =< g, \varphi_l > /\lambda_l$.
The index V^X is then given by

$$V^X = \mathrm{Var}(\mathbb{E}(Y|X)) = \mathbb{E}(\mathbb{E}(Y|X)^2) = \mathbb{E}(< \beta, X >< \beta, X >) =< \beta, \Gamma(\beta) > .$$

which expends $V^X = \sum_{l=1}^{\infty} \lambda_l \gamma_l^2$.

We consider the empirical unbiased estimator of γ_l :

$$\widehat{\gamma}_l = \frac{1}{\lambda_l}\frac{1}{n}\sum_{i=1}^{n} < X_i, \varphi_l > Y_i.$$

For all $m \in \mathbb{N}^*$, we introduce the U-statistics of order 2 :

$$\widehat{V}_m^X = \sum_{l=1}^{m} \frac{1}{\lambda_l}\frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} < X_i, \varphi_l > Y_i < X_j, \varphi_l > Y_j. \tag{3}$$

We have $\mathbb{E}\widehat{V}_m^X = \sum_{l=1}^{m} \lambda_l \gamma_l^2$, hence $\widehat{V}_m^X$ is a biased estimator of V^X, with bias $\mathbb{B}_m = \sum_{l=m+1}^{\infty} \lambda_l \gamma_l^2$.

In the following section, we study the asymptotic behaviour of $\widehat{V}_m^X$, for a suitable choice of the parameter m.

2.1 Central limit theorem

Using Hoeffding's decompsition of the U-statistics $\widehat{V}_m^X$ (see [7]), we straightforwardly get the following result.

Proposition 1. *$\widehat{V}_m^X$ can be rewritten as the sum of a totally degenerated U-statistics of order 2, a centered linear term and a deterministic term in the following way :*

$$\widehat{V}_m^X = U_n K + P_n L + \sum_{l=1}^{m} \gamma_l^2 \lambda_l \tag{4}$$

where

$$U_n K := \sum_{l=1}^{m} \frac{1}{\lambda_l} \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} \left(Y_i < X_i, \varphi_l > -\gamma_l \lambda_l\right) \left(Y_j < X_j, \varphi_l > -\gamma_l \lambda_l\right)$$

$$P_n L := \frac{2}{n} \sum_{l=1}^{m} \sum_{i=1}^{n} \gamma_l \left(Y_i < X_i, \varphi_l > -\gamma_l \lambda_l\right).$$

As a consequence, we have

$$\widehat{V}_m^X - V^X = U_n K + P_n L - \sum_{l>m} \gamma_l^2 \lambda_l = U_n K + P_n L - \mathbb{B}_m.$$

Theorem 2. *Let $(X_i, Y_i)_{1 \le i \le n}$ be i.i.d. observations with the same distribution as (X, Y) from Model (2). We assume that $\mathbb{E}(\|X\|^4) < +\infty$ and that $\mathbb{E}(\varepsilon^4) < +\infty$. We consider the Karhunen-Loève expansion of X :*

$$X = \sum_{l \ge 1} \sqrt{\lambda_l} \xi_l \varphi_l.$$

We assume

$$\sup_{l \ge 1} \mathbb{E}(\xi_l^4) < +\infty. \tag{5}$$

We consider the estimator V_m^X of V_X defined by (3) with $m = m(n) = \sqrt{n} h(n)$, where $h(n)$ satisfies : $h(n) \to 0$ and $\forall \alpha > 0$, $n^\alpha h(n) \to +\infty$ as $n \to +\infty$.
We assume that there exist $C > 0$ and $\delta > 1$ such that

$$\forall l \ge 1, \quad \lambda_l \le C l^{-\delta}$$

Then we have

$$\sqrt{n}(\widehat{V}_m^X - V^X) \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, 4\mathrm{Var}(Y < X, \beta >)). \tag{6}$$

Comments :

1. We may assume that $h(n) = 1/\log(n)$, and hence $m(n) = \sqrt{n}/\log n$, to fill the condition
$$\forall \alpha > 0, \lim_{n \to \infty} n^\alpha h(n) = +\infty.$$
The estimator $\widehat{V}_m^X$ converges at the parametric rate $1/\sqrt{n}$, for any β.

2. The nature of the problem of estimating the quadratic functional V_m^X is completely different from the estimation of the signal β, where nonparametric rates are obtained (see for example [2] for the estimation of β in a circular functional linear model).

3. We will prove in the next section the asymptotic efficiency of the estimator $\widehat{V}_m^X$.

4. Note that the condition (5) is verified when X is a Gaussian process, since in the case, the variables $(\xi_l)_{l\geq 1}$ are i.i.d. standard Gaussian variables.

5. In Theorem 2 we have assumed that there exist $C > 0$ and $\delta > 1$ such that
$$\forall l \geq 1, \quad \lambda_l \leq C l^{-\delta}.$$
Let us recall that since $\mathbb{E}\left(\|X\|^2\right) < +\infty$, $\sum_{l\geq 1} \lambda_l < +\infty$ hence this assumption is not very strong.

Proof. In order to prove (6), we will show that
$$\begin{cases} \mathbb{B}_m^2 = o\left(1/n\right) \\ \mathbb{E}\left((U_n K)^2\right) = o\left(1/n\right) \\ \sqrt{n} P_n L \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0, 4\mathrm{Var}(Y < X, \beta >)) \end{cases}.$$

a) Bias term

Since we have assumed that $\forall l \geq 1, \quad \lambda_l \leq C l^{-\delta}$ for some $C > 0$ and $\delta > 1$, and since $\sum_{l\geq 1} \gamma_l^2 < +\infty$, we get
$$\mathbb{B}_m = \sum_{l>m} \gamma_l^2 \lambda_l \leq C \sum_{l>m} \gamma_l^2 l^{-\delta} \leq m^{-\delta} \sum_{l>m} \gamma_l^2.$$
Recalling the definition of $m = m(n)$, we obtain
$$\mathbb{B}_m^2 = o\left(\frac{1}{n}\right).$$

b) Term $U_n K$

One has $\mathbb{E}((U_n K)^2) = \mathbb{E}_1 + \mathbb{E}_2 + \mathbb{E}_3$ where
$$\begin{aligned}
\mathbb{E}_1 &= \frac{2}{(n(n-1))^2} \sum_{l_1,l_2=1}^{m} \frac{1}{\lambda_{l_1}\lambda_{l_2}} \sum_{1\leq i_1\neq j_1\leq n} \mathbb{E}\Big[(Y_{i_1} < X_{i_1}, \varphi_{l_1} > -\lambda_{l_1}\gamma_{l_1}) \\
&\quad (Y_{i_1} < X_{i_1}, \varphi_{l_2} > -\lambda_{l_2}\gamma_{l_2})(Y_{j_1} < X_{j_1}, \varphi_{l_1} > -\lambda_{l_1}\gamma_{l_1})(Y_{j_1} < X_{j_1}, \varphi_{l_2} > -\lambda_{l_2}\gamma_{l_2})\Big], \\
\mathbb{E}_2 &= \frac{4}{(n(n-1))^2} \sum_{l_1,l_2=1}^{m} \frac{1}{\lambda_{l_1}\lambda_{l_2}} \sum_{i_1,j_1,j_2 all\neq} \mathbb{E}\Big[(Y_{i_1} < X_{i_1}, \varphi_{l_1} > -\lambda_{l_1}\gamma_{l_1}) \\
&\quad (Y_{i_1} < X_{i_1}, \varphi_{l_2} > -\lambda_{l_2}\gamma_{l_2})(Y_{j_1} < X_{j_1}, \varphi_{l_1} > -\lambda_{l_1}\gamma_{l_1})(Y_{j_2} < X_{j_2}, \varphi_{l_2} > -\lambda_{l_2}\gamma_{l_2})\Big], \\
\mathbb{E}_3 &= \frac{1}{(n(n-1))^2} \sum_{l_1,l_2=1}^{m} \frac{1}{\lambda_{l_1}\lambda_{l_2}} \sum_{i_1,j_1,i_2,j_2 all\neq} \mathbb{E}\Big[(Y_{i_1} < X_{i_1}, \varphi_{l_1} > -\lambda_{l_1}\gamma_{l_1}) \\
&\quad (Y_{i_2} < X_{i_2}, \varphi_{l_2} > -\lambda_{l_2}\gamma_{l_2})(Y_{j_1} < X_{j_1}, \varphi_{l_1} > -\lambda_{l_1}\gamma_{l_1})(Y_{j_2} < X_{j_2}, \varphi_{l_2} > -\lambda_{l_2}\gamma_{l_2})\Big].
\end{aligned}$$

One easily see that

$$\mathbb{E}_2 = 0 \quad \text{and} \quad \mathbb{E}_3 = 0,$$

since the variables $(Y < X, \varphi_l > -\lambda_l \gamma_l)$ are centered and independent.

Let us now compute $\mathbb{E}_1$. Denote $\bar{Z}_{j,k} := Y_j < X_j, \varphi_k > -\lambda_k \gamma_k$.

$$\begin{aligned}
\mathbb{E}_1 &= \frac{2}{(n(n-1))^2} \sum_{l,k=1}^{m} \frac{1}{\lambda_l \lambda_k} \sum_{1 \le i \ne j \le n} \mathbb{E}\left[\bar{Z}_{1,l} \bar{Z}_{2,l} \bar{Z}_{1,k} \bar{Z}_{2,k}\right] \\
&= \frac{2}{n(n-1)} \sum_{l,k=1}^{m} \frac{1}{\lambda_l \lambda_k} \mathbb{E}\left[\bar{Z}_{1,l} \bar{Z}_{2,l} \bar{Z}_{1,k} \bar{Z}_{2,k}\right] = \frac{2}{n(n-1)} \sum_{l,k=1}^{m} \frac{1}{\lambda_l \lambda_k} \mathbb{E}^2\left[\bar{Z}_{1,l} \bar{Z}_{1,k}\right] \\
&\le \frac{2}{n(n-1)} \sum_{l,k=1}^{m} \frac{1}{\lambda_l \lambda_k} \mathbb{E}\left[(Y < X, \varphi_l >)^2\right] \mathbb{E}\left[(Y < X, \varphi_k >)^2\right] \\
&= \frac{2}{n(n-1)} \sum_{l,k=1}^{m} \mathbb{E}\left[Y^2 \xi_l^2\right] \mathbb{E}\left[Y^2 \xi_k^2\right] \le \frac{2}{n(n-1)} \mathbb{E}\left[Y^4\right] \sum_{l,k=1}^{m} \mathbb{E}\left[\xi_l^4\right]^{1/2} \mathbb{E}\left[\xi_k^4\right]^{1/2} \\
&\le \frac{2}{n(n-1)} \mathbb{E}\left[Y^4\right] \left(\sum_{l=1}^{m} \mathbb{E}\left[\xi_l^4\right]^{1/2} \right)^2
\end{aligned}$$

By assumption (5) we know that $\sup_k \mathbb{E}(\xi_k^4) \le K$, hence we have

$$\mathbb{E}_1 \le \frac{2Km^2}{n(n-1)} \mathbb{E}\left(Y^4\right)$$

and we still obtain that $\mathbb{E}\left((U_n K)^2\right) = o\left(\frac{1}{n}\right)$.

c) Term $P_n L$

First, let

$$\begin{aligned}
P_n L' &:= \frac{2}{n} \sum_{i=1}^{n} \left[Y_i < X_i, \beta > -\mathbb{E}(Y < X, \beta >) \right] \qquad (7) \\
&= \frac{2}{n} \sum_{i=1}^{n} \left[Y_i < X_i, \beta > - \sum_{l \ge 1} \gamma_l^2 \lambda_l \right] \qquad (8)
\end{aligned}$$

since

$$\mathbb{E}(Y < X, \beta >) = \mathbb{E}(< X, \beta >^2) = \mathbb{E}\left[\left(\sum_{l=1}^{\infty} \gamma_l \sqrt{\lambda_l} \xi_l \right)^2 \right] = \sum_{l=1}^{\infty} \gamma_l^2 \lambda_l,$$

we write

$$\begin{aligned} P_nL &= \frac{2}{n}\sum_{l=1}^{m}\sum_{i=1}^{n}\gamma_l\left(Y_i < X_i, \varphi_l > -\gamma_l\lambda_l\right) \\ &= \frac{2}{n}\sum_{i=1}^{n}\left(Y_i < X_i, \beta >_m - \sum_{l=1}^{m}\gamma_l^2\lambda_l\right) \\ &= P_nL' + (P_nL - P_nL'), \end{aligned}$$

where $< X_i, \beta >_m$ denotes $\sum_{l=1}^{m}\gamma_l < X_i, \varphi_l >$. In the one hand,

$$\begin{aligned} P_nL' - P_nL &= \frac{2}{n}\sum_{i=1}^{n}Y_i\left[< X_i, \beta > - < X_i, \beta >_m\right] - 2\sum_{l>m}\gamma_l^2\lambda_l \\ &= \frac{2}{n}\sum_{i=1}^{n}\sum_{l>m}\gamma_l\left[Y_i < X_i, \varphi_l > -\gamma_l\lambda_l\right] \end{aligned}$$

Let $Z := \sum_{l>m}\gamma_l\left[Y < X, \varphi_l > -\gamma_l\lambda_l\right]$ and let us give an upper bound for $\mathrm{Var}(Z)$.

$$\begin{aligned} \mathrm{Var}(Z) &= \mathrm{Var}\left(\sum_{l>m}\gamma_l\left[Y < X, \varphi_l > -\gamma_l\lambda_l\right]\right) \\ &= \mathrm{Var}\left(\sum_{l>m}\gamma_l Y < X, \varphi_l >\right) \\ &\leq \mathbb{E}\left(Y^2\left(\sum_{l>m}\gamma_l < X, \varphi_l >\right)^2\right) \\ &= \mathbb{E}\left(Y^2 < X, \beta >_{m^\perp}^2\right) \\ &\leq \mathbb{E}\left(Y^2||X||_{m^\perp}^2\right)||\beta||_{m^\perp}^2 \quad \text{by Cauchy-Schwarz inequality} \\ &\leq \mathbb{E}\left(Y^2||X||^2\right)||\beta||_{m^\perp}^2, \end{aligned}$$

where $< X, \beta >_{m^\perp}$ and $|\beta||_{m^\perp}$ respectively denote $\sum_{l>m}\gamma_l < X_i, \varphi_l >$ and $\sum_{l>m}\gamma_l^2$. The first expectation does not depend on n and the second term tends to 0 when n (and therefore m) tends to ∞. As a consequence,

- $\mathrm{Var}(Z) = o(1)$,
- $\mathrm{Var}(P_nL - P_nL') = o\left(\frac{1}{n}\right)$,
- $P_nL - P_nL' = o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$.

In the other hand, we establish a central limit theorem for P_nL', since it is an empirical sum of iid centered variables. As a conclusion,

$$\sqrt{n}P_nL' \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0, 4\mathrm{Var}(Y < X, \beta >)) \tag{9}$$

and by Slutsky theorem since $P_nL - P_nL' = o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$,

$$\sqrt{n}P_nL \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0, 4\mathrm{Var}(Y < X, \beta >)). \tag{10}$$

This concludes the proof. □

Our aim is to estimate the Sobol index S^X which is defined by $S^X = V^X/\mathrm{Var}(Y)$. It is therefore natural to introduce the estimator of S^X defined by $\widehat{V}_m^X/\widehat{V}$, where $\widehat{V}$ is the empirical variance estimating $\mathrm{Var}(Y)$:

$$\widehat{V} := \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y}_n)^2.$$

In the following theorem, we obtain the asymptotic behaviour of this estimator.

Theorem 3. *Under the same assumptions as in Theorem 2, we have*

$$\sqrt{n}\left(\frac{\widehat{V}_m^X}{\widehat{V}} - S^X\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, \frac{\mathrm{Var}(U)}{(\mathrm{Var}(Y))^2}\right) \tag{11}$$

where $U := 2Y < X, \beta > -S^X(Y - \mathbb{E}(Y))^2$.

Proof. (i) First, let $W_i = (2\left[Y_i < X_i, \beta > -\mathbb{E}(Y < X, \beta >)\right], Y_i^2, Y_i)^t$ $(i = 1, \ldots, n)$ and $W = (2\left[Y < X, \beta > -\mathbb{E}(Y < X, \beta >)\right], Y^2, Y)^t$. Then

$$\overline{W}_n = \frac{1}{n}\sum_{i=1}^{n} W_i = (P_nL', \overline{Y^2}_n, \overline{Y}_n) \quad \text{and} \quad \mathbb{E}_W := \mathbb{E}(W) = (0, \mathbb{E}(Y^2), \mathbb{E}(Y)).$$

Now from the Central Limit Theorem,

$$\sqrt{n}\left(\overline{W}_n - \mathbb{E}_W\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_3\left(0, \Sigma\right)$$

with

$$\Sigma = \begin{pmatrix} 4\mathrm{Var}(Y < X, \beta >) & 2\mathrm{Cov}(Y_i < X, \beta >, Y^2) & 2\mathrm{Cov}(Y < X, \beta >, Y) \\ 2\mathrm{Cov}(Y < X, \beta >, Y^2) & \mathrm{Var}(Y^2) & \mathrm{Cov}(Y^2, Y) \\ 2\mathrm{Cov}(Y < X, \beta >, Y) & \mathrm{Cov}(Y^2, Y) & \mathrm{Var}(Y) \end{pmatrix}$$

(ii) Then let $W_i' = W_i + (U_nK + (P_nL - P_nL') + \sum_{l=1}^{m}\gamma_l^2\lambda_l, 0, 0)$. We have

$$\overline{W}_n' = \frac{1}{n}\sum_{i=1}^{n} W_i' = (\widehat{V}_m^X, \overline{Y^2}_n, \overline{Y}_n).$$

Since, $U_nK + (P_nL - P_nL') + \sum_{l>m}\gamma_l^2\lambda_l = o_{\mathbb{P}}\left(\frac{1}{n}\right)$, we still have a Central Limit Theorem

$$\sqrt{n}\left(\overline{W}_n' - \mathbb{E}_{W'}\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_3\left(0, \Sigma\right)$$

where $\mathbb{E}_{W'} = (\mathbb{E}(Y < X, \beta >), \mathbb{E}(Y^2), \mathbb{E}(Y))$.
(iii) Let Φ the mapping from $\mathbb{R}^3$ to $\mathbb{R}$ defined by $\Phi(x, y, z) = \frac{x}{y - z^2}$. We want to apply the Delta method (cf. [24]) to W' and Φ. Easily,

$$\begin{cases} \frac{\partial\Phi}{\partial x}(x, y, z) = \frac{1}{y-z^2} & \frac{\partial\Phi}{\partial x}(\mathbb{E}_{W'}) = \frac{1}{\mathrm{Var}(Y)} \\ \frac{\partial\Phi}{\partial y}(x, y, z) = -\frac{x}{(y-z^2)^2} & \frac{\partial\Phi}{\partial y}(\mathbb{E}_{W'}) = -\frac{\mathbb{E}(Y<X,\beta>)}{(\mathrm{Var}(Y))^2} \\ \frac{\partial\Phi}{\partial z}(x, y, z) = \frac{2xz}{(y-z^2)^2} & \frac{\partial\Phi}{\partial x}(\mathbb{E}_{W'}) = \frac{2\mathbb{E}(Y<X,\beta>)\mathbb{E}(Y)}{(\mathrm{Var}(Y))^2} \end{cases}$$

so that

$$\sqrt{n}\left(\Phi(\overline{W}'_n) - \Phi(\mathbb{E}_{W'})\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}\left(0, \Phi'(\mathbb{E}_{W'})\Sigma\Phi'(\mathbb{E}_{W'})^t\right).$$

But $\Phi(\overline{W}_n) = \frac{\widehat{V}_m^X}{\widehat{V}}$, $\Phi(\mathbb{E}_{W'}) = S^X$ and

$$\begin{aligned}\Phi'(\mathbb{E}_{W'})\Sigma\Phi'(\mathbb{E}_{W'})^t &= 4\frac{\mathrm{Var}(Y<X,\beta>)}{(\mathrm{Var}(Y))^2} + 2\frac{\mathbb{E}(Y<X,\beta>)}{(\mathrm{Var}(Y))^3}\Big[4\mathbb{E}(Y)\mathrm{Cov}(Y<X,\beta>,Y)\\ &\quad -2\mathrm{Cov}(Y<X,\beta>,Y^2) + 2(\mathbb{E}(Y))^2\mathbb{E}(Y<X,\beta>)\Big]\\ &\quad +\frac{(\mathbb{E}(Y<X,\beta>))^2}{(\mathrm{Var}(Y))^4}\left[\mathrm{Var}(Y^2) - 4\mathbb{E}(Y)\mathrm{Cov}(Y^2,Y)\right]\end{aligned}$$

One can check that

$$\Phi'(\mathbb{E}_{W'})\Sigma\Phi'(\mathbb{E}_{W'})^t = \frac{\mathrm{Var}(U)}{(\mathrm{Var}(Y))^2}.$$

□

Remark 2.1. *1. In the case where the variance of Y is known and plug in the estimator, the result is*

$$\sqrt{n}\left(\frac{\widehat{V}_m^X}{\mathrm{Var}(Y)} - S^X\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}(0, 4\frac{\mathrm{Var}(Y<X,\beta>)}{(\mathrm{Var}(Y))^2}). \tag{12}$$

2. In the general case where the variance of Y is unknown and Y is centered, we thus have an extra term (due to the estimation of the variance), namely the difference between the two asymptotic variances is

$$\frac{\mathrm{Var}(U) - 4\mathrm{Var}(Y<X,\beta>)}{(\mathrm{Var}(Y))^2} =$$

$$\frac{\mathbb{E}(Y<X,\beta>)}{(\mathrm{Var}(Y))^4}\Big[\mathbb{E}(Y<X,\beta>)\mathrm{Var}(Y^2) - 4\mathrm{Var}(Y)\mathrm{Cov}(Y<X,\beta>,Y^2)\Big]$$

It is worth to notice that this term is not always positive. Namely, if $<X,\beta>$ is a centered variable with second moment s^2 and fourth moment $k_1 s^4$ and if ε is a centered variable with second moment η^2 and fourth moment $k_2\eta^4$, then the latter extra term is

$$-\frac{6s^4}{(s^2+\eta^2)^2} + \frac{s^4}{(s^2+\eta^2)^4}((3-k_1)(3s^4+4\eta^2s^2) - (3-k_2)\eta^4).$$

In particular, when $<X,\beta> \overset{\mathcal{L}}{\sim} \mathcal{N}(0,s^2)$ and $\epsilon \overset{\mathcal{L}}{\sim} \mathcal{N}(0,\eta^2)$, $k_1 = k_2 = 3$ and one gets an extra term equal to

$$\frac{-6s^4(s^2+\eta^2)^2}{(\mathrm{Var}(Y))^4}.$$

In this particular case, the asymptotic variance of the estimator of the Sobol index is smaller when the variance of Y is estimated that when it is known and plug in the estimator.

2.2 Asymptotic efficiency

In this section we prove that $\widehat{V}_m^X/\widehat{V}$ is asymptotically efficient for estimating the Sobol index S^X (see [24], Section 25 for the definition of asymptotic efficiency). This notion somewhat extends the notion of Cramér-Rao bound and enables to define a criterion of optimality for estimators, called asymptotic efficiency.

Proposition 2.1 (Asymptotic efficiency)**.** *Under the same assumptions as in Theorem 2, $\widehat{V}_m^X/\widehat{V}$ is asymptotically efficient for estimating S.*

Proof. Note that

$$S = \Phi(V^X, \mathrm{Var}(Y)) \quad \text{and} \quad \frac{\widehat{V}_m^X}{\widehat{V}} = \Phi(\widehat{V}_m^X, \widehat{V})$$

where Φ is defined by $\Phi(x, y) = x/y$.
First let us prove that $\widehat{V}_m^X$ is asymptotically efficient for estimating V^X. The result directly comes from decomposition (4) and the Hoeffding's decomposition $\widehat{V}_m^X - V^X = U_n K + P_n L - \mathbb{B}_m$, $P_n L$ is asymptotically efficient (example 25.24 [efficiency of the empirical distribution]), the other terms are $o_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right)$. From which we conclude the asymptotic efficiency of $\widehat{V}_m^X$.

Second using again example 25.24, we get the asymptotic efficiency of $\widehat{V}$ for estimating $\mathrm{Var}(Y)$.
Then theorem 25.50 (efficiency in product space) in [24] gives that $(\widehat{V}_m^X, \widehat{V})$ is asymptotically efficient for estimating $(V^X, \mathrm{Var}(Y))$.
Now since Φ is differentiable in $\mathbb{R}^2 \setminus \{y = 0\}$ then by theorem 25.47 (efficiency and Delta method), $(\Phi(\widehat{V}_m^X, \widehat{V}))_n$ is asymptotically efficient for estimating $\Phi(V^X, \mathrm{Var}(Y))$. □

3 Multiple functional linear regression model

We recall the model (1):

$$Y = \mu + \sum_{k=1}^{p} < \beta^k, X^k > + \varepsilon.$$

In this setting we observe a n-sample $(Y_i, X_i^1, \ldots, X_i^p)$ of $(Y, X^1, \ldots, X^p)$.

3.1 Estimation by U-statistics of order 2

We assume that the processes $\left(X^k\right)_{1 \le k \le p}$ are independent. Under this assumption, the previous section whole applies and the Sobol indices can be estimated from the U statistics of order 2. For all $m \ge 1$, let

$$\widehat{V}_m^k = \sum_{l=1}^{m} \frac{1}{\lambda_l^k} \frac{1}{n(n-1)} \sum_{1 \le i \ne j \le n} < X_i^k, \varphi_l > Y_i < X_j^k, \varphi_l > Y_j.$$

For a suitable choice of $m = m(n)$, $\widehat{V}_m^k$ is a consistent estimators of V^k. Actually by independence

$$\mathbb{E}(Y | X^k) =< \beta^k, X^k >,$$

and denoting $\tilde{\varepsilon} = \sum_{j=1, j\neq k}^{p} < \beta^j, X^j > +\varepsilon$, we retrieve the simple regression model.
In fact we can obtain a central limit theorem for $\widehat{V}_m$ where

$$\widehat{V}_m = (\widehat{V}_m^1, \widehat{V}_m^2, \ldots, \widehat{V}_m^p)^T. \tag{13}$$

Let $V^j = \mathrm{Var}(\mathbb{E}(Y|X^j))$, $j = 1 \ldots p$ and $V = (V^1, V^2, \ldots, V^p)^T$.

Theorem 4. *Let $(Y_i, X_i^1, \ldots, X_i^p)$ be i.i.d. observations from Model (1). We assume that for all $1 \leq j \leq p$, $E(\|X^j\|^4) < +\infty$ and that $E(\varepsilon^4) < +\infty$. We consider the Karhunen-Loève expansion of X^j :*

$$X^j = \sum_{l \geq 1} \sqrt{\lambda_l^j} \xi_l^j \varphi_l^j.$$

We assume

$$\forall 1 \leq j \leq p, \sup_{l \geq 1} E((\xi_l^j)^4) < +\infty. \tag{14}$$

We consider the estimator $\widehat{V}_m$ of V defined by (13) with $m = m(n) = \sqrt{n}h(n)$, where $h(n)$ satisfies : $h(n) \to 0$ as $n \to +\infty$ and $\forall \alpha > 0$, $n^\alpha h(n) \to +\infty$ as $n \to +\infty$.
We assume that

$$\forall\ 1 \leq j \leq p,\ \forall l \geq 1,\ \lambda_l^j \leq C_j l^{-\delta_j}$$

for some $C_j > 0$ and $\delta_j > 1$.
The following result holds :

$$\sqrt{n}(\widehat{V}_m - V) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_p\left(0, 4\left(\mathrm{Cov}(Y < X^k, \beta^k >, Y < X^l, \beta^l >)\right)_{k,l=1\ldots p}\right). \tag{15}$$

Proof. Using the same decomposition

$$\widehat{V}_m - V = U_n K + P_n L - \mathbb{B}_m$$

and the fact that $\|U_n K\|$, $\|P_n L - P_n L'\|$ and $\|\mathbb{B}_m\|^2$ are $o_{\mathbb{P}}\left(\frac{1}{n}\right)$ and $o\left(\frac{1}{n}\right)$, we just need to check that $P_n L' = (P_n L'^1, P_n L'^2, \ldots, P_n L'^p)^T$ converges in distribution towards a $\mathcal{N}_p\left(0, 4\left(\mathrm{Cov}(Y < X^k, \beta^k >, Y < X^l, \beta^l >)\right)_{k,l=1\ldots p}\right)$.
We have

$$P_n L' = \frac{2}{n} \sum_{i=1}^{n} W_i$$

where $W_i = (W_i^1, \ldots, W_i^p)^T$ and $W_i^k = Y_i < X_i^k, \beta^k > -\mathbb{E}(Y_i < X_i^k, \beta^k >)$. Since the W_i's are i.i.d., the result follows from the central limit theorem in $\mathbb{R}^p$. □

Now define $\widehat{S}_m$ as $\widehat{V}_m/\widehat{V} = (\widehat{V}_m^1/\widehat{V}, \ldots, \widehat{V}_m^p/\widehat{V})^T$, where $\widehat{V}$) denotes the empirical estimator of the $\mathrm{Var}(Y)$. Let us remind that $S = V/\mathrm{Var}(Y) = (S^1, \ldots, S^p)^T$. Using once again the Delta method, we easily get the following result.

Proposition 3.1. *Under the same assumptions as in Theorem 4, we have*

$$\sqrt{n}\left(\widehat{S}_m - S\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_p\left(0, \left(\frac{\mathrm{Cov}(U^k, U^l)}{(\mathrm{Var}(Y))^2}\right)_{k,l=1\ldots p}\right). \tag{16}$$

where $U^k := 2(Y - \mathbb{E}(Y)) < X^k, \beta^k > -S^k(Y - \mathbb{E}(Y))^2$.

3.2 Comparison with Sobol estimators

Sobol [21] proposed an empirical method, based on a particular design of experiments, to estimate Sobol indices. This method called Sobol Pick and Freeze (SPF) is also studied in [18]. In [9, 5], the authors prove asymptotic normality and efficiency of Sobol estimators. In this section, our aim is to compare the method proposed in this paper, which is based on U statistics of order 2, and the method proposed by Sobol in order to see which experimental set up should be used by the practitionner. It is important to notice that the two procedures are not based on the same design of experiments. Both methods lead to asymptotically normal and efficient estimators, but the asymptotic variances are different due to the designs of experiments. We therefore want to compare the asymptotic variances obtained by the two methods, for a similar number of experiments in both cases. Let us first recall the design of experiments and the estimators proposed by Sobol.
Let (X, Y) obey to the model (1), where $X = (X^1, \dots, X^p)$. Let $X' = (X'^1, \dots, X'^p)$ an i.i.d. copy of X. For all $k \in \{1, \dots, p\}$, let Y^k be defined by

$$Y^k =< X^k, \beta^k > + \sum_{j=1, j\neq k}^{p} < X'^j, \beta^j > +\varepsilon',$$

where ε' is an i.i.d. copy of ε. Let $(Y_1, \dots, Y_n)$ be i.i.d. copies of Y and for all $k \in \{1, \dots, p\}$ let $(Y_1^k, \dots, Y_n^k)$ be i.i.d. copies of Y^k. The estimator proposed by Sobol to estimate $V^k = \mathrm{Var}(\mathbb{E}(Y|X^k))$ is defined by

$$\widehat{V}_{SPF}^k = \frac{1}{n}\sum_{i=1}^{n} Y_i Y_i^k - \left(\frac{1}{n(p+1)}\sum_{i=1}^{n}\left[Y_i + Y_i^1 + \ldots + Y_i^p\right]\right)^2.$$

Now let $\widehat{S}_{SPF} = \widehat{V}_{SPF}/\widetilde{V} = (\widehat{V}_{SPF}^1/\widetilde{V}, \dots, \widehat{V}_{SPF}^p/\widetilde{V})^T$, where $\widetilde{V}$ is the estimator of $\mathrm{Var}(Y)$ in the S.P.F. method defined by

$$\widetilde{V} = \frac{1}{n(p+1)}\sum_{i=1}^{n}\left((Y_i)^2 + \left(Y_i^1\right)^2 + \ldots + (Y_i^p)^2\right) - \left(\frac{1}{n(p+1)}\sum_{i=1}^{n}\left[Y_i + Y_i^1 + \ldots + Y_i^p\right]\right)^2.$$

Theorem 5. *We have (see [9, 5])*

$$\sqrt{n}\left(\widehat{S}_{SPF} - S\right) \underset{n\to\infty}{\overset{\mathcal{L}}{\to}} \mathcal{N}_p\left(0, \left(\frac{\mathrm{Cov}(V^k, V^l)}{(\mathrm{Var}(Y))^2}\right)_{k,l=1\dots p}\right). \tag{17}$$

where $V^k := YY^k - S^k\left[Y^2 + \sum_{k=1}^{p}(Y^k)^2\right]/(p+1)$.

The asymptotic variance appearing in the central limit theorem for the estimator $\widehat{S}_{SPF} = (\widehat{S}_{SPF}^1, \dots, \widehat{S}_{SPF}^p)$ (resp. $\widehat{S}_m = (\widehat{S}_m^1, \dots, \widehat{S}_m^p)$) is denoted by $(\mathrm{Var}(Y))^{-2}\Gamma_{SPF}$ (resp. $(\mathrm{Var}(Y))^{-2}\Gamma$ (cf. (15)). Our aim is to compare Γ and Γ_{SPF}.

For sake of simplicity we assume $\varepsilon = 0$. Then we define $W_i =< X^i, \beta^i >$, $\sigma_i^2 = \mathrm{Var}(W_i)$, $\sigma^2 = \sum_{i=1}^{p}\sigma_i^2$ and $S^i = \sigma_i^2/\sigma^2$. After computations, one obtains

that Γ is defined by

$$\begin{aligned}
\forall k,\ \Gamma(k,k) &= \frac{\sigma_k^4}{\sigma^4}\sum_{i=1}^{p}\left[\mathbb{E}(W_i^4)-3\sigma_i^4\right]+4\frac{\sigma^2-\sigma_k^2}{\sigma^2}\mathbb{E}(W_k^4)+4\sigma^2\sigma_k^2+12\frac{\sigma_k^6}{\sigma^2}-14\sigma_k^4\\
\forall k\neq l,\ \Gamma(k,l) &= -\frac{2}{\sigma^2}\left[\mathbb{E}(W_k^4)\sigma_l^2+\mathbb{E}(W_l^4)\sigma_i^2\right]+2\sigma_k^2\sigma_l^2\left[3\frac{\sigma_k^2+\sigma_l^2}{\sigma^2}-1\right]\\
&\quad+\frac{\sigma_k^2\sigma_l^2}{\sigma^4}\sum_{i=1}^{p}\left[\mathbb{E}(W_i^4)-3\sigma_i^4\right]
\end{aligned}$$

and Γ_{SPF} by

$$\begin{aligned}
\forall k,\ \Gamma_{SPF}(k,k) &= \left(1-\frac{4}{p+1}\left(\frac{\sigma_k^2}{\sigma^2}\right)\right)\mathbb{E}(W_k^4)+\frac{\sigma_k^4}{\sigma^4}\left(\frac{p^2-2p+5}{(p+1)^2}\right)\sum_{i=1}^{p}\mathbb{E}(W_i^4)\\
&\quad+\left(\frac{\sigma_k^2}{(p+1)\sigma^2}\right)\left(4+\frac{\sigma_k^2}{\sigma^2}\left(\frac{-2p^2+5p-13}{p+1}\right)\right)\sum_{i=1}^{p}\sigma_i^4\\
&\quad+\sigma^4+\left(\frac{-p^2-7p+6}{(p+1)^2}\right)\sigma_k^4+\frac{4}{p+1}\frac{\sigma_k^6}{\sigma^2}-\frac{4}{p+1}\sigma_k^2\sigma^2\\
\forall k\neq l,\ \Gamma_{SPF}(k,l) &= -\frac{2}{p+1}\left(\frac{\sigma_k^2\mathbb{E}(W_l^4)+\sigma_l^2\mathbb{E}(W_k^4)}{\sigma^2}\right)\\
&\quad+\left(\frac{\sigma_k^2\sigma_l^2}{\sigma^4}\right)\left(\frac{p^2-2p+5}{(p+1)^2}\right)\sum_{i=1}^{p}\mathbb{E}(W_i^4)\\
&\quad+\left(2(\sigma_k^2+\sigma_l^2)+\frac{-2p^2+5p-13}{(p+1)^2\sigma^4}\right)\sum_{i=1}^{p}\sigma_i^4\\
&\quad+\sigma^4+\sigma_k^2\sigma_l^2\left(\frac{2p^2-p+9}{(p+1)^2}\right)-\frac{p+3}{p+1}\sigma^2(\sigma_k^2+\sigma_l^2)\\
&\quad+\frac{2}{p+1}\frac{\sigma_k^2\sigma_l^2}{\sigma^2}(\sigma_k^2+\sigma_l^2).
\end{aligned}$$

Sobol experiment requires $(p+1)n$ observations (or computations of the black-box code) to estimate the p indices. In order to have a fair comparison of both methods, we consider that we have $n(p+1)$ i.i.d. observations from Model (1) to estimate the Sobol indices by our methods. With $n(p+1)$ observations instead of n, the asymptotic variance matrix Γ_m is devided by $p+1$, hence, we have to study the matrix

$$D := (p+1)\Gamma_{SPF}-\Gamma.$$

In order to compare the two methods, we evaluate the eigenvalues of the matrix D in order to determine whether it is positive-definite or not. If not we study only the sign of the diagonal terms of D that correspond to the difference of the asymptotic variances obtained with both methods.

First example: We asssume that for $i=1,\ldots,p$ with $p\geq 2$

$$< X^i,\beta^i > \sim \mathcal{N}(0,1).$$

We have

$$D_{i,i} = \frac{p^5 + 2p^4 - 5p^3 - 2p^2 + 6p - 6}{p(p+1)} > 0 \qquad \forall \quad 1 \leq i \leq p,$$
$$D_{i,j} = \frac{p^5 + 4p^4 + 3p^3 + 2p - 6}{p(p+1)} \qquad \forall \quad 1 \leq i \neq j \leq p.$$

Let $\alpha = p^5 + 2p^4 - 5p^3 - 2p^2 + 6p - 6$, $\beta = p^5 + 4p^4 + 3p^3 + 2p - 6$. The eigenvalues of D are $[\alpha - \beta]/[p(p+1)]$ of order $p-1$ and $[\alpha + (p-1)\beta]/[p(p+1)]$ of order 1. But

$$\alpha - \beta = 2p(-p^3 - 4p^2 - p + 2) < 0 \quad \forall p \geq 2,$$

and

$$\alpha - \beta + p\beta = p(p^5 + 4p^4 + p^3 - 8p^2 - 2) > 0 \quad \forall p \geq 2.$$

For $p \geq 2$, the matrix D is not positive nor negative. Hence, as mentioned before, we simply study the sign of the diagonal terms of the matrix D. Since $D_{i,i} > 0$ for any $p \geq 2$ and $i = 1 \dots p$, the asymptotic variance of each estimator of Sobol indices with SPF method is larger than the one obtained by the method proposed in this paper using the same number of observations.

Second example : In the case $p = 2$ and $\sigma^2 = \sigma_1^2 + \sigma_2^2 = 1$, the model is

$$Y = < X^1, \beta^1 > + < X^2, \beta^2 > .$$

Then if $W_1 = < X^1, \beta^1 >$ and $W_2 = < X^2, \beta^2 >$,

$$D_{1,1} = \mathbb{E}(W_1^4)\left(-1 + \frac{2}{3}\sigma_1^4\right) + \frac{2}{3}\sigma_1^4\mathbb{E}(W_2^4) + 3 - 8\sigma_1^6 + 10\sigma_1^4 - 8\sigma_1^2 + \sigma_1^2(\sigma_1^4 + \sigma_2^4)\left(4 - \frac{2}{3}\sigma_1^2\right)$$
$$D_{2,2} = \mathbb{E}(W_2^4)\left(-1 + \frac{2}{3}\sigma_2^4\right) + \frac{2}{3}\sigma_2^4\mathbb{E}(W_1^4) + 3 - 8\sigma_2^6 + 10\sigma_2^4 - 8\sigma_2^2 + \sigma_2^2(\sigma_1^4 + \sigma_2^4)\left(4 - \frac{2}{3}\sigma_2^2\right)$$
$$D_{2,1} = \sigma_1^2\sigma_2^2\left(\frac{2}{3}(\mathbb{E}(W_1^4) + \mathbb{E}(W_2^4)) + 3\right) + \left(\sigma_1^4 + \sigma_2^4\right)\left(\frac{7}{3} + 3\sigma_1^2\sigma_2^2\right) - 2$$

In the particular case where $W_1 \sim \mathcal{N}(0, x)$ and $W_2 \sim \mathcal{N}(0, 1-x)$, for some $x \in]0, 1[$, we obtain

$$D_{1,1}(x) = 3 - 4x + \frac{1}{3}\left(x^2 - 8x^3 + 8x^4\right),$$
$$D_{2,2}(x) = 3 - 4(1-x) + \frac{1}{3}\left((1-x)^2 - 8(1-x)^3 + 8(1-x)^4\right),$$
$$D_{1,2}(x) = \frac{1}{3} + \frac{10}{3}x - \frac{40}{3}x^2 + 20x^3 - 10x^4.$$

Let $\lambda_1(x)$ and $\lambda_2(x)$ be the eigenvalues of $D(x)$ and x_1 and $1 - x_1$ the real zeros of its determinant ($x_1 \approx 0.6701$). Then

x	0		$1 - x_1$		x_1		1
$\lambda_1(x) + \lambda_2(x)$	+	|	+	|	+		
$\lambda_1(x)\lambda_2(x)$	-		0	+	0		-

For $x \in]x_1, x_2[$, the matrix $D(x)$ is positive-definite. We can also study the signs of the diagonal terms of the matrix $D(x)$ for $x \in]0, 1[$.

Now let x_0 the zero of $D_{1,1}(x)$ ($x_0 \approx 0.6738$). The folowing tabular gives the signs of the diagonal terms of the matrix $D(x)$:

x	0		$1-x_0$		x_0	1
$D_{1,1}(x)$	+	|		+	0	-
$D_{2,2}(x)$	-		0	+	|	+

We deduce from this tabular that for $x \in]1-x_0, x_0[$, our method leads to smaller variances for both estimators.

4 Numerical experiments

We perform a simulation study to evaluate the performances of the procedure proposed in this paper for estimating Sobol indices in a functional linear model, and to compare this procedure with Sobol's procedure.
We consider the model :

$$Y_i = \sum_{k=1}^{p} < \beta^k, X_i^k > + \varepsilon_i, \tag{18}$$

where for all $k = 1, \ldots, p$, X^k and β^k are defined from the coefficients of their expansions onto an orthonormal basis of $\mathbb{L}^2([0,1])$. For all k, we consider the basis corresponding to the eigenfunctions of the Karhunen-Loève expansion of the process X^k and we define the function β^k by the coefficients of its expansion onto this basis. For the simulations, the basis that we consider is either the one associated to the Karhunen-Loève expansion of the Brownian motion (that has been recalled in Section 2) or the one associated to the fractional Brownian motion. It is important to recall that the processes X^k and the related functions β^k are expanded onto the same basis and that the estimators proposed in this paper will perform well in the case where the coefficients of the functions β^k in those basis are non-increasing. The variables ε_i are i.i.d. Gaussian centered variables with variance σ^2.

4.1 First example

4.1.1 Using Karhunen-Loève expansion of the Brownian motion

We consider the model (18) with $p = 2$ and $\varepsilon_i = 0$ for all i. The processes X^1 and X^2 are given by a truncated expansion of the Brownian motion onto its Karhunen-Loève basis, with coefficients:

$$\lambda_l = \frac{1}{\pi(l-1/2)^2}, \ 1 \leq l \leq L;$$
$$\lambda_l = 0, \ l > L.$$

Concerning the functions β^1, β^2, they are respectively caracterized by the coefficients $(\gamma_l)_{l \geq 1}$ of their expansion onto the same basis:

1. First Model: let $(\gamma_l^1)_l$ and $(\gamma_l^2)_l$ be defined as
$$\gamma_l^1 = l^{-1/2-1/100}, \quad 1 \le l \le L, \quad \gamma_l^1 = 0, \quad l > L;$$
$$\gamma_l^2 = l^{-1}, \quad 1 \le l \le L, \quad \gamma_l^2 = 0, \quad l > L.$$

2. Second Model: we suppress the two first coefficients of γ defined as in the first model.
Let $(\gamma_l'^1)_l$ and $(\gamma_l'^2)_l$ be defined by $\gamma_l''^1 = \gamma_{l+2}^1$ and $\gamma_l''^2 = \gamma_{l+2}^2$.

3. Third Model: we replace the fourth first coefficients of γ defined as in the first model by 0.
Let $(\gamma_l''^1)_l$ and $(\gamma_l''^2)_l$ be defined by $\gamma_i''^1 = \gamma_i''^2 = 0$ for $i = 1, \dots 4$ and for $l \ge 5$ by $\gamma_l''^1 = \gamma_l^1$ and $\gamma_l''^2 = \gamma_l^2$.

We denote by $S = (S^1, S^2)$ the vector of Sobol indices. We perform $N_{sim} = 5000$ simulations, we set $L = 100$ and we study the influence of the parameter n with $n = 10^2$ or 10^3. We compare the estimator $\hat{S}_m$ of the vector S defined in Section 3.1 with the SPF estimator defined in Section 3.2. Both estimators are based on $3n$ observations. The observations are i.i.d. for the estimator $\hat{S}_m$ and obey to the design described in Section 3.2 for $\hat{S}_{SPF}$. We set $m = \lfloor \sqrt{3n/\log 3n} \rfloor$ (i.e. $m = 7$ or 19) in the definition of the estimator $\hat{S}_m$.

In the following tabulars we report an empirical estimator of the bias ($\mathrm{Bias}(\hat{S}_m)$ and $\mathrm{Bias}(\hat{S}_{SPF})$), of the standard deviation ($\mathrm{Std}(\hat{S}_m)$ and $\mathrm{Std}(\hat{S}_{SPF})$) and of the square root of the Mean Squared Error ($RMSE(\hat{S}_m)$ and $RMSE(\hat{S}_{SPF})$) for each estimator.

First Model: $S = [0.5107, 0.4893]$				
n	$\mathrm{Bias}(\hat{S}_m)$	$\mathrm{Bias}(\hat{S}_{SPF})$	$\mathrm{Std}(\hat{S}_m)$	$\mathrm{Std}(\hat{S}_{SPF})$
10^2	$10^{-3}[3.65, 3.98]$	$10^{-3}[5.42, 6.34]$	$10^{-2}[7.16, 7.20]$	$10^{-2}[8.94, 9.12]$
10^3	$10^{-4}[6.80, 0.55]$	$10^{-4}[5.38, 3.79]$	$10^{-2}[2.26, 2.20]$	$10^{-2}[2.79, 2.83]$
Second Model: $S = [0.7535, 0.2465]$				
n	$\mathrm{Bias}(\hat{S}_m)$	$\mathrm{Bias}(\hat{S}_{SPF})$	$\mathrm{Std}(\hat{S}_m)$	$\mathrm{Std}(\hat{S}_{SPF})$
10^2	$10^{-3}[7.27, 0.37]$	$10^{-3}[6.07, 3.61]$	$10^{-2}[8.04, 5.45]$	$10^{-2}[7.78, 9.90]$
10^3	$10^{-4}[7.76, 7.04]$	$10^{-4}[7.58, 3.52]$	$10^{-2}[2.52, 1.71]$	$10^{-2}[2.42, 3.13]$
Third Model: $S = [0.8655, 0.1345]$				
n	$\mathrm{Bias}(\hat{S}_m)$	$\mathrm{Bias}(\hat{S}_{SPF})$	$\mathrm{Std}(\hat{S}_m)$	$\mathrm{Std}(\hat{S}_{SPF})$
10^2	$10^{-1}[2.91, 0.29]$	$10^{-3}[2.44, 7.73]$	$10^{-2}[7.56, 3.89]$	$10^{-2}[7.11, 9.94]$
10^3	$10^{-2}[3.93, 0.17]$	$10^{-4}[5.71, 6.25]$	$10^{-2}[2.52, 1.27]$	$10^{-2}[2.24, 3.17]$

First Model: $S = [0.5107, 0.4893]$		
n	$\mathrm{RMSE}(\hat{S}_m)$	$\mathrm{RMSE}(\hat{S}_{SPF})$
10^2	$10^{-2}[7.17, 7.21]$	$10^{-2}[8.95, 9.14]$
10^3	$10^{-2}[2.26, 2.20]$	$10^{-2}[2.79, 2.83]$
Second Model: $S = [0.7535, 0.2465]$		
n	$\mathrm{RMSE}(\hat{S}_m)$	$\mathrm{RMSE}(\hat{S}_{SPF})$
10^2	$10^{-2}[8.07, 5.45]$	$10^{-2}[7.80, 9.90]$
10^3	$10^{-2}[2.52, 1.71]$	$10^{-2}[2.41, 3.13]$
Third Model: $S = [0.8655, 0.1345]$		
n	$\mathrm{RMSE}(\hat{S}_m)$	$\mathrm{RMSE}(\hat{S}_{SPF})$
10^2	$10^{-1}[3.01, 0.48]$	$10^{-2}[7.12, 9.97]$
10^3	$10^{-2}[4.67, 1.28]$	$10^{-2}[2.24, 3.17]$

Those simulations show that as expected the method of the paper provides better MSE as soon as the signal is concentrated on the first terms of the basis

which is considered. If it is not the case as in the third model, the SPF can be more efficient. Models 2 and 3 are unfavorable to our estimator, but it is worth to notice that in many cases, considering the linear functional regression model, when a signal has weak components on the first elements of the karhunen-Loève basis it tends to make the influence on the model (Sobol index) smaller.

4.1.2 Using Karhunen-Loève expansion of the fractional Brownian motion

We consider the model (18) with $p = 2$ and $\varepsilon_i = 0$ for all i. The processes X^1 and X^2 are defined by the coefficients of their expansion onto the Karhunen-Loève basis of the fractional Brownian motion (we use an approximation of the eigenvalues) :

$$\lambda_l = \frac{\sin(\pi H)\Gamma(2H+1)}{(l\pi)^{2H+1}},\ 1 \le l \le L;$$
$$\lambda_l = 0,\ l > L.$$

Note that no close form of the Karhunen-Loève basis of the fractional Brownian motion is known but we don't need it for the simulations, we simply use the fact that $\langle X^i, \varphi_l \rangle = \sqrt{\lambda_l}\xi_l^i$, where the variables $(\xi_l^i)_{l\ge 1}$, $i = 1, 2$ are i.i.d. standard normal variables.
Concerning the functions β^1, β^2, they are respectively caracterized by the coefficients of their expansion onto the same basis:

1. First Model: let $(\gamma_l^1)_l$ and $(\gamma_l^2)_l$ be defined as
$$\gamma_l^1 = l^{-1/2-1/100}, \quad 1 \le l \le L, \quad \gamma_l^1 = 0, \quad l > L;$$
$$\gamma_l^2 = l^{-1}, \quad 1 \le l \le L, \quad \gamma_l^2 = 0, \quad l > L.$$

2. Second Model: we suppress the two first coefficients of γ defined as in the first model.
Let $(\gamma_l'^1)_l$ and $(\gamma_l'^2)_l$ be defined by $\gamma_l''^1 = \gamma_{l+2}^1$ and $\gamma_l''^2 = \gamma_{l+2}^2$.

We perform $N_{sim} = 5000$ simulations, $L = 100$ and study the influence of the parameter n. We set $H = 1/8$, $m = \lfloor \sqrt{3n/\log(3n)} \rfloor$ in the definition of $\hat{S}_m$ and denote by $S = (S^1, S^2)$ the vector of Sobol indices.

In the following tabulars we report the simulation results.

First Model: $S = [0.5551, 0.4448]$				
n	Bias($\hat{S}_m$)	Bias($\hat{S}_{SPF}$)	Std($\hat{S}_m$)	Std($\hat{S}_{SPF}$)
10^2	$10^{-2}[2.55, 0.39]$	$10^{-3}[5.55, 6.94]$	$10^{-2}[7.28, 6.91]$	$10^{-2}[8.79, 9.32]$
10^3	$10^{-3}[5.78, 0.15]$	$10^{-4}[8.22, 2.65]$	$10^{-2}[2.30, 2.16]$	$10^{-2}[2.74, 2.91]$
Second Model: $S = [0.7873, 0.2127]$				
n	Bias($\hat{S}_m$)	Bias($\hat{S}_{SPF}$)	Std($\hat{S}_m$)	Std($\hat{S}_{SPF}$)
10^2	$10^{-2}[6.84, 0.56]$	$10^{-3}[3.15, 5.15]$	$10^{-2}[7.89, 5.04]$	$10^{-2}[7.54, 9.87]$
10^3	$10^{-2}[2.06, 0.07]$	$10^{-4}[3.59, 4.08]$	$10^{-2}[2.50, 1.60]$	$10^{-2}[2.37, 3.13]$

First Model: $S = [0.5551, 0.4448]$		
n	RMSE($\hat{S}_m$)	RMSE($\hat{S}_{SPF}$)
10^2	$10^{-2}[7.71, 6.92]$	$10^{-2}[8.81, 9.35]$
10^3	$10^{-2}[2.37, 2.16]$	$10^{-2}[2.74, 2.92]$

Second Model: $S = [0.7873, 0.2127]$		
n	RMSE($\hat{S}_m$)	RMSE($\hat{S}_{SPF}$)
10^2	$10^{-1}[1.04, 5.07]$	$10^{-2}[7.54, 9.89]$
10^3	$10^{-2}[3.23, 1.60]$	$10^{-2}[2.37, 3.13]$

The same conclusion holds: if the signal is not concentrated on the first terms of the basis which is considered, then the SPF can be more efficient. When this is not the case, our estimator has better performances than the SPF for a similar number of observations.

4.2 Second example

4.2.1 Using Karhunen-Loève expansion of the Brownian motion

We consider the model (18) with $p = 4$ and $\varepsilon_i = 0$ for all i. The processes X_i^k for $1 \leq k \leq p$ and $1 \leq i \leq n$ are i.i.d. and defined as X^1 and X^2 in Section 4.1.1.

1. First Model: concerning the functions β^k for $k = 1, \ldots, 4$, they are respectively caracterized by their coefficients $(\gamma_l^k)_l$'s $k = 1, \ldots, 4$ of their expansion onto the same basis:

$$\begin{array}{llll} \gamma_l^1 = (l+1)^{-1/2-1/100}, & 1 \leq l \leq L, & \gamma_l^1 = 0, & l > L; \\ \gamma_l^2 = (l+1)^{-1}, & 1 \leq l \leq L, & \gamma_l^2 = 0, & l > L; \\ \gamma_l^3 = (l+1)^{-2}, & 1 \leq l \leq L, & \gamma_l^3 = 0, & l > L; \\ \gamma_l^4 = (l+1)^{-3/2}, & 1 \leq l \leq L, & \gamma_l^4 = 0, & l > L. \end{array}$$

2. Second Model: we replace the first coefficient of γ defined as in the first model by 0.
   Let for $k = 1, \ldots, 4$, $(\gamma_l'^k)_l$ be defined by $\gamma_1'^k = 0$ and for $l \geq 2$ by $\gamma_l'^k = \gamma_l^k$.

3. Third Model: we suppress the two first coefficients of γ defined as in the first model.
   Let for $k = 1, \ldots, 4$, $(\gamma_l''^k)_l$ be defined by $\gamma_l''^k = \gamma_{l+2}^k$.

We perform $N_{sim} = 5000$ simulations, $L = 100$ and we study the influence of the parameter n, where $5n$ observations are used for both methods. We set $m = \lfloor \sqrt{5n}/\log(5n) \rfloor$ in the definition of $\hat{S}_m$ and denote by $S = (S^1, S^2, S^3, S^4)$ the vector of Sobol indices.

In the following tabulars we report the simulation results. We see here that the method based on U-statistics always leads to better results than the SPF method. This is due to the fact that the SPF method needs $5n$ simulations to estimate the four Sobol indices. With the same number of simulations, the method proposed in the paper performs better, even in the second and third model. Let us recall that a drawback of the method proposed in this paper is that the distribution of the input processes $X^1, \ldots, X^p$ is assumed to be known which is not necessary for the method SPF.

First Model: $S = [0.5438, 0.2639, 0.0635, 0.1288]$		
n	Bias($\hat{S}_m$)	Bias($\hat{S}_{SPF}$)
10^2	$10^{-3}[3.39, 1.30, 0.56, 0.08]$	$10^{-3}[5.32, 5.41, 8.87, 4.70]$
10^3	$10^{-4}[7.66, 3.60, 0.33, 2.07]$	$10^{-4}[3.84, 0.59, 0.35, 5.65]$

Second Model: $S = [0.7561, 0.1871, 0.0112, 0.0456]$		
n	Bias($\hat{S}_m$)	Bias($\hat{S}_{SPF}$)
10^2	$10^{-3}[5.72, 0.81, 0.17, 0.18]$	$10^{-3}[1.01, 5.64, 5.68, 05.44]$
10^3	$10^{-4}[6.27, 1.05, 0.08, 0.61]$	$10^{-4}[5.38, 2.92, 7.14, 7.56]$

First Model: $S = [0.5438, 0.2639, 0.0635, 0.1288]$		
n	Std($\hat{S}_m$)	Std($\hat{S}_{SPF}$)
10^2	$10^{-2}[5.54, 4.28, 2.35, 3.23]$	$10^{-2}[9.91, 9.79, 9.71, 9.62]$
10^3	$10^{-2}[1.82, 1.36, 0.72, 0.99]$	$10^{-2}[3.13, 3.12, 3.11, 3.06]$

Second Model: $S = [0.7561, 0.1871, 0.0112, 0.0456]$		
n	Std($\hat{S}_m$)	Std($\hat{S}_{SPF}$)
10^2	$10^{-2}[6.11, 3.72, 1.22, 2.01]$	$10^{-1}[1.07, 1.00, 1.01, 0.99]$
10^3	$10^{-2}[1.97, 1.17, 0.33, 0.60]$	$10^{-2}[3.36, 3.16, 3.14, 3.13]$

First Model: $S = [0.5438, 0.2639, 0.0635, 0.1288]$		
n	RMSE($\hat{S}_m$)	RMSE($\hat{S}_{SPF}$)
10^2	$10^{-2}[5.55, 4.29, 2.35, 3.22]$	$10^{-2}[9.92, 9.80, 9.75, 9.63]$
10^3	$10^{-2}[1.82, 1.36, .72, 0.99]$	$10^{-2}[3.13, 3.12, 3.11, 3.06]$

Second Model: $S = [0.7561, 0.1871, 0.0112, 0.0456]$		
n	RMSE($\hat{S}_m$)	RMSE($\hat{S}_{SPF}$)
10^2	$10^{-2}[6.14, 3.72, 1.22, 2.01]$	$10^{-1}[1.07, 1.00, 1.01, 0.99]$
10^3	$10^{-2}[1.97, 1.17, 0.33, 0.60]$	$10^{-2}[3.36, 3.16, 3.14, 3.13]$

5 Conclusion

We have proposed an estimator of the sobol indices in a simple functional regression model. When the entries of the model are independent our estimator is asymptotically efficient. This is also the case of the so-called SPF estimator. As the designs underlying the two estimators are not the same, we numerically compared the two estimators in the same conditions, that is with the same number of calls to the model. The results are strongly dependent on the behavior of the regressors. When the regressors are well represented onto the Karhune-Loève basis associated to the entries then our estimator behave well and takes advantage of the information given by the knowledge of the entries, it performs better than the SPF. A contrario, if the regressors have small coefficients on the first elements of the Karhunen-Loève basis, the SPF is more precise than our estimator, since the knowledge of the entries is of less use.

These conclusions are brought in the case of a model with independent entries. One possible advantage of our estimator is that we can generalize it to the case of dependent entries, which is not the case of the SPF which deeply relies on the independence of the entries. This will be the topic of a forthcoming research paper.

Acknowledgements This work has been partially supported by the French National Research Agency (ANR) through COSINUS program (project COSTA-BRAVA no ANR-09-COSI-015).

References

[1] J. C. Bronski. Small ball constants and tight eigenvalue asymptotics for fractional Brownian motions. *J. Theoret. Probab.*, 16(1):87–100, 2003.

[2] F. Comte and J. Johannes. Adaptive estimation in circular functional linear models. *Math. Methods Statist.*, 19(1):42–63, 2010.

[3] RI Cukier, HB Levine, and KE Shuler. Nonlinear sensitivity analysis of multiparameter model systems. *Journal of computational physics*, 26(1):1–42, 1978.

[4] S. Da Veiga and F. Gamboa. Efficient estimation of nonlinear conditional functionals of a density. *Submitted*, 2008.

[5] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Statistical inference for sobol pick freeze monte carlo method. *In preparation*, 2012.

[6] J.C. Helton, J.D. Johnson, C.J. Sallaberry, and C.B. Storlie. Survey of sampling-based methods for uncertainty and sensitivity analysis. *Reliability Engineering & System Safety*, 91(10-11):1175–1209, 2006.

[7] W. Hoeffding. A class of statistics with asymptotically normal distribution. *Ann. Math. Statistics*, 19:293–325, 1948.

[8] I.A. Ibragimov and RZ Has' Minskii. *Statistical estimation–asymptotic theory*, volume 16. Springer, 1981.

[9] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of a two sobol index estimators. *Preprint, http://hal.inria.fr/docs/00/66/50/48/PDF/ArtAsymptSobol.pdf*, 2012.

[10] K. Karhunen. Über lineare Methoden in der Wahrscheinlichkeitsrechnung. *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.*, 1947(37):79, 1947.

[11] Jack P.C. Kleijnen. Sensitivity analysis and related analyses: A review of some statistical techniques. *Journal of Statistical Computation and Simulation*, 57(1-4):111–142, 1997.

[12] B. Laurent. Efficient estimation of integral functionals of a density. *Ann. Statist.*, 24(2):659–681, 1996.

[13] B. Laurent. Adaptive estimation of a quadratic functional of a density by model selection. *ESAIM Probab. Stat.*, 9:1–18 (electronic), 2005.

[14] M. Loève. Fonctions aléatoires de second ordre. *Revue Sci.*, 84:195–206, 1946.

[15] M. Loève. *Probability theory. I.* Springer-Verlag, New York, fourth edition, 1977. Graduate Texts in Mathematics, Vol. 45.

[16] H. Luschgy and G. Pagès. Functional quantization of Gaussian processes. *J. Funct. Anal.*, 196(2):486–531, 2002.

[17] Jeremy E. Oakley and Anthony O'Hagan. Probabilistic sensitivity analysis of complex models: a Bayesian approach. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 66(3):751–769, 2004.

[18] A. Saltelli. Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, 145:280–297, 2002.

[19] A. Saltelli, K. Chan, and E.M. Scott. Sensitivity analysis, 2000.

[20] I. M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Math. Modeling Comput. Experiment*, 1(4):407–414 (1995), 1993.

[21] I.M. Sobol. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiments*, 1:407–414, 1993.

[22] I.M. Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, 2001.

[23] J.Y. Tissot and C. Prieur. A bias correction method for the estimation of sensitivity indices based on random balance designs. *hal-00507526v1*.

[24] A.W. Van der Vaart. *Asymptotic statistics*. Cambridge Univ Press, 2000.