A tradeoff between explorations and repetitions for estimators of two global sensitivity indices in stochastic models induced by probability measures

Gildas Mazo

A TRADEOFF BETWEEN EXPLORATIONS AND REPETITIONS FOR ESTIMATORS OF TWO GLOBAL SENSITIVITY INDICES IN STOCHASTIC MODELS INDUCED BY PROBABILITY MEASURES*

GILDAS MAZO†

Abstract. Sobol sensitivity indices assess how the output of a given mathematical model is sensitive to its inputs. If the model is stochastic then it cannot be represented as a function of the inputs, thus raising questions as how to do a sensitivity analysis in those models. Practitioners have been using a method that exploits the availability of softwares for deterministic models. For each input, the stochastic model is repeated and the outputs are averaged. These averages are seen as if they came from a deterministic model and hence Sobol's method can be used. However, in the context of limited computational resources, one must ensure that the number of repetitions of the stochastic model multiplied by the number of explorations of the input space is less than a fixed threshold. The problem of finding an optimal tradeoff between the repetitions and the explorations is addressed. A bound on an error criterion that penalizes bad rankings of the inputs' sensitivities is minimized. The estimator induced by the empirical method described above is shown to be asymptotically biased if the number of repetitions goes to infinity too slowly. A functional relationship between the output, the input and some random noise is derived from minimal distributional assumptions, which leads to a new sensitivity index with better statistical properties. The theory is illustrated on numerical experiments.

Key words. asymptotic normality, Sobol indices, tradeoff, sensitivity analysis, stochastic model.

AMS subject classifications. 62G20, 60H99, 65C05

1. Introduction.

The goal of sensitivity analysis is to assess how the output of a given physical or mathematical model is sensitive to its inputs [26, 27]. Classically, the model of interest is deterministic. To each input there corresponds an output given by the model. Thus, in this case, the model is in fact a function, say f. To assess the sensitivity of the model to its inputs, the probabilistic/statistical framework is often employed. One draws at random a large number of inputs and observe how the corresponding outputs vary. From a statistical perspective, at each draw, one observes a random pair (X, Y) such that $Y = f(X)$, where $X = (X_1, \dots, X_p)$ is the input vector and Y is the output.

Sobol's idea [29, 30] was to notice that, if $X_1, \dots, X_p$ are drawn independently then $f(X)$ can be decomposed into a sum of lower-dimensional functions and that this decomposition can be used to allocate the variance of the output to the individual components of the decomposition. More precisely, we have

$$
\begin{aligned}
f(X) - f_0 = {} & f_1(X_1) + \cdots + f_p(X_p) \\
& + f_{1,2}(X_1, X_2) + \cdots + f_{p-1,p}(X_{p-1}, X_p) \\
& + \cdots \\
& + f_{1,\dots,p}(X_1, \dots, X_p),
\end{aligned}
\tag{1.1}
$$

where $f_0 = \mathrm{E}\, f(X)$, $f_j(X_j) = \mathrm{E}(f(X) - f_0 | X_j)$, $j = 1, \dots, p$, and $f_{1,2}, \dots, f_{1,\dots,p}$ are some functions defined iteratively; see [29] and [33, p. 157] for more details. In the field of uncertainty quantification the above decomposition is known as the Sobol-Hoeffding decomposition in reference to [11, 29]. The expectations and the covariances

*Submitted to the editors 2019/03/07.

†MaIAGE, INRA, Université Paris-Saclay, 78350, Jouy-en-Josas, France (gildas.mazo@inra.fr, http://genome.jouy.inra.fr/~gmazo/).

of the individual components in the right-hand side of (1.1) are zero and hence we have the variance decomposition

$$\operatorname{Var} f(X) = \operatorname{Var} f_1(X_1) + \cdots + \operatorname{Var} f_p(X_p) + \cdots ,$$

which leads to the so-called Sobol indices

$$S_j = \frac{\operatorname{Var} f_j(X_j)}{\operatorname{Var} f(X)} = \frac{\operatorname{Var} \operatorname{E}(f(X)|X_j)}{\operatorname{Var} f(X)} = \frac{\operatorname{E} f(X) f(\widetilde{X}_{-j}) - (\operatorname{E} f(X))^2}{\operatorname{E} f(X)^2 - (\operatorname{E} f(X))^2},$$

$j = 1, \ldots, p$; here $\widetilde{X}_{-j}$ stands for an independent copy of X where the jth component has been replaced by that of X. Thus the Sobol index associated with the jth argument of f is defined as the proportion of the total variance associated with the lower-dimensional function that depends on the jth argument only. Sobol indices are interpreted as sensitivity measures and used to achieve various goals in uncertainty quantification [27].

If the model is nonadditive (it is said that the inputs "interact" with each other) then the Sobol indices may be inadequate. To account for interactions, the so-called total sensitivity indices [12] are often computed along with Sobol indices. The total sensitivity index associated with the jth argument of f is given by

$$S_{\mathrm{T}_j} = 1 - \frac{\operatorname{Var} \operatorname{E}(f(X)|X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p)}{\operatorname{Var} f(X)}.$$

The total sensitivity index quantifies the sensitivity of the output of f to its jth argument through the interactions it may have with the other inputs.

There are numerous methods to estimate the sensitivity indices. For simplicity, we describe below Sobol's original method to estimate S_j through Monte Carlo sampling [29]. For a review of the many other methods, see [23] or the package `sensitivity` [16] of the `R` software for an up-to-date list of many methods, with references. Thus, draw two independent sets of inputs $\{X^{(i)}, i = 1, \ldots, n\}$, $\{\widetilde{X}^{(i)} := (\widetilde{X}_1, \ldots, \widetilde{X}_p), i = 1, \ldots, n\}$ and make p more sets by combining the first two: $\{\widetilde{X}^{(i)}_{-j}, i = 1, \ldots, n\}$, $j = 1, \ldots, p$, where

$$\widetilde{X}^{(i)}_{-j} := (\widetilde{X}^{(i)}_1, \ldots, \widetilde{X}^{(i)}_{j-1}, X^{(i)}_j, \widetilde{X}^{(i)}_{j+1}, \ldots, \widetilde{X}^{(i)}_p). \tag{1.2}$$

The first and the p last sets are passed on to the function f which produces the outputs $\{Y^{(i)}, i = 1, \ldots, n\}$ (for the first set) and $\{Y^{(i)}_j, i = 1, \ldots, n\}$, $j = 1, \ldots, p$ (for the p last sets), which in turn make up the so-called pick-freeze estimator

$$\widehat{S}_j = \frac{\frac{1}{n}\sum_{i=1}^n Y^{(i)} Y^{(i)}_j - \left(\frac{1}{n}\sum_{i=1}^n Y^{(i)}\right)^2}{\frac{1}{n}\sum_{i=1}^n Y^{(i)2} - \left(\frac{1}{n}\sum_{i=1}^n Y^{(i)}\right)^2}. \tag{1.3}$$

This gives a simple procedure to estimate all the Sobol indices $S_1, \ldots, S_p$ with $(p+1)n$ runs of the model. The pick-freeze estimator is asymptotically normal [7, 17]. The above formula can be improved in many ways [12, 17, 21]. Many versions of this estimator exist, the goal being always to get the most efficient estimator with the least computations. Sobol indices for multivariate, functional outputs [6, 18] or functional inputs [15] have been proposed as well.

The big difference between a deterministic model and a stochastic model is that the stochastic model is not a function anymore. To a particular value of the input there does not correspond any particular value for the output. Instead, there corresponds a range of possible values, assumed to come from a probability distribution depending on the input. Examples can be found in epidemiology [2, 3, 24, 28] or ecology [31], to name a few.

To do the sensitivity analysis of a stochastic model, several approaches have been investigated. In [19], to the best of my understanding, the authors carry out the sensitivity analysis of a stochastic model based on a joint metamodel. In [10], a stochastic model is seen as a functional relation of the form $Y(\vartheta, \omega) = f(X(\vartheta), \omega)$, where the X is a random vector on some probability space, ω is a point in some probability space distinct from that on which X is defined, f is some function and $Y(\vartheta, \omega)$ is a random variable on the induced product probability space. The quantity $f(X(\vartheta), \omega)$ represents the output of the stochastic model run with input $X(\vartheta)$; the point ω represents the intrinsic randomness. The idea is then to decompose the function $\vartheta \mapsto f(X(\vartheta), \omega)$ for each ω and estimate the associated sensitivity indices, which depend on ω. The estimates are then averaged over ω to make the final sensitivity estimates. In [1], to the best of my understanding, the stochastic model is represented as a deterministic mapping which with an input associates a probability density function. The Sobol-Hoeffding decomposition is applied to the mapping which with an input associates the entropy of the output evaluated at that input. Here the entropy is the Kullback-Leibler divergence of the output density. In [34], the output of the stochastic model is seen as a semiparametric statistical model—the generalized lambda distribution—with parameters depending on the inputs. These parameters have a polynomial chaos expansion which is estimated by maximum likelihood. Once the law of the output conditionally on the input has been estimated, its inverse cumulative distribution function is used to turn the stochastic model into a deterministic model to which standard methods are applied. In [5], the stochastic model is seen as a mapping that goes from the input space to a space of probability measures equipped with the Wasserstein distance. Following [8, 9], the Wasserstein space is mapped to the real line $\mathbb{R}$ with some family of test functions, thus allowing for a standard Sobol-Hoeffding decomposition which is then averaged over all possible test functions. In more specific contexts, global sensitivity analysis methods also have been proposed. For instance, there are methods for stochastic differential equations [4] and chemical reaction networks [22].

In practice, although it has not been formally defined in the literature, another method has been used for some time [2, 24, 28, 31]. The idea is simple: at each draw of the input $X^{(i)}$, one produces as many outputs $Y^{(i,1)}, \dots, Y^{(i,m)}$ as possible, makes the average $m^{-1} \sum_{k=1}^m Y^{(i,k)}$ and does *as if* it were the output of some deterministic model. The same is done with the inputs $\widetilde{X}_{-j}^{(i)}$ (1.2) to produce the outputs $m^{-1} \sum_{k=1}^m Y_j^{(i,k)}$. The obtained estimator is then the same as that in (1.3) but with $Y^{(i)}$ replaced by $m^{-1} \sum_{k=1}^m Y^{(i,k)}$ and $Y_j^{(i)}$ replaced by $m^{-1} \sum_{k=1}^m Y_j^{(i,k)}$, yielding

$$\widehat{S}_j = \frac{\frac{1}{n} \sum_{i=1}^n m^{-1} \sum_{k=1}^m Y^{(i,k)} m^{-1} \sum_{k=1}^m Y_j^{(i,k)} - \left(\frac{1}{n} \sum_{i=1}^n m^{-1} \sum_{k=1}^m Y^{(i,k)} \right)^2}{\frac{1}{n} \sum_{i=1}^n \left(m^{-1} \sum_{k=1}^m Y^{(i,k)} \right)^2 - \left(\frac{1}{n} \sum_{i=1}^n m^{-1} \sum_{k=1}^m Y^{(i,k)} \right)^2}. \tag{1.4}$$

The big advantage for practitioners is that they can use the numerous available and ready-to-use softwares for deterministic models.

To build the estimator (1.4), the stochastic model must be run $mn(p+1)$ times. The number m is called the number of repetitions and the number n is called the number of explorations. If the stochastic model is computationally intensive—that is, each model run is time-consuming—, then the estimator is built with limited resources. In this context, an increase of m must go along with a decrease of n, and conversely. What is then a good balance between m and n? How to choose m and n such that the estimator (1.4) will be the most efficient? This question was asked by [31].

We address this problem by minimizing a bound on the missranking error. The missranking error penalizes bad rankings of the Sobol indices associated with the inputs. This type of error leads to an explicit solution of the induced minimization problem and hence the "optimal" pair (m, n) can be estimated. A two-step procedure can then be implemented to get efficient estimators. We show that, asymptotically, this two-step procedure is always better than any choice (m, n), except the optimal one. Futhermore, we show that the estimator (1.4) is asymptotically normal but biased, as $n \to \infty$. To remove the bias, we also must have $m \to \infty$ fast enough. The sensitivity index to which the estimator (1.4) converges is called *the sensitivity index of the second kind*. Finally, although we only assume that a stochastic model is a set of probability measures that capture how the outputs are drawn, we show that a function linking the output, the inputs and some random "noise" can be extracted and be the object to which the Sobol-Hoeffding decomposition can be applied, thus yielding a new sensitivity index, called *the sensitivity index of the first kind*. One advantage is that this index leads to estimators with better statistical properties. The indices of the first and of the second kinds are complementary as they offer distinct pieces of information. Interestingly, these indices can be estimated jointly with no additional cost, the joint estimator is asymptotically normal and the two kinds of sensitivity indices lead to the same solution for the tradeoff problem.

This paper is organized as follows. Section 2 gives a definition of stochastic models in terms of probability measures and shows how one can construct a functional representation linking the output, the input and some random noise. Section 3 defines the indices of both kinds and their estimators. The asymptotic properties are deferred to Section 5. Section 4 introduces the tradeoff problem, gives a procedure to attack it and gives some theoretical guarantees. Section 6 illustrates the theory on numerical simulations. A Conclusion closes the paper.

2. Representations of stochastic models. The concept of stochastic models is intuitive and shared by many people but there are different mathematical routes to describe them. One is given in Section 2.1. It makes minimal distributional assumptions to get to a representation in terms of random variables and establishes the existence of a function to which the Sobol-Hoeffding decomposition can be applied. Section 2.2 makes connections with the stochastic models of [10].

2.1. Representing stochastic models from minimal distributional assumptions. A stochastic model is some mechanism that produces outputs at random given some inputs. Thus, a stochastic model can be seen as family of probability measures $\{Q_x, x \in \mathcal{X}\}$ indexed by some input space $\mathcal{X} \subset \mathbb{R}^p$. We assume that each probability measure Q_x is defined on the measurable space $(\mathbb{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-field induced by $\mathbb{R}$. The law Q_x governs how the outputs are produced given the input x. Since $\mathcal{X}$ is a subset of $\mathbb{R}^p$ we can endow the later with its product Borel σ-field $\mathcal{B}^p$. Let P^* be a product probability measure on the measurable space $(\mathbb{R}^p, \mathcal{B}^p)$. Assume that the support of P^* and $\mathcal{X}$ coincide. The probability measure P^* represents how the inputs are drawn by the practitioner. In particular, since P^* is a product probability measure, the inputs are drawn independently.

The stochastic experiment that consists of drawing inputs at random according to P^* and observing the corresponding outputs is fully characterized by the family $\{Q_x\}$ and the probability measure P^*. This leads us to Definition 2.1.

DEFINITION 2.1. *If P^* and $\{Q_x\}$ are probability measures as described above then the pair $(P^*, \{Q_x\})$ is called the complete stochastic model.*

Now we look for a representation in terms of random variables that will allow us to use the Sobol-Hoeffding decomposition later on.

LEMMA 2.2. *If $(P^*, \{Q_x\})$ is a complete stochastic model then there exist a probability space $(\Omega, \mathcal{F}, P)$, a random vector $(X, Z) : \Omega \to \mathbb{R}^{p+1}$ and a function $f : \mathbb{R}^{p+1} \to \mathbb{R}$ such that*

(i) $f(x, Z)$ is measurable for every $x \in \mathcal{X}$,

(ii) $P(f(x, Z) \in B) = Q_x(B)$ for every $x \in \mathcal{X}$ and every $B \in \mathcal{B}$,

(iii) $P(X \in A, Z \in B) = P^*(A)P(Z \in B)$ for every $A \in \mathcal{B}^p$ and $B \in \mathcal{B}$.

Moreover, if $(X, f(X, Z))$ and $(X', f'(X', Z'))$ are two joint vectors that satisfy the conditions (i), (ii) and (iii) then $(X, f(X, Z)) \overset{d}{=} (X', f'(X', Z'))$ where $\overset{d}{=}$ means equality in distribution.

Note that the conditions in Lemma 2.2 do not determine the law of Z; see the example below.

EXAMPLE 1 (The law of Z is not determined). *Let $p = 1$. Let P^* be the standard uniform distribution and Q_x be the Gaussian distribution with mean $x \in \mathbb{R}$ and variance 1. Let $\Omega = (0, 1)^2$ endowed with its Borel σ-field and set P to be the product Lebesgue measure. Let $X_1(\omega) = \omega_1$ for $\omega = (\omega_1, \omega_2) \in \Omega$. Let Φ denote the distribution function of the standard Gaussian distribution and denote by Φ^{-1} the inverse of Φ. If $Z(\omega) = \omega_2$ and $f(x, z) = \Phi^{-1}(z) + x$, $x \in \mathbb{R}$, $z \in (0, 1)$, then it is easy to see that (X, Z) and f satisfy the conditions of Lemma 2.2 and the law of Z is the standard uniform distribution. But the conditions of Lemma 2.2 are also satisfied with $Z(\omega) = \sqrt{\omega_2}$ and $f(x, z) = \Phi^{-1}(z^2) + x$, in which case, $P(Z \le t) = t^2$, $t \in (0, 1)$, that is, the law of Z is the beta distribution with parameter $(2, 1)$.*

The indeterminacy of the law of Z is symptomatic of the lack of control of the intrinsic randomness assumed in our definition of stochastic models. But this is not an issue because our interest lies in the joint vector $(X, f(X, Z))$, the law of which is fully characterized by the conditions in Lemma 2.2. To each complete stochastic model there corresponds a unique law that all vectors $(X, f(X, Z))$ must have, regardless of the chosen representation. Therefore, the pair $(X, f(X, Z))$ can be used to define the pair (input, output) of a complete stochastic model, as done in Definition 2.3.

DEFINITION 2.3. *If (X, Z) and f satisfy the conditions in Lemma 2.2 then the pair $(X, f(X, Z))$ is called an observation of the complete stochastic model $(P^*, \{Q_x\})$; the random variable X is called the input and $f(X, Z)$ is called the output.*

In sum, we have established the existence of random variables on a common probability space and a function f that characterize the statistical experiment that consists of drawing inputs and observing the outputs of a stochastic model. The set of assumptions used to represent outputs and inputs of a stochastic model is minimal: all we need is a family $\{Q_x\}$ and a probability measure P^*. We remark that the above formalism of stochastic models can be used to represent physical models [32] as well.

2.2. Links with the stochastic models and the sensitivity indices in [10]. In [10], the authors consider the model $(X'(\omega'), \varphi(X'(\omega'), \omega''))$, $\omega' \in \Omega'$, $\omega'' \in \Omega''$, where $(\Omega', \mathcal{F}', P')$ and $(\Omega'', \mathcal{F}'', P'')$ are probability spaces, $X' = (X'_1, \ldots, X'_p)$ is a random vector on Ω' and φ is some function. They consider the sensitivity indices

$$S_j^{\mathrm{HAG}} = \int_{\Omega''} S_j(\omega'')\, P''(\mathrm{d}\omega''),$$

where

$$S_j(\omega'') = \frac{\mathrm{Var}[\mathrm{E}(\varphi(X', \omega'')|X'_j)]}{\mathrm{Var}[\varphi(X', \omega'')]};$$

above the variances and the expectation are to be understood as integrals on Ω' with respect to P'.

One can choose a representation in Lemma 2.2 that corresponds to the models in [10]. In particular, one can recover the sensitivity indices S_j^{HAG}, $j = 1, \ldots, p$. Let us illustrate this with an example. Let $(P^*, \{Q_x\})$ be a complete stochastic model and let $X = (X_1, \ldots, X_p)$, Z and f be as in Lemma 2.2. Define

$$\widetilde{S}_j^{\mathrm{HAG}} = \mathrm{E}\left(\frac{\mathrm{Var}(\mathrm{E}[f(X,Z)|X_j, Z]|Z)}{\mathrm{Var}(f(X,Z)|Z)}\right), \quad j = 1, \ldots, p.$$

Consider the model in Example 1.1 of [10], given by

$$\varphi(X'(\omega'), \omega'') = X_1(\omega') + X_2(\omega')\omega'', \tag{2.1}$$

where the law of X'_1 is the uniform distribution on $(0,1)$, the law of X'_2 is the uniform distribution on $(1, L+1)$, $L > 0$, and P'' is the standard normal distribution on $\Omega'' = \mathbb{R}$. The indices in Example 1.1 of [10] are given by

$$S_1^{\mathrm{HAG}} = \int_{\Omega''} \frac{1}{1 + L^2\omega''}\, P''(\mathrm{d}\omega'') = \int_{\mathbb{R}} \frac{1}{1 + L^2 w} \exp\left(-\frac{w^2}{2}\right) \frac{1}{\sqrt{2\pi}}\, \mathrm{d}w$$

and $S_2^{\mathrm{HAG}} = 1 - S_2^{\mathrm{HAG}}$.

We can build a probability space $(\Omega, \mathcal{F}, P)$, a random vector (X, Z) and a function f such that $\widetilde{S}_1^{\mathrm{HAG}} = S_1^{\mathrm{HAG}}$, as shown in Example 2 below.

Example 2. *Let us first extract the induced complete stochastic model. Set* $P^*((0, t_1] \times (1, t_2]) = t_1(t_2 - 1)/L$ *for all* $0 < t_1 < 1$, $1 < t_2 < L + 1$, $L > 0$ *and* $Q_x(-\infty, t] = \Phi((t - x_1)/x_2)$ *for all* $t \in \mathbb{R}$, *where* $\Phi(t) = \int_{-\infty}^{t} (2\pi)^{-1/2} e^{-s^2/2}\, \mathrm{d}s$ *and* $x = (x_1, x_2) \in \mathbb{R} \times (0, \infty)$. *Now it remains to choose a representation that fulfills the conditions in Lemma 2.2 and ensures that* $S_1^{HAG} = \widetilde{S}_1^{HAG}$. *Such a representation can easily be found. For instance, take* $\Omega = (0,1)^3$ *endowed with the product Lebesgue measure and put* $Z(\omega) = \omega_3$, $X_1(\omega) = F_1^{-1}(\omega_1)$ *and* $X_2(\omega) = F_2(\omega_2)^{-1}$ *for* $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega$, *where* $F_1(t_1) = t_1$ *for* $0 < t_1 < 1$ *and* $F_2(t_2) = (t_2 - 1)/L$ *for* $1 < t_2 < L + 1$. *Finally take* $f(x, z) = \Phi^{-1}(z)x_2 + x_1$ *for* $x_1 \in \mathbb{R}$, $x_2 > 0$ *and* $z \in (0,1)$. *Then the conditions of Lemma 2.2 are fulfilled by construction and the detailed calculations in Appendix A show that* $S_1^{HAG} = \widetilde{S}_1^{HAG}$.

In sum, the stochastic models in [10] can be expressed with the framework of Section 2.1. There is however a difference between [10] and Section 2.1. In [10], the function φ is fixed. It is given as being a part of the stochastic model. In our side, the function f is constructed from the probability measures that we are given in the first place. It is not unique. Consequently, it is unclear whether or not the indices $\widetilde{S}_1^{\mathrm{HAG}}$ are uniquely determined.

3. The sensitivity indices and their estimators. In view of Section 2, we can assume that there are a random vector $(X, Z) \in \mathbb{R}^p \times \mathbb{R}$ with mutually independent $p+1$ components on some probability space $(\Omega, \mathcal{F}, P)$ and a real function f such that the pair $(X, f(X, Z)) \in \mathbb{R}^p \times \mathbb{R}$ represents a random observation (input, output) of the stochastic model of interest. To ensure the existence of the sensitivity indices and later to derive theoretical results for the estimators, we need to assume the following: there exists some function F with $\mathrm{E}\, F(X)^8 < \infty$ such that

$$|f(X, Z)| \le F(X) \tag{3.1}$$

almost surely. This assumption appears to be mild. In particular every stochastic model with bounded outputs fulfills the condition.

3.1. Definition of the sensitivity indices. We define two kinds of sensitivity indices. The sensitivity indices of the first kind exploit the existence of the function f by applying the Sobol-Hoeffding decomposition to it directly. The sensitivity indices of the second kind result from an application of the Sobol-Hoeffding decomposition to the conditional expectation of $f(X, Z)$ given X, which is a function of X alone. The indices of the second kind are those to which the estimators (1.4) mentioned in the Introduction converge.

3.1.1. Indices of the first kind. Applying the Sobol-Hoeffding decomposition to f yields

$$f(X, Z) - f_0 = f_1(X_1) + \cdots + f_p(X_p) + f_{p+1}(Z) + \cdots\,, \tag{3.2}$$

where $f_0 = \mathrm{E}\, f(X, Z)$, $f_j(X_j) = \mathrm{E}(f(X, Z) - f_0 | X_j)$, $j = 1, \ldots, p$, $f_{p+1}(Z) = \mathrm{E}(f(X, Z) - f_0 | Z)$ and $+\cdots$ stands for the interaction terms. Since X and Z are independent, we have $\mathrm{Var}\, f(X, Z) = \mathrm{Var}\, f_1(X_1) + \cdots + \mathrm{Var}\, f_p(X_p) + \mathrm{Var}\, f_{p+1}(Z) + \cdots$, which leads us to the indices in Definition 3.1.

Definition 3.1 (Sobol indices of the first kind). *The Sobol indices of the first kind are defined as*

$$S'_j = \frac{\mathrm{Var}\, \mathrm{E}(f(X, Z) | X_j)}{\mathrm{Var}\, f(X, Z)}, \quad j = 1, \ldots, p.$$

It is important to notice that the indices of the first kind depend on the law of $(X, f(X, Z))$ only and hence are uniquely determined. Note that total sensitivity indices could be defined as well but it is unclear whether or not they depend on the chosen representation.

3.1.2. Indices of the second kind. Let $g(X) := \mathrm{E}[f(X, Z) | X]$ be the conditional expectation of the output of the stochastic model given the input. The object g is a function and the Sobol-Hoeffding decomposition can be applied to it, yielding

$$g(X) - g_0 = g_1(X_1) + \cdots + g_p(X_p) + \cdots\,,$$

where $g_0 = \mathrm{E}\, g(X)$, $g_j(X_j) = \mathrm{E}(g(X) - g_0 | X_j)$, $j = 1, \ldots, p$ and $+\cdots$ stands for the interaction terms. Since the components of X are independent, we have $\mathrm{Var}\, g(X) = \mathrm{Var}\, g_1(X_1) + \cdots + \mathrm{Var}\, g_p(X_p) + \cdots$, leading to the indices in Definition 3.2.

Definition 3.2 (Sobol indices of the second kind). *The Sobol indices of the second kind are defined as*

$$S''_j = \frac{\mathrm{Var}\, \mathrm{E}(g(X) | X_j)}{\mathrm{Var}\, g(X)} = \frac{\mathrm{Var}\, \mathrm{E}(\mathrm{E}[f(X, Z) | X] | X_j)}{\mathrm{Var}\, \mathrm{E}[f(X, Z) | X]}, \quad j = 1, \ldots, p.$$

The total sensitivity indices, defined by

$$S''_{Tj} = 1 - \frac{\operatorname{Var} \operatorname{E} (g(X)|X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_p)}{\operatorname{Var} g(X)}, \tag{3.3}$$

$j = 1, \ldots, p$, are uniquely determined.

3.1.3. Comparison of the definitions. The sensitivity indices of the first kind provide more refined "first-order" information than the indices of the second kind. Example 3 and 4 illustrate this.

EXAMPLE 3. *Let $f(X, Z) = aX_1 + cX_2\phi(Z)$, where X_1, X_2, Z are independent standard normal variables, a, c are real coefficients and ϕ is a function such that* $\operatorname{E}\phi(Z) = 0$. *Then*

$$S'_1 = \frac{a^2}{a^2 + c^2 \operatorname{E} \phi(Z)^2}, \ S'_2 = 0, \ S''_1 = 1 \text{ and } S''_2 = 0.$$

According to the sensitivity indices of the second kind, X_1 has the same importance regardless of the value of its coefficient a, while the sensitivity indices of the first kind acknowledge that the importance of X_1 should depend on its coefficient. However, the sensitivity indices of the first kind cannot provide insight into the interactions between the inputs. For instance, if a is small then the sum $S'_1 + S'_2$ will be small and hence the contribution to the variance of the output must come from elsewhere. Perhaps it comes from the intrinsic stochasticity of the model or from the interactions.

Example 4 returns to the model (2.1).

EXAMPLE 4. *Let $f(X, Z) = \Phi^{-1}(Z)X_2 + X_1$ such that the law of X_1 and that of Z are the uniform distribution on $(0, 1)$, the law of X_2 is the uniform distribution on $(1, L + 1)$, $L > 0$, and Φ^{-1} denotes the inverse distribution function of the standard normal distribution. It is easy to see that $S''_1 = 1$ and $S''_2 = 0$. The detailed calculations in Appendix A show that $S'_2 = 0$ and*

$$S'_1 = \frac{1}{4(L^2 + 3(L + 1)) + 1}.$$

As in Example 3, the sensitivity indices of the second kind do not depend on the coefficient L. The sensitivity indices of the first kind do depend on L but note that $S'_1 + S'_2 \leq 1/13$, indicating that most of the contribution to the output comes from the intrinsic randomness or the interactions.

In sum, both kinds of sensitivity indices provide useful insights although neither kind is perfect. The sensitivity indices of the second kind are good indices for doing a sensitivity analysis of the model averaged over the intrinsic randomness but by doing so information may be lost. The sensitivity indices of the first kind provide more refined information into the individual contributions of the inputs but the information is only partial because the knowledge of the interactions and the intrinsic randomness are lacking.

3.2. Construction of the estimators. We construct estimators for the indices in Definition 3.1 and 3.2 by Monte-Carlo simulation. The input space is "explored" n times; at each exploration two independent input vectors are drawn, combined and passed to the stochastic model which is run m times. The integer n is called the number of explorations and the integer m is called the number of repetitions.

The couple (n,m) is called the *design* of the Monte-Carlo sampling scheme. The total number of calls to the stochastic model is $mn(p+1)$. The details are given in Algorithm 3.1.

Algorithm 3.1 Generate a Monte-Carlo sample

for $i=1$ to n **do**
  draw two independent copies $X^{(i)}=(X_1^{(i)},\dots,X_p^{(i)})$, $\widetilde{X}^{(i)}=(\widetilde{X}_1^{(i)},\dots\widetilde{X}_p^{(i)})$
  for $j=0,1,\dots,p$ **do**
    for $k=1$ to m **do**
      run the stochastic model at $\widetilde{X}_{-j}^{(i)}:=(\widetilde{X}_1^{(i)},\dots,\widetilde{X}_{j-1}^{(i)},X_j^{(i)},\widetilde{X}_{j+1}^{(i)},\dots,\widetilde{X}_p^{(i)})$ to get an output $Y_j^{(i,k)}$
    end for
  end for
end for

In the algorithm above, $\widetilde{X}_{-0}^{(i)}=X^{(i)}$ by convention. By assumption, the objects $\widetilde{X}^{(i)}$, $\widetilde{X}_{-j}^{(i)}$ and $Y_j^{(i,k)}$, $j=0,\dots,p$, $k=1,\dots,m$, $i=1,\dots,n$, are random vectors such that the sets $\{\widetilde{X}^{(i)},\widetilde{X}_{-j}^{(i)},Y_j^{(i,k)}\ :\ j=0,\dots,p;\ k=1,\dots,m\}$, $i=1,\dots,n$, are i.i.d., $X^{(i)}$ and $\widetilde{X}^{(i)}$ are independent and $P(\cap_{j=0}^{p}\cap_{k=1}^{m}\{Y_j^{(i,k)}\in B_j^{(k)}\}|X^{(i)},\widetilde{X}^{(i)})=\prod_{j=0}^{p}\prod_{k=1}^{m}P(Y_j^{(i,k)}\in B_j^{(k)}|X^{(i)},\widetilde{X}^{(i)})$ for all Borel sets $B_j^{(k)}\in\mathcal{B}$. It is easy to see that these conditions characterize the joint law of the set $\{\widetilde{X}_{-j}^{(i)},Y_j^{(i,k)}\ :\ j=0,\dots,p;\ k=1,\dots,m;\ i=1,\dots,n\}$, that is, the inputs and the outputs of Algorithm 3.1.

In view of Section 2, assume without loss of generality that there is some function f and some random variables $Z_j^{(i,k)}$, $j=0,\dots,p$, $k=1,\dots,m$, $i=1,\dots,n$, such that $Y_j^{(i,k)}=f(\widetilde{X}_{-j}^{(i)},Z_j^{(i,k)})$, where all of the random vectors in the sets $\{\widetilde{X}^{(i)},X^{(i)},Z_j^{(i,k)}\ :\ j=0,\dots,p;\ k=1,\dots,m\}$, $i=1,\dots,n$, are mutually independent and all of these sets are i.i.d. We shall use both the notations Y and $f(X,Z)$ to denote the outputs.

With the above notation, the estimators (1.4) of the indices of the second kind are rewritten

$$\widehat{S}''_{j;n,m}=\frac{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{k=1}^{m}Y_0^{(i,k)}\frac{1}{m}\sum_{k'=1}^{m}Y_j^{(i,k')}-\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{k=1}^{m}Y_0^{(i,k)}\right)^2}{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{k=1}^{m}Y_0^{(i,k)}\right)^2-\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{k=1}^{m}Y_0^{(i,k)}\right)^2},\tag{3.4}$$

$j=1,\dots,p$, which are indeed the empirical versions of the indices S''_j, since

$$\begin{aligned}S''_j&=\frac{\mathrm{E}\,g(X^{(1)})g(\widetilde{X}_{-j}^{(1)})-(\mathrm{E}\,g(X^{(1)}))^2}{\mathrm{E}\,g(X^{(1)})^2-(\mathrm{E}\,g(X^{(1)}))^2}\\&=\frac{\mathrm{E}\,\mathrm{E}[f(X^{(1)},Z_0^{(1,1)})|X^{(1)}]\,\mathrm{E}[f(\widetilde{X}_{-j}^{(1)},Z_j^{(1,1)})|\widetilde{X}_{-j}^{(1)}]-(\mathrm{E}\,\mathrm{E}[f(X^{(1)},Z_0^{(1,1)})|X^{(1)}])^2}{\mathrm{E}\,\mathrm{E}[f(X^{(1)},Z_0^{(1,1)})|X^{(1)}]^2-(\mathrm{E}\,\mathrm{E}[f(X^{(1)},Z_0^{(1,1)})|X^{(1)}])^2}.\end{aligned}\tag{3.5}$$

As said in the Introduction, this estimator is used implicitly by practitioners but has not been formally studied in the literature. A simplified version with $m=n$ appears in [13, 14].

To estimate the sensitivity indices of the first kind, we exploit a formula similar to (3.5). Indeed, we have

$$
\begin{aligned}
S'_j =& \frac{\mathrm{E}\, f(X^{(1)}, Z_0^{(1,1)}) f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) - (\mathrm{E}\, f(X^{(1)}, Z_0^{(1,1)}))^2}{\mathrm{E}\, f(X^{(1)}, Z_0^{(1,1)})^2 - (\mathrm{E}\, f(X^{(1)}, Z_0^{(1,1)}))^2} \\
& \qquad (3.6) \\
=& \frac{\mathrm{E}\,\mathrm{E}[f(X^{(1)}, Z_0^{(1,1)})|X^{(1)}]\,\mathrm{E}[f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)})|\widetilde{X}_{-j}^{(1)}] - \Big(\mathrm{E}\,\mathrm{E}[f(X^{(1)}, Z_0^{(1,1)})|X^{(1)}]\Big)^2}{\mathrm{E}\,\mathrm{E}[f(X^{(1)}, Z_0^{(1,1)})^2|X^{(1)}] - \Big(\mathrm{E}\,\mathrm{E}[f(X^{(1)}, Z_0^{(1,1)})|X^{(1)}]\Big)^2}.
\end{aligned}
$$

Notice that the upper left, upper right and the lower right terms are identical to the upper left, upper right and the lower right terms in (3.5) respectively. The upper left term is the only term that depends on j and, therefore, it is the only term that permits to discriminate between any two indices of the same kind. For this reason, it is called the discriminator, denoted by D_j. Formula (3.6) yields the estimator

$$
(3.7) \quad \widehat{S}'_{j;n,m} = \frac{\frac{1}{n}\sum_{i=1}^n \frac{1}{m}\sum_{k=1}^m Y_0^{(i,k)} \frac{1}{m}\sum_{k'=1}^m Y_j^{(i,k')} - \Big(\frac{1}{n}\sum_{i=1}^n \frac{1}{m}\sum_{k=1}^m Y_0^{(i,k)}\Big)^2}{\frac{1}{n}\sum_{i=1}^n \frac{1}{m}\sum_{k=1}^m Y_0^{(i,k)2} - \Big(\frac{1}{n}\sum_{i=1}^n \frac{1}{m}\sum_{k=1}^m Y_0^{(i,k)}\Big)^2}.
$$

Since the estimators for the discriminator are identical, both kinds of sensitivity indices lead to the same estimated ranking of the inputs. All of the $2p$ estimators can be computed with $mn(p+1)$ runs of the stochastic model. In (3.7) and (3.4), if $m = 1$ and if the function f does in fact not depend on Z, then the estimators reduce to Sobol estimators [29, 30] for deterministic models.

4. Choosing between Monte-Carlo designs. The estimators in Section 3 depend on the design (n, m) of the Monte-Carlo sampling scheme. To estimate the sensitivity indices in Definition 3.1 and Definition 3.2, the stochastic model has to be called $(p + 1)mn$ times.

It is reasonable to think of a sensitivity analysis as done the following way. The total number of calls is set to a limit, say T. Then n and m are chosen so that $T = (p + 1)mn$. For instance, suppose that one cannot afford more than 150 calls to a model with two inputs. Then $T = 150$, $p = 2$ and one can choose either one of the columns in the following table

n	50	25	10	5	2	1
m	1	2	5	10	25	50.

Denote by $\mathrm{div}_p(T)$ the set of all divisors of $T/(p + 1)$ between 1 and $T/(p + 1)$. In the example above, $\mathrm{div}_2(150) = \{1, 2, 5, 10, 25, 50\}$. There are as many designs as there are elements in the set $\mathrm{div}_p(T)$. Each one of those elements corresponds to a possible combination for n and m which Algorithm 3.1 can be run with. The resulting estimators require the same number of calls but do not perform equally well. The goal of this section is to find the "best" way to estimate the sensitivity indices.

4.1. Introducing the miss-ranking error and its bound. To compare the estimators, a measure of performance has to be defined. We shall consider the miss-ranking error (MRE), defined by

$$
\mathrm{MRE} = \mathrm{E} \sum_{j=1}^p |\widehat{R}_{j;n,m} - R_j|,
$$

where R_j is the rank of D_j among $D_1, \ldots, D_p$, that is, $R_j = \sum_{i=1}^p \mathbf{1}(D_i \leq D_j)$, and $\widehat{R}_{j;n,m}$ is an estimator of R_j. Recall that $D_1, \ldots, D_p$ are the upper-left terms in (3.6) and (3.5). They determine the ranks of the sensitivity indices. Recall that the ranks of the sensitivity indices of the first kind coincide with the ranks of the sensitivity indices of the second kind. Thus, the MRE permits to find a unique solution for both kinds of sensitivity indices. The MRE is small when one succeeds in ranking the inputs from the most to the least important, a task which is called "factors prioritization" in [27, p. 52].

The MRE has a bound with interesting mathematical properties. Denote by $\mathrm{MRE}(T, m)$ the MRE based on T number of calls and m repetitions, so that the number of explorations is $T/(p+1)/m$. To shorten the notation, let $(X^{(1)}, \widetilde{X}^{(1)}) = \mathbf{X}$, $f(X^{(1)}, Z_0^{(1,1)}) = Y_0$ and $f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) = Y_j$.

Proposition 4.1. *Let $\widehat{D}_{j;n,m}$, $j = 1, \ldots, p$, be the upper-left term in (3.7) or (3.4) and put $\widehat{R}_{j;n,m} = \sum_{i=1}^p \mathbf{1}(\widehat{D}_{i;n,m} \leq \widehat{D}_{j;n,m})$. If $D_1, \ldots, D_p$ are all distincts then*

$$
\begin{aligned}
\mathrm{MRE}(T, m) \leq \frac{L}{nm} \Big(m \sum_{j=1}^p & \mathrm{Var}(\mathrm{E}[Y_0 Y_j | \mathbf{X}]) \\
& + \sum_{j=1}^p \mathrm{E}(\mathrm{Var}[Y_0 Y_j | \mathbf{X}] - \mathrm{Var}[Y_0 | \mathbf{X}] \, \mathrm{Var}[Y_j | \mathbf{X}]) \\
& + \frac{1}{m} \sum_{j=1}^p \mathrm{E}(\mathrm{Var}[Y_0 | \mathbf{X}] \, \mathrm{Var}[Y_j | \mathbf{X}]) \Big),
\end{aligned}
$$

where

$$
L = \frac{4(p-1)}{\min\limits_{j<j'} (|D_j - D_{j'}|^2)}.
$$

The constant L tells us that the bound is smaller when the indices are well separated. The bound goes to zero when the number of explorations goes to infinity. This is true even if the number of repetitions is fixed. Most interestingly, the bound separates T and m: substituting $T/(p+1)$ for nm in the denominator of the bound, we get

$$
\mathrm{MRE}(T, m) \leq \frac{1}{T} v(m), \quad m \in \mathrm{div}_p(T), \tag{4.1}
$$

where

$$
\begin{aligned}
v(m) = L(p+1) \Big(m \sum_{j=1}^p \mathrm{Var}(\mathrm{E}[Y_0 Y_j | \mathbf{X}]) + \sum_{j=1}^p \mathrm{E}(\mathrm{Var}[Y_0 Y_j | \mathbf{X}] - \mathrm{Var}[Y_0 | \mathbf{X}] \, \mathrm{Var}[Y_j | \mathbf{X}]) \\
+ \frac{1}{m} \sum_{j=1}^p \mathrm{E}(\mathrm{Var}[Y_0 | \mathbf{X}] \, \mathrm{Var}[Y_j | \mathbf{X}]) \Big).
\end{aligned}
$$

Denote by $m_T^\dagger$ the element $m \in \mathrm{div}_p(T)$ that minimizes $v(m)$. Taking $m = m_T^\dagger$ in (4.1), we get the bound

$$
\mathrm{MRE}(T, m_T^\dagger) \leq \frac{v(m_T^\dagger)}{T} \leq \frac{v(m)}{T}, \quad \text{for all } m \in \mathrm{div}_p(T).
$$

Thus choosing $m = m_T^\dagger$ and $n = T/(p+1)/m_T^\dagger$ in Algorithm 3.1 ensures that the MRE cannot exceed the least possible bound. The least possible bound $v(m_T^\dagger)/T$ is also called *the best obtainable guarantee*. However, $m_T^\dagger$ is unknown and must be estimated.

Remark 4.2. The choice of T, through the specification of $\operatorname{div}_p(T)$, will influence the quality of the bound. It is clear that choosing $T/(p+1)$ a prime number may not be a good idea because $v(m_T^\dagger)$ will be either $v(1)$ or $v(T/(p+1))$. On the opposite, choosing $T/(p+1)$ a factorial number ensures many more choices (in fact, all).

4.2. A two-stage procedure to estimate the sensitivity indices. The results in Section 4.1 suggest a two-stage procedure to estimate the sensitivity indices. The procedure is given in Algorithm 4.1. The computational budget is split into two parts K and $T-K$. Denote by $m_{T-K}^\dagger$ the element $m \in \operatorname{div}_p(T-K)$ that minimizes the function $v(m)$. The first K calls to the model are used to estimate $m_{T-K}^\dagger$. The last $T-K$ calls to the model are used to estimate the sensitivity indices.

Algorithm 4.1 Estimate the sensitivity indices by a two-stage procedure

Stage 1. Choose an integer K such that $K/(p+1)$ and $(T-K)/(p+1)$ are integers also. Choose integers m_0 and n_0 such that $K = m_0 n_0 (p+1)$. Run Algorithm 3.1 with $m = m_0$ and $n = n_0$. Estimate $m_{T-K}^\dagger$ by an estimator $\widehat{m}_{T-K}^\dagger \in \operatorname{div}_p(T-K)$.

Stage 2. Run Algorithm 3.1 with $m = \widehat{m}_{T-K}^\dagger$ and

$$n = \frac{T-K}{(p+1)\widehat{m}_{T-K}^\dagger}.$$

Compute the sensitivity indices estimators (3.7) and (3.4).

In Algorithm 4.1 we need $\widehat{m}_{T-K}^\dagger$ an estimator of $m_{T-K}^\dagger$. Let us build one. Let m^* be the minimizer of v seen as a function on the positive reals. Since v is convex, the minimizer is unique. It follows from (4.1) and Proposition 4.1 that

$$m^* := \sqrt{\frac{\sum_{j=1}^p \operatorname{E}\operatorname{Var}[Y_0|\mathbf{X}]\operatorname{Var}[Y_j|\mathbf{X}]}{\sum_{j=1}^p \operatorname{Var}\operatorname{E}[Y_0Y_j|\mathbf{X}]}} = \sqrt{\frac{\sum_{j=1}^p \zeta_{3,j}}{\sum_{j=1}^p \zeta_{1,j}}}, \tag{4.2}$$

where $\zeta_{3,j} = \operatorname{E}\operatorname{Var}[Y_0|\mathbf{X}]\operatorname{Var}[Y_j|\mathbf{X}]$ and $\zeta_{1,j} = \operatorname{Var}\operatorname{E}[Y_0Y_j|\mathbf{X}]$, $j = 1, \dots, p$.

Let $\varphi_T : (0,\infty) \longrightarrow \operatorname{div}_p(T)$, be the function defined by $\varphi_T(x) = 1$ if $0 < x < 1$, $\varphi_T(x) = T/(p+1)$ if $x > T/(p+1)$, and

$$\varphi_T(x) = \begin{cases} \llcorner x \lrcorner_T & \text{if } \sqrt{\llcorner x \lrcorner_T \ulcorner x \urcorner_T} > x \geq 1 \\ \ulcorner x \urcorner_T & \text{if } \sqrt{\llcorner x \lrcorner_T \ulcorner x \urcorner_T} \leq x \leq \frac{T}{p+1} \end{cases}$$

where

$$\llcorner x^* \lrcorner_T = \max\{m \in \operatorname{div}_p(T),\, m \leq x\}, \quad \ulcorner x^* \urcorner_T = \min\{m \in \operatorname{div}_p(T),\, m \geq x\}.$$

The function φ_T is piecewise constant with discontinuity points at $\sqrt{ij}$, where i and j are two consecutive elements of $\operatorname{div}_p(T)$.

Proposition 4.3. *If $m^* > 0$ then $m_{T-K}^\dagger = \varphi_{T-K}(m^*)$. If, moreover, $\llcorner m^* \lrcorner_{T-K} \ulcorner m^* \urcorner_{T-K}$ is not equal to m^{*2} then the minimizer of $v(m)$, $m \in \operatorname{div}_p(T-K)$, is unique.*

Proposition 4.3 suggests that $m^\dagger_{T-K}$ can be estimated by applying the function φ_{T-K} to an estimate of m^*. Thus, our problem of estimating $m^\dagger_{T-K}$ boils down to the problem of estimating m^*. Let us find an estimator of m^*. Remember that it has to be based on the first $K = m_0 n_0(p+1)$ calls to the model. In view of (4.2), put

$$\widehat{m}^*_K := \sqrt{\frac{\sum_{j=1}^p \widehat{\zeta}_{3,j}}{\sum_{j=1}^p \widehat{\zeta}_{1,j}}}, \tag{4.3}$$

where

$$\begin{aligned}
&\hat{\zeta}_{3,j} = \\
&\frac{1}{n_0}\sum_{i=1}^n \frac{1}{m_0}\sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})^2 \frac{1}{m_0}\sum_{k_2=1}^{m_0} f(\widetilde{X}^{(i)}_{-j}, Z_j^{(i,k_2)})^2 && (4.4)\\
&+\frac{1}{n_0}\sum_{i=1}^n \left(\frac{1}{m_0}\sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})\right)^2 \left(\frac{1}{m_0}\sum_{k_2=1}^{m_0} f(\widetilde{X}^{(i)}_{-j}, Z_j^{(i,k_2)})\right)^2 && (4.5)\\
&-\frac{1}{n_0}\sum_{i=1}^n \left(\frac{1}{m_0}\sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})\right)^2 \frac{1}{m_0}\sum_{k_2=1}^{m_0} f(\widetilde{X}^{(i)}_{-j}, Z_j^{(i,k_2)})^2 && (4.6)\\
&-\frac{1}{n_0}\sum_{i=1}^n \frac{1}{m_0}\sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})^2 \left(\frac{1}{m_0}\sum_{k_2=1}^{m_0} f(\widetilde{X}^{(i)}_{-j}, Z_j^{(i,k_2)})\right)^2, && (4.7)
\end{aligned}$$

and

$$\begin{aligned}
&\hat{\zeta}_{1,j} = \\
&\frac{1}{n_0}\sum_{i=1}^n \left(\frac{1}{m_0}\sum_{k=1}^{m_0} f(X^{(i)}, Z_0^{(i,k)}) f(\widetilde{X}^{(i)}_{-j}, Z_j^{(i,k)})\right)^2 && (4.8)\\
&-\left(\frac{1}{n_0}\sum_{i=1}^n \frac{1}{m_0}\sum_{k=1}^{m_0} f(X^{(i)}, Z_0^{(i,k)}) f(\widetilde{X}^{(i)}_{-j}, Z_j^{(i,k)})\right)^2. && (4.9)
\end{aligned}$$

Notice that $\widehat{\zeta}_{1,j} \geq 0$ and $\widehat{\zeta}_{3,j} \geq 0$ so that $\widehat{m}^*_K \geq 0$. If $m_0 = 1$ then $\widehat{\zeta}_{3,j} = 0$ and hence $\widehat{m}^*_K = 0$.

The estimator $\widehat{m}^*_K$ is consistent and asymptotically normal on some conditions on the rates of n_0 and m_0.

THEOREM 4.4. *Assume (3.1) holds. Let $n_0 \to \infty$. If m_0 is fixed then*

$$\sqrt{n_0}\left(\widehat{m}^*_K - \left[m^* + \frac{C}{m_0} + \epsilon_{m_0}\right]\right) \xrightarrow{d} \mathcal{N}(0, \sigma^2_{m_0}),$$

where C is some constant, $\epsilon_{m_0} = C_1/m_0^2 + \cdots + C_N/m_0^{N+1}$ for some constants $C_1, \ldots, C_N$ and $\sigma^2_{m_0}$ is some variance depending on m_0. If $m_0 \to \infty$ then the above display with $\epsilon_{m_0} = o(1/m_0)$ and σ_{m_0} replaced by $\lim_{m_0\to\infty} \sigma_{m_0}$ is true.

Theorem 4.4 shows that $\widehat{m}^*_K$ is asymptotically biased. The bias is polynomial in $1/m_0$. Corollary 4.5 shows that letting $m_0 \to \infty$ suffices to get the consistency of $\widehat{m}^*_K$ but to get a central limit theorem centered around m^*, it is furthermore needed that $\sqrt{n_0}/m_0 \to 0$.

Corollary 4.5. *Assume* (3.1) *holds. Let* $n_0 \to \infty$ *and* $m_0 \to \infty$. *Then* $\widehat{m}_K^* \xrightarrow{P} m^*$. *If, moreover,* $\sqrt{n_0}/m_0 \to 0$, *then*

$$\sqrt{n_0}(\widehat{m}_K^* - m^*) \xrightarrow{d} \mathcal{N}(0, \lim_{m_0 \to \infty} \sigma_{m_0}^2).$$

Now we have everything that is needed to estimate $m_{T-K}^\dagger$. Put $\widehat{m}_{T-K}^\dagger = \varphi_{T-K}(\widehat{m}_K^*)$. Proposition 4.6 states that $\widehat{m}_{T-K}^\dagger$ and $m_{T-K}^\dagger$ are equal with probability going to one.

Proposition 4.6. *Assume* (3.1) *holds. Let* $n_0 \to \infty$ *and* $m_0 \to \infty$. *Then*

$$P\left(\widehat{m}_{T-K}^\dagger = m_{T-K}^\dagger\right) \to 1.$$

All the details of Algorithm 4.1 have been given.

4.3. Performance. To get some insight into the performance of the procedure given in Algorithm 4.1, we look at the performance of the sensitivity indices estimators produced in Stage 2. Since they are built with $T-K$ calls to the model with $\widehat{m}_{T-K}^\dagger$ repetitions, equation (4.1) yields

$$\text{MRE}(T-K, \widehat{m}_{T-K}^\dagger) \leq \frac{1}{T-K} v(\widehat{m}_{T-K}^\dagger), \tag{4.10}$$

where the left-hand side is the conditional expectation of the MRE, given the outputs produced in Stage 1. The estimator $\widehat{m}_{T-K}^\dagger$ is computed with K calls only.

It is difficult to compare the guarantee above with that got by choosing an arbitrary number of repetitions, say m. In the later case $K = 0$ and hence the guarantee is $v(m)/T$, see (4.1). The denominator in (4.10) is smaller than T but we expect that the numerator $v(\widehat{m}_{T-K}^\dagger)$, which should be close to $v(m_{T-K}^\dagger)$, will be less than $v(m)$ for many values of m because the number of repetitions $m_{T-K}^\dagger$ has been optimized in some way. Note that the numerator and the denominator in (4.10) cannot be good at the same time and K determines the balance. Under conditions on the rates of K and T, we have the following result.

Theorem 4.7. *Assume that the conditions of Proposition 4.6 are fulfilled. Suppose furthermore that* $K \to \infty$ *such that* $K/T \to 0$. *Then*

$$\frac{1}{T-K} v(\widehat{m}_{T-K}^\dagger) = \frac{1}{T} v(m_{T-K}^\dagger)(1 + o_P(1)).$$

Theorem 4.7 says that, roughly speaking, our garantee is $v(m_{T-K}^\dagger)/T$, which may or may not be better than the best obtainable guarantee $v(m_T^\dagger)/T$, depending on the potential difference between $m_{T-K}^\dagger$ and $m_T^\dagger$. This problem stems from the fact that the sets $\text{div}_p(T-K)$ and $\text{div}_p(T)$ may be different even if T and $T-K$ are of the same order of magnitude. For instance, to ensure that our guarantee is the best possible, we may choose T and K large enough and such that $\text{div}_p(T-K) = \{1, 2, \ldots, (T-K)/(p+1)\}$ and $\text{div}_p(T) = \{1, 2, \ldots, T/(p+1)\}$. Then $v(m_{T-K}^\dagger)$ and $v(m_T^\dagger)$ will be equal. Corollary 4.8 below puts this formally.

Corollary 4.8. *If* $\text{div}_p(T) \cap \text{div}_p(T-K) = \text{div}_p(T-K)$ *then for every fixed* $m \neq m_T^\dagger$, *it holds that* $P(T^{-1}v(m_{T-K}^\dagger)(1 + o_P(1)) \leq T^{-1}v(m)) \to 1$.

Thus, roughly speaking, it is always better, in terms of obtainable guarantees, to use the two-step procedure rather than to choose the number of repetitions arbitrarily, except for the lucky case $m = m_T^\dagger$.

5. Asymptotic normality of the sensitivity indices estimators. The sensitivity indices estimators of Section 3.2 depend on both m and n. It is clear that n should go to infinity to get central limit theorems. It may be less clear, however, whether or not m should go to infinity as well. The answer depends on the kind of the sensitivity index we are looking at.

Two frameworks are considered:

- $n \to \infty$ and m is fixed;
- $n \to \infty$ and $m \to \infty$.

In the second framework $m = m_n$ is a sequence indexed by n that goes to infinity as n goes to infinity. Denote by $\mathbf{S}'$ (resp. $\mathbf{S}''$) the (column) vector with coordinates S'_j (resp. S''_j), $j = 1, \dots, p$, and denote by $\widehat{\mathbf{S}}'_{n,m}$ (resp. $\widehat{\mathbf{S}}''_{n,m}$) the vector with coordinates $\widehat{S}'_{j;n,m}$ given in (3.7) (resp. $\widehat{S}''_{j;n,m}$ given in (3.4)).

THEOREM 5.1. *Assume* (3.1) *holds. Let $n \to \infty$. If m is fixed then*

$$\sqrt{n} \begin{pmatrix} \widehat{\mathbf{S}}'_{n,m} - \mathbf{S}' \\ \widehat{\mathbf{S}}''_{n,m} - \mathbf{S}'' \left[1 - \frac{\mathrm{E}\operatorname{Var}[f(X,Z)|X]}{\mathrm{E}\operatorname{Var}[f(X,Z)|X] + m \operatorname{Var}\mathrm{E}[f(X,Z)|X]} \right] \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, \Xi_m),$$

for some nonnegative matrix Ξ_m of size $2p \times 2p$. If $m \to \infty$ then, elementwise, $\lim_{m\to\infty} \Xi_m$ exists and the above display with Ξ_m replaced by $\lim_{m\to\infty} \Xi_m$ is true.

Theorem 5.1 predicts the behavior of the joint vector $(\widehat{\mathbf{S}}'^{\top}_{n,m}, \widehat{\mathbf{S}}''^{\top}_{n,m})$. However the behaviors of $\widehat{\mathbf{S}}'_{n,m}$ and $\widehat{\mathbf{S}}''_{n,m}$ are different. The estimator $\widehat{\mathbf{S}}'^{\top}_{n,m}$ is asymptotically normal around $\mathbf{S}'$, even if m is kept fixed. The estimator $\widehat{\mathbf{S}}''^{\top}_{n,m}$ is also asymptotically normal, but not around $\mathbf{S}''$.

The estimator $\widehat{\mathbf{S}}''_{n,m}$ under-estimates $\mathbf{S}''$. The bias, given by

$$\mathbf{S}'' \frac{\mathrm{E}\operatorname{Var}[f(X,Z)|X]}{\mathrm{E}\operatorname{Var}[f(X,Z)|X] + m \operatorname{Var}\mathrm{E}[f(X,Z)|X]},$$

is null whenever f actually does not depend on Z, and large whenever the stochastic model is highly stochastic. As Theorem 5.1 shows, the bias is still present even if m goes to infinity. Corollary 5.2 shows that m must go to infinity fast enough to avoid the estimator to be tightly concentrated around the wrong target.

COROLLARY 5.2. *Assume* (3.1) *holds. Let $n \to \infty$. If $m \to \infty$ such that $\sqrt{n}/m \to 0$ then*

$$\sqrt{n} \left(\widehat{\mathbf{S}}''_{n,m} - \mathbf{S}'' \right) \xrightarrow{d} \mathcal{N}(0, \Xi_{22}),$$

where Ξ_{22} is the lower-right block of the matrix $\lim_{m\to\infty} \Xi_m$ given in Theorem 5.1.

The difference between $\widehat{\mathbf{S}}'_{n,m}$ and $\widehat{\mathbf{S}}''_{n,m}$ is due to the difference between the lower-left terms in (3.7) and (3.4). While the lower-left term in (3.7) is unbiased for all n and m, the lower-left term in (3.4) has a bias depending on m which propagates to the estimator of the sensitivity indices. (The calculations are carried out in Appendix D.)

From a statistical perspective, it is more difficult to estimate the sensitivity indices of the second kind than to estimate the sensitivity indices of the first kind. To estimate the former, one needs to repeat the model many times. To estimate the later, this is not necessary.

6. Numerical tests. Section 6.1 illustrates how the MRE responds to a change in the Monte-Carlo design. In Section 6.1 the total budget T is kept fixed. Section 6.2

illustrates how the sensitivity indices estimators behave asymptotically. In Section 6.2 the total budget T increases.

6.1. Comparison of Monte-Carlo designs. The effect of the number of repetitions on the sensitivity indices estimators and the effect of the calibration in the two-stage procedure are examined in two kinds of experiments: the "direct" experiments and the "calibration" experiments.

In the direct experiments, the sensitivity indices are estimated directly with the given number of repetitions. Increasing numbers of repetitions m are tested. (Since the budget is fixed, this goes with decreasing numbers of explorations.) For each m, the mean squared errors (MSEs), given by $\mathrm{E}\sum_{j=1}^{p}(\widehat{S}'_{j;n,m}-S'_j)^2$ and $\mathrm{E}\sum_{j=1}^{p}(\widehat{S}''_{j;n,m}-S''_j)^2$, are estimated with replications. They are also split into the sum of the squared biases and the sum of the variances to get further insight about the behavior of the estimators. The MREs are estimated as well. A normalized version is considered: it is the MRE divided by the number of variables. For models with two inputs, the normalized MRE is interpreted directly as the probability that the two inputs are ranked incorrectly.

In the calibration experiments, the sensitivity indices are estimated with the two-stage procedure, the results of which depend on the calibration parameters K and m_0. Various calibration parameters are tested to see their effect on the MRE. The budgets for the direct experiments and the calibration experiments are the same so that the numbers can be compared. In particular, the direct experiments correspond to the case $K = 0$ in the calibration experiments.

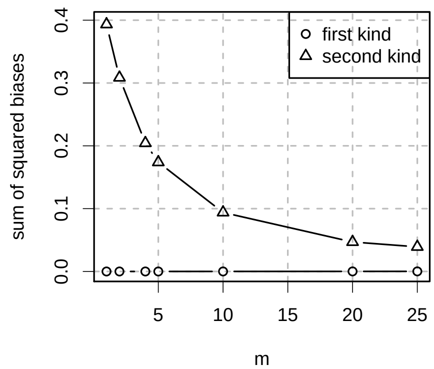
A linear model of the form $Y = X_1 + \beta X_2 + \sigma Z$, where X_1, X_2, Z, are standard normal random variables and β, σ are real coefficients, has been considered because the sensitivity indices are explicit and hence the performance of the estimators can be evaluated easily. The quantity m^* is explicit: the formula is given in Appendix E.

6.1.1. High noise context. The coefficients are $\beta = 1.2$ and $\sigma = 4$. The sensitivity indices are $S'_1 = 0.05$, $S'_2 = 0.08$, $S''_1 = 0.41$ and $S''_2 = 0.59$. The real m^* is about 5.8. The total budget is $T = 3 \times 500 = 1500$ and hence $\mathrm{div}_2(1500) = \{1, 2, 4, 5, 10, 20, 25, 50, 100, 125, 250, 500\}$. The integer $m^\dagger_{1500}$ is equal to $\varphi_{1500}(m^*) = 5$. Since the budget is kept fixed, the numbers of explorations are, respectively, $500, 250, 125, 100, 50, 25, 20, 10, 5, 4, 2, 1$. The number of replications is 1500.

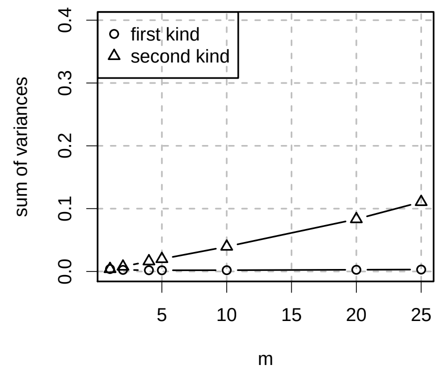
The results of the direct experiment are given in Figure 1 for m = 1, 2, 4, 5, 10, 20, 25. The MSE of first kind does not vary with the number of repetitions and is much lower than the MSE of second kind, see (c). The estimators of the second kind are highly biased for small numbers of repetitions (a) and they have a higher variance for larger numbers of repetitions (b). The fact that the bias is high for small numbers of repetitions agrees with the theory, according to which the bias should vanish as m goes to infinity. Overall, the sensitivity indices of the second kind seem to be much harder to estimate than the indices of the first kind, the estimators of which have a negligible bias and a very small variance whatever the number of repetitions.

According to Figure 1(c), the normalized MRE curve has a banana shape with a minimum of about slightly less than 30% reached around $m \in \{5, 10\}$ and endpoints with a value of about 35%. A value of 30% means that the probability of ranking the inputs correctly is about 70%. The region of observed optimal performance $m \in \{5, 10\}$ coincides with $m^\dagger_{1500} = 5$, the point at which the bound is minimal.

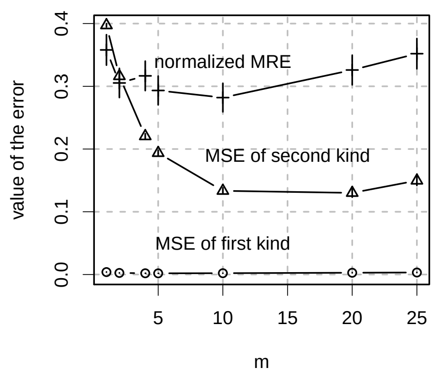
The results of the calibration experiment is given in Table 1 for the normalized MRE. The lowest MREs are reached at the bottom right of the table, with values



(a) Squared bias



(b) Variance



(c) Errors

Fig. 1: Sum of squared biases (a), sum of variances (b) and errors (c) of the sensitivity indices estimators for the linear model in the high noise setting. Confidence intervals of level 95% are added in (c).

corresponding to $2 \leq m \leq 10$ in Figure 1 (c). Optimal performance is reached with very few explorations in the first stage of the two-stage procedure. In this case, the estimator $\widehat{m}_K^*$ has a small bias but a high variance. It seems to be better than an estimator with a small variance but a large bias. This might be explained by the low curvature of the MRE curve.

6.1.2. Low noise context. The coefficients are $\beta = 1.2$ and $\sigma = 0.9$. The sensitivity indices are $S_1' = 0.31$, $S_2' = 0.44$, $S_1'' = 0.41$ and $S_2'' = 0.59$. The real m^* is about 0.30 and hence the integer $m_{1500}^\dagger$ is equal to 1. As expected, these numbers are smaller than the ones found in the high noise context. The total budget

	m_0				n_0			
$K/3$	2	5	10	20	20	10	5	2
400	0.43	0.42	0.42	-	-	0.42	0.39	0.40
200	0.38	0.39	0.37	-	-	0.35	0.35	0.34
100	0.36	0.37	-	-	-	-	0.32	0.30
50	0.39	0.33	-	-	-	-	0.33	0.31

Table 1: Normalized MRE in the linear model with high noise for various calibrations: $K/(p+1) = 50, 100, 200, 400$ and $m_0 = 2, 5, 10, 20, \ldots$ For instance, for $K/(p+1) = 200 = m_0 n_0$, the normalized MRE is available for $m_0 = 2, 5, 10, 20, 40, 100$.

	m_0				n_0			
$K/3$	2	5	10	20	20	10	5	2
400	0.18	0.15	0.17	-	-	0.16	0.18	0.20
200	0.05	0.04	0.04	-	-	0.06	0.05	0.07
100	0.02	0.04	-	-	-	-	0.04	0.04
50	0.03	0.02	-	-	-	-	0.02	0.04

Table 2: Normalized MRE in the linear model with low noise for various calibrations: $K/(p+1) = 50, 100, 200, 400$ and $m_0 = 2, 5, 10, 20, \ldots$ For instance, for $K/(p+1) = 200 = m_0 n_0$, the normalized MRE is available for $m_0 = 2, 5, 10, 20, 40, 100$.

is $T = 3 \times 500 = 1500$. The number of replications is 500.

The results for the direct experiment are given in Figure 2. The MSE of first kind increases with the number of repetitions, see (c): this is due to the increase of the variance (b), while the bias is negligible (a). As in the high noise context, the estimators of the second kind have a decreasing bias and an increasing variance, although the decrease of the bias is of much less magnitude. This agrees with the theory, where we have seen that, for the sensitivity indices of the second kind, the biases of the estimators are small when the noise of the model is low.

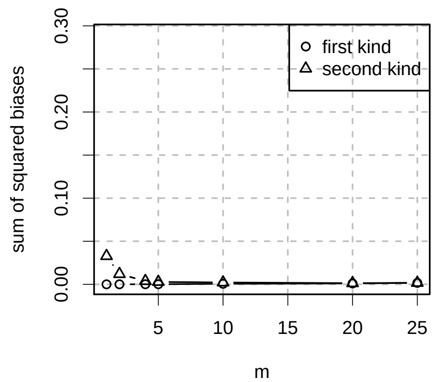
In Figure 2 (c), the normalized MRE varies a lot. It increases from about 2% at $m = 1$ to 30% at $m = 25$. Thus, unlike in the high noise setting, choosing a good number of repetitions is important. The best performance is achieved at $m = 1$, which coincides with the minimizer $m^\dagger_{1500} = 1$ of the bound.

The results of the calibration experiment for the normalized MRE is given in Table 2. The best performance is reached at the bottom left of the table with numbers that correspond to the optimal performance in Figure 2 (c). Moreover, notice that a large spectrum of calibration parameters (K, m_0) yield low errors.

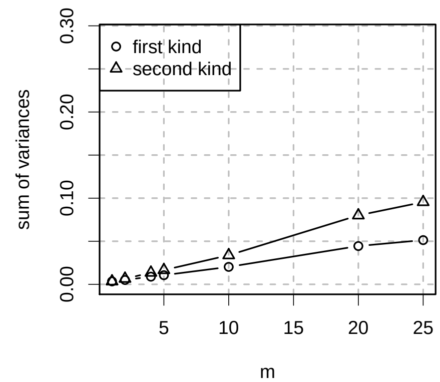
6.2. Asymptotic behavior of the sensitivity indices estimators. To illustrate the asymptotic behavior of the sensitivity indices estimators, Sobol's g-function, a benchmark in sensitivity analysis [25, 20], is considered. Sobol's g-function is given by

$$g(U_1, \ldots, U_{p+1}) = \prod_{j=1}^{p+1} \frac{|4U_j - 2| + a_j}{1 + a_j},$$

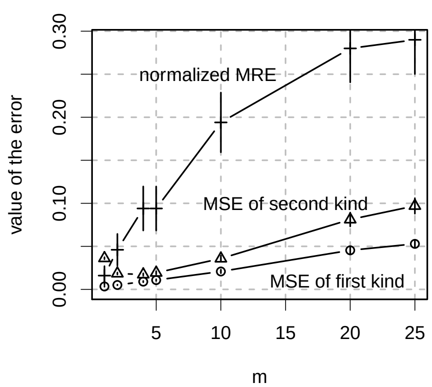
where the a_j are nonnegative and the U_j are independent standard uniform random variables. The less a_j the more U_j is important. Elementary calculations show that



(a) Squared bias



(b) Variance



(c) Error

Fig. 2: Sum of squared biases (a), sum of variances (b) and errors (c) of the sensitivity indices estimators for the linear model in the low noise context. Confidence intervals of level 95% are added in (c).

the first-order Sobol index associated with U_j is given by

$$S_j^{(a_1,\ldots,a_{p+1})} = \frac{1}{3(1+a_j)^2}\left(-1+\prod_{j=1}^{p+1}\frac{(4/3+a_j^2+2a_j)}{(1+a_j)^2}\right)^{-1}.$$

To build a stochastic model out of Sobol's g-function, we let one of the U_j play the role of Z. For instance if U_i, $1 \le i \le p+1$, were to play this role, then the stochastic model would be

$$(6.1)\qquad Y = f(X_1,\ldots,X_p,Z) = g(X_1,\ldots,X_{i-1},Z,X_i,\ldots,X_p).$$

Of course Y and f above depend on i. In the rest of this section we choose arbitrarily $i = 2$ and $p = 4$.

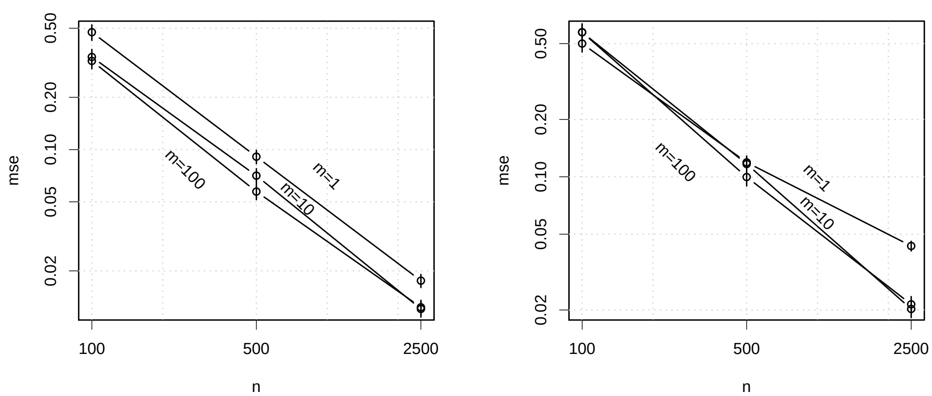
The Sobol indices of the first and of the second kind (in the sense of Definition 3.1 and 3.2) are then easily seen to be

$$S'_j = \begin{cases} S_j^{(a_1,\ldots,a_{p+1})} & \text{if } 1 \le j \le i-1 \\ S_{j+1}^{(a_1,\ldots,a_{p+1})} & \text{if } i \le j \le p \end{cases}$$

and $S''_j = S_j^{(b_{i1},\ldots,b_{ip})}$, where

$$b_{ij} = \begin{cases} a_j & \text{if } 1 \le j \le i-1, \\ a_{j+1} & \text{if } i \le j \le p. \end{cases}$$

For each kind of Sobol index, we produced 500 estimates of the p Sobol indices and computed the values of the mean squared error (MSE) by averaging over the 500 replications and summing over the p indices. We tested $n = 100, 500, 2500$ and $m = 1, 10, 100$.



(a) first kind (b) second kind

Fig. 3: MSEs for the Sobol index estimators of the first and second kind (logarithmic scale).

The MSEs are shown in Figure 3. Let us look at 3a. As n increases, the decrease is linear for each m. This indicates that the MSEs go to zero at a polynomial rate, even if m is fixed (look at the line $m = 1$). This agrees with the theoretical results of Section 5. The picture is different for the estimator of Sobol indices of the second kind. In 3b, the curve for $m = 1$ is not a straight line, indicating that the MSE may not go to zero. Indeed, the MSE for m fixed is not expected to go to zero because of the bias depending on m. To make the MSE go to zero, one has to force m go to infinity.

Figure 4, which shows the distribution of the estimates for the index associated to X_1, better explains this phenomenon. Here the bias is apparent for $m = 1$ and vanishes as m goes to infinity. The bias for the indices associated with the other inputs is not as large (not shown here).

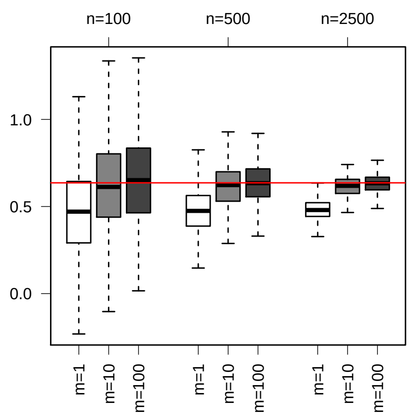


Fig. 4: Boxplots of the estimates for the Sobol index of the second kind associated with X_1. The red horizontal line is the truth.

7. Conclusion. The practical method that consists of repeating the stochastic model at each exploration of the input space was analysed in the context of global sensitivity analysis. A method to find a tradeoff between the number of explorations n and the number of repetitions m was proposed. The missranking error (MRE) was used to measure the performance of the sensitivity analysis because it is both sensible and fruitful, as it allows to explicitly find a pair (n, m) that is expected to be approximately optimal. A two-step procedure was given to build an estimator that, asymptotically, does always better than any pair (n, m), except the optimal one. Two sensitivity indices were analysed. The sensitivity index of the first kind results from the existence of a function that links the output, the inputs and some random noise, which were shown to represent stochastic models defined through probability measures. The sensitivity index of the second kind is the index to which the estimator (1.4) converges. An asymptotic analysis was conducted. It was found that the estimators for the indices of the second kind are asymptotically biased, while the estimators for the indices of the first kind are not. To test the theory, simulation experiments were conducted, where the bias of the sensitivity estimator of the second kind was confirmed. Optimal compromises between repetitions and explorations have been identified and compared with the output of the two-stage procedure for different values of the tuning parameters.

This work opens many research directions. First, the sensitivity estimators of the two stages could be aggregated to build estimators with a lower variance. Second, other methods might be developed to optimize the Monte-Carlo sampling scheme. For instance the MSE might be approximated or asymptotic variance-covariance matrices might be minimized. Third, multilevel Monte-Carlo sampling schemes might be considered to alleviate the bias issue. Fourth, a finite-sample analysis could be conducted to get insight into the tradeoff K is subjected to. Fifth, since the bias is known, it could be estimated to build bias-corrected sensitivity indices estimators. Sixth, the problem of choosing a number of calls with many divisors must be addressed. It may be worth to call the model a bit less if this permits to have a better set $\mathrm{div}_p(T)$. Seventh, the connection between our representation of stochastic models and that of [10] could be investigated further.

Appendix A. Calculations of some sensitivity indices.

A.1. Calculations for $\widetilde{S}_1^{\mathbf{HAG}}$ in Example 2. We have

$$\widetilde{S}_1^{\mathrm{HAG}} = \mathrm{E}\left(\frac{\mathrm{Var}(\mathrm{E}[f(X,Z)|X_j,Z]|Z)}{\mathrm{Var}(f(X,Z)|Z)}\right) = \int_\Omega \frac{\mathrm{Var}(\mathrm{E}[f(X,Z)|X_j,Z]|Z)}{\mathrm{Var}(f(X,Z)|Z)}\,\mathrm{d}P.$$

Since the term inside the integral is a function of Z and the law of Z is the standard uniform distribution, a change of measures yields

$$\widetilde{S}_1^{\mathrm{HAG}} = \int_{(0,1)} \frac{\mathrm{Var}(\mathrm{E}[f(X,z)|X_j,Z=z]|Z=z)}{\mathrm{Var}(f(X,z)|Z=z)}\,\mathrm{d}z = \int_{(0,1)} \frac{\mathrm{Var}(\mathrm{E}[f(X,z)|X_1])}{\mathrm{Var}(f(X,z))}\,\mathrm{d}z.$$

It remains to know what the ratio inside the integral is. We have

$$\begin{aligned}\mathrm{Var}(f(X,z)) = \mathrm{Var}(\Phi^{-1}(z)X_2 + X_1) &= \Phi^{-1}(z)^2\,\mathrm{Var}(X_2) + \mathrm{Var}(X_1)\\ &= \Phi^{-1}(z)^2\frac{L^2}{12} + \frac{1}{12},\end{aligned}$$

and

$$\begin{aligned}\mathrm{Var}(\mathrm{E}[f(X,z)|X_1]) &= \mathrm{Var}(\mathrm{E}[\Phi^{-1}(z)X_2 + X_1|X_1])\\ &= \mathrm{Var}(\Phi^{-1}(z)\,\mathrm{E}[X_2|X_1] + \mathrm{E}[X_1|X_1])\\ &= \mathrm{Var}(\Phi^{-1}(z)\,\mathrm{E}[X_2] + X_1)\\ &= \mathrm{Var}(X_1)\\ &= \frac{1}{12}\end{aligned}$$

and hence

$$\widetilde{S}_1^{\mathrm{HAG}} = \int_{(0,1)} \frac{1}{\Phi^{-1}(z)^2L^2+1}\,\mathrm{d}z = \int_{-\infty}^{\infty} \frac{1}{z^2L+1}\frac{1}{\sqrt{2\pi}}e^{-z^2/2}\,\mathrm{d}z.$$

A.2. Calculations for S_1' in Example 4. The sensitivity index of the first kind associated with the first input is given by

$$S_1' = \frac{\mathrm{Var}[\mathrm{E}(X_1 + X_2\Phi^{-1}(Z)|X_1)]}{\mathrm{Var}[X_1 + X_2\Phi^{-1}(Z)]}.$$

The numerator is given by $\mathrm{Var}[\mathrm{E}(X_1 + X_2\Phi^{-1}(Z)|X_1)] = \mathrm{Var}[X_1 + \mathrm{E}(X_1\Phi^{-1}(Z))] = \mathrm{Var}[X_1] = 1/12$. The denominator is given by $\mathrm{Var}[X_1 + X_2\Phi^{-1}(Z)] = \mathrm{Var}[X_1] + \mathrm{Var}[X_2\Phi^{-1}(Z)] = 1/12 + \mathrm{Var}[X_2\Phi^{-1}(Z)]$, where

$$\begin{aligned}\mathrm{Var}[X_2\Phi^{-1}(Z)] &= \mathrm{Var}[\mathrm{E}(X_2\Phi^{-1}(Z)|Z)] + \mathrm{E}(\mathrm{Var}[X_2\Phi^{-1}(Z)|Z])\\ &= \mathrm{Var}\left[\Phi^{-1}(Z)\left(\frac{L}{2}+1\right)\right] + \int_0^1 \Phi^{-1}(z)^2\,\mathrm{Var}[X_2]\,\mathrm{d}z\\ &= \left(\frac{L}{2}+1\right)^2 + \frac{L^2}{12},\end{aligned}$$

so that

$$S_1' = \frac{1/12}{1/12 + (L/2+1)^2 + L^2/12} = \frac{1}{4(L^2 + 3(L+1)) + 1}.$$

Appendix B. Proofs.

B.1. Proof of Lemma 2.2. Since P^* is a product probability measure, we can write $P^* = \otimes_{j=1}^p P_j^*$. Let $\Omega = (0,1)^{p+1}$ endowed with its Borel σ-field and let P be the product Lebesgue measure $\lambda^{\otimes_{j=1}^{p+1}}$. If F_j denotes the distribution function corresonding to P_j^* then, for $\omega = (\omega_1, \dots, \omega_{p+1}) \in \Omega$, put $X_j(\omega) = F_j^{\leftarrow}(\omega_j) := \inf\{x_j \in \mathbb{R} : F_j(x_j) \geq \omega_j\}$ for all $j = 1, \dots, p$ and $Z(\omega) = \omega_{p+1}$. Take $f(x,z) = F_x^{\leftarrow}(z) := \inf\{t \in \mathbb{R} : F_x(t) \geq z\}$, $z \in (0,1)$, where F_x is the cumulative distribution function associated with Q_x. Standard probability techniques show that $f(x,Z)$ is measurable for every x. Moreover, for every $t \in \mathbb{R}$,

$$\begin{aligned} & P(f(x,Z) \leq t) \\ =& P(Z \leq F_x(t)) = \lambda^{\otimes_{j=1}^{p+1}}\{\omega \in \Omega : \omega_{p+1} \leq F_x(t)\} = \lambda(0, F_x(t)] = F_x(t). \end{aligned}$$

Finally, by the same token,

$$\begin{aligned} & P(X_1 \leq t_1, \dots, X_p \leq t_p, Z \leq t_{p+1}) \\ =& P\{\omega : \omega_1 \leq F_1(t_1), \dots, \omega_p \leq F_p(t_p), \omega_{p+1} \leq t_{p+1}\} = t_{p+1} \prod_{j=1}^p F_j(t_j). \end{aligned}$$

The proof is complete.

Proof of Proposition 4.1. Assume without loss of generality that $D_1 < \cdots < D_p$. We first prove the following Lemma. For convenience, the subscripts n and m are left out.

LEMMA B.1. *Let $i < j$. Then*

$$P(\widehat{D}_i - \widehat{D}_j \geq 0) \leq \frac{\operatorname{Var} \widehat{D}_i + \operatorname{Var} \widehat{D}_j}{\frac{1}{2}|D_i - D_j|^2}.$$

Proof. *We have*

$$\begin{aligned} P(\widehat{D}_i - \widehat{D}_j \geq 0) \leq & P(|\widehat{D}_i - D_i| + |\widehat{D}_j - D_j| \geq D_j - D_i) \\ \leq & P(|\widehat{D}_i - D_i|^2 + |\widehat{D}_j - D_j|^2 \geq \frac{1}{2}|D_j - D_i|^2) \end{aligned}$$

and the claim follows from Markov's inequality.

We now prove Proposition 4.1. Recall that $D_1 < \cdots < D_p$. We have

$$\begin{aligned} \sum_{i=1}^p \operatorname{E}|\widehat{R}_i - R_i| \leq & \sum_{i=1}^p \sum_{j=1}^p \operatorname{E}|\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i) - \mathbf{1}(D_j \leq D_i)| \\ \leq & \sum_{i=1}^p \sum_{j \neq i} \frac{\operatorname{Var} \widehat{D}_i + \operatorname{Var} \widehat{D}_j}{\frac{1}{2}|D_i - D_j|^2} \\ \leq & \frac{4(p-1)}{\min_{j<j'} |D_j - D_{j'}|^2} \sum_{i=1}^p \operatorname{Var} \widehat{D}_i, \end{aligned}$$

where the second inequality holds by Lemma B.1 and because

$$\mathrm{E}\,|\mathbf{1}(\widehat{D}_j \le \widehat{D}_i) - \mathbf{1}(D_j \le D_i)| = \begin{cases} \mathrm{E}\,|\mathbf{1}(\widehat{D}_j > \widehat{D}_i)| & \text{if } j < i, \\ 0 & \text{if } j = i, \\ \mathrm{E}\,|\mathbf{1}(\widehat{D}_j \le \widehat{D}_i)| & \text{if } j > i. \end{cases}$$

It remains to calculate the variances. But this is done in Lemma D.3 in Appendix D, where it is found that

$$\begin{aligned} \operatorname{Var} \widehat{D}_j =& \frac{1}{n}\{\operatorname{Var}\mathrm{E}[Y_0 Y_j|\mathbf{X}] + \frac{1}{m}(\mathrm{E}\operatorname{Var}[Y_0 Y_j|\mathbf{X}] - \operatorname{Var}[Y_0|\mathbf{X}]\operatorname{Var}[Y_j|\mathbf{X}]) \\ &+ \frac{1}{m^2}\,\mathrm{E}\operatorname{Var}[Y_0|\mathbf{X}]\operatorname{Var}[Y_j|\mathbf{X}]\}. \end{aligned}$$

Proof of Proposition 4.3. We distinguish between three cases: $0 < m^* < 1$, $m^* > (T-K)/(p+1)$ and $1 \le m^* \le (T-K)/(p+1)$. Recall that $m^\dagger_{T-K}$ is the minimizer of $v(m)$, m in $\operatorname{div}_p(T-K)$.

If $0 < m^* < 1$ then by definition $\varphi_{T-K}(m^*) = 1$ and by convexity $v(m^*) \le v(1) \le v(m)$ for all m in $\operatorname{div}_p(T-K)$. Therefore $m^\dagger_{T-K} = 1$.

If $m^* > (T-K)/(p+1)$ then by definition $\varphi_{T-K}(m^*) = (T-K)/(p+1)$ and by convexity $v(m^*) \le v((T-K)/(p+1)) \le v(m)$ for all m in $\operatorname{div}_p(T-K)$. Therefore $m^\dagger_{T-K} = (T-K)/(p+1)$.

If $1 \le m^* \le (T-K)/(p+1)$ then by definition

$$\varphi_{T-K}(m^*) = \begin{cases} \llcorner m^* \lrcorner_{T-K} & \text{if } \sqrt{\llcorner m^* \lrcorner_{T-K} \ulcorner m^* \urcorner_{T-K}} > m^* \\ \ulcorner m^* \urcorner_{T-K} & \text{if } \sqrt{\llcorner m^* \lrcorner_{T-K} \ulcorner m^* \urcorner_{T-K}} \le m^*. \end{cases}$$

By convexity $m^\dagger_{T-K}$ must be $\llcorner m^* \lrcorner_{T-K}$ or $\ulcorner m^* \urcorner_{T-K}$. If $\llcorner m^* \lrcorner_{T-K} = \ulcorner m^* \urcorner_{T-K}$ then $m^\dagger_{T-K} = \ulcorner m^* \urcorner_{T-K} = \varphi_{T-K}(m^*)$. Otherwise, since $v(x) = \zeta_1 x + \zeta_2 + \zeta_3/x$, $x > 0$, for some constants ζ_1, ζ_2 and ζ_3 such that $\zeta_3/\zeta_1 = m^*$, we have

$$v(\llcorner m^* \lrcorner_{T-K}) < v(\ulcorner m^* \urcorner_{T-K}) \text{ iff } \sqrt{\llcorner m^* \lrcorner_{T-K} \ulcorner m^* \urcorner_{T-K}} > \frac{\zeta_3}{\zeta_1} = m^*.$$

Therefore $\varphi_{T-K}(m^*) = m^\dagger_{T-K}$.

Let us prove that the minimizer of $v(m)$, $m \in \operatorname{div}_p(T-K)$, is unique if $m^* \neq \sqrt{\llcorner m^* \lrcorner_{T-K} \ulcorner m^* \urcorner_{T-K}}$. If it were not, then we would have $v(\llcorner m^* \lrcorner_{T-K}) = v(\ulcorner m^* \urcorner_{T-K})$. Bus this implies $m^* = \sqrt{\llcorner m^* \lrcorner_{T-K} \ulcorner m^* \urcorner_{T-K}}$, which is a contradiction.

Proof of Theorem 4.4. In this proof m_0 and n_0 are denoted by m and n, respectively. In view of (4.3) and (4.4)–(4.9), we have

$$\widehat{m}^*_K = \sqrt{\frac{\sum_{j=1}^p \widehat{\zeta}_{3,j}}{\sum_{j=1}^p \widehat{\zeta}_{1,j}}} = \sqrt{\frac{\sum_{j=1}^p \frac{1}{n}\sum_{i=1}^n \xi^{(4.4)}_{j;m,i} + \xi^{(4.5)}_{j;m,i} - \xi^{(4.6)}_{j;m,i} - \xi^{(4.7)}_{j;m,i}}{\sum_{j=1}^p \frac{1}{n}\sum_{i=1}^n \xi^{(4.8)}_{j;m,i} - \left(\frac{1}{n}\sum_{i=1}^n \xi^{(4.9)}_{j;m,i}\right)^2}},$$

where the $\xi^{(e)}_{j;m,i}$, $i = 1,\ldots,n$, $j = 1,\ldots,p$, $e = 4.4,\ldots,4.9$, are implicitly defined through (4.4)–(4.9). Let

$$\begin{aligned}
\overline{\boldsymbol{\xi}} &= \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\xi}_{m,i},\\
\boldsymbol{\xi}_{m,i} &= (\boldsymbol{\xi}^\top_{1;m,i},\ldots,\boldsymbol{\xi}^\top_{p;m,i})^\top, \quad i = 1,\ldots,n,\\
\boldsymbol{\xi}_{j;m,i} &= (\xi^{(4.4)}_{j;m,i},\ldots,\xi^{(4.9)}_{j;m,i})^\top, \quad j = 1,\ldots,p, \quad i = 1,\ldots,n.
\end{aligned}$$

Let s be the function defined by

$$s(\mathbf{x}) = \sqrt{\frac{\sum_{j=1}^{p} x^{(4.4)}_j + x^{(4.5)}_j - x^{(4.6)}_j - x^{(4.7)}_j}{\sum_{j=1}^{p} x^{(4.8)}_j - x^{(4.9)2}_j}},$$

where $\mathbf{x} = (\mathbf{x}^\top_1,\ldots,\mathbf{x}^\top_p)^\top$, $\mathbf{x}_j = (x^{(4.4)}_j,\ldots,x^{(4.9)}_j)^\top$, $j = 1,\ldots,p$. With the above notation we have $\widehat{m}^*_K = s(\overline{\boldsymbol{\xi}})$. Moreover, elementary calculations show that

$$\mathrm{E}\,\boldsymbol{\xi}_{m,1} = \boldsymbol{\theta} + \sum_{\nu=1}^{4}\frac{\mathbf{C}_\nu}{m^\nu}, \tag{B.1}$$

where the $\mathbf{C}_\nu$ are vectors of constants, $\boldsymbol{\theta} = (\boldsymbol{\theta}^\top_1,\ldots,\boldsymbol{\theta}^\top_p)^\top$ and

$$\boldsymbol{\theta}_j = \mathrm{E}\begin{pmatrix} Y^{(1,1)2}_0 Y^{(1,1)2}_j \\ Y^{(1,1)}_0 Y^{(1,2)}_0 Y^{(1,1)}_j Y^{(1,2)}_j \\ Y^{(1,1)}_0 Y^{(1,2)}_0 Y^{(1,1)2}_j \\ Y^{(1,1)}_j Y^{(1,2)}_j Y^{(1,1)2}_0 \\ Y^{(1,1)}_0 Y^{(1,2)}_0 Y^{(1,1)}_j Y^{(1,2)}_j \\ Y^{(1,1)}_j Y^{(1,1)}_0 \end{pmatrix}.$$

Check that $m^* = s(\boldsymbol{\theta})$. A concatenation of two Taylor expansions yield

$$\begin{aligned}
&\sqrt{n}(\overline{\boldsymbol{\xi}} - \mathrm{E}\,\boldsymbol{\xi}_{m,1})^\top \dot{s}(\mathrm{E}\,\boldsymbol{\xi}_{m,1}) + \frac{1}{2}(\overline{\boldsymbol{\xi}} - \mathrm{E}\,\boldsymbol{\xi}_{m,1})^\top \ddot{s}_{n,m}(\overline{\boldsymbol{\xi}} - \mathrm{E}\,\boldsymbol{\xi}_{m,1})\\
&= \sqrt{n}(s(\overline{\boldsymbol{\xi}}) - s(\mathrm{E}\,\boldsymbol{\xi}_{m,1})) \qquad \text{(B.2)}\\
&= \sqrt{n}(s(\overline{\boldsymbol{\xi}}) - s(\boldsymbol{\theta}) - (\mathrm{E}\,\boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \dot{s}(\boldsymbol{\theta}) - \frac{1}{2}(\mathrm{E}\,\boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \ddot{s}_m(\mathrm{E}\,\boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})),
\end{aligned}$$

where $\dot{s}$ is the gradient of s, $\ddot{s}_{n,m}$ is the Hessian matrix of s at a point between $\overline{\boldsymbol{\xi}}$ and $\boldsymbol{\theta}_m$, and, $\ddot{s}_m$ is the Hessian matrix of s at a point between $\mathrm{E}\,\boldsymbol{\xi}_{m,1}$ and $\boldsymbol{\theta}$. It follows from (B.1) that $(\mathrm{E}\,\boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \dot{s}(\boldsymbol{\theta})$ is clearly of the form $\sum_{\nu=1}^{4} C_\nu/m^\nu$ for some constants C_ν. Putting

$$\epsilon_m = \frac{1}{2}(\mathrm{E}\,\boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \ddot{s}_m(\mathrm{E}\,\boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})) + \sum_{\nu=2}^{4}\frac{C_\nu}{m^\nu},$$

it follows from (B.2) that

$$\begin{aligned}
&\sqrt{n}(\overline{\boldsymbol{\xi}} - \mathrm{E}\,\boldsymbol{\xi}_{m,1})^\top \dot{s}(\mathrm{E}\,\boldsymbol{\xi}_{m,1}) + \frac{1}{2}(\overline{\boldsymbol{\xi}} - \mathrm{E}\,\boldsymbol{\xi}_{m,1})^\top \ddot{s}_{n,m}(\overline{\boldsymbol{\xi}} - \mathrm{E}\,\boldsymbol{\xi}_{m,1}) \qquad \text{(B.3)}\\
&\qquad = \sqrt{n}(\widehat{m}^*_K - m^* - \frac{C_1}{m} - \epsilon_m).
\end{aligned}$$

If m is fixed then Lemma C.2 in Appendix C yields

$$\sqrt{n}(\overline{\boldsymbol{\xi}} - \mathrm{E}\,\boldsymbol{\xi}_{m,1}) \to \mathcal{N}(0, \Sigma_m),$$

for some variance-covariance matrix Σ_m of size $6p \times 6p$. Moreover, the second term in the left-hand side of (B.3) is $o_P(1)$ by Cauchy-Schwartz's inequality and the continuity of the second derivatives of s. The first term goes to $\mathcal{N}(0, \dot{s}(\mathrm{E}\,\boldsymbol{\xi}_{m,1})^\top \Sigma_m \dot{s}(\mathrm{E}\,\boldsymbol{\xi}_{m,1}))$ and hence the claim follows with $\sigma_m^2 = \dot{s}(\mathrm{E}\,\boldsymbol{\xi}_{m,1})^\top \Sigma_m \dot{s}(\mathrm{E}\,\boldsymbol{\xi}_{m,1})$ and $C = C_1$.

If $m \to \infty$ then again Lemma C.2 in Appendix C applies: we have

$$\sqrt{n}(\overline{\boldsymbol{\xi}} - \mathrm{E}\,\boldsymbol{\xi}_{m,1}) \to \mathcal{N}(0, \lim_{m\to\infty} \Sigma_m).$$

Since $\epsilon_m - \sum_{\nu=2}^4 C_\nu/m^\nu = o(m^{-1})$, $\dot{s}$ is continuous and $\mathrm{E}\,\boldsymbol{\xi}_{m,1} \to \boldsymbol{\theta}$, the claim follows. The proof is complete.

Proof of Proposition 4.6. By definition, $\widehat{m}^\dagger_{T-K} = \varphi_{T-K}(\widehat{m}^*_K)$ and $m^\dagger_{T-K} = \varphi_{T-K}(m^*)$. The function φ_{T-K} is piecewise constant and has $|\mathrm{div}_p(T-K)|-1$ points of discontinuity of the form $\sqrt{ij}$, where i and j are two consecutive members of

$$\mathrm{div}_p(T-K) \setminus \left\{1, \frac{T-K}{p+1}\right\}.$$

Denote the set of discontinuity points by $\mathcal{D}_{T-K}$. Clearly,

$$\mathcal{D}_{T-K} \subset \{\sqrt{ij} : i \text{ and } j \text{ are two consecutive integers}\} = \mathcal{E}.$$

There exists an open interval that contains m^* but does not contain any points of $\mathcal{E}$ and hence does not contain any points of $\mathcal{D}_{T-K}$, whatever T and K. If $\widehat{m}^*_K$ is in this interval then there are no discontinuity points between m^* and $\widehat{m}^*_K$ and hence $\widehat{m}^\dagger_{T-K} = \varphi_{T-K}(\widehat{m}^*_K) = \varphi_{T-K}(m^*) = m^\dagger_{T-K}$. By Corollary 4.5, the probability of that event goes to one as m_0 and n_0 go to infinity.

Proof of Theorem 4.7. Let $\varepsilon > 0$. An obvious algebraic manipulation and Taylor's expansion yield

$$\begin{aligned}
&P\left(\left|\frac{\frac{1}{T-K}v(\widehat{m}^\dagger_{T-K}) - \frac{1}{T}v(m^\dagger_{T-K})}{\frac{1}{T}v(m^\dagger_{T-K})} > \varepsilon\right|\right) \\
&\le P\left(\left|\frac{T}{T-K}(\widehat{m}^\dagger_{T-K} - m^\dagger_{T-K})v'(\widetilde{m}) + \frac{K}{T-K}v(m^\dagger_{T-K})\right| > v(m^\dagger_{T-K})\varepsilon\right),
\end{aligned}$$

where $\widetilde{m}$ denotes a real between $\widehat{m}^\dagger_{T-K}$ and $m^\dagger_{T-K}$. A decomposition of the probability above according to whether $\widehat{m}^\dagger_{T-K} - m^\dagger_{T-K} \neq 0$ or $\widehat{m}^\dagger_{T-K} - m^\dagger_{T-K} = 0$ yields the bound

$$P\left(\widehat{m}^\dagger_{T-K} - m^\dagger_{T-K} \neq 0\right) + P\left(\frac{K}{T-K} > \varepsilon\right).$$

The first term goes to zero by Proposition 4.6. The second term goes to zero because $K/T \to 0$.

Proof of Theorem 5.1. The proof is based on the results in Appendix C. The Sobol estimators in (3.7) and (3.4) are of the form

$$\widehat{S}'_{j;n,m} = \frac{\frac{1}{n}\sum_{i=1}^{n} \xi^{\mathrm{UL}}_{j;m,i} - \left(\frac{1}{n}\sum_{i=1}^{n} \xi^{\mathrm{UR}}_{m,i}\right)^2}{\frac{1}{n}\sum_{i=1}^{n} \xi'^{\mathrm{LL}}_{m,i} - \left(\frac{1}{n}\sum_{i=1}^{n} \xi^{\mathrm{UR}}_{m,i}\right)^2}, \quad j = 1, \dots, p,$$

and

$$\widehat{S}''_{j;n,m} = \frac{\frac{1}{n}\sum_{i=1}^{n} \xi^{\mathrm{UL}}_{j;m,i} - \left(\frac{1}{n}\sum_{i=1}^{n} \xi^{\mathrm{UR}}_{m,i}\right)^2}{\frac{1}{n}\sum_{i=1}^{n} \xi''^{\mathrm{LL}}_{m,i} - \left(\frac{1}{n}\sum_{i=1}^{n} \xi^{\mathrm{UR}}_{m,i}\right)^2}, \quad j = 1, \dots, p,$$

where the notation is obvious. Denote $\boldsymbol{\xi}_{m,i} := (\xi^{\mathrm{UL}}_{1;m,i}, \dots, \xi^{\mathrm{UL}}_{p;m,i}, \xi^{\mathrm{UR}}_{m,i}, \xi'^{\mathrm{LL}}_{m,i}, \xi''^{\mathrm{LL}}_{m,i})^\top$. Elementary but burdensome calculations show that

$$\mathrm{E}\,\boldsymbol{\xi}_{m,1} = \begin{pmatrix} \mathrm{E}\,\mathrm{E}[f(X,Z)|X]\,\mathrm{E}[f(\widetilde{X}_{-1},Z)|\widetilde{X}_{-1}] \\ \vdots \\ \mathrm{E}\,\mathrm{E}[f(X,Z)|X]\,\mathrm{E}[f(\widetilde{X}_{-p},Z)|\widetilde{X}_{-p}] \\ \mathrm{E}\,f(X,Z) \\ \mathrm{E}\,f(X,Z)^2 \\ \mathrm{E}\,\mathrm{E}[f(X,Z)|X]^2 + \frac{\mathrm{E}\,\mathrm{Var}[f(X,Z)|X]}{m} \end{pmatrix}.$$

(Some calculations are carried out in Appendix D.) Define the function

$$\begin{aligned} &s(x_1, \dots, x_p, x_{p+1}, x_{p+2}, x_{p+3}) \\ &\quad = \left(\frac{x_1 - x_{p+1}^2}{x_{p+2} - x_{p+1}^2}, \dots, \frac{x_p - x_{p+1}^2}{x_{p+2} - x_{p+1}^2}, \frac{x_1 - x_{p+1}^2}{x_{p+3} - x_{p+1}^2}, \dots, \frac{x_p - x_{p+1}^2}{x_{p+3} - x_{p+1}^2} \right). \end{aligned}$$

Clearly, we have

$$s\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\xi}_{m,i}\right) = \begin{pmatrix} \widehat{\mathbf{S}}'_{n,m} \\ \widehat{\mathbf{S}}''_{n,m} \end{pmatrix}$$

and

$$s(\mathrm{E}\,\boldsymbol{\xi}_{m,1}) = \begin{pmatrix} \mathbf{S}' \\ \mathbf{S}'' \left[1 - \frac{\mathrm{E}\,\mathrm{Var}[f(X,Z)|X]}{\mathrm{E}\,\mathrm{Var}[f(X,Z)|X] + m\,\mathrm{Var}\,\mathrm{E}[f(X,Z)|X]}\right] \end{pmatrix}.$$

If m is fixed then Lemma C.2 in Appendix C yields

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\xi}_{m,i} - \mathrm{E}\,\boldsymbol{\xi}_{m,1}\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_m),$$

for some nonnegative matrix Σ_m of size $(p+3) \times (p+3)$ and the result follows by the delta-method.

If $m \to \infty$, Lemma C.2 still holds with the variance-covariance matrix replaced by its limit. Taylor's expansion yields

$$\begin{aligned}
&\sqrt{n}\left(s\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\xi}_{m,i}\right)-s(\mathrm{E}\,\boldsymbol{\xi}_{m,1})\right)\\
&\quad=\sqrt{n}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\xi}_{m,i}-\mathrm{E}\,\boldsymbol{\xi}_{m,1}\right)\dot{s}_m\right.\\
&\qquad\left.+\frac{1}{2}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\xi}_{m,i}-\mathrm{E}\,\boldsymbol{\xi}_{m,1}\right)^{\top}\ddot{s}_{n,m}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{\xi}_{m,i}-\mathrm{E}\,\boldsymbol{\xi}_{m,1}\right)\right),
\end{aligned}$$

where $\dot{s}_m$ is the gradient of s at $\mathrm{E}\,\boldsymbol{\xi}_{m,1}$ and $\ddot{s}_{n,m}$ is the Hessian matrix of s at a point between $n^{-1}\sum_i \boldsymbol{\xi}_{m,i}$ and $\mathrm{E}\,\boldsymbol{\xi}_{m,1}$. Since that point goes to a constant and s has continuous second derivatives, it holds that $\ddot{s}_{n,m}$ goes to a constant as well. So does $\dot{s}_m$ and the claim follows by Slutsky's lemma.

Appendix C. A unified treatment of the asymptotics. All estimators in this paper have a common form, given by

$$\frac{1}{n}\sum_{i=1}^{n}\xi_{m,i}, \tag{C.1}$$

with

$$\xi_{m,i}=\prod_{l=1}^{L}\frac{1}{m}\sum_{k=1}^{m}\prod_{j=0}^{p}Y_j^{(i,k)b_{j;l}}, \tag{C.2}$$

where $Y_0^{(i,k)}=Y^{(i,k)}=f(X^{(i)},Z_0^{(i,k)})$, $Y_j^{(i,k)}=f(\widetilde{X}_{-j}^{(i)},Z_j^{(i,k)})$ for $j=1,\ldots,p$, and $b_{j;l}$, $j=0,\ldots,p$, $l=1,\ldots,L$, are nonnegative coefficients. The coefficients are arranged in a matrix $(b_{j;l})$ with L rows and $p+1$ columns, where $b_{j;l}$ is the element in the lth row and $(j+1)$th column. This way, all estimators of the form (C.1) and (C.2), or, equivalently, all summands (C.2), can be represented by a matrix. We sometimes write $\xi_{m,i}\simeq(b_{j;l})$, where $(b_{j;l})$ is the matrix of size $L\times(p+1)$ with coefficients $b_{j;l}$, $j=0,\ldots,p$, $l=1,\ldots,L$.

C.1. Examples. The estimator

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{k=1}^{m}Y_0^{(i,k)}\frac{1}{m}\sum_{k'=1}^{m}Y_j^{(i,k')}$$

is of the form (C.1) and (C.2) with $L=2$ and coefficients

$$\begin{pmatrix}1&0&\cdots&0&0&0&\cdots&0\\0&0&\cdots&0&1&0&\cdots&0\end{pmatrix},$$

where the non-null columns are the first and the $(j+1)$th ones. The estimators

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{k=1}^{m}Y_0^{(i,k)},\quad \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{k=1}^{m}Y_0^{(i,k)2},\\
&\frac{1}{n}\sum_{i=1}^{n}\left(\frac{1}{m}\sum_{k=1}^{m}Y_0^{(i,k)}\right)^2
\end{aligned}$$

are of the form (C.1) and (C.2) with $L = 2$ and coefficients

$$\begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

respectively.

The estimators of Section 4. In view of (4.4)–(4.9), the estimators $\widehat{\zeta}_{3,j}$ and $\widehat{\zeta}_{1,j}$ can be expressed in terms of estimators of the form (C.1) and (C.2): we have

$$\widehat{\zeta}_{3,j} = \frac{1}{n}\sum_{i=1}^{n} \xi_{j;m,i}^{(4.4)} + \xi_{j;m,i}^{(4.5)} - \xi_{j;m,i}^{(4.6)} - \xi_{j;m,i}^{(4.7)}, \quad \text{and,}$$
$$\widehat{\zeta}_{1,j} = \frac{1}{n}\sum_{i=1}^{n} \xi_{j;m,i}^{(4.8)} - \left(\frac{1}{n}\sum_{i=1}^{n} \xi_{j;m,i}^{(4.9)}\right)^2,$$

where

$$\begin{aligned} &\xi_{j;m,i}^{(4.4)}, \quad \xi_{j;m,i}^{(4.5)} \\ &\xi_{j;m,i}^{(4.6)}, \quad \xi_{j;m,i}^{(4.7)}, \\ &\xi_{j;m,i}^{(4.8)}, \quad \xi_{j;m,i}^{(4.9)} \end{aligned}$$

are all of the form (C.2) with $L = 4$ and coefficients

$$\begin{pmatrix} 2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$
$$\begin{pmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

respectively. In the matrices above, the first and $j+1$th columns are nonnull.

The estimators of Section 5. The Sobol estimators in (3.7) and (3.4) are of the form (C.1) and (C.2) with $L = 2$ and coefficients

$$\xi_{1;m,i}^{\mathrm{UL}} \simeq \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \end{pmatrix}, \dots, \xi_{p;m,i}^{\mathrm{UL}} \simeq \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

for the upper left (UL) terms,

$$\xi_{m,i}^{\mathrm{UR}} \simeq \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

for the upper right (UR) term,

$$\xi_{m,i}'^{\mathrm{LL}} \simeq \begin{pmatrix} 2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

for the lower left (LL) term of $\widehat{S}'_{j;n,m}$ and

$$\xi_{m,i}''^{\mathrm{LL}} \simeq \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

for the lower left (LL) term of $\widehat{S}''_{j;n,m}$.

C.2. A central limit theorem. For each n, the random variables $\xi_{m,1}, \dots, \xi_{m,n}$ are independent and identically distributed. Denote by $\mathcal{E}_{m,i}(L)$ the set of all summands (C.2). In other words, $\mathcal{E}_{m,i}(L)$ is the set of all nonnegative matrices of size $L \times (p+1)$. This set has useful properties, gathered in Proposition C.1 for subsequent use.

PROPOSITION C.1. *Let ξ be an element of $\mathcal{E}_{m,i}(L)$ with coefficients $(b_{j;l})$. The following statements are true.*

(i) If ξ' is an element of $\mathcal{E}_{m,i}(L)$ with coefficients $(b'_{j;l})$ then $\xi\xi'$ is an element of $\mathcal{E}_{m,i}(2L)$ with coefficients

$$\begin{pmatrix} b_{0;1} & \cdots & b_{p;1} \\ \vdots & & \vdots \\ b_{0;L} & \cdots & b_{p;L} \\ b'_{0;1} & \cdots & b'_{p;1} \\ \vdots & & \vdots \\ b'_{0;L} & \cdots & b'_{p;L} \end{pmatrix}.$$

(ii) The limit of $\mathrm{E}\,\xi$ exists as $m \to \infty$.

(iii) If there exists some function F such that $|f(x,z)| \le F(x)$ for all x and z in the domain of definition of f then

$$|\xi| \le \left(\bigvee_{j=0}^{p} F_j(\mathbf{X}^{(i)}) \right)^{\sum_{j=0}^{p} \sum_{l=1}^{L} b_{j;l}},$$

where $F_j(\mathbf{X}^{(i)})$ is $F(X^{(i)})$ if $j = 0$ and $F(\widetilde{X}^{(i)}_{-j})$ if $j \ge 1$.

Proof. The proof of (i) is trivial. Let us prove (ii). We have

$$\begin{aligned} \mathrm{E}\,\xi &= \frac{1}{m^L} \sum_{(k_1,\dots,k_L)\in\{1,\dots,m\}^L} \mathrm{E} \prod_{l=1}^{L} \prod_{j=0}^{p} Y_j^{(1,k_l) b_{j;l}} \\ &= \frac{1}{m^L} \sum_{(k_1,\dots,k_L)\in\{1,\dots,m\}^L} \mathrm{E}\,\mathrm{E} \left(\prod_{l=1}^{L} \prod_{j=0}^{p} Y_j^{(1,k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right) \\ &= \frac{1}{m^L} \sum_{(k_1,\dots,k_L)\in\{1,\dots,m\}^L} \mathrm{E} \prod_{j=0}^{p} \mathrm{E} \left(\prod_{l=1}^{L} Y_j^{(1,k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right). \end{aligned} \tag{C.3}$$

Since (i) $\mathbf{X}^{(1)}$ and $\{\mathbf{Z}^{(1,k)},\ k=1,\dots,m\}$ are independent and (ii) the law of

$$(\mathbf{Z}^{(1,k_1)},\dots,\mathbf{Z}^{(1,k_L)})$$

is invariant through any permutation of distinct $k_1,\dots,k_L$, all the inner expectations in (C.3) are equal to some others. For if $k_1,\dots,k_L$ are distinct then

$$\mathrm{E}\left(\prod_{l=1}^{L} Y_j^{(1,k_l)b_{j;l}} \middle| \mathbf{X}^{(1)}\right) = \mathrm{E}\left(\prod_{l=1}^{L} Y_j^{(1,l)b_{j;l}} \middle| \mathbf{X}^{(1)}\right)$$

for all $j=0,\dots,p$. The number of inner expectations equal to the one above is $m(m-1)\cdots(m-L+1)$, a polynomial in m with degree L. If some components of the tuple $(k_1,\dots,k_L)$ are equal, then we can always write

$$\mathrm{E}\left(\prod_{l=1}^{L} Y_j^{(1,k_l)b_{jl}} \middle| \mathbf{X}^{(1)}\right) = \mathrm{E}\left(\prod_{l=1}^{L'} Y_j^{(1,l)\beta_{j;l}} \middle| \mathbf{X}^{(1)}\right)$$

for some $L' \le L$ and coefficients β_{jl} It is easy to see that the number of inner expectations equal to the one above is a polynomial in m with degree at most L. (Looking at examples helps to see this; see e.g. the proof of Lemma D.2 in Appendix D.) Therefore, the sum in (C.3) is also a polynomial in m with degree at most L and the claim follows ($\mathrm{E}\,\xi$ can be zero). To prove (iii), simply remember that, by assumption, $|Y^{(1,k)}| \le F(X^{(1)})$ and $|Y_j^{(1,k)}| \le F(\widetilde{X}_{-j}^{(1)})$ for all k and all j. ∎

Two frameworks are considered:

- $n\to\infty$ and m is fixed;
- $n\to\infty$ and $m\to\infty$.

In the second framework m_n is a sequence indexed by n that goes to infinity as n goes to infinity.

Lemma C.2. *Let $\xi_{m,i}^{(I)}$, $I=1,\dots,N$, be elements of $\mathcal{E}_{m,i}(L)$ with coefficients $(b_{j;l}^{(I)})$. Assume*

$$\mathrm{E}\,F(X^{(1)})^{2\sum_{j=0}^{p}\sum_{l=1}^{L} b_{j;l}^{(I)}} < \infty$$

for all $I=1,\dots,N$. Let $n\to\infty$. If m is fixed then

$$\sqrt{n}\left[\frac{1}{n}\sum_{i=1}^{n}\xi_{m,i}^{(1)} - \mathrm{E}\,\xi_{m,1}^{(1)},\dots,\frac{1}{n}\sum_{i=1}^{n}\xi_{m,i}^{(N)} - \mathrm{E}\,\xi_{m,1}^{(N)}\right]^{\top} \xrightarrow{d} \mathcal{N}(0,\Sigma_m),$$

where Σ_m is the variance-covariance matrix of $\boldsymbol{\xi}_{m,i} = (\xi_{m,i}^{(1)},\dots,\xi_{m,i}^{(N)})^{\top}$. If $m\to\infty$ then $\lim_{m\to\infty}\Sigma_m$ exists elementwise and the above display with Σ_m replaced by $\lim_{m\to\infty}\Sigma_m$ is true.

Proof. Let m be fixed. By Proposition C.1 (i), $\xi_{m,i}^{(I)2}$, $I=1,\dots,N$, belongs to $\mathcal{E}_{m,i}(2L)$ and has coefficients

$$\xi_{m,i}^{(I)2} \simeq \begin{pmatrix} b_{0;1}^{(I)} & \cdots & b_{p;1}^{(I)} \\ \vdots & & \vdots \\ b_{0;L}^{(I)} & \cdots & b_{p;L}^{(I)} \\ b_{0;1}^{(I)} & \cdots & b_{p;1}^{(I)} \\ \vdots & & \vdots \\ b_{0;L}^{(I)} & \cdots & b_{p;L}^{(I)} \end{pmatrix}.$$

Thus, denoting $\sum_{j=0}^{p}\sum_{l=1}^{L} b_{j;l}^{(I)}$ by β, Proposition C.1 (iii) yields

$$\xi_{m,i}^{(I)2} \le \bigvee_{j=0}^{p} F_j(\mathbf{X}^{(i)})^{2\beta} \tag{C.4}$$

and hence

$$\mathrm{E}\,\xi_{m,i}^{(I)2} \le \mathrm{E}\bigvee_{j=0}^{p} F_j(\mathbf{X}^{(1)})^{2\beta} \le (p+1)\,\mathrm{E}\left(F(X^{(1)})\right)^{2\beta} < \infty.$$

Therefore we can apply the central limit theorem to finish the proof for m fixed.

Let $m\to\infty$. According to Lindeberg-Feller's central limit theorem (see e.g. [33]), it suffices to show

(i) for all $\epsilon>0$,

$$\sum_{i=1}^{n}\mathrm{E}\left\|\frac{1}{\sqrt{n}}\boldsymbol{\xi}_{m,i}\right\|^2 \mathbf{1}\left\{\left\|\frac{1}{\sqrt{n}}\boldsymbol{\xi}_{m,i}\right\| > \epsilon\right\} \to 0,$$

and

(ii) the limit $\sum_{i=1}^{n}\mathrm{Cov}(\boldsymbol{\xi}_{m,i}/\sqrt{n})$ exists and is finite.

Let us show (i). Denoting $\mathbf{X}=(X^{(1)},\widetilde{X}^{(1)})$, we have

$$\begin{aligned}
\sum_{i=1}^{n}\mathrm{E}\left\|\frac{\boldsymbol{\xi}_{m,i}}{\sqrt{n}}\right\|^2 \mathbf{1}\left\{\|\boldsymbol{\xi}_{m,i}\| > \sqrt{n}\epsilon\right\} &= \mathrm{E}\,\|\boldsymbol{\xi}_{m,1}\|^2\mathbf{1}\{\|\boldsymbol{\xi}_{m,1}\|>\sqrt{n}\epsilon\}\\
&= \mathrm{E}\sum_{I=1}^{N}\xi_{m,1}^{(I)2}\mathbf{1}\{\|\boldsymbol{\xi}_{m,1}\|>\sqrt{n}\epsilon\}\\
&= \sum_{I=1}^{N}\mathrm{E}\left[\mathrm{E}\left(\xi_{m,1}^{(I)2}\mathbf{1}\{\|\boldsymbol{\xi}_{m,1}\|>\sqrt{n}\epsilon\}|\mathbf{X}\right)\right].
\end{aligned}$$

By (C.4), we have

$$\begin{aligned}
\mathrm{E}\left(\xi_{m,1}^{(I)2}\mathbf{1}\{\|\boldsymbol{\xi}_{m,1}\|>\sqrt{n}\epsilon\}|\mathbf{X}\right) &\le \bigvee_{j=0}^{p} F_j(\mathbf{X}^{(1)})^{2\beta} P\left(\|\boldsymbol{\xi}_{m,1}\|>\sqrt{n}\epsilon|\mathbf{X}\right)\\
&\le \bigvee_{j=0}^{p} F_j(\mathbf{X}^{(1)})^{2\beta}\frac{\sum_{I=1}^{N}\mathrm{E}\left(\xi_{m,1}^{(I)2}|\mathbf{X}\right)}{n\varepsilon^2}\\
&\le \frac{N\bigvee_{j=0}^{p}F_j(\mathbf{X}^{(1)})^{4\beta}}{n\varepsilon^2},
\end{aligned}$$

where the last inequality holds by using (C.4) once more. The upper bound goes to zero and is dominated by an integrable function. Thus, we can apply the dominated convergence theorem to complete the proof.

Let us show that (ii) holds. We have $\sum_{i=1}^{n}\mathrm{Cov}(\boldsymbol{\xi}_{m,i}/\sqrt{n}) = \mathrm{Cov}(\boldsymbol{\xi}_{m,1})$. The element (I,J) in this matrix is given by $\mathrm{E}\,\xi_{m,1}^{(I)}\xi_{m,1}^{(J)} - \mathrm{E}\,\xi_{m,1}^{(I)}\,\mathrm{E}\,\xi_{m,1}^{(J)}$. Remember that $\mathrm{E}\,\xi_{m,1}^{(I)2}<\infty$, $I=1,\dots,N$, and hence $\mathrm{E}\,\xi_{m,1}^{(I)}\xi_{m,1}^{(J)} \le \mathrm{E}\,\xi_{m,1}^{(I)2}/2+\xi_{m,1}^{(J)2}/2<\infty$. Therefore the limit of $\mathrm{Cov}\,\boldsymbol{\xi}_{m,1}$ exists and is finite. The proof is complete. ∎

Appendix D. Explicit moment calculations. Explicit moment calculations are given for the summands in the proof of Theorem 5.1. In this section, $\mathrm{E}\, f(X,Z)$ and $\mathrm{E}\,\mathrm{E}[f(X,Z)|X]^2$ are denoted by μ and D, respectively. Recall that the upper-left term in (3.6) and (3.5) is denoted by D_j. The moments are given in Lemma D.1 and Lemma D.2. The variances and covariances are given in Lemma D.3. Let $\mathbf{X} = (X^{(1)}, \widetilde{X}^{(1)})$. Whenever there is a superscript $\mathbf{X}$ added to the expectation symbol E or the variance symbol Var, this means that these operators are to be understood conditionally on $\mathbf{X}$. An integral with respect to $\mathbf{P}^*(\mathrm{d}\mathbf{x})$ means that we integrate with respect to the law of $\mathbf{X}$.

LEMMA D.1 (Moments of order 1). *The moments of order 1 are given by*

$$\begin{aligned}
\mathrm{E}\,\xi^{UL}_{j;m1} &= D_j,\\
\mathrm{E}\,\xi^{UR}_{m1} &= \mu,\\
\mathrm{E}\,\xi''^{LL}_{m1} &= \frac{1}{m}\,\mathrm{E}\,\mathrm{Var}^X f(X^{(1)}, Z^{(1,1)}) + D.
\end{aligned}$$

Proof. One has

$$\begin{aligned}
\mathrm{E}\,\xi^{\mathrm{UL}}_{j;m1} =& \frac{1}{m^2}\sum_{k,k'} \mathrm{E}\, f(X^{(1)}, Z^{(1,k)}) f(\widetilde{X}^{(1)}_{-j}, Z^{(1,k')}_j)\\
=& \frac{1}{m^2}\sum_{k,k'} \int \mathrm{E}\, f(x, Z^{(1,k)}) f(\tilde{x}_{-j}, Z^{(1,k')}_j)\,\mathbf{P}^*(\mathrm{d}\mathbf{x})\\
=& \,\mathrm{E}\, f(X^{(1)}, Z^{(1,1)}) f(\widetilde{X}^{(1)}_{-j}, Z^{(1,1)}_j)\\
=& D_j,
\end{aligned}$$

where the integral is taken with respect to the law of $\mathbf{x} = (x, \tilde{x})$, and,

$$\begin{aligned}
\mathrm{E}\,\xi''^{\mathrm{LL}}_{m1} =& \frac{1}{m^2}\sum_{k,k'} \mathrm{E}\, f(X^{(1)}, Z^{(1,k)}) f(X^{(1)}, Z^{(1,k')})\\
=& \frac{1}{m}\,\mathrm{E}\,\mathrm{Var}^X f(X,Z) + \mathrm{E}(\mathrm{E}^X f(X,Z))^2\\
=& \frac{1}{m}\,\mathrm{E}\,\mathrm{Var}^X f(X,Z) + D.
\end{aligned}$$

The proof for ξ^{UR}_{m1} is similar. ∎

LEMMA D.2 (Moments of order 2). *The moments of order 2 are given by*

$$
\begin{aligned}
\mathrm{E}\,\xi_{j;m1}^{(UL)2} &= \operatorname{Var}\mathrm{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) + D_j^2 \\
&\quad + \frac{1}{m}[\mathrm{E}\operatorname{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
&\quad - \operatorname{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \operatorname{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})] \\
&\quad + \frac{1}{m^2}\mathrm{E}\operatorname{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \operatorname{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}), \\
\mathrm{E}\,\xi_{m1}^{(UR)2} &= \frac{1}{m}\mathrm{E}\operatorname{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) + \mathrm{E}(\mathrm{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}))^2, \\
\mathrm{E}\,\xi_{m1}^{\prime\prime(LL)2} &= \frac{m(m-1)(m-2)(m-3)}{m^4} \\
&\quad \mathrm{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) f(X^{(1)}, Z^{(1,4)}) \\
&\quad + \frac{\binom{4}{2} m(m-1)(m-2)}{m^4}\mathrm{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) \\
&\quad + \frac{\binom{4}{3} m(m-1)}{m^4}\mathrm{E} f(X^{(1)}, Z^{(1,1)})^3 f(X^{(1)}, Z^{(1,2)}) \\
&\quad + \frac{m}{m^4}\mathrm{E} f(X^{(1)}, Z^{(1,1)})^4 \\
&\quad + \frac{\binom{4}{2} m(m-1)/2}{m^4}\mathrm{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)})^2
\end{aligned}
$$

Proof. Let us first deal with $\xi_{j;m1}^{\mathrm{UL}}$. We have

$$
\mathrm{E}\,\xi_{j;m1}^{(\mathrm{UL})2} = \frac{1}{m^4}\sum_{k_1,k_2,k_3,k_4}\mathrm{E} f(X^{(1)}, Z^{(1,k_1)}) f(X^{(1)}, Z^{(1,k_2)}) \\
f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,k_3)}) f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,k_4)})
$$

where, in the sum, the indices run over $1,\ldots,m$. We split the sum into four parts. The first contains the $m^2(m-1)^2$ terms that satisfy $k_1 \neq k_2$ and $k_3 \neq k_4$. In this part, all the terms are equal to

$$
\text{(term 1)} \qquad \mathrm{E}\left(\mathrm{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})\right)^2.
$$

The second part contains the $m^2(m-1)$ terms that satisfy $k_1 \neq k_2$ and $k_3 = k_4$ and that are equal to

$$
\text{(term 2)} \qquad \mathrm{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)})^2.
$$

The third part contains the $m^2(m-1)$ terms that satisfy $k_1 = k_2$ and $k_3 \neq k_4$ and that are equal to

$$
\text{(term 3)} \qquad \mathrm{E} f(X^{(1)}, Z^{(1,1)})^2 f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,2)}).
$$

Finally, the fourth part contains the m^2 terms that satisfy $k_1 = k_2$ and $k_3 = k_4$ and are equal to

$$
\text{(term 4)} \qquad \mathrm{E} f(X^{(1)}, Z^{(1,1)})^2 f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)})^2.
$$

(One can see that the number of terms is m^4.) Thus,

$$\begin{aligned}
\mathrm{E}\,\xi_{m1}^{(\mathrm{UL})2} =&(\text{term } 1)\\
&+\frac{1}{m}[(\text{term } 2)+(\text{term } 3)-2(\text{term } 1)]\\
&+\frac{1}{m^2}[(\text{term } 1)-(\text{term } 2)-(\text{term } 3)+(\text{term } 4)].
\end{aligned}$$

Furthermore, [(term 1) - (term 2) - (term 3) + (term 4)] is equal to

$$\begin{aligned}
&\int \Big(\mathrm{E}^{\mathbf{X}} f(x,Z) f(\tilde{x}_{-j}, Z_j)\Big)^2\\
&\qquad - \mathrm{E}^{\mathbf{X}} f(x, Z^{(1,1)}) f(x, Z^{(1,2)}) f(\tilde{x}_{-j}, Z_j^{(1,1)})^2\\
&\qquad - \mathrm{E}^{\mathbf{X}} f(x, Z^{(1,1)})^2 f(\tilde{x}_{-j}, Z_j^{(1,1)}) f(\tilde{x}_{-j}, Z_j^{(1,2)})\\
&\qquad + \mathrm{E}^{\mathbf{X}} f(x, Z^{(1,1)})^2 f(\tilde{x}_{-j}, Z_j^{(1,1)})^2 \,\mathrm{d}\mathbf{P}^*(\mathbf{x})\\
=&\int \Big(\mathrm{E}^{\mathbf{X}} f(x,Z)\Big)^2 \Big(\mathrm{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)\Big)^2\\
&\qquad - \Big(\mathrm{E}^{\mathbf{X}} f(x,Z)\Big)^2 \mathrm{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)^2\\
&\qquad - \mathrm{E}^{\mathbf{X}} f(x,Z)^2 \Big(\mathrm{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)\Big)^2\\
&\qquad + \mathrm{E}^{\mathbf{X}} f(x,Z)^2\, \mathrm{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)^2 \,\mathrm{d}\mathbf{P}^*(\mathbf{x})\\
=&\int \mathrm{Var}^{\mathbf{X}} f(X,Z)\, \mathrm{Var}^{\mathbf{X}} f(\tilde{X}_{-j}, Z_j)\,\mathrm{d}\mathbf{P}^*(\mathbf{x}).
\end{aligned}$$

Likewise, we find that [(term 2)+(term 3)-2(term 1)] is equal to

$$\mathrm{E}\,\mathrm{Var}^{\mathbf{X}} f(X,Z) f(\tilde{X}_{-j}, Z_j) - \mathrm{Var}^{\mathbf{X}} f(X,Z)\,\mathrm{Var}^{\mathbf{X}} f(\tilde{X}_{-j}, Z_j),$$

and term 1 is $\mathrm{Var}\,\mathrm{E}^{\mathbf{X}} f(X,Z) f(\tilde{X}_{-j}, \tilde{Z}) + D_j^2$.

We now deal with $\xi_{m1}^{\prime\prime\mathrm{LL}}$. We have

$$\begin{aligned}
\mathrm{E}\,\xi_{m1}^{\prime\prime(\mathrm{LL})2} = \frac{1}{m^4} \sum_{k_1,k_2,k_3,k_4} \mathrm{E}\, f(X^{(1)}, Z^{(1,k_1)}) f(X^{(1)}, Z^{(1,k_2)})&\\
f(X^{(1)}, Z^{(1,k_3)}) f(X^{(1)}, Z^{(1,k_4)}).&
\end{aligned}$$

The sum is split into five parts. The first part consists of the $m(m-1)(m-2)(m-3)$ terms with different indices; those terms are equal to

$$\mathrm{E}\, f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) f(X^{(1)}, Z^{(1,4)}).$$

The second part consists of the $\binom{4}{2}m(m-1)(m-2)$ terms with exactly two equal indices; those terms are equal to

$$\mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}).$$

The third part consists of the $\binom{4}{3}m(m-1)$ terms with exactly three equal indices; those terms are equal to

$$\mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^3 f(X^{(1)}, Z^{(1,2)}).$$

The fourth part consists of the m terms with exactly four equal indices; those terms are equal to

$$\mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^4.$$

The fifth and last part consists of the $\binom{4}{2} m(m-1)/2$ terms with exactly two pairs of equal indices; those terms are equal to

$$\mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)})^2.$$

(One can check that the total number of terms is m^4.) ◻

LEMMA D.3 (Variances and covariances).

$$\begin{aligned}
&\text{(i)} && \operatorname{Var} \xi_{m1}^{UL} = \operatorname{Var} \mathrm{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
& && \quad + \frac{1}{m} [\mathrm{E} \operatorname{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
& && \quad - \operatorname{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \operatorname{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})] \\
& && \quad + \frac{1}{m^2} \mathrm{E} \operatorname{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \operatorname{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}), \\
&\text{(ii)} && \operatorname{Cov}(\xi_{m1}^{UL}, \xi_{m1}^{UR}) = \frac{m-1}{m} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
& && \quad + \frac{1}{m} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^2 f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) - D_j \mu \\
&\text{(iii)} && \operatorname{Cov}(\xi_{m1}^{UL}, f(X,Z)^2) = \frac{1}{m} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^3 f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
&\text{(iii)} && \quad + \frac{m-1}{m} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) - D_j \kappa \\
&\text{(iv)} && \operatorname{Var} \xi_{m1}^{UR} = \frac{1}{m} \operatorname{Var} f(X, Z) \\
&\text{(v)} && \operatorname{Cov}(\xi_{m1}^{UR}, f(X,Z)^2) = \frac{1}{m} f(X,Z)^3 \\
& && \quad + \frac{m-1}{m} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) - \mu\kappa \\
&\text{(vi)} && \operatorname{Cov}(\xi_{m_n 1}^{UL}, \xi_{m_n 1}^{\prime\prime LL}) = \frac{m}{m^3} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^3 f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
& && \quad + \frac{3m(m-1)}{m^3} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
& && \quad + \frac{m(m-1)(m-2)}{m^3} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) \\
& && \quad f(X^{(1)}, Z^{(1,3)}) f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
& && \quad - \mathrm{E}\, f(X^{(1)}, Z^{(1,1)}) f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
& && \quad \left\{ \frac{1}{m} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)})^2 + \frac{m-1}{m} \mathrm{E}\, f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) \right\}
\end{aligned}$$

Proof. The proof follows from direct calculations. ◻

Appendix E. Calculations for the linear model.

LEMMA E.1. *Suppose that* $f(X,Z) = \beta_0 + \beta_{p+1} Z + \sum_{j=1}^{p} \beta_j X_j$ *where* $X = (X_1, \ldots, X_p), Z_k, \tilde{Z}_{ik}$ *are independent,* $\mathrm{E}\, X_j = \mathrm{E}\, Z = 0$, $\mathrm{E}\, X_j^2 = \mathrm{E}\, Z^2 = 1$, $\mathrm{E}\, X_j^3 = 0$,

$\mathrm{E}\, X_j^4 = 3$. *Then the squared optimal number of repetitions is given by*

$$(m_i^*)^2 = \frac{\beta_{p+1}^4}{(\beta_0+\beta_i)^2 - 2\beta_0^4 + (\sum_{j=0}^p \beta_j^2)^2}$$

and the discriminator (the upper-left term in (3.6) *and* (3.5)*) is*

$$\beta_0^2 + \beta_i^2.$$

Proof. We have

$$m_i^* = \frac{A_i + B_i + C_i + D_i}{E_i},$$

with

$$\begin{aligned}
A_i &= \mathrm{E}\, f(X,Z_1)^2 f(\tilde{X}_{-i},\tilde{Z}_{i1})^2 \\
B_i &= \mathrm{E}\, f(X,Z_1) f(\tilde{X}_{-i},\tilde{Z}_{i1}) f(X,Z_2) f(\tilde{X}_{-i},\tilde{Z}_{i2}) \\
C_i &= -\,\mathrm{E}\, f(X,Z_1)^2 f(\tilde{X}_{-i},\tilde{Z}_{i1}) f(\tilde{X}_{-i},\tilde{Z}_{i2}) \\
D_i &= -\,\mathrm{E}\, f(\tilde{X}_{-i},\tilde{Z}_{i1})^2 f(X,Z_1) f(X,Z_2) \\
E_i &= B - [\mathrm{E}\, f(X,Z_1) f(\tilde{X}_{-i},\tilde{Z}_{i1})]^2
\end{aligned}$$

where $X = (X_1,\ldots,X_p)$, $Z_k, \tilde{Z}_{ik}$ are independent, $\mathrm{E}\, X_j = \mathrm{E}\, Z = 0$, $\mathrm{E}\, X_j^2 = \mathrm{E}\, Z^2 = 1$, $\mathrm{E}\, X_j^3 = 0$, $\mathrm{E}\, X_j^4 = 3$. We deal with the case

$$f(X,Z) = \beta_0 + \beta_{p+1} Z + \sum_{j=1}^p \beta_j X_j.$$

We calculate the terms one by one as follows. We have

$$\begin{aligned}
A_j =\, &\mathrm{E}\left(\beta_0 + \sum_{j=1}^p \beta_j X_j\right)^2 \left(\beta_0 + \beta_i X_i + \sum_{j:1\le j\ne i} \beta_j \tilde{X}_j\right)^2 \\
&+ \left(\beta_0 + \sum_{j=1}^p \beta_j X_j\right)^2 \beta_{p+1}^2 \tilde{Z}_{i1}^2 + \beta_{p+1}^4 Z_1^2 \tilde{Z}_{i1}^2 \\
&+ \beta_{p+1}^2 Z_1^2 \left(\beta_0 + \beta_i X_i + \sum_{j:1\le j\ne i} \beta_j \tilde{X}_j\right)^2 \\
=\, & A_{j1} + A_{j2} + A_{j3},
\end{aligned}$$

where $\mathrm{E}\,(\mathrm{A2}) = \beta_{p+1}^4 + \beta_{p+1}^2 \sum_{j=0}^p \beta_j^2$, $\mathrm{E}\,(\mathrm{A3}) = \beta_{p+1}^2 \sum_{j=0}^p \beta_j^2$. Elementary but somewhat tedious calculations yield

$$\mathrm{E}\,(\mathrm{A1}) = \beta_0^4 + 3\beta_i^4 + 6\beta_0^2\beta_i^2 + 2(\beta_0^2+\beta_i^2)\sum_{j:1\le j\ne i}\beta_j^2 + \left(\sum_{j:1\le j\ne i}\beta_j^2\right)^2.$$

Similar calculations show that $B_j = A_{j1}$, $C_j = -A_{j1} - A_{j3}$, $D_j = -A_{j1} - A_{j3}$, $E_j = A_{j1} - (\beta_0^2 + \beta_i^2)^2$. Thus,

$$(m_i^*)^2 = \frac{\beta_{p+1}^4}{(\beta_0 + \beta_i)^2 - 2\beta_0^4 + (\sum_{j=0}^p \beta_j^2)^2}. \qquad \square$$

Acknowledgments. The author thanks two anonymous referees who made useful suggestions. The author thanks Hervé Monod and Elisabeta Vergu for stimulating discussions from which this work arose and Bertrand Iooss for the references [10, 13].

REFERENCES

[1] S. Azzi, B. Sudret, and J. Wiart, *Sensitivity analysis for stochastic simulators using differential entropy*, International Journal for Uncertainty Quantification, 10 (2020).

[2] A. Courcoul, H. Monod, M. Nielen, D. Klinkenberg, L. Hogerwerf, F. Beaudeau, and E. Vergu, *Modelling the effect of heterogeneity of shedding on the within herd coxiella burnetii spread and identification of key parameters by sensitivity analysis*, Journal of Theoretical Biology, 284 (2011), pp. 130–141.

[3] D. J. Daley and J. Gani, *Epidemic Modelling*, Cambridge, 1999.

[4] P. Etoré, C. Prieur, D. K. Pham, and L. Li, *Global sensitivity analysis for models described by stochastic differential equations*, Methodology and Computing in Applied Probability, (2019), pp. 1–29.

[5] J.-C. Fort, T. Klein, and A. Lagnoux, *Global sensitivity analysis and Wasserstein spaces*, arXiv preprint arXiv:2007.12378, (2020).

[6] F. Gamboa, A. Janon, T. Klein, and A. Lagnoux, *Sensitivity analysis for multidimensional and functional outputs*, Electron. J. Stat., 8 (2014), pp. 575–603, https://doi.org/10.1214/14-EJS895.

[7] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur, *Statistical inference for Sobol pick-freeze Monte Carlo method*, Statistics, 50 (2016), pp. 881–902.

[8] F. Gamboa, T. Klein, and A. Lagnoux, *Sensitivity analysis based on cramér–von mises distance*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 522–548.

[9] F. Gamboa, T. Klein, A. Lagnoux, and L. Moreno, *Sensitivity analysis in general metric spaces*, arXiv preprint arXiv:2002.04465, (2020).

[10] J. L. Hart, A. Alexanderian, and P. A. Gremaud, *Efficient computation of Sobol'indices for stochastic models*, SIAM Journal on Scientific Computing, 39 (2017), pp. A1514–A1530.

[11] W. Hoeffding, *A class of statistics with asymptotically normal distribution*, The Annals of Mathematical Statistics, 19 (1948), pp. 293–325.

[12] T. Homma and A. Saltelli, *Importance measures in global sensitivity analysis of nonlinear models*, Reliability Engineering & System Safety, 52 (1996), pp. 1–17, https://doi.org/https://doi.org/10.1016/0951-8320(96)00002-6, http://www.sciencedirect.com/science/article/pii/0951832096000026.

[13] B. Iooss, L. L. Gratiet, A. Lagnoux, and T. Klein, *Sensitivity analysis for stochastic computer codes: Theory and estimation methods*, tech. report, EDF R&D, 2014.

[14] B. Iooss, T. Klein, and A. Lagnoux, *Sobol' sensitivity analysis for stochastic numerical codes*, in 8th International ConferenceSensitivity Analysis of Model Output, 2016, pp. 48–49.

[15] B. Iooss and M. Ribatet, *Global sensitivity analysis of computer models with functional inputs*, Reliability Engineering & System Safety, 94 (2009), pp. 1194 – 1204, https://doi.org/https://doi.org/10.1016/j.ress.2008.09.010, http://www.sciencedirect.com/science/article/pii/S0951832008002299.

[16] B. Iooss, S. D. Veiga, A. Janon, G. Pujol, with contribu-tions from Baptiste Broto, K. Boumhaout, T. Delage, R. E. Amri, J. Fruth, L. rent Gilquin, J. Guillaume, L. L. Gratiet, P. Lemaitre, A. Marrel, A. Mey-naoui, B. L. Nelson, F. ippo Monari, R. Oomen, O. Rakovec, B. Ramos, O. Roustant, E. hye Song, J. Staum, R. Sueur, T. Touati, and F. Weber, *sensitivity*, 2020, https://cran.r-project.org/package=sensitivity. R package version 1.23.0.

[17] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur, *Asymptotic normality and efficiency of two Sobol index estimators*, ESAIM: Probability and Statistics, 18 (2014), pp. 342–364, https://doi.org/10.1051/ps/2013040.

[18] M. Lamboni, H. Monod, and D. Makowski, *Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models*, Reliability Engineering and System Safety, 96 (2010), pp. 450–459.

[19] A. Marrel, B. Iooss, S. Da Veiga, and M. Ribatet, *Global sensitivity analysis of stochastic computer models with joint metamodels*, Statistics and Computing, 22 (2012), pp. 833–847, https://doi.org/10.1007/s11222-011-9274-8, https://doi.org/10.1007/s11222-011-9274-8.

[20] A. Marrel, B. Iooss, B. Laurent, and O. Roustant, *Calculations of sobol indices for the gaussian process metamodel*, Reliability Engineering & System Safety, 94 (2009), pp. 742 – 751, https://doi.org/https://doi.org/10.1016/j.ress.2008.07.008, http://www.sciencedirect.com/science/article/pii/S0951832008001981.

[21] H. Monod, C. Naud, and D. Makowski, *Working with dynamic crop models: Evaluation, analysis, parameterization, and applications*, in Uncertainty and sensitivity analysis for crop models, Elsevier, 2006, pp. 55–100.

[22] M. Navarro Jimenez, O. P. Le Maître, and O. M. Knio, *Global sensitivity analysis in stochastic simulators of uncertain reaction networks*, The Journal of Chemical Physics, 145 (2016), p. 244106, https://doi.org/10.1063/1.4971797, https://doi.org/10.1063/1.4971797, https://arxiv.org/abs/https://doi.org/10.1063/1.4971797.

[23] C. Prieur and S. Tarantola, *Variance-based sensitivity analysis: theory and estimation algorithms*, in Handbook of Uncertainty Quantification, Springer, 2015, pp. 1–23.

[24] L. Rimbaud, C. Bruchou, S. Dallot, D. R. J. Pleydell, E. Jacquot, S. Soubeyrand, and G. Thébaud, *Using sensitivity analysis to identify key factors for the propagation of a plant epidemic*, Open Science, 5 (2018), https://doi.org/10.1098/rsos.171435, http://rsos.royalsocietypublishing.org/content/5/1/171435, https://arxiv.org/abs/http://rsos.royalsocietypublishing.org/content/5/1/171435.full.pdf.

[25] A. Saltelli and I. M. Sobol', *About the use of rank transformation in sensitivity analysis of model output*, Reliability Engineering & System Safety, 50 (1995), pp. 225 – 239, https://doi.org/https://doi.org/10.1016/0951-8320(95)00099-2, http://www.sciencedirect.com/science/article/pii/0951832095000992.

[26] A. Saltelli, S. Tarantola, and F. Campolongo, *Sensitivity anaysis as an ingredient of modeling*, Statistical Science, 15 (2000), pp. 377–395.

[27] A. Saltelli, S. Tarantola, F. Campolongo, and M. Ratto, *Sensitivity analysis in practice*, Wiley, 2004.

[28] J. F. Savall, C. Bidot, M. Leblanc-Maridor, C. Belloc, and S. Touzeau, *Modelling salmonella transmission among pigs from farm to slaughterhouse: interplay between management variability and epidemiological uncertainty*, International Journal of Food Microbiology, 229 (2016), pp. 33–43.

[29] I. M. Sobol, *Sensitivity estimates for nonlinear mathematical models*, Mathematical modelling and computational experiments, 1 (1993), pp. 407–414.

[30] I. M. Sobol, *Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates*, Mathematics and computers in simulation, 55 (2001), pp. 271–280.

[31] M. Spence, *Statistical issues in ecological simulation models*, PhD thesis, University of Sheffield, 2015. http://etheses.whiterose.ac.uk/10517.

[32] L. Szewczyk, J.-L. Grimaud, and I. Cojan, *Experimental evidence for bifurcation angles control on abandoned channel fill geometry*, Earth Surface Dynamics, 8 (2020), pp. 275–288, https://doi.org/10.5194/esurf-8-275-2020, https://hal.archives-ouvertes.fr/hal-02562670.

[33] A. W. van der Vaart, *Asymptotic Statistics*, Cambridge University Press, 1998.

[34] X. Zhu and B. Sudret, *Global sensitivity analysis for stochastic simulators based on generalized lambda surrogate models*, arXiv preprint arXiv:2005.01309, (2020).