



**HAL**  
open science

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

Gildas Mazo

► **To cite this version:**

Gildas Mazo. A tradeoff between explorations and repetitions for estimators of two global sensitivity indices in stochastic models induced by probability measures. 2021. hal-02113448v6

**HAL Id: hal-02113448**

**<https://hal.science/hal-02113448v6>**

Preprint submitted on 2 Jun 2021 (v6), last revised 7 Jun 2021 (v7)

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

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

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

4 GILDAS MAZO<sup>†</sup>

5 **Abstract.** Sobol sensitivity indices assess how the output of a given mathematical model is  
 6 sensitive to its inputs. If the model is stochastic then it cannot be represented as a function of the  
 7 inputs, thus raising questions as how to do a sensitivity analysis in those models. Practitioners have  
 8 been using an approach that exploits the availability of methods for deterministic models. For each  
 9 input, the stochastic model is repeated and the outputs are averaged. These averages are seen as  
 10 if they came from a deterministic model and hence Sobol's method can be used. We show that  
 11 the estimator so obtained is asymptotically biased if the number of repetitions goes to infinity too  
 12 slowly. With limited computational resources, the number of repetitions of the stochastic model  
 13 and the number of explorations of the input space cannot be large together and hence some balance  
 14 must be found. We find the pair of numbers that minimizes a bound on some rank-based error  
 15 criterion, penalizing bad rankings of the inputs' sensitivities. Also, under minimal distributional  
 16 assumptions, we derive a functional relationship between the output, the input and some random  
 17 noise; the Sobol-Hoeffding decomposition can be applied to it to define a new sensitivity index, which  
 18 asymptotically is estimated without bias even though the number of repetitions remains fixed. The  
 19 theory is illustrated on numerical experiments.

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

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

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

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

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

39 where  $f_0 = E f(X)$ ,  $f_j(X_j) = E(f(X) - f_0 | X_j)$ ,  $j = 1, \dots, p$ , and  $f_{1,2}, \dots, f_{1,\dots,p}$   
 40 are some functions defined iteratively; see [29] and [33, p. 157] for more details. In  
 41 the field of uncertainty quantification the above decomposition is known as the Sobol-  
 42 Hoeffding decomposition in reference to [11, 29]. The expectations and the covariances  
 43 of the individual components in the right-hand side of (1.1) are zero and hence we

---

\*Submitted to the editors 2019/03/07.

<sup>†</sup>MaIMAGE, INRA, Université Paris-Saclay, 78350, Jouy-en-Josas, France ([gildas.mazo@inra.fr](mailto:gildas.mazo@inra.fr), <http://genome.jouy.inra.fr/~gmazo/>).

44 have the variance decomposition

$$45 \quad \text{Var } f(X) = \text{Var } f_1(X_1) + \cdots + \text{Var } f_p(X_p) + \cdots ,$$

47 which leads to the so-called Sobol indices

$$48 \quad S_j = \frac{\text{Var } f_j(X_j)}{\text{Var } f(X)} = \frac{\text{Var } \mathbb{E}(f(X)|X_j)}{\text{Var } f(X)} = \frac{\mathbb{E} f(X)f(\tilde{X}_{-j}) - (\mathbb{E} f(X))^2}{\mathbb{E} f(X)^2 - (\mathbb{E} f(X))^2},$$

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

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

$$60 \quad S_{T_j} = 1 - \frac{\text{Var } \mathbb{E}(f(X)|X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p)}{\text{Var } f(X)}.$$

62 The total sensitivity index quantifies the sensitivity of the output of  $f$  to its  $j$ th  
 63 argument through the interactions it may have with the other inputs.

64 There are numerous methods to estimate the sensitivity indices. For simplic-  
 65 ity, we describe below Sobol’s original method to estimate  $S_j$  through Monte Carlo  
 66 sampling [29]. For a review of the many other methods, see [23] or the package  
 67 **sensitivity** [16] of the R software for an up-to-date list of many methods, with refer-  
 68 ences. Thus, draw two independent sets of inputs  $\{X^{(i)}, i = 1, \dots, n\}$ ,  $\{\tilde{X}^{(i)} :=$   
 69  $(\tilde{X}_1, \dots, \tilde{X}_p), i = 1, \dots, n\}$  and make  $p$  more sets by combining the first two:  $\{\tilde{X}_{-j}^{(i)},$   
 70  $i = 1, \dots, n\}$ ,  $j = 1, \dots, p$ , where

$$71 \quad (1.2) \quad \tilde{X}_{-j}^{(i)} := (\tilde{X}_1^{(i)}, \dots, \tilde{X}_{j-1}^{(i)}, X_j^{(i)}, \tilde{X}_{j+1}^{(i)}, \dots, \tilde{X}_p^{(i)}).$$

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

$$76 \quad (1.3) \quad \hat{S}_j = \frac{\frac{1}{n} \sum_{i=1}^n Y^{(i)} Y_j^{(i)} - \left(\frac{1}{n} \sum_{i=1}^n Y^{(i)}\right)^2}{\frac{1}{n} \sum_{i=1}^n Y^{(i)2} - \left(\frac{1}{n} \sum_{i=1}^n Y^{(i)}\right)^2}.$$

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

84 The big difference between a deterministic model and a stochastic model is that  
 85 the stochastic model is not a function anymore. To a particular value of the input there  
 86 does not correspond any particular value for the output. Instead, there corresponds a

87 range of possible values, assumed to come from a probability distribution depending  
 88 on the input. Examples can be found in epidemiology [2, 3, 24, 28] or ecology [31], to  
 89 name a few.

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

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

(1.4)

$$126 \quad \hat{S}_j = \frac{\frac{1}{n} \sum_{i=1}^n m^{-1} \sum_{k=1}^m Y^{(i,k)} m^{-1} \sum_{k=1}^m Y_j^{(i,k)} - \left( \frac{1}{n} \sum_{i=1}^n m^{-1} \sum_{k=1}^m Y^{(i,k)} \right)^2}{127 \quad \frac{1}{n} \sum_{i=1}^n \left( m^{-1} \sum_{k=1}^m Y^{(i,k)} \right)^2 - \left( \frac{1}{n} \sum_{i=1}^n m^{-1} \sum_{k=1}^m Y^{(i,k)} \right)^2}.$$

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

130 To build the estimator (1.4), the stochastic model must be run  $mn(p+1)$  times.  
 131 The number  $m$  is called the number of repetitions and the number  $n$  is called the  
 132 number of explorations. If the stochastic model is computationally intensive—that

133 is, each model run is time-consuming—, then the estimator is built with limited  
 134 resources. In this context, an increase of  $m$  must go along with a decrease of  $n$ , and  
 135 conversely. What is then a good balance between  $m$  and  $n$ ? How to choose  $m$  and  
 136  $n$  such that the estimator (1.4) will be the most efficient? This question was asked  
 137 by [31].

138 We address this problem by finding a pair  $(m, n)$  that minimizes a bound on  
 139 the so-called missranking error. The missranking error penalizes bad rankings of the  
 140 Sobol indices associated with the inputs. The estimated minimizer is used to build  
 141 well-balanced estimators, leading to a two-step procedure to estimate the sensitivity  
 142 indices. This two-step procedure is shown to have good asymptotic properties. We  
 143 show that the estimator (1.4) is asymptotically normal but may be biased if  $m$  goes  
 144 to infinity too slowly. If  $m/\sqrt{n} \rightarrow \infty$  then it is asymptotically unbiased and converges  
 145 to the so-called *sensitivity index of the second kind*. Under the minimal assumption  
 146 that a stochastic model is a set of probability measures that capture how the outputs  
 147 are drawn, we show that the output, the inputs and some random “noise” are linked  
 148 through a function, which the Sobol-Hoeffding decomposition can be applied to. This  
 149 yields a new sensitivity index, called *the sensitivity index of the first kind*, with the  
 150 advantage that asymptotically unbiased estimators can be built even though  $m$  re-  
 151 mains fixed. The indices of the first and of the second kinds are complementary as  
 152 they offer distinct pieces of information. Interestingly, these indices can be estimated  
 153 jointly with no additional cost, the joint estimator is asymptotically normal and the  
 154 two kinds of sensitivity indices lead to the same solution for the tradeoff problem.

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

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

168 **2.1. Representing stochastic models from minimal distributional as-**  
 169 **sumptions.** A stochastic model is some mechanism that produces outputs at ran-  
 170 dom given some inputs. Thus, a stochastic model can be seen as family of probability  
 171 measures  $\{Q_x, x \in \mathcal{X}\}$  indexed by some input space  $\mathcal{X} \subset \mathbb{R}^p$ . We assume that each  
 172 probability measure  $Q_x$  is defined on the measurable space  $(\mathbb{R}, \mathcal{B})$ , where  $\mathcal{B}$  is the  
 173 Borel  $\sigma$ -field induced by  $\mathbb{R}$ . The law  $Q_x$  governs how the outputs are produced given  
 174 the input  $x \in \mathcal{X} \subset \mathbb{R}^p$ . Let us endow  $\mathbb{R}^p$  with its product Borel  $\sigma$ -field  $\mathcal{B}^p$  and let  
 175  $P^*$  be a product probability measure on the measurable space  $(\mathbb{R}^p, \mathcal{B}^p)$  such that  
 176  $P^*(\mathcal{X}) = 1$ . (Thus we assume that  $\mathcal{X} \in \mathcal{B}^p$ .) The probability measure  $P^*$  represents  
 177 how the inputs are drawn by the practitioner. In particular, since  $P^*$  is a product  
 178 probability measure, the inputs are drawn independently.

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

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

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

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

- 189 (i)  *$f(x, Z)$  is measurable for every  $x \in \mathcal{X}$ ,*
  - 190 (ii)  *$P(f(x, Z) \in B) = Q_x(B)$  for every  $x \in \mathcal{X}$  and every  $B \in \mathcal{B}$ ,*
  - 191 (iii)  *$P(X \in A, Z \in B) = P^*(A)P(Z \in B)$  for every  $A \in \mathcal{B}^p$  and  $B \in \mathcal{B}$ .*
- 192 *Moreover, if  $(X, f(X, Z))$  and  $(X', f'(X', Z'))$  are two joint vectors that satisfy the*  
 193 *conditions (i), (ii) and (iii) then  $(X, f(X, Z)) \stackrel{d}{=} (X', f'(X', Z'))$  where  $\stackrel{d}{=}$  means*  
 194 *equality in distribution.*

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

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

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

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

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

223 **2.2. Links with the stochastic models and the sensitivity indices in [10].**  
 224 In [10], the authors consider the model  $(X'(\omega'), \varphi(X'(\omega'), \omega''))$ ,  $\omega' \in \Omega'$ ,  $\omega'' \in \Omega''$ ,  
 225 where  $(\Omega', \mathcal{F}', P')$  and  $(\Omega'', \mathcal{F}'', P'')$  are probability spaces,  $X' = (X'_1, \dots, X'_p)$  is a

226 random vector on  $\Omega'$  and  $\varphi$  is some function. They consider the sensitivity indices

$$227 \quad S_j^{\text{HAG}} = \int_{\Omega''} S_j(\omega'') P''(d\omega''),$$

229 where

$$230 \quad S_j(\omega'') = \frac{\text{Var E}(\varphi(X', \omega'') | X'_j)}{\text{Var } \varphi(X', \omega'')};$$

232 above the variances and the expectation are to be understood as integrals on  $\Omega'$  with  
233 respect to  $P'$ .

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

$$238 \quad \tilde{S}_j^{\text{HAG}} = \text{E} \left( \frac{\text{Var}(\text{E}[f(X, Z) | X_j, Z] | Z)}{\text{Var}(f(X, Z) | Z)} \right), \quad j = 1, \dots, p.$$

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

$$241 \quad (2.1) \quad \varphi(X'(\omega'), \omega'') = X_1(\omega') + X_2(\omega')\omega'',$$

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

$$246 \quad S_1^{\text{HAG}} = \int_{\Omega''} \frac{1}{1 + L^2\omega''} P''(d\omega'') = \int_{\mathbb{R}} \frac{1}{1 + L^2w} \exp\left(-\frac{w^2}{2}\right) \frac{1}{\sqrt{2\pi}} dw$$

248 and  $S_2^{\text{HAG}} = 1 - S_1^{\text{HAG}}$ .

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

251 **EXAMPLE 2.** *Let us first extract the induced complete stochastic model. Set  $P^*((0, t_1] \times$   
252  $(1, t_2]) = t_1(t_2 - 1)/L$  for all  $0 < t_1 < 1$ ,  $1 < t_2 < L + 1$ ,  $L > 0$  and  $Q_x(-\infty, t] =$   
253  $\Phi((t - x_1)/x_2)$  for all  $t \in \mathbb{R}$ , where  $\Phi(t) = \int_{-\infty}^t (2\pi)^{-1/2} e^{-s^2/2} ds$  and  $x = (x_1, x_2) \in$   
254  $\mathbb{R} \times (0, \infty)$ . Now it remains to choose a representation that fulfills the conditions in  
255 Lemma 2.2 and ensures that  $S_1^{\text{HAG}} = \tilde{S}_1^{\text{HAG}}$ . Such a representation can easily be  
256 found. For instance, take  $\Omega = (0, 1)^3$  endowed with the product Lebesgue measure and  
257 put  $Z(\omega) = \omega_3$ ,  $X_1(\omega) = F_1^{-1}(\omega_1)$  and  $X_2(\omega) = F_2(\omega_2)^{-1}$  for  $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega$ ,  
258 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  
259 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  
260 of Lemma 2.2 are fulfilled by construction and the detailed calculations in Appendix A  
261 show that  $S_1^{\text{HAG}} = \tilde{S}_1^{\text{HAG}}$ .*

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

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

$$275 \quad (3.1) \quad |f(X, Z)| \leq F(X)$$

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

278 **3.1. Definition of the sensitivity indices.** We define two kinds of sensitivity  
 279 indices. The sensitivity indices of the first kind exploit the existence of the function  $f$   
 280 by applying the Sobol-Hoeffding decomposition to it directly. The sensitivity indices  
 281 of the second kind result from an application of the Sobol-Hoeffding decomposition  
 282 to the conditional expectation of  $f(X, Z)$  given  $X$ , which is a function of  $X$  alone.  
 283 The indices of the second kind are those to which the estimators (1.4) mentioned in  
 284 the Introduction converge.

285 **3.1.1. Indices of the first kind.** Applying the Sobol-Hoeffding decomposition  
 286 to  $f$  yields

$$287 \quad (3.2) \quad f(X, Z) - f_0 = f_1(X_1) + \cdots + f_p(X_p) + f_{p+1}(Z) + \cdots,$$

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

293 **DEFINITION 3.1** (Sobol indices of the first kind). *The Sobol indices of the first*  
 294 *kind are defined as*

$$295 \quad S'_j = \frac{\text{Var } E(f(X, Z) | X_j)}{\text{Var } f(X, Z)}, \quad j = 1, \dots, p.$$

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

300 **3.1.2. Indices of the second kind.** Let  $g(X) := E(f(X, Z) | X)$  be the condi-  
 301 tional expectation of the output of the stochastic model given the input. The object  
 302  $g$  is a function and the Sobol-Hoeffding decomposition can be applied to it, yielding

$$303 \quad g(X) - g_0 = g_1(X_1) + \cdots + g_p(X_p) + \cdots,$$

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

308 **DEFINITION 3.2** (Sobol indices of the second kind). *The Sobol indices of the*  
 309 *second kind are defined as*

$$310 \quad S''_j = \frac{\text{Var } E(g(X) | X_j)}{\text{Var } g(X)} = \frac{\text{Var } E[E(f(X, Z) | X) | X_j]}{\text{Var } E(f(X, Z) | X)}, \quad j = 1, \dots, p.$$



311 The total sensitivity indices, defined by

$$312 \quad (3.3) \quad S''_{Tj} = 1 - \frac{\text{Var E}(g(X)|X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p)}{\text{Var } g(X)},$$

313  
314  $j = 1, \dots, p$ , are uniquely determined.

315 **3.1.3. Comparison of the definitions.** The sensitivity indices of the first kind  
316 provide more refined “first-order” information than the indices of the second kind.  
317 Example 3 and 4 illustrate this.

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

$$321 \quad S'_1 = \frac{a^2}{a^2 + c^2 E\phi(Z)^2}, S'_2 = 0, S''_1 = 1 \text{ and } S''_2 = 0.$$

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

331 Example 4 returns to the model (2.1).

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

$$337 \quad S'_1 = \frac{1}{4(L^2 + 3(L + 1)) + 1}.$$

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

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

350 **3.2. Construction of the estimators.** We construct estimators for the indices  
351 in Definition 3.1 and 3.2 by Monte-Carlo simulation. The input space is “explored”  
352  $n$  times; at each exploration two independent input vectors are drawn, combined  
353 and passed to the stochastic model which is run  $m$  times. The integer  $n$  is called  
354 the number of explorations and the integer  $m$  is called the number of repetitions.

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

---

**Algorithm 3.1** Generate a Monte-Carlo sample

---

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

---

358 In the algorithm above,  $\tilde{X}_{-0}^{(i)} = X^{(i)}$  by convention. By assumption, the objects  
 359  $\tilde{X}^{(i)}$ ,  $\tilde{X}_{-j}^{(i)}$  and  $Y_j^{(i,k)}$ ,  $j = 0, \dots, p$ ,  $k = 1, \dots, m$ ,  $i = 1, \dots, n$ , are random vectors  
 360 such that the sets  $\{\tilde{X}^{(i)}, \tilde{X}_{-j}^{(i)}, Y_j^{(i,k)} : j = 0, \dots, p; k = 1, \dots, m\}$ ,  $i = 1, \dots, n$ , are  
 361 i.i.d.,  $X^{(i)}$  and  $\tilde{X}^{(i)}$  are independent and  $P(\cap_{j=0}^p \cap_{k=1}^m \{Y_j^{(i,k)} \in B_j^{(k)}\} | X^{(i)}, \tilde{X}^{(i)}) =$   
 362  $\prod_{j=0}^p \prod_{k=1}^m P(Y_j^{(i,k)} \in B_j^{(k)} | X^{(i)}, \tilde{X}^{(i)})$  for all Borel sets  $B_j^{(k)} \in \mathcal{B}$ . It is easy to see that  
 363 these conditions characterize the joint law of the set  $\{\tilde{X}_{-j}^{(i)}, Y_j^{(i,k)} : j = 0, \dots, p; k =$   
 364  $1, \dots, m; i = 1, \dots, n\}$ , that is, the inputs and the outputs of Algorithm 3.1.

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

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

$$372 \quad (3.4) \quad \hat{S}_{j;n,m}'' = \frac{\frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \frac{1}{m} \sum_{k'=1}^m Y_j^{(i,k')} - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2}{\frac{1}{n} \sum_{i=1}^n \left( \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2 - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2},$$

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

$$374 \quad S_j'' = \frac{\mathbb{E} g(X^{(1)}) g(\tilde{X}_{-j}^{(1)}) - (\mathbb{E} g(X^{(1)}))^2}{\mathbb{E} g(X^{(1)})^2 - (\mathbb{E} g(X^{(1)}))^2}$$

$$(3.5)$$

$$375 \quad = \frac{\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}] \mathbb{E}[f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) | \tilde{X}_{-j}^{(1)}] - (\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}])^2}{\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}]^2 - (\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}])^2}.$$

376

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

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

$$\begin{aligned}
382 \quad S'_j &= \frac{\mathbb{E} f(X^{(1)}, Z_0^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) - (\mathbb{E} f(X^{(1)}, Z_0^{(1,1)}))^2}{\mathbb{E} f(X^{(1)}, Z_0^{(1,1)})^2 - (\mathbb{E} f(X^{(1)}, Z_0^{(1,1)}))^2} \\
(3.6) \quad &= \frac{\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}] \mathbb{E}[f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) | \tilde{X}_{-j}^{(1)}] - \left( \mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}] \right)^2}{\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)})^2 | X^{(1)}] - \left( \mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}] \right)^2}.
\end{aligned}$$

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

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

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

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

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

405	$n$	50	25	10	5	2	1
	$m$	1	2	5	10	25	50.

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

412 **4.1. Introducing the miss-ranking error and its bound.** To compare the  
413 estimators, a measure of performance has to be defined. We shall consider the miss-

414 ranking error (MRE), defined by

$$415 \quad \text{MRE} = \mathbb{E} \left( \sum_{j=1}^p |\widehat{R}_{j;n,m} - R_j| \right),$$

416 where  $R_j$  is the rank of  $D_j$  among  $D_1, \dots, D_p$ , that is,  $R_j = \sum_{i=1}^p \mathbf{1}(D_i \leq D_j)$ , and  
 417  $\widehat{R}_{j;n,m}$  is an estimator of  $R_j$ . Recall that  $D_1, \dots, D_p$  are the upper-left terms in (3.6)  
 418 and (3.5). They determine the ranks of the sensitivity indices. The MRE is small  
 419 when one succeeds in ranking the inputs from the most to the least important, a task  
 420 which is called ‘‘factors prioritization’’ in [27, p. 52]. For simplicity, we assume that  
 421 the sensitivity indices, and hence the quantities  $D_1, \dots, D_p$ , are distincts, so that they  
 422 can be ranked unambiguously. This assumption is mild: even if  $D_j$  and  $D_{j'}$  only differ  
 423 by a small amount, it is still satisfied. (It would be really bad luck to have a model  
 424 and a distribution for the inputs such that  $D_j = D_{j'}$  exactly.) Recall that the ranks  
 425 of the sensitivity indices of the first kind coincide with the ranks of the sensitivity  
 426 indices of the second kind. Thus, the MRE permits to find a unique solution for both  
 427 kinds of sensitivity indices.

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

432 **PROPOSITION 4.1.** *Let  $\widehat{D}_{j;n,m}$ ,  $j = 1, \dots, p$ , be the upper-left term in (3.7) or (3.4)*  
 433 *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  $D_1, \dots, D_p$  are all distincts then*

$$434 \quad \text{MRE}(T, m) \leq \frac{L}{nm} \left( m \sum_{j=1}^p \text{Var}(\mathbb{E}[Y_0 Y_j | \mathbf{X}]) \right. \\
 435 \quad \quad \quad \left. + \sum_{j=1}^p \mathbb{E}[\text{Var}(Y_0 Y_j | \mathbf{X}) - \text{Var}(Y_0 | \mathbf{X}) \text{Var}(Y_j | \mathbf{X})] \right. \\
 436 \quad \quad \quad \left. + \frac{1}{m} \sum_{j=1}^p \mathbb{E}[\text{Var}(Y_0 | \mathbf{X}) \text{Var}(Y_j | \mathbf{X})] \right),$$

437 where

$$438 \quad L = \frac{4(p-1)}{\min_{j < j'} (|D_j - D_{j'}|^2)}.$$

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

$$444 \quad (4.1) \quad \text{MRE}(T, m) \leq \frac{1}{T} v(m), \quad m \in \text{div}_p(T),$$

448 where  
449

$$450 \quad v(m) = L(p+1) \left( m \sum_{j=1}^p \text{Var}(\mathbb{E}[Y_0 Y_j | \mathbf{X}]) + \sum_{j=1}^p \mathbb{E}[\text{Var}(Y_0 Y_j | \mathbf{X}) - \text{Var}(Y_0 | \mathbf{X}) \text{Var}(Y_j | \mathbf{X})] \right. \\ 451 \quad \quad \quad \left. + \frac{1}{m} \sum_{j=1}^p \mathbb{E}[\text{Var}(Y_0 | \mathbf{X}) \text{Var}(Y_j | \mathbf{X})] \right). \\ 452$$

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

$$455 \quad \text{MRE}(T, m_T^\dagger) \leq \frac{v(m_T^\dagger)}{T} \leq \frac{v(m)}{T}, \quad \text{for all } m \in \text{div}_p(T).$$

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

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

464 **4.2. A two-stage procedure to estimate the sensitivity indices.** The re-  
465 sults in Section 4.1 suggest a two-stage procedure to estimate the sensitivity indices.  
466 The procedure is given in Algorithm 4.1. The computational budget is split into two  
467 parts  $K$  and  $T - K$ . Denote by  $m_{T-K}^\dagger$  the element  $m \in \text{div}_p(T - K)$  that minimizes  
468 the function  $v(m)$ . The first  $K$  calls to the model are used to estimate  $m_{T-K}^\dagger$ . The  
469 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  $\hat{m}_{T-K}^\dagger \in \text{div}_p(T - K)$ .

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

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

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

---

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

$$473 \quad (4.2) \quad m^* := \sqrt{\frac{\sum_{j=1}^p \mathbb{E}[\text{Var}(Y_0 | \mathbf{X}) \text{Var}(Y_j | \mathbf{X})]}{\sum_{j=1}^p \text{Var}(\mathbb{E}[Y_0 Y_j | \mathbf{X}])}} = \sqrt{\frac{\sum_{j=1}^p \zeta_{3,j}}{\sum_{j=1}^p \zeta_{1,j}}}, \\ 474$$

475 where  $\zeta_{3,j} = \mathbb{E}[\text{Var}(Y_0|\mathbf{X}) \text{Var}(Y_j|\mathbf{X})]$  and  $\zeta_{1,j} = \text{Var}(\mathbb{E}[Y_0 Y_j|\mathbf{X}])$ ,  $j = 1, \dots, p$ .

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

$$478 \quad \varphi_T(x) = \begin{cases} \lfloor x \rfloor_T & \text{if } \sqrt{\lfloor x \rfloor_T \lceil x \rceil_T} > x \geq 1 \\ \lceil x \rceil_T & \text{if } \sqrt{\lfloor x \rfloor_T \lceil x \rceil_T} \leq x \leq \frac{T}{p+1} \end{cases}$$

480 where

$$481 \quad \lfloor x \rfloor_T = \max\{m \in \text{div}_p(T), m \leq x\}, \quad \lceil x \rceil_T = \min\{m \in \text{div}_p(T), m \geq x\}.$$

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

485 **PROPOSITION 4.3.** *If  $m^* > 0$  then  $m_{T-K}^\dagger = \varphi_{T-K}(m^*)$ . If, moreover,*  
486  $\lfloor m^* \rfloor_{T-K} \lceil m^* \rceil_{T-K}$  *is not equal to  $m^{*2}$  then the minimizer of  $v(m)$ ,  $m \in \text{div}_p(T-K)$ ,*  
487 *is unique.*

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

$$492 \quad (4.3) \quad \hat{m}_K^* := \sqrt{\frac{\sum_{j=1}^p \hat{\zeta}_{3,j}}{\sum_{j=1}^p \hat{\zeta}_{1,j}}},$$

493 where

$$494 \quad \hat{\zeta}_{3,j} =$$

$$495 \quad (4.4) \quad \frac{1}{n_0} \sum_{i=1}^n \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})^2 \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)})^2$$

$$496 \quad (4.5) \quad + \frac{1}{n_0} \sum_{i=1}^n \left( \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)}) \right)^2 \left( \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)}) \right)^2$$

$$497 \quad (4.6) \quad - \frac{1}{n_0} \sum_{i=1}^n \left( \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)}) \right)^2 \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)})^2$$

$$498 \quad (4.7) \quad - \frac{1}{n_0} \sum_{i=1}^n \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})^2 \left( \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)}) \right)^2,$$

500 and

$$501 \quad \hat{\zeta}_{1,j} =$$

$$502 \quad (4.8) \quad \frac{1}{n_0} \sum_{i=1}^n \left( \frac{1}{m_0} \sum_{k=1}^{m_0} f(X^{(i)}, Z_0^{(i,k)}) f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k)}) \right)^2$$

$$503 \quad (4.9) \quad - \left( \frac{1}{n_0} \sum_{i=1}^n \frac{1}{m_0} \sum_{k=1}^{m_0} f(X^{(i)}, Z_0^{(i,k)}) f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k)}) \right)^2.$$

504

505 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  
 506  $\widehat{m}_K^* = 0$ .

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

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

$$510 \quad \sqrt{n_0} \left( \widehat{m}_K^* - \left[ m^* + \frac{C}{m_0} + \epsilon_{m_0} \right] \right) \xrightarrow{d} \mathcal{N}(0, \sigma_{m_0}^2),$$

511 where  $C$  is some constant,  $\epsilon_{m_0} = C_1/m_0^2 + \dots + C_N/m_0^{N+1}$  for some constants  
 512  $C_1, \dots, C_N$  and  $\sigma_{m_0}^2$  is some variance depending on  $m_0$ . If  $m_0 \rightarrow \infty$  then the above  
 513 display with  $\epsilon_{m_0} = o(1/m_0)$  and  $\sigma_{m_0}$  replaced by  $\lim_{m_0 \rightarrow \infty} \sigma_{m_0}$  is true.

514 Theorem 4.4 shows that  $\widehat{m}_K^*$  is asymptotically biased. The bias is polynomial in  
 515  $1/m_0$ . Corollary 4.5 shows that letting  $m_0 \rightarrow \infty$  suffices to get the consistency of  $\widehat{m}_K^*$   
 516 but to get a central limit theorem centered around  $m^*$ , it is furthermore needed that  
 517  $\sqrt{n_0}/m_0 \rightarrow 0$ .

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

$$520 \quad \sqrt{n_0}(\widehat{m}_K^* - m^*) \xrightarrow{d} \mathcal{N}(0, \lim_{m_0 \rightarrow \infty} \sigma_{m_0}^2).$$

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

524 **PROPOSITION 4.6.** *Assume (3.1) holds. Let  $n_0 \rightarrow \infty$  and  $m_0 \rightarrow \infty$ . Then*

$$525 \quad P \left( \widehat{m}_{T-K}^\dagger = m_{T-K}^\dagger \right) \rightarrow 1.$$

526  
 527 All the details of Algorithm 4.1 have been given.

528 **4.3. Performance.** To get some insight into the performance of Algorithm 4.1,  
 529 we look at the estimators produced in Stage 2, which are built with  $T - K$  calls  
 530 to the model with  $\widehat{m}_{T-K}^\dagger$  repetitions and  $(T - K)/(p + 1)/\widehat{m}_{T-K}^\dagger$  explorations. Let  
 531  $\widehat{D}_{j;\hat{n},\widehat{m}}$ ,  $j = 1, \dots, p$ , be the discriminators of those estimators, so that  $\hat{n} = (T -$   
 532  $K)/(p + 1)/\widehat{m}_{T-K}^\dagger$  and  $\widehat{m} = \widehat{m}_{T-K}^\dagger$ . Let  $\widehat{R}_{j;\hat{n},\widehat{m}}$  be the rank of  $\widehat{D}_{j;\hat{n},\widehat{m}}$  among  
 533  $\widehat{D}_{1;\hat{n},\widehat{m}}, \dots, \widehat{D}_{p;\hat{n},\widehat{m}}$ . The conditional expectation of  $\sum_{j=1}^p |\widehat{R}_{j;\hat{n},\widehat{m}} - R_j|$  given the  
 534 outputs produced in Stage 1 is equal to  $\text{MRE}(T - K, \widehat{m}_{T-K}^\dagger)$ , the MRE based on  
 535  $T - K$  calls to the model with  $\widehat{m}_{T-K}^\dagger$  repetitions and  $(T - K)/(p + 1)/\widehat{m}_{T-K}^\dagger$  explo-  
 536 rations. Recall that the estimator  $\widehat{m}_{T-K}^\dagger$  is computed with  $K$  calls to the model and  
 537 by construction  $\widehat{m}_{T-K}^\dagger \in \text{div}_p(T - K)$ . Equation (4.1) yields

$$538 \quad (4.10) \quad \text{MRE}(T - K, \widehat{m}_{T-K}^\dagger) \leq \frac{1}{T - K} v(\widehat{m}_{T-K}^\dagger),$$

539  
 540 where the inequality above is to be understood almost surely, since both sides of  
 541 equation are random variables.

542 We shall show that, under conditions on  $T$  and  $K$ , the bound in (4.10) is in  
 543 fact close to the best obtainable guarantee  $v(m_T^\dagger)/T$ . Suppose for simplicity that

544  $\text{div}_p(T - K) \subset \text{div}_p(T)$ . Since  $\widehat{m}_{T-K}^\dagger \in \text{div}_p(T - K)$  by construction, we have  
 545  $\widehat{m}_{T-K}^\dagger \in \text{div}_p(T)$  and therefore  $v(m_T^\dagger)/T \leq v(\widehat{m}_{T-K}^\dagger)/(T - K)$  by definition of  $m_T^\dagger$ .  
 546 The left hand side is the best guarantee we can hope for and corresponds to the case  
 547 where  $K = 0$  in Algorithm 4.1. It can be interpreted as the guarantee obtained if we  
 548 would know the value of  $m^*$  in (4.2), for in that case we would not need to spend  
 549 budget in Stage 1 of Algorithm 4.1. We proceed in two steps.

550 **THEOREM 4.7.** *Assume that the conditions of Proposition 4.6 are fulfilled. Suppose*  
 551 *furthermore that  $K \rightarrow \infty$  such that  $K/T \rightarrow 0$ . Then*

$$552 \frac{1}{T - K} v(\widehat{m}_{T-K}^\dagger) = \frac{1}{T} v(m_{T-K}^\dagger) (1 + o_P(1)).$$

554 Theorem 4.7 holds without the condition  $\text{div}_p(T - K) \subset \text{div}_p(T)$ . Imposing this  
 555 condition, we get Corollary 4.8 below.

556 **COROLLARY 4.8.** *If, in addition to the conditions of Theorem 4.7,  $\text{div}_p(T - K) \subset$*   
 557  *$\text{div}_p(T)$  then*

$$558 \text{MRE}(T - K, \widehat{m}_{T-K}^\dagger) \leq \frac{v(m_T^\dagger)}{T} (1 + o_P(1)).$$

560 The result of Corollary 4.8 easily follows from (4.10) and Theorem 4.7 because  
 561  $m_{T-K}^\dagger = m_T^\dagger$  as soon as  $(T - K)/(p + 1) > m_{T-K}^\dagger$ , which happens eventually as  $T$   
 562 and  $K$  go to infinity because the function  $v$  is convex.

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

568 Two frameworks are considered:

- 569 •  $n \rightarrow \infty$  and  $m$  is fixed;
- 570 •  $n \rightarrow \infty$  and  $m \rightarrow \infty$ .

571 In the second framework  $m = m_n$  is a sequence indexed by  $n$  that goes to infinity as  
 572  $n$  goes to infinity. Denote by  $\mathbf{S}'$  (resp.  $\mathbf{S}''$ ) the column vector with coordinates  $S'_j$   
 573 (resp.  $S''_j$ ),  $j = 1, \dots, p$ , and denote by  $\widehat{\mathbf{S}}'_{n,m}$  (resp.  $\widehat{\mathbf{S}}''_{n,m}$ ) the column vector with  
 574 coordinates  $\widehat{S}'_{j;n,m}$  given in (3.7) (resp.  $\widehat{S}''_{j;n,m}$  given in (3.4)). Theorem 5.1 below  
 575 predicts that the joint vector  $(\widehat{\mathbf{S}}'_{n,m}, \widehat{\mathbf{S}}''_{n,m})^\top$  is asymptotically normal.

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

$$577 \sqrt{n} \left( \widehat{\mathbf{S}}'_{n,m} - \mathbf{S}' \left[ 1 - \frac{\widehat{\mathbf{S}}'_{n,m} - \mathbf{S}'}{\text{E Var}(f(X, Z)|X) + m \text{Var E}(f(X, Z)|X)} \right] \right) \xrightarrow{d} \mathcal{N}(0, \Xi_m),$$

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

580 A blockwise reading of Theorem 5.1 shows that the behaviors of  $\widehat{\mathbf{S}}'_{n,m}$  and  $\widehat{\mathbf{S}}''_{n,m}$   
 581 differ. While  $\widehat{\mathbf{S}}'_{n,m}$  is asymptotically unbiased even if  $m$  is kept fixed,  $\widehat{\mathbf{S}}''_{n,m}$  is asymptot-  
 582 ically biased in general even if  $m$  goes to infinity. The estimator  $\widehat{\mathbf{S}}''_{n,m}$  under-estimates  
 583  $\mathbf{S}''$ . The bias, given by

$$584 \mathbf{S}'' \frac{\text{E Var}(f(X, Z)|X)}{\text{E Var}(f(X, Z)|X) + m \text{Var E}(f(X, Z)|X)},$$

585



586 is null whenever  $f$  actually does not depend on  $Z$ , and large whenever the stochastic  
587 model is highly stochastic.

588 Corollary 5.2 below shows that  $m$  must go to infinity fast enough to avoid the  
589 estimator to be concentrated around the wrong target.

590 COROLLARY 5.2. *Assume (3.1) holds. Let  $n \rightarrow \infty$ . If  $m \rightarrow \infty$  such that  $\sqrt{n}/m \rightarrow$   
591  $0$  then*

$$592 \quad \sqrt{n} \left( \widehat{\mathbf{S}}''_{n,m} - \mathbf{S}'' \right) \xrightarrow{d} \mathcal{N}(0, \Xi_{22}),$$

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

594 The difference between  $\widehat{\mathbf{S}}'_{n,m}$  and  $\widehat{\mathbf{S}}''_{n,m}$  is due to the difference between the lower-  
595 left terms in (3.7) and (3.4). While the lower-left term in (3.7) is unbiased for all  $n$   
596 and  $m$ , the lower-left term in (3.4) has a bias depending on  $m$  which propagates to the  
597 estimator of the sensitivity indices. (The calculations are carried out in Appendix D.)  
598 From a statistical perspective, it is more difficult to estimate the sensitivity indices of  
599 the second kind than to estimate the sensitivity indices of the first kind. To estimate  
600 the former, one needs to repeat the model many times. To estimate the later, this is  
601 not necessary.

602 **6. Numerical tests.** Section 6.1 illustrates how the MRE responds to a change  
603 in the Monte-Carlo design. In Section 6.1 the total budget  $T$  is kept fixed. Section 6.2  
604 illustrates how the sensitivity indices estimators behave asymptotically. In Section 6.2  
605 the total budget  $T$  increases.

606 **6.1. Comparison of Monte-Carlo designs.** The effect of the number of rep-  
607 etitions on the sensitivity indices estimators and the effect of the calibration in the  
608 two-stage procedure are examined in two kinds of experiments: the “direct” experi-  
609 ments and the “calibration” experiments.

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

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

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

630 **6.1.1. High noise context.** The coefficients are  $\beta = 1.2$  and  $\sigma = 4$ . The  
631 sensitivity indices are  $S'_1 = 0.05$ ,  $S'_2 = 0.08$ ,  $S''_1 = 0.41$  and  $S''_2 = 0.59$ . The real

$K/3$	$m_0$				$n_0$			
	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  $m_0 = 2, 5, 10, 20, \dots$ . For instance, for  $K/(p+1) = 200 = m_0 n_0$ , the normalized MRE is available for  $m_0 = 2, 5, 10, 20, 40, 100$ .

632  $m^*$  is about 5.8. The total budget is  $T = 3 \times 500 = 1500$  and hence  $\text{div}_2(1500) =$   
633  $\{1, 2, 4, 5, 10, 20, 25, 50, 100, 125, 250, 500\}$ . The integer  $m_{1500}^\dagger$  is equal to  $\varphi_{1500}(m^*) =$   
634 5. Since the budget is kept fixed, the numbers of explorations are, respectively,  
635 500, 250, 125, 100, 50, 25, 20, 10, 5, 4, 2, 1. The number of replications is 1500.

636 The results of the direct experiment are given in Figure 1 for  $m = 1, 2, 4, 5, 10,$   
637 20, 25. The MSE of first kind does not vary with the number of repetitions and is  
638 much lower than the MSE of second kind, see (c). The estimators of the second kind  
639 are highly biased for small numbers of repetitions (a) and they have a higher variance  
640 for larger numbers of repetitions (b). The fact that the bias is high for small numbers  
641 of repetitions agrees with the theory, according to which the bias should vanish as  $m$   
642 goes to infinity. Overall, the sensitivity indices of the second kind seem to be much  
643 harder to estimate than the indices of the first kind, the estimators of which have a  
644 negligible bias and a very small variance whatever the number of repetitions.

645 According to Figure 1(c), the normalized MRE curve has a banana shape with a  
646 minimum of about slightly less than 30% reached around  $m \in \{5, 10\}$  and endpoints  
647 with a value of about 35%. A value of 30% means that the probability of ranking the  
648 inputs correctly is about 70%. The region of observed optimal performance  $m \in$   
649  $\{5, 10\}$  coincides with  $m_{1500}^\dagger = 5$ , the point at which the bound is minimal.

650 The results of the calibration experiment is given in Table 1 for the normalized  
651 MRE. The lowest MREs are reached at the bottom right of the table, with values  
652 corresponding to  $2 \leq m \leq 10$  in Figure 1 (c). Optimal performance is reached with  
653 very few explorations in the first stage of the two-stage procedure. In this case, the  
654 estimator  $\hat{m}_K^*$  has a small bias but a high variance. It seems to be better than an  
655 estimator with a small variance but a large bias. This might be explained by the low  
656 curvature of the MRE curve.

657 **6.1.2. Low noise context.** The coefficients are  $\beta = 1.2$  and  $\sigma = 0.9$ . The  
658 sensitivity indices are  $S'_1 = 0.31$ ,  $S'_2 = 0.44$ ,  $S''_1 = 0.41$  and  $S''_2 = 0.59$ . The real  
659  $m^*$  is about 0.30 and hence the integer  $m_{1500}^\dagger$  is equal to 1. As expected, these  
660 numbers are smaller than the ones found in the high noise context. The total budget  
661 is  $T = 3 \times 500 = 1500$ . The number of replications is 500.

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

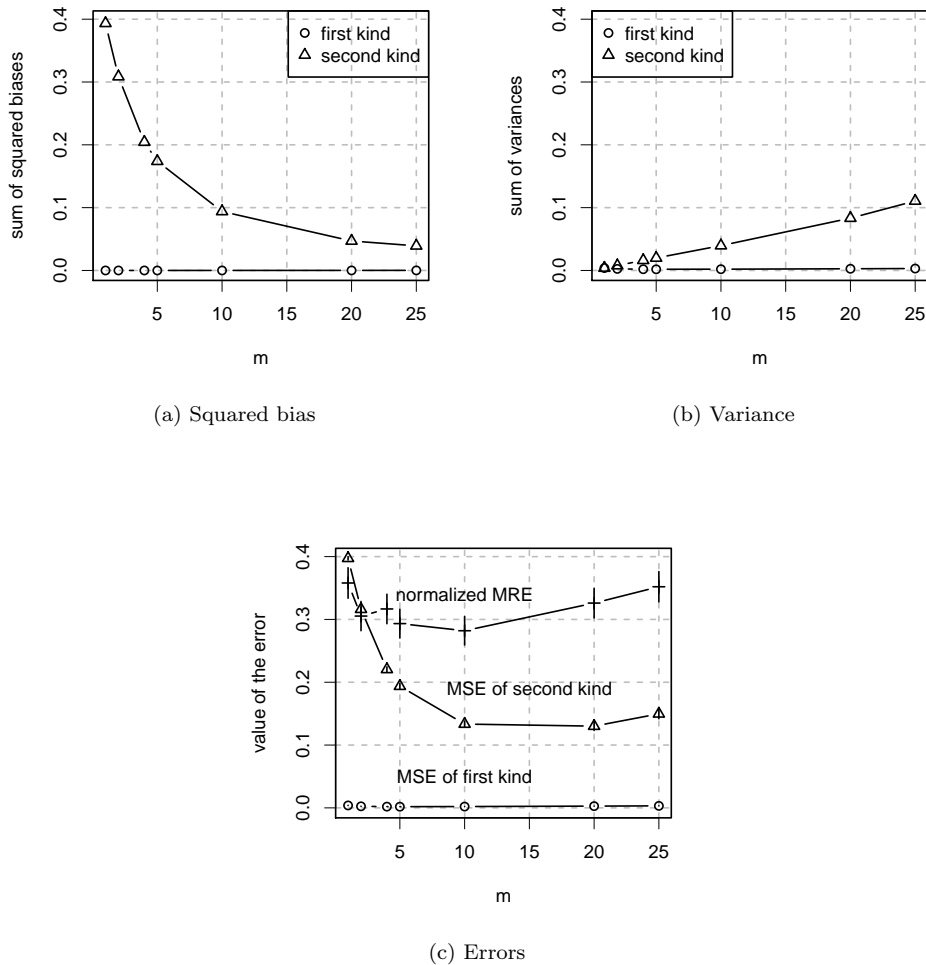


Fig. 1: Sum of squared biases (a), sum of variances (b) and errors (c) of the sensitivity indices estimators for the linear model in the high noise setting. Confidence intervals of level 95% are added in (c).

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

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

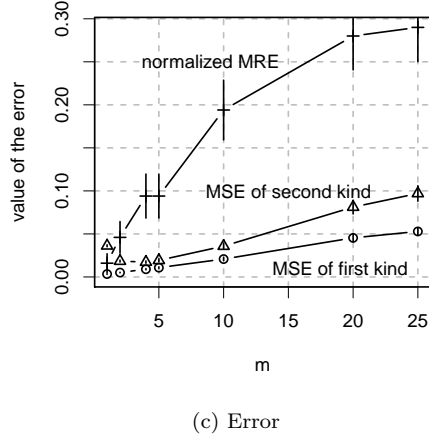
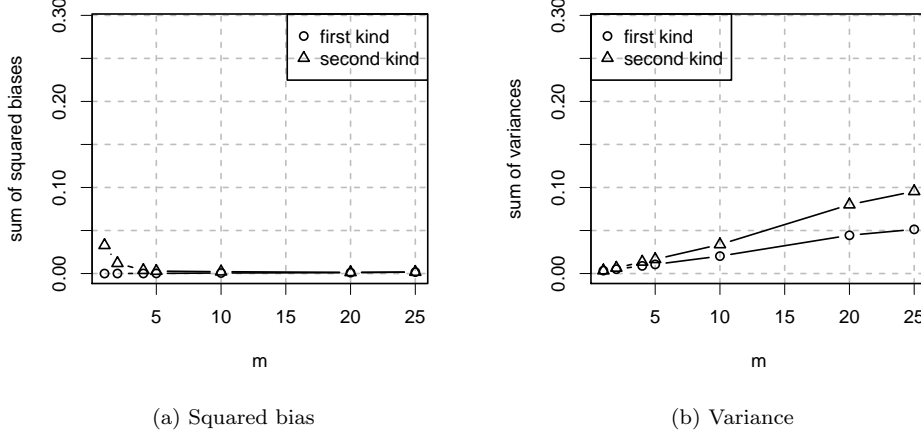


Fig. 2: Sum of squared biases (a), sum of variances (b) and errors (c) of the sensitivity indices estimators for the linear model in the low noise context. Confidence intervals of level 95% are added in (c).

$K/3$	$m_0$				$n_0$			
	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  $m_0 = 2, 5, 10, 20, \dots$ . For instance, for  $K/(p + 1) = 200 = m_0 n_0$ , the normalized MRE is available for  $m_0 = 2, 5, 10, 20, 40, 100$ .

677 **6.2. Asymptotic behavior of the sensitivity indices estimators.** To illus-  
 678 trate the asymptotic behavior of the sensitivity indices estimators, Sobol's g-function,  
 679 a benchmark in sensitivity analysis [25, 20], is considered. Sobol's g-function is given  
 680 by

$$681 \quad g(U_1, \dots, U_{p+1}) = \prod_{j=1}^{p+1} \frac{|4U_j - 2| + a_j}{1 + a_j},$$

683 where the  $a_j$  are nonnegative and the  $U_j$  are independent standard uniform random  
 684 variables. The less  $a_j$  the more  $U_j$  is important. Elementary calculations show that  
 685 the first-order Sobol index associated with  $U_j$  is given by

$$686 \quad S_j^{(a_1, \dots, a_{p+1})} = \frac{1}{3(1 + a_j)^2} \left( -1 + \prod_{j=1}^{p+1} \frac{(4/3 + a_j^2 + 2a_j)}{(1 + a_j)^2} \right)^{-1}.$$

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

$$691 \quad (6.1) \quad Y = f(X_1, \dots, X_p, Z) = g(X_1, \dots, X_{i-1}, Z, X_i, \dots, X_p).$$

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

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

$$697 \quad S'_j = \begin{cases} S_j^{(a_1, \dots, a_{p+1})} & \text{if } 1 \leq j \leq i - 1 \\ S_{j+1}^{(a_1, \dots, a_{p+1})} & \text{if } i \leq j \leq p \end{cases}$$

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

$$700 \quad b_{ij} = \begin{cases} a_j & \text{if } 1 \leq j \leq i - 1, \\ a_{j+1} & \text{if } i \leq j \leq p. \end{cases}$$

702 For each kind of Sobol index, we produced 500 estimates of the  $p$  Sobol indices  
 703 and computed the values of the mean squared error (MSE) by averaging over the  
 704 500 replications and summing over the  $p$  indices. We tested  $n = 100, 500, 2500$  and  
 705  $m = 1, 10, 100$ .

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

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

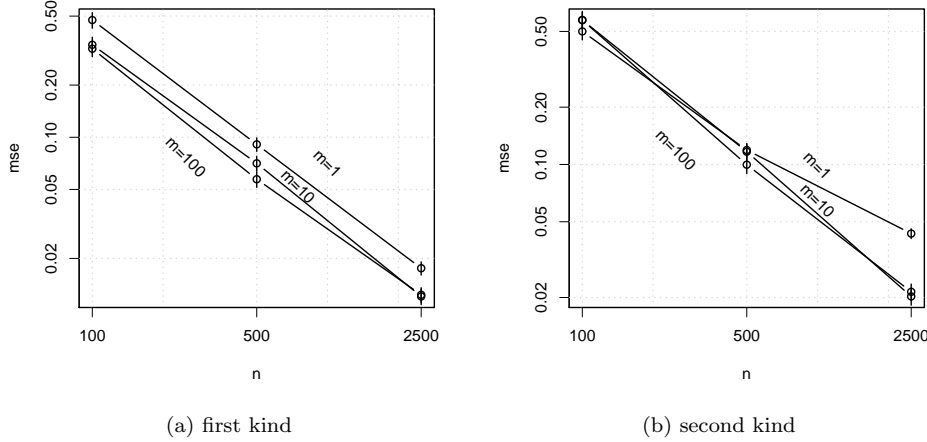


Fig. 3: MSEs for the Sobol index estimators of the first and second kind (logarithmic scale).

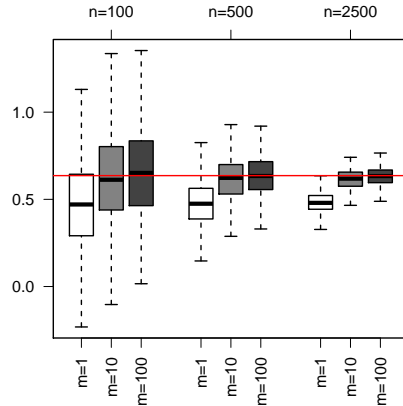


Fig. 4: Boxplots of the estimates for the Sobol index of the second kind associated with  $X_1$ . The red horizontal line is the truth.

718 **7. Conclusion.** The practical method that consists of repeating the stochastic  
 719 model at each exploration of the input space was analysed in the context of global  
 720 sensitivity analysis. To find a tradeoff between the number of explorations  $n$  and the  
 721 number of repetitions  $m$ , a bound on the missranking error (MRE) was found and  
 722 minimized, leading to a solution in closed form. A two-step procedure was imple-  
 723 mented to estimate the sensitivity indices. It was shown to have good asymptotic  
 724 properties. Two sensitivity indices were considered. The sensitivity index of the first  
 725 kind results from the existence of a function that links the output, the inputs and

726 some random noise in stochastic models defined through probability measures. The  
 727 sensitivity index of the second kind is the population version of the estimator (1.4). An  
 728 asymptotic analysis of the estimators was conducted. It was found that the estimators  
 729 for the indices of the second kind may be asymptotically biased if  $m$  goes to infinity  
 730 too slowly, while the estimators for the indices of the first kind are asymptotically  
 731 unbiased even if  $m$  remains fixed. To test the theory, simulation experiments were  
 732 conducted and the bias of the sensitivity estimator of the second kind was confirmed.  
 733 Optimal compromises between repetitions and explorations have been identified and  
 734 compared with the output of the two-stage procedure.

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

## 747 Appendix A. Calculations of some sensitivity indices.

748 **A.1. Calculations for  $\tilde{S}_1^{\text{HAG}}$  in Example 2.** We have

$$749 \quad \tilde{S}_1^{\text{HAG}} = \mathbb{E} \left( \frac{\text{Var}(\mathbb{E}[f(X, Z)|X_j, Z]|Z)}{\text{Var}(f(X, Z)|Z)} \right) = \int_{\Omega} \frac{\text{Var}(\mathbb{E}[f(X, Z)|X_j, Z]|Z)}{\text{Var}(f(X, Z)|Z)} dP.$$

750

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

$$754 \quad \tilde{S}_1^{\text{HAG}} = \int_{(0,1)} \frac{\text{Var}(\mathbb{E}[f(X, z)|X_j, Z = z]|Z = z)}{\text{Var}(f(X, z)|Z = z)} dz = \int_{(0,1)} \frac{\text{Var}(\mathbb{E}[f(X, z)|X_1])}{\text{Var}(f(X, z))} dz.$$

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

$$757 \quad \begin{aligned} \text{Var}(f(X, z)) &= \text{Var}(\Phi^{-1}(z)X_2 + X_1) = \Phi^{-1}(z)^2 \text{Var}(X_2) + \text{Var}(X_1) \\ &= \Phi^{-1}(z)^2 \frac{L^2}{12} + \frac{1}{12}, \end{aligned}$$

760 and

$$761 \quad \begin{aligned} \text{Var}(\mathbb{E}[f(X, z)|X_1]) &= \text{Var}(\mathbb{E}[\Phi^{-1}(z)X_2 + X_1|X_1]) \\ 762 &= \text{Var}(\Phi^{-1}(z) \mathbb{E}[X_2|X_1] + \mathbb{E}[X_1|X_1]) \\ 763 &= \text{Var}(\Phi^{-1}(z) \mathbb{E}[X_2] + X_1) \\ 764 &= \text{Var}(X_1) \\ &= \frac{1}{12} \end{aligned}$$

765

766

767 and hence

$$768 \quad \widetilde{S}_1^{\text{HAG}} = \int_{(0,1)} \frac{1}{\Phi^{-1}(z)^2 L^2 + 1} dz = \int_{-\infty}^{\infty} \frac{1}{z^2 L + 1} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

770 **A.2. Calculations for  $S'_1$  in Example 4.** The sensitivity index of the first  
771 kind associated with the first input is given by

$$772 \quad S'_1 = \frac{\text{Var}[\mathbb{E}(X_1 + X_2\Phi^{-1}(Z)|X_1)]}{\text{Var}[X_1 + X_2\Phi^{-1}(Z)]}.$$

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

$$777 \quad \begin{aligned} \text{Var}[X_2\Phi^{-1}(Z)] &= \text{Var}[\mathbb{E}(X_2\Phi^{-1}(Z)|Z)] + \mathbb{E}(\text{Var}[X_2\Phi^{-1}(Z)|Z]) \\ 778 \quad &= \text{Var}\left[\Phi^{-1}(Z)\left(\frac{L}{2} + 1\right)\right] + \int_0^1 \Phi^{-1}(z)^2 \text{Var}[X_2] dz \\ 779 \quad &= \left(\frac{L}{2} + 1\right)^2 + \frac{L^2}{12}, \end{aligned}$$

781 so that

$$782 \quad S'_1 = \frac{1/12}{1/12 + (L/2 + 1)^2 + L^2/12} = \frac{1}{4(L^2 + 3(L + 1)) + 1}.$$

## 784 Appendix B. Proofs.

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

$$793 \quad \begin{aligned} P(f(x, Z) \leq t) \\ 794 \quad &= P(Z \leq F_x(t)) = \lambda^{\otimes_{j=1}^{p+1}}\{\omega \in \Omega : \omega_{p+1} \leq F_x(t)\} = \lambda(0, F_x(t)) = F_x(t). \end{aligned}$$

796 Finally, by the same token,

$$797 \quad \begin{aligned} P(X_1 \leq t_1, \dots, X_p \leq t_p, Z \leq t_{p+1}) \\ 798 \quad &= P\{\omega : \omega_1 \leq F_1(t_1), \dots, \omega_p \leq F_p(t_p), \omega_{p+1} \leq t_{p+1}\} = t_{p+1} \prod_{j=1}^p F_j(t_j). \end{aligned}$$

800 The proof is complete.

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



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

$$805 \quad P(\widehat{D}_i - \widehat{D}_j \geq 0) \leq \frac{\text{Var } \widehat{D}_i + \text{Var } \widehat{D}_j}{\frac{1}{2}|D_i - D_j|^2}.$$

806  
807 *Proof. We have*

$$808 \quad P(\widehat{D}_i - \widehat{D}_j \geq 0) \leq P(|\widehat{D}_i - D_i| + |\widehat{D}_j - D_j| \geq D_j - D_i)$$

$$809 \quad \leq P(|\widehat{D}_i - D_i|^2 + |\widehat{D}_j - D_j|^2 \geq \frac{1}{2}|D_j - D_i|^2)$$

810

811 *and the claim follows from Markov's inequality.*

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

$$814 \quad \sum_{i=1}^p \mathbb{E} |\widehat{R}_i - R_i| \leq \sum_{i=1}^p \sum_{j=1}^p \mathbb{E} |\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i) - \mathbf{1}(D_j \leq D_i)|$$

$$815 \quad \leq \sum_{i=1}^p \sum_{j \neq i} \frac{\text{Var } \widehat{D}_i + \text{Var } \widehat{D}_j}{\frac{1}{2}|D_i - D_j|^2}$$

$$816 \quad \leq \frac{4(p-1)}{\min_{j < j'} |D_j - D_{j'}|^2} \sum_{i=1}^p \text{Var } \widehat{D}_i,$$

817  
818 where the second inequality holds by Lemma B.1 and because

$$819 \quad \mathbb{E} |\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i) - \mathbf{1}(D_j \leq D_i)| = \begin{cases} \mathbb{E} |\mathbf{1}(\widehat{D}_j > \widehat{D}_i)| & \text{if } j < i, \\ 0 & \text{if } j = i, \\ \mathbb{E} |\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i)| & \text{if } j > i. \end{cases}$$

820  
821 It remains to calculate the variances. But this is done in Lemma D.3 in Appendix D,  
822 where it is found that

$$823 \quad \text{Var } \widehat{D}_j = \frac{1}{n} \{ \text{Var } \mathbb{E}[Y_0 Y_j | \mathbf{X}] + \frac{1}{m} (\mathbb{E} \text{Var}[Y_0 Y_j | \mathbf{X}] - \text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}])$$

$$824 \quad + \frac{1}{m^2} \mathbb{E} \text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}] \}.$$

825  
826 **Proof of Proposition 4.3.** We distinguish between three cases:  $0 < m^* < 1$ ,  
827  $m^* > (T-K)/(p+1)$  and  $1 \leq m^* \leq (T-K)/(p+1)$ . Recall that  $m_{T-K}^\dagger$  is the  
828 minimizer of  $v(m)$ ,  $m$  in  $\text{div}_p(T-K)$ .

829 If  $0 < m^* < 1$  then by definition  $\varphi_{T-K}(m^*) = 1$  and by convexity  $v(m^*) \leq$   
830  $v(1) \leq v(m)$  for all  $m$  in  $\text{div}_p(T-K)$ . Therefore  $m_{T-K}^\dagger = 1$ .

831 If  $m^* > (T-K)/(p+1)$  then by definition  $\varphi_{T-K}(m^*) = (T-K)/(p+1)$  and by  
832 convexity  $v(m^*) \leq v((T-K)/(p+1)) \leq v(m)$  for all  $m$  in  $\text{div}_p(T-K)$ . Therefore  
833  $m_{T-K}^\dagger = (T-K)/(p+1)$ .

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

$$835 \quad \varphi_{T-K}(m^*) = \begin{cases} \lrcorner m^* \lrcorner_{T-K} & \text{if } \sqrt{\lrcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}} > m^* \\ \lrcorner m^* \lrcorner_{T-K} & \text{if } \sqrt{\lrcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}} \leq m^*. \end{cases}$$

836

837 By convexity  $m_{T-K}^\dagger$  must be  $\lrcorner m^* \lrcorner_{T-K}$  or  $\lrcorner m^* \lrcorner_{T-K}$ . If  $\lrcorner m^* \lrcorner_{T-K} = \lrcorner m^* \lrcorner_{T-K}$  then  
 838  $m_{T-K}^\dagger = \lrcorner m^* \lrcorner_{T-K} = \varphi_{T-K}(m^*)$ . Otherwise, since  $v(x) = \zeta_1 x + \zeta_2 + \zeta_3/x$ ,  $x > 0$ ,  
 839 for some constants  $\zeta_1, \zeta_2$  and  $\zeta_3$  such that  $\zeta_3/\zeta_1 = m^*$ , we have

$$840 \quad v(\lrcorner m^* \lrcorner_{T-K}) < v(\lrcorner m^* \lrcorner_{T-K}) \text{ iff } \sqrt{\lrcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}} > \frac{\zeta_3}{\zeta_1} = m^*.$$

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

843 Let us prove that the minimizer of  $v(m)$ ,  $m \in \text{div}_p(T-K)$ , is unique if  $m^* \neq$   
 844  $\sqrt{\lrcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}}$ . If it were not, then we would have  $v(\lrcorner m^* \lrcorner_{T-K})$   
 845  $= v(\lrcorner m^* \lrcorner_{T-K})$ . But this implies  $m^* = \sqrt{\lrcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}}$ , which is a contra-  
 846 diction.

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

$$849 \quad \widehat{m}_K^* = \sqrt{\frac{\sum_{j=1}^p \widehat{\zeta}_{3,j}}{\sum_{j=1}^p \widehat{\zeta}_{1,j}}} = \sqrt{\frac{\sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.4)} + \xi_{j;m,i}^{(4.5)} - \xi_{j;m,i}^{(4.6)} - \xi_{j;m,i}^{(4.7)}}{\sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.8)} - \left(\frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.9)}\right)^2}},$$

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

$$853 \quad \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_{m,i},$$

$$854 \quad \xi_{m,i} = (\xi_{1;m,i}^\top, \dots, \xi_{p;m,i}^\top)^\top, \quad i = 1, \dots, n,$$

$$855 \quad \xi_{j;m,i} = (\xi_{j;m,i}^{(4.4)}, \dots, \xi_{j;m,i}^{(4.9)})^\top, \quad j = 1, \dots, p, \quad i = 1, \dots, n.$$

857 Let  $s$  be the function defined by

$$858 \quad s(\mathbf{x}) = \sqrt{\frac{\sum_{j=1}^p x_j^{(4.4)} + x_j^{(4.5)} - x_j^{(4.6)} - x_j^{(4.7)}}{\sum_{j=1}^p x_j^{(4.8)} - x_j^{(4.9)2}},$$

860 where  $\mathbf{x} = (\mathbf{x}_1^\top, \dots, \mathbf{x}_p^\top)^\top$ ,  $\mathbf{x}_j = (x_j^{(4.4)}, \dots, x_j^{(4.9)})^\top$ ,  $j = 1, \dots, p$ . With the above  
 861 notation we have  $\widehat{m}_K^* = s(\bar{\xi})$ . Moreover, elementary calculations show that

$$862 \quad (\text{B.1}) \quad \mathbb{E} \xi_{m,1} = \boldsymbol{\theta} + \sum_{\nu=1}^4 \frac{\mathbf{C}_\nu}{m^\nu},$$

864 where the  $\mathbf{C}_\nu$  are vectors of constants,  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^\top, \dots, \boldsymbol{\theta}_p^\top)^\top$  and

$$865 \quad \boldsymbol{\theta}_j = \mathbb{E} \begin{pmatrix} Y_0^{(1,1)2} Y_j^{(1,1)2} \\ Y_0^{(1,1)} Y_0^{(1,2)} Y_j^{(1,1)} Y_j^{(1,2)} \\ Y_0^{(1,1)} Y_0^{(1,2)} Y_j^{(1,1)2} \\ Y_j^{(1,1)} Y_j^{(1,2)} Y_0^{(1,1)2} \\ Y_0^{(1,1)} Y_0^{(1,2)} Y_j^{(1,1)} Y_j^{(1,2)} \\ Y_j^{(1,1)} Y_0^{(1,1)} \end{pmatrix}.$$

866

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

$$\begin{aligned}
868 \quad & \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1})^\top \dot{s}(\mathbb{E} \boldsymbol{\xi}_{m,1}) + \frac{1}{2}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1})^\top \ddot{s}_{n,m}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1}) \\
869 \quad (\text{B.2}) \quad & = \sqrt{n}(s(\bar{\boldsymbol{\xi}}) - s(\mathbb{E} \boldsymbol{\xi}_{m,1})) \\
870 \quad & = \sqrt{n}(s(\bar{\boldsymbol{\xi}}) - s(\boldsymbol{\theta}) - (\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \dot{s}(\boldsymbol{\theta}) - \frac{1}{2}(\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \ddot{s}_m(\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})), \\
871
\end{aligned}$$

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

$$\begin{aligned}
876 \quad \epsilon_m &= \frac{1}{2}(\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \ddot{s}_m(\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta}) + \sum_{\nu=2}^4 \frac{C_\nu}{m^\nu}, \\
877
\end{aligned}$$

878 it follows from (B.2) that

$$\begin{aligned}
880 \quad (\text{B.3}) \quad & \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1})^\top \dot{s}(\mathbb{E} \boldsymbol{\xi}_{m,1}) + \frac{1}{2}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1})^\top \ddot{s}_{n,m}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1}) \\
881 \quad & = \sqrt{n}(\widehat{m}_K^* - m^* - \frac{C_1}{m} - \epsilon_m). \\
882
\end{aligned}$$

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

$$\begin{aligned}
884 \quad & \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1}) \rightarrow \mathcal{N}(0, \Sigma_m), \\
885
\end{aligned}$$

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

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

$$\begin{aligned}
891 \quad & \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1}) \rightarrow \mathcal{N}(0, \lim_{m \rightarrow \infty} \Sigma_m). \\
892
\end{aligned}$$

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

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

$$\begin{aligned}
898 \quad & \text{div}_p(T-K) \setminus \left\{ 1, \frac{T-K}{p+1} \right\}. \\
899
\end{aligned}$$

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

$$\begin{aligned}
901 \quad & \mathcal{D}_{T-K} \subset \{\sqrt{ij} : i \text{ and } j \text{ are two consecutive integers}\} = \mathcal{E}. \\
902
\end{aligned}$$

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

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

$$910 \quad P \left( \left| \frac{\frac{1}{T-K} v(\widehat{m}_{T-K}^\dagger) - \frac{1}{T} v(m_{T-K}^\dagger)}{\frac{1}{T} v(m_{T-K}^\dagger)} > \varepsilon \right| \right)$$

$$911 \quad \leq P \left( \left| \frac{T}{T-K} (\widehat{m}_{T-K}^\dagger - m_{T-K}^\dagger) v'(\widetilde{m}) + \frac{K}{T-K} v(m_{T-K}^\dagger) \right| > v(m_{T-K}^\dagger) \varepsilon \right),$$

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

$$916 \quad P \left( \widehat{m}_{T-K}^\dagger - m_{T-K}^\dagger \neq 0 \right) + P \left( \frac{K}{T-K} > \varepsilon \right).$$

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

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

$$922 \quad \widehat{S}'_{j;n,m} = \frac{\frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{\text{UL}} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}} \right)^2}{\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{LL}} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}} \right)^2}, \quad j = 1, \dots, p,$$

924 and

$$925 \quad \widehat{S}''_{j;n,m} = \frac{\frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{\text{UL}} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}} \right)^2}{\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{LL}} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}} \right)^2}, \quad j = 1, \dots, p,$$

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

$$929 \quad \mathbb{E} \boldsymbol{\xi}_{m,1} = \begin{pmatrix} \mathbb{E} \mathbb{E}[f(X, Z)|X] \mathbb{E}[f(\widetilde{X}_{-1}, Z)|\widetilde{X}_{-1}] \\ \vdots \\ \mathbb{E} \mathbb{E}[f(X, Z)|X] \mathbb{E}[f(\widetilde{X}_{-p}, Z)|\widetilde{X}_{-p}] \\ \mathbb{E} f(X, Z) \\ \mathbb{E} f(X, Z)^2 \\ \mathbb{E} \mathbb{E}[f(X, Z)|X]^2 + \frac{\mathbb{E} \text{Var}[f(X, Z)|X]}{m} \end{pmatrix}.$$

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

$$931 \quad s(x_1, \dots, x_p, x_{p+1}, x_{p+2}, x_{p+3})$$

$$932 \quad = \left( \frac{x_1 - x_{p+1}^2}{x_{p+2} - x_{p+1}^2}, \dots, \frac{x_p - x_{p+1}^2}{x_{p+2} - x_{p+1}^2}, \frac{x_1 - x_{p+1}^2}{x_{p+3} - x_{p+1}^2}, \dots, \frac{x_p - x_{p+1}^2}{x_{p+3} - x_{p+1}^2} \right).$$

935 Clearly, we have

$$936 \quad s \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} \right) = \begin{pmatrix} \widehat{S}'_{n,m} \\ \widehat{S}''_{n,m} \end{pmatrix}$$

937 and

$$938 \quad s(\mathbf{E} \boldsymbol{\xi}_{m,1}) = \left( \mathbf{S}'' \left[ 1 - \frac{\mathbf{S}'}{\mathbf{E} \operatorname{Var}[f(X,Z)|X] + m \operatorname{Var} \mathbf{E}[f(X,Z)|X]} \right] \right).$$

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

$$940 \quad \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} - \mathbf{E} \boldsymbol{\xi}_{m,1} \right) \xrightarrow{d} \mathcal{N}(0, \Sigma_m),$$

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

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

$$\begin{aligned} 945 \quad & \sqrt{n} \left( s \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} \right) - s(\mathbf{E} \boldsymbol{\xi}_{m,1}) \right) \\ 946 \quad & = \sqrt{n} \left( \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} - \mathbf{E} \boldsymbol{\xi}_{m,1} \right) \dot{s}_m \right. \\ 947 \quad & \quad \left. + \frac{1}{2} \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} - \mathbf{E} \boldsymbol{\xi}_{m,1} \right)^\top \ddot{s}_{n,m} \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} - \mathbf{E} \boldsymbol{\xi}_{m,1} \right) \right), \end{aligned}$$

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

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

$$955 \quad (\text{C.1}) \quad \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i},$$

956 with

$$957 \quad (\text{C.2}) \quad \boldsymbol{\xi}_{m,i} = \prod_{l=1}^L \frac{1}{m} \sum_{k=1}^m \prod_{j=0}^p Y_j^{(i,k) b_{j;l}},$$

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

966 **C.1. Examples.** The estimator

$$967 \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \frac{1}{m} \sum_{k'=1}^m Y_j^{(i,k')}$$

968

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

$$970 \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 971 & 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix},$$

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

$$973 \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)}, \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)2},$$

$$974 \quad \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2$$

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

$$977 \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 978 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 979 & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

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

980 respectively.

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

$$983 \quad \widehat{\zeta}_{3,j} = \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.4)} + \xi_{j;m,i}^{(4.5)} - \xi_{j;m,i}^{(4.6)} - \xi_{j;m,i}^{(4.7)} \quad \text{and,}$$

$$984 \quad \widehat{\zeta}_{1,j} = \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.8)} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.9)} \right)^2,$$

985 where

$$987 \quad \xi_{j;m,i}^{(4.4)}, \quad \xi_{j;m,i}^{(4.5)}$$

$$988 \quad \xi_{j;m,i}^{(4.6)}, \quad \xi_{j;m,i}^{(4.7)}$$

$$989 \quad \xi_{j;m,i}^{(4.8)}, \quad \xi_{j;m,i}^{(4.9)}$$

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

$$992 \quad \begin{pmatrix} 2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix},$$

$$993 \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$994 \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

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

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

$$999 \quad \xi_{1;m,i}^{\text{UL}} \simeq \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \end{pmatrix}, \dots, \xi_{p;m,i}^{\text{UL}} \simeq \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}$$

1000 for the upper left (UL) terms,

$$1001 \quad \xi_{m,i}^{\text{UR}} \simeq \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

1002 for the upper right (UR) term,

$$1003 \quad \xi_{m,i}^{\text{LL}} \simeq \begin{pmatrix} 2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

1004 for the lower left (LL) term of  $\widehat{S}'_{j;n,m}$  and

$$1005 \quad \xi_{m,i}^{\text{LL}} \simeq \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

1006 for the lower left (LL) term of  $\widehat{S}''_{j;n,m}$ .

1007 **C.2. A central limit theorem.** For each  $n$ , the random variables  $\xi_{m,1}, \dots, \xi_{m,n}$   
 1008 are independent and identically distributed. Denote by  $\mathcal{E}_{m,i}(L)$  the set of all sum-  
 1009 mands (C.2). In other words,  $\mathcal{E}_{m,i}(L)$  is the set of all nonnegative matrices of size  
 1010  $L \times (p+1)$ . This set has useful properties, gathered in Proposition C.1 for subsequent  
 1011 use.

1012 **PROPOSITION C.1.** *Let  $\xi$  be an element of  $\mathcal{E}_{m,i}(L)$  with coefficients  $(b_{j;l})$ . The*  
 1013 *following statements are true.*

1014 (i) *If  $\xi'$  is an element of  $\mathcal{E}_{m,i}(L)$  with coefficients  $(b'_{j;l})$  then  $\xi\xi'$  is an element of*  
 1015  *$\mathcal{E}_{m,i}(2L)$  with coefficients*

$$1016 \quad \begin{pmatrix} b_{0;1} & \cdots & b_{p;1} \\ \vdots & & \vdots \\ b_{0;L} & \cdots & b_{p;L} \\ b'_{0;1} & \cdots & b'_{p;1} \\ \vdots & & \vdots \\ b'_{0;L} & \cdots & b'_{p;L} \end{pmatrix}.$$

1018 (ii) *The limit of  $\mathbb{E}\xi$  exists as  $m \rightarrow \infty$ .*

1019 (iii) *If there exists some function  $F$  such that  $|f(x, z)| \leq F(x)$  for all  $x$  and  $z$  in the*  
 1020 *domain of definition of  $f$  then*

$$1021 \quad |\xi| \leq \left( \bigvee_{j=0}^p F_j(\mathbf{X}^{(i)}) \right)^{\sum_{j=0}^p \sum_{l=1}^L b_{j;l}},$$

1022 *where  $F_j(\mathbf{X}^{(i)})$  is  $F(X^{(i)})$  if  $j = 0$  and  $F(\widetilde{X}_{-j}^{(i)})$  if  $j \geq 1$ .*

1023 *Proof.* The proof of (i) is trivial. Let us prove (ii). We have

$$\begin{aligned}
1024 \quad \mathbb{E} \xi &= \frac{1}{m^L} \sum_{(k_1, \dots, k_L) \in \{1, \dots, m\}^L} \mathbb{E} \prod_{l=1}^L \prod_{j=0}^p Y_j^{(1, k_l) b_{j;l}} \\
1025 \quad &= \frac{1}{m^L} \sum_{(k_1, \dots, k_L) \in \{1, \dots, m\}^L} \mathbb{E} \mathbb{E} \left( \prod_{l=1}^L \prod_{j=0}^p Y_j^{(1, k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right) \\
1026 \quad (\text{C.3}) \quad &= \frac{1}{m^L} \sum_{(k_1, \dots, k_L) \in \{1, \dots, m\}^L} \mathbb{E} \prod_{j=0}^p \mathbb{E} \left( \prod_{l=1}^L Y_j^{(1, k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right). \\
1027
\end{aligned}$$

1028 Since (i)  $\mathbf{X}^{(1)}$  and  $\{\mathbf{Z}^{(1, k)}, k = 1, \dots, m\}$  are independent and (ii) the law of

$$1029 \quad (\mathbf{Z}^{(1, k_1)}, \dots, \mathbf{Z}^{(1, k_L)})$$

1030 is invariant through any permutation of distinct  $k_1, \dots, k_L$ , all the inner expectations  
1031 in (C.3) are equal to some others. For if  $k_1, \dots, k_L$  are distinct then

$$1032 \quad \mathbb{E} \left( \prod_{l=1}^L Y_j^{(1, k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right) = \mathbb{E} \left( \prod_{l=1}^L Y_j^{(1, l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right)$$

1033  
1034 for all  $j = 0, \dots, p$ . The number of inner expectations equal to the one above is  
1035  $m(m-1) \cdots (m-L+1)$ , a polynomial in  $m$  with degree  $L$ . If some components of  
1036 the tuple  $(k_1, \dots, k_L)$  are equal, then we can always write

$$1037 \quad \mathbb{E} \left( \prod_{l=1}^L Y_j^{(1, k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right) = \mathbb{E} \left( \prod_{l=1}^{L'} Y_j^{(1, l) \beta_{j;l}} \middle| \mathbf{X}^{(1)} \right)$$

1038  
1039 for some  $L' \leq L$  and coefficients  $\beta_{j;l}$ . It is easy to see that the number of inner expecta-  
1040 tions equal to the one above is a polynomial in  $m$  with degree at most  $L$ . (Looking  
1041 at examples helps to see this; see e.g. the proof of Lemma D.2 in Appendix D.)  
1042 Therefore, the sum in (C.3) is also a polynomial in  $m$  with degree at most  $L$  and the  
1043 claim follows ( $\mathbb{E} \xi$  can be zero). To prove (iii), simply remember that, by assumption,  
1044  $|Y^{(1, k)}| \leq F(X^{(1)})$  and  $|Y_j^{(1, k)}| \leq F(\tilde{X}_{-j}^{(1)})$  for all  $k$  and all  $j$ .  $\square$

1045 Two frameworks are considered:

- 1046 •  $n \rightarrow \infty$  and  $m$  is fixed;
- 1047 •  $n \rightarrow \infty$  and  $m \rightarrow \infty$ .

1048 In the second framework  $m_n$  is a sequence indexed by  $n$  that goes to infinity as  $n$  goes  
1049 to infinity.

1050 LEMMA C.2. Let  $\xi_{m,i}^{(I)}$ ,  $I = 1, \dots, N$ , be elements of  $\mathcal{E}_{m,i}(L)$  with coefficients  
1051  $(b_{j;l}^{(I)})$ . Assume

$$1052 \quad \mathbb{E} F(X^{(1)})^2 \sum_{j=0}^p \sum_{l=1}^L b_{j;l}^{(I)} < \infty$$

1053 for all  $I = 1, \dots, N$ . Let  $n \rightarrow \infty$ . If  $m$  is fixed then

$$1054 \quad \sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{(1)} - \mathbb{E} \xi_{m,1}^{(1)}, \dots, \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{(N)} - \mathbb{E} \xi_{m,1}^{(N)} \right]^\top \xrightarrow{d} \mathcal{N}(0, \Sigma_m),$$

1055



1056 where  $\Sigma_m$  is the variance-covariance matrix of  $\boldsymbol{\xi}_{m,i} = (\xi_{m,i}^{(1)}, \dots, \xi_{m,i}^{(N)})^\top$ . If  $m \rightarrow$   
 1057  $\infty$  then  $\lim_{m \rightarrow \infty} \Sigma_m$  exists elementwise and the above display with  $\Sigma_m$  replaced by  
 1058  $\lim_{m \rightarrow \infty} \Sigma_m$  is true.

1059 *Proof.* Let  $m$  be fixed. By Proposition C.1 (i),  $\xi_{m,i}^{(I)2}$ ,  $I = 1, \dots, N$ , belongs to  
 1060  $\mathcal{E}_{m,i}(2L)$  and has coefficients

$$1061 \quad \xi_{m,i}^{(I)2} \simeq \begin{pmatrix} b_{0;1}^{(I)} & \cdots & b_{p;1}^{(I)} \\ \vdots & & \vdots \\ b_{0;L}^{(I)} & \cdots & b_{p;L}^{(I)} \\ b_{0;1}^{(I)} & \cdots & b_{p;1}^{(I)} \\ \vdots & & \vdots \\ b_{0;L}^{(I)} & \cdots & b_{p;L}^{(I)} \end{pmatrix}.$$

1063 Thus, denoting  $\sum_{j=0}^p \sum_{l=1}^L b_{j;l}^{(I)}$  by  $\beta$ , Proposition C.1 (iii) yields

$$1064 \quad (\text{C.4}) \quad \xi_{m,i}^{(I)2} \leq \bigvee_{j=0}^p F_j(\mathbf{X}^{(i)})^{2\beta}$$

1066 and hence

$$1067 \quad \mathbb{E} \xi_{m,i}^{(I)2} \leq \mathbb{E} \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{2\beta} \leq (p+1) \mathbb{E} \left( F(X^{(1)}) \right)^{2\beta} < \infty.$$

1069 Therefore we can apply the central limit theorem to finish the proof for  $m$  fixed.

1070 Let  $m \rightarrow \infty$ . According to Lindeberg-Feller's central limit theorem (see e.g. [33]),  
 1071 it suffices to show

1072 (i) for all  $\epsilon > 0$ ,

$$1073 \quad \sum_{i=1}^n \mathbb{E} \left\| \frac{1}{\sqrt{n}} \boldsymbol{\xi}_{m,i} \right\|^2 \mathbf{1} \left\{ \left\| \frac{1}{\sqrt{n}} \boldsymbol{\xi}_{m,i} \right\| > \epsilon \right\} \rightarrow 0,$$

1075 and

1076 (ii) the limit  $\sum_{i=1}^n \text{Cov}(\boldsymbol{\xi}_{m,i}/\sqrt{n})$  exists and is finite.

1077 Let us show (i). Denoting  $\mathbf{X} = (X^{(1)}, \tilde{X}^{(1)})$ , we have

$$1078 \quad \sum_{i=1}^n \mathbb{E} \left\| \frac{\boldsymbol{\xi}_{m,i}}{\sqrt{n}} \right\|^2 \mathbf{1} \{ \|\boldsymbol{\xi}_{m,i}\| > \sqrt{n}\epsilon \} = \mathbb{E} \|\boldsymbol{\xi}_{m,1}\|^2 \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \}$$

$$1079 \quad = \mathbb{E} \sum_{I=1}^N \xi_{m,1}^{(I)2} \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \}$$

$$1080 \quad = \sum_{I=1}^N \mathbb{E} \left[ \mathbb{E} \left( \xi_{m,1}^{(I)2} \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \} \mid \mathbf{X} \right) \right].$$

1081

1082 By (C.4), we have

$$\begin{aligned}
1083 \quad \mathbb{E} \left( \xi_{m,1}^{(I)2} \mathbf{1} \{ \|\xi_{m,1}\| > \sqrt{n}\epsilon \} | \mathbf{X} \right) &\leq \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{2\beta} P(\|\xi_{m,1}\| > \sqrt{n}\epsilon | \mathbf{X}) \\
1084 &\leq \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{2\beta} \frac{\sum_{I=1}^N \mathbb{E} \left( \xi_{m,1}^{(I)2} | \mathbf{X} \right)}{n\epsilon^2} \\
1085 &\leq \frac{N \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{4\beta}}{n\epsilon^2}, \\
1086
\end{aligned}$$

1087 where the last inequality holds by using (C.4) once more. The upper bound goes to  
1088 zero and is dominated by an integrable function. Thus, we can apply the dominated  
1089 convergence theorem to complete the proof.

1090 Let us show that (ii) holds. We have  $\sum_{i=1}^n \text{Cov}(\xi_{m,i}/\sqrt{n}) = \text{Cov}(\xi_{m,1})$ . The  
1091 element  $(I, J)$  in this matrix is given by  $\mathbb{E} \xi_{m,1}^{(I)} \xi_{m,1}^{(J)} - \mathbb{E} \xi_{m,1}^{(I)} \mathbb{E} \xi_{m,1}^{(J)}$ . Remember that  
1092  $\mathbb{E} \xi_{m,1}^{(I)2} < \infty$ ,  $I = 1, \dots, N$ , and hence  $\mathbb{E} \xi_{m,1}^{(I)} \xi_{m,1}^{(J)} \leq \mathbb{E} \xi_{m,1}^{(I)2}/2 + \mathbb{E} \xi_{m,1}^{(J)2}/2 < \infty$ . Therefore  
1093 the limit of  $\text{Cov} \xi_{m,1}$  exists and is finite. The proof is complete.  $\square$

1094 **Appendix D. Explicit moment calculations.** Explicit moment calculations  
1095 are given for the summands in the proof of Theorem 5.1. In this section,  $\mathbb{E} f(X, Z)$   
1096 and  $\mathbb{E} \mathbb{E}[f(X, Z) | X]^2$  are denoted by  $\mu$  and  $D$ , respectively. Recall that the upper-left  
1097 term in (3.6) and (3.5) is denoted by  $D_j$ . The moments are given in Lemma D.1  
1098 and Lemma D.2. The variances and covariances are given in Lemma D.3. Let  $\mathbf{X} =$   
1099  $(X^{(1)}, \tilde{X}^{(1)})$ . Whenever there is a superscript  $\mathbf{X}$  added to the expectation symbol  $\mathbb{E}$   
1100 or the variance symbol  $\text{Var}$ , this means that these operators are to be understood  
1101 conditionally on  $\mathbf{X}$ . An integral with respect to  $\mathbf{P}^*(d\mathbf{x})$  means that we integrate with  
1102 respect to the law of  $\mathbf{X}$ .

1103 LEMMA D.1 (Moments of order 1). *The moments of order 1 are given by*

$$\begin{aligned}
1104 \quad \mathbb{E} \xi_{j;m1}^{UL} &= D_j, \\
1105 \quad \mathbb{E} \xi_{m1}^{UR} &= \mu, \\
1106 \quad \mathbb{E} \xi_{m1}^{ULL} &= \frac{1}{m} \mathbb{E} \text{Var}^X f(X^{(1)}, Z^{(1,1)}) + D. \\
1107
\end{aligned}$$

1108

1109 *Proof.* One has

$$\begin{aligned}
1110 \quad \mathbb{E} \xi_{j;m1}^{UL} &= \frac{1}{m^2} \sum_{k,k'} \mathbb{E} f(X^{(1)}, Z^{(1,k)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,k')}) \\
1111 &= \frac{1}{m^2} \sum_{k,k'} \int \mathbb{E} f(x, Z^{(1,k)}) f(\tilde{x}_{-j}, Z_j^{(1,k')}) \mathbf{P}^*(d\mathbf{x}) \\
1112 &= \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1113 &= D_j,
\end{aligned}$$

1115 where the integral is taken with respect to the law of  $\mathbf{x} = (x, \tilde{x})$ , and,

$$\begin{aligned}
1116 \quad \mathbb{E} \xi_{m1}^{\prime\prime LL} &= \frac{1}{m^2} \sum_{k, k'} \mathbb{E} f(X^{(1)}, Z^{(1,k)}) f(X^{(1)}, Z^{(1,k')}) \\
1117 &= \frac{1}{m} \mathbb{E} \text{Var}^X f(X, Z) + \mathbb{E} (\mathbb{E}^X f(X, Z))^2 \\
1118 &= \frac{1}{m} \mathbb{E} \text{Var}^X f(X, Z) + D. \\
1119
\end{aligned}$$

1120 The proof for  $\xi_{m1}^{\text{UR}}$  is similar. □

1121 LEMMA D.2 (Moments of order 2). *The moments of order 2 are given by*

$$\begin{aligned}
1122 \quad \mathbb{E} \xi_{j;m1}^{(UL)2} &= \text{Var} \mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) + D_j^2 \\
1123 &\quad + \frac{1}{m} [\mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1124 &\quad - \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})] \\
1125 &\quad + \frac{1}{m^2} \mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}), \\
1126 \quad \mathbb{E} \xi_{m1}^{(UR)2} &= \frac{1}{m} \mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) + \mathbb{E} (\mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}))^2, \\
1127 \quad \mathbb{E} \xi_{m1}^{\prime\prime (LL)2} &= \frac{m(m-1)(m-2)(m-3)}{m^4} \\
1128 &\quad \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) f(X^{(1)}, Z^{(1,4)}) \\
1129 &\quad + \frac{\binom{4}{2} m(m-1)(m-2)}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) \\
1130 &\quad + \frac{\binom{4}{3} m(m-1)}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^3 f(X^{(1)}, Z^{(1,2)}) \\
1131 &\quad + \frac{m}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^4 \\
1132 &\quad + \frac{\binom{4}{2} m(m-1)/2}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)})^2 \\
1133 \\
1134
\end{aligned}$$

1135 *Proof.* Let us first deal with  $\xi_{j;m1}^{\text{UL}}$ . We have

$$\begin{aligned}
1137 \quad \mathbb{E} \xi_{j;m1}^{\text{UL}2} &= \frac{1}{m^4} \sum_{k_1, k_2, k_3, k_4} \mathbb{E} f(X^{(1)}, Z^{(1, k_1)}) f(X^{(1)}, Z^{(1, k_2)}) \\
1138 &\quad f(\tilde{X}_{-j}^{(1)}, Z_j^{(1, k_3)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1, k_4)})
\end{aligned}$$

1140 where, in the sum, the indices run over  $1, \dots, m$ . We split the sum into four parts.  
1141 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  
1142 part, all the terms are equal to

$$1143 \quad (\text{term 1}) \quad \mathbb{E} \left( \mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \right)^2.$$

1145 The second part contains the  $m^2(m-1)$  terms that satisfy  $k_1 \neq k_2$  and  $k_3 = k_4$  and  
1146 that are equal to

$$1147 \quad (\text{term 2}) \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})^2.$$

1149 The third part contains the  $m^2(m-1)$  terms that satisfy  $k_1 = k_2$  and  $k_3 \neq k_4$  and  
1150 that are equal to

$$1151 \quad (\text{term 3}) \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,2)}).$$

1153 Finally, the fourth part contains the  $m^2$  terms that satisfy  $k_1 = k_2$  and  $k_3 = k_4$  and  
1154 are equal to

$$1155 \quad (\text{term 4}) \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})^2.$$

1157 (One can see that the number of terms is  $m^4$ .) Thus,

$$\begin{aligned} 1158 \quad \mathbb{E} \xi_{m1}^{(\text{UL})2} &= (\text{term 1}) \\ 1159 \quad &+ \frac{1}{m} [(\text{term 2}) + (\text{term 3}) - 2(\text{term 1})] \\ 1160 \quad &+ \frac{1}{m^2} [(\text{term 1}) - (\text{term 2}) - (\text{term 3}) + (\text{term 4})]. \end{aligned}$$

1162 Furthermore,  $[(\text{term 1}) - (\text{term 2}) - (\text{term 3}) + (\text{term 4})]$  is equal to

$$\begin{aligned} 1163 \quad &\int \left( \mathbb{E}^{\mathbf{X}} f(x, Z) f(\tilde{x}_{-j}, Z_j) \right)^2 \\ 1164 \quad &- \mathbb{E}^{\mathbf{X}} f(x, Z^{(1,1)}) f(x, Z^{(1,2)}) f(\tilde{x}_{-j}, Z_j^{(1,1)})^2 \\ 1165 \quad &- \mathbb{E}^{\mathbf{X}} f(x, Z^{(1,1)})^2 f(\tilde{x}_{-j}, Z_j^{(1,1)}) f(\tilde{x}_{-j}, Z_j^{(1,2)}) \\ 1166 \quad &+ \mathbb{E}^{\mathbf{X}} f(x, Z^{(1,1)})^2 f(\tilde{x}_{-j}, Z_j^{(1,1)})^2 d\mathbf{P}^*(\mathbf{x}) \\ 1167 \quad &= \int \left( \mathbb{E}^{\mathbf{X}} f(x, Z) \right)^2 \left( \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j) \right)^2 \\ 1168 \quad &- \left( \mathbb{E}^{\mathbf{X}} f(x, Z) \right)^2 \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)^2 \\ 1169 \quad &- \mathbb{E}^{\mathbf{X}} f(x, Z)^2 \left( \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j) \right)^2 \\ 1170 \quad &+ \mathbb{E}^{\mathbf{X}} f(x, Z)^2 \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)^2 d\mathbf{P}^*(\mathbf{x}) \\ 1171 \quad &= \int \text{Var}^{\mathbf{X}} f(X, Z) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}, Z_j) d\mathbf{P}^*(\mathbf{x}). \\ 1172 \end{aligned}$$

1173 Likewise, we find that  $[(\text{term 2}) + (\text{term 3}) - 2(\text{term 1})]$  is equal to

$$1174 \quad \mathbb{E} \text{Var}^{\mathbf{X}} f(X, Z) f(\tilde{X}_{-j}, Z_j) - \text{Var}^{\mathbf{X}} f(X, Z) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}, Z_j),$$

1176 and term 1 is  $\text{Var} \mathbb{E}^{\mathbf{X}} f(X, Z) f(\tilde{X}_{-j}, \tilde{Z}) + D_j^2$ .

1177 We now deal with  $\xi_{m1}^{\text{LL}}$ . We have

$$\begin{aligned} 1179 \quad \mathbb{E} \xi_{m1}^{\text{LL}2} &= \frac{1}{m^4} \sum_{k_1, k_2, k_3, k_4} \mathbb{E} f(X^{(1)}, Z^{(1, k_1)}) f(X^{(1)}, Z^{(1, k_2)}) \\ 1180 \quad &f(X^{(1)}, Z^{(1, k_3)}) f(X^{(1)}, Z^{(1, k_4)}). \end{aligned}$$

1182 The sum is split into five parts. The first part consists of the  $m(m-1)(m-2)(m-3)$   
1183 terms with different indices; those terms are equal to

$$1184 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) f(X^{(1)}, Z^{(1,4)}).$$

1185 The second part consists of the  $\binom{4}{2}m(m-1)(m-2)$  terms with exactly two equal  
 1186 indices; those terms are equal to

$$1187 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}).$$

1188 The third part consists of the  $\binom{4}{3}m(m-1)$  terms with exactly three equal indices;  
 1189 those terms are equal to

$$1190 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^3 f(X^{(1)}, Z^{(1,2)}).$$

1191 The fourth part consists of the  $m$  terms with exactly four equal indices; those terms  
 1192 are equal to

$$1193 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^4.$$

1194 The fifth and last part consists of the  $\binom{4}{2}m(m-1)/2$  terms with exactly two pairs of  
 1195 equal indices; those terms are equal to

$$1196 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)})^2.$$

1197 (One can check that the total number of terms is  $m^4$ .)

□

1198 LEMMA D.3 (Variances and covariances).

$$\begin{aligned}
1199 \quad (i) \quad & \text{Var } \xi_{m1}^{UL} = \text{Var } \mathbf{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1200 & + \frac{1}{m} [\mathbf{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1201 & - \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})] \\
1202 & + \frac{1}{m^2} \mathbf{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}), \\
1203 \quad (ii) \quad & \text{Cov}(\xi_{m1}^{UL}, \xi_{m1}^{UR}) = \frac{m-1}{m} \mathbf{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1204 & + \frac{1}{m} \mathbf{E} f(X^{(1)}, Z^{(1,1)})^2 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) - D_j \mu \\
1205 \quad (iii) \quad & \text{Cov}(\xi_{m1}^{UL}, f(X, Z)^2) = \frac{1}{m} \mathbf{E} f(X^{(1)}, Z^{(1,1)})^3 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1206 \quad (iii) \quad & + \frac{m-1}{m} \mathbf{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) - D_j \kappa \\
1207 \quad (iv) \quad & \text{Var } \xi_{m1}^{UR} = \frac{1}{m} \text{Var } f(X, Z) \\
1208 \quad (v) \quad & \text{Cov}(\xi_{m1}^{UR}, f(X, Z)^2) = \frac{1}{m} f(X, Z)^3 \\
1209 & + \frac{m-1}{m} \mathbf{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) - \mu \kappa \\
1210 \quad (vi) \quad & \text{Cov}(\xi_{m_{n1}}^{UL}, \xi_{m_{n1}}^{ULL}) = \frac{m}{m^3} \mathbf{E} f(X^{(1)}, Z^{(1,1)})^3 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1211 & + \frac{3m(m-1)}{m^3} \mathbf{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1212 & + \frac{m(m-1)(m-2)}{m^3} \mathbf{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) \\
1213 & f(X^{(1)}, Z^{(1,3)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1214 & - \mathbf{E} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1215 & \left\{ \frac{1}{m} \mathbf{E} f(X^{(1)}, Z^{(1,1)})^2 + \frac{m-1}{m} \mathbf{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) \right\} \\
1216 & \\
1217 &
\end{aligned}$$

1218 *Proof.* The proof follows from direct calculations.  $\square$

### 1219 Appendix E. Calculations for the linear model.

1220 LEMMA E.1. Suppose that  $f(X, Z) = \beta_0 + \beta_{p+1}Z + \sum_{j=1}^p \beta_j X_j$  where  $X =$   
1221  $(X_1, \dots, X_p)$ ,  $Z_k, \tilde{Z}_{ik}$  are independent,  $\mathbf{E} X_j = \mathbf{E} Z = 0$ ,  $\mathbf{E} X_j^2 = \mathbf{E} Z^2 = 1$ ,  $\mathbf{E} X_j^3 = 0$ ,  
1222  $\mathbf{E} X_j^4 = 3$ . Then the squared optimal number of repetitions is given by

$$1223 \quad (m_i^*)^2 = \frac{\beta_{p+1}^4}{(\beta_0 + \beta_i)^2 - 2\beta_0^4 + (\sum_{j=0}^p \beta_j^2)^2}$$

1224 and the discriminator (the upper-left term in (3.6) and (3.5)) is

$$1225 \quad \beta_0^2 + \beta_i^2.$$

1226

1227 *Proof.* We have

$$1228 \quad m_i^* = \frac{A_i + B_i + C_i + D_i}{E_i},$$

1229 with

$$\begin{aligned} 1230 \quad A_i &= \mathbf{E} f(X, Z_1)^2 f(\tilde{X}_{-i}, \tilde{Z}_{i1})^2 \\ 1231 \quad B_i &= \mathbf{E} f(X, Z_1) f(\tilde{X}_{-i}, \tilde{Z}_{i1}) f(X, Z_2) f(\tilde{X}_{-i}, \tilde{Z}_{i2}) \\ 1232 \quad C_i &= -\mathbf{E} f(X, Z_1)^2 f(\tilde{X}_{-i}, \tilde{Z}_{i1}) f(\tilde{X}_{-i}, \tilde{Z}_{i2}) \\ 1233 \quad D_i &= -\mathbf{E} f(\tilde{X}_{-i}, \tilde{Z}_{i1})^2 f(X, Z_1) f(X, Z_2) \\ 1234 \quad E_i &= B - [\mathbf{E} f(X, Z_1) f(\tilde{X}_{-i}, \tilde{Z}_{i1})]^2 \end{aligned}$$

1236 where  $X = (X_1, \dots, X_p)$ ,  $Z_k, \tilde{Z}_{ik}$  are independent,  $\mathbf{E} X_j = \mathbf{E} Z = 0$ ,  $\mathbf{E} X_j^2 = \mathbf{E} Z^2 = 1$ ,  
1237  $\mathbf{E} X_j^3 = 0$ ,  $\mathbf{E} X_j^4 = 3$ . We deal with the case

$$1238 \quad f(X, Z) = \beta_0 + \beta_{p+1} Z + \sum_{j=1}^p \beta_j X_j.$$

1239 We calculate the terms one by one as follows. We have

$$\begin{aligned} 1240 \quad A_j &= \mathbf{E} \left( \beta_0 + \sum_{j=1}^p \beta_j X_j \right)^2 \left( \beta_0 + \beta_i X_i + \sum_{j:1 \leq j \neq i} \beta_j \tilde{X}_j \right)^2 \\ 1241 \quad &+ \left( \beta_0 + \sum_{j=1}^p \beta_j X_j \right)^2 \beta_{p+1}^2 \tilde{Z}_{i1}^2 + \beta_{p+1}^4 Z_1^2 \tilde{Z}_{i1}^2 \\ 1242 \quad &+ \beta_{p+1}^2 Z_1^2 \left( \beta_0 + \beta_i X_i + \sum_{j:1 \leq j \neq i} \beta_j \tilde{X}_j \right)^2 \\ 1243 \quad &= A_{j1} + A_{j2} + A_{j3}, \end{aligned}$$

1245 where  $\mathbf{E} (A2) = \beta_{p+1}^4 + \beta_{p+1}^2 \sum_{j=0}^p \beta_j^2$ ,  $\mathbf{E} (A3) = \beta_{p+1}^2 \sum_{j=0}^p \beta_j^2$ . Elementary but some-  
1246 what tedious calculations yield

$$1247 \quad \mathbf{E} (A1) = \beta_0^4 + 3\beta_i^4 + 6\beta_0^2 \beta_i^2 + 2(\beta_0^2 + \beta_i^2) \sum_{j:1 \leq j \neq i} \beta_j^2 + \left( \sum_{j:1 \leq j \neq i} \beta_j^2 \right)^2.$$

1249 Similar calculations show that  $B_j = A_{j1}$ ,  $C_j = -A_{j1} - A_{j3}$ ,  $D_j = -A_{j1} - A_{j3}$ ,  
1250  $E_j = A_{j1} - (\beta_0^2 + \beta_i^2)^2$ . Thus,

$$1251 \quad (m_i^*)^2 = \frac{\beta_{p+1}^4}{(\beta_0 + \beta_i)^2 - 2\beta_0^4 + (\sum_{j=0}^p \beta_j^2)^2}. \quad \square$$

1252 **Acknowledgments.** The author thanks two anonymous referees who made use-  
1253 ful suggestions. The author thanks Hervé Monod and Elisabeta Vergu for stimulating  
1254 discussions from which this work arose and Bertrand Iooss for the references [10, 13].

1255

## REFERENCES

- 1256 [1] S. AZZI, B. SUDRET, AND J. WIART, *Sensitivity analysis for stochastic simulators using differ-*  
 1257 *ential entropy*, International Journal for Uncertainty Quantification, 10 (2020).
- 1258 [2] A. COURCOUL, H. MONOD, M. NIELEN, D. KLINKENBERG, L. HOGERWERF, F. BEAUDEAU, AND  
 1259 E. VERGU, *Modelling the effect of heterogeneity of shedding on the within herd coxiella*  
 1260 *burnetii spread and identification of key parameters by sensitivity analysis*, Journal of  
 1261 Theoretical Biology, 284 (2011), pp. 130–141.
- 1262 [3] D. J. DALEY AND J. GANI, *Epidemic Modelling*, Cambridge, 1999.
- 1263 [4] P. ETORÉ, C. PRIEUR, D. K. PHAM, AND L. LI, *Global sensitivity analysis for models described*  
 1264 *by stochastic differential equations*, Methodology and Computing in Applied Probability,  
 1265 (2019), pp. 1–29.
- 1266 [5] J.-C. FORT, T. KLEIN, AND A. LAGNOUX, *Global sensitivity analysis and Wasserstein spaces*,  
 1267 arXiv preprint arXiv:2007.12378, (2020).
- 1268 [6] F. GAMBOA, A. JANON, T. KLEIN, AND A. LAGNOUX, *Sensitivity analysis for multidimensional*  
 1269 *and functional outputs*, Electron. J. Stat., 8 (2014), pp. 575–603, [https://doi.org/10.1214/](https://doi.org/10.1214/14-EJS895)  
 1270 [14-EJS895](https://doi.org/10.1214/14-EJS895).
- 1271 [7] F. GAMBOA, A. JANON, T. KLEIN, A. LAGNOUX, AND C. PRIEUR, *Statistical inference for Sobol*  
 1272 *pick-freeze Monte Carlo method*, Statistics, 50 (2016), pp. 881–902.
- 1273 [8] F. GAMBOA, T. KLEIN, AND A. LAGNOUX, *Sensitivity analysis based on cramér–von mises*  
 1274 *distance*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 522–548.
- 1275 [9] F. GAMBOA, T. KLEIN, A. LAGNOUX, AND L. MORENO, *Sensitivity analysis in general metric*  
 1276 *spaces*, arXiv preprint arXiv:2002.04465, (2020).
- 1277 [10] J. L. HART, A. ALEXANDERIAN, AND P. A. GREMAUD, *Efficient computation of Sobol’indices*  
 1278 *for stochastic models*, SIAM Journal on Scientific Computing, 39 (2017), pp. A1514–A1530.
- 1279 [11] W. HOEFFDING, *A class of statistics with asymptotically normal distribution*, The Annals of  
 1280 Mathematical Statistics, 19 (1948), pp. 293–325.
- 1281 [12] T. HOMMA AND A. SALTELLI, *Importance measures in global sensitivity analysis of non-*  
 1282 *linear models*, Reliability Engineering & System Safety, 52 (1996), pp. 1–17, [https://doi.org/https://doi.org/10.1016/0951-8320\(96\)00002-6](https://doi.org/https://doi.org/10.1016/0951-8320(96)00002-6), [http://www.sciencedirect.com/](http://www.sciencedirect.com/science/article/pii/0951832096000026)  
 1283 [science/article/pii/0951832096000026](http://www.sciencedirect.com/science/article/pii/0951832096000026).
- 1284 [13] B. IOOSS, L. L. GRATIET, A. LAGNOUX, AND T. KLEIN, *Sensitivity analysis for stochastic*  
 1285 *computer codes: Theory and estimation methods*, tech. report, EDF R&D, 2014.
- 1286 [14] B. IOOSS, T. KLEIN, AND A. LAGNOUX, *Sobol’ sensitivity analysis for stochastic numerical*  
 1287 *codes*, in 8th International Conference Sensitivity Analysis of Model Output, 2016, pp. 48–  
 1288 49.
- 1289 [15] B. IOOSS AND M. RIBATET, *Global sensitivity analysis of computer models with functional in-*  
 1290 *puts*, Reliability Engineering & System Safety, 94 (2009), pp. 1194 – 1204, [https://doi.org/](https://doi.org/https://doi.org/10.1016/j.res.2008.09.010)  
 1291 <https://doi.org/10.1016/j.res.2008.09.010>, [http://www.sciencedirect.com/science/article/](http://www.sciencedirect.com/science/article/pii/S0951832008002299)  
 1292 [pii/S0951832008002299](http://www.sciencedirect.com/science/article/pii/S0951832008002299).
- 1293 [16] B. IOOSS, S. D. VEIGA, A. JANON, G. PUJOL, WITH CONTRIBUTIONS FROM BAPTISTE BROTO,  
 1294 K. BOUMHAOUT, T. DELAGE, R. E. AMRI, J. FRUTH, L. RENT GILQUIN, J. GUILLAUME,  
 1295 L. L. GRATIET, P. LEMAITRE, A. MARREL, A. MEY-NAOUI, B. L. NELSON, F. IPPO  
 1296 MONARI, R. OOMEN, O. RAKOVEC, B. RAMOS, O. ROUSTANT, E. HYE SONG, J. STAUM,  
 1297 R. SUEUR, T. TOUATI, AND F. WEBER, *sensitivity*, 2020, [https://cran.r-project.org/](https://cran.r-project.org/package=sensitivity)  
 1298 [package=sensitivity](https://cran.r-project.org/package=sensitivity). R package version 1.23.0.
- 1299 [17] A. JANON, T. KLEIN, A. LAGNOUX, M. NODET, AND C. PRIEUR, *Asymptotic normality and*  
 1300 *efficiency of two Sobol index estimators*, ESAIM: Probability and Statistics, 18 (2014),  
 1301 pp. 342–364, <https://doi.org/10.1051/ps/2013040>.
- 1302 [18] M. LAMBONI, H. MONOD, AND D. MAKOWSKI, *Multivariate sensitivity analysis to measure*  
 1303 *global contribution of input factors in dynamic models*, Reliability Engineering and System  
 1304 Safety, 96 (2010), pp. 450–459.
- 1305 [19] A. MARREL, B. IOOSS, S. DA VEIGA, AND M. RIBATET, *Global sensitivity analysis of stochastic*  
 1306 *computer models with joint metamodels*, Statistics and Computing, 22 (2012), pp. 833–847,  
 1307 <https://doi.org/10.1007/s11222-011-9274-8>, <https://doi.org/10.1007/s11222-011-9274-8>.
- 1308 [20] A. MARREL, B. IOOSS, B. LAURENT, AND O. ROUSTANT, *Calculations of sobol indices for*  
 1309 *the gaussian process metamodel*, Reliability Engineering & System Safety, 94 (2009),  
 1310 pp. 742 – 751, <https://doi.org/https://doi.org/10.1016/j.res.2008.07.008>, [http://www.](http://www.sciencedirect.com/science/article/pii/S0951832008001981)  
 1311 [sciencedirect.com/science/article/pii/S0951832008001981](http://www.sciencedirect.com/science/article/pii/S0951832008001981).
- 1312 [21] H. MONOD, C. NAUD, AND D. MAKOWSKI, *Working with dynamic crop models: Evaluation,*  
 1313 *analysis, parameterization, and applications*, in Uncertainty and sensitivity analysis for  
 1314 crop models, Elsevier, 2006, pp. 55–100.
- 1315



- 1316 [22] M. NAVARRO JIMENEZ, O. P. LE MAÎTRE, AND O. M. KNIO, *Global sensitivity analysis in*  
1317 *stochastic simulators of uncertain reaction networks*, The Journal of Chemical Physics, 145  
1318 (2016), p. 244106, <https://doi.org/10.1063/1.4971797>, <https://doi.org/10.1063/1.4971797>,  
1319 <https://arxiv.org/abs/https://doi.org/10.1063/1.4971797>.
- 1320 [23] C. PRIEUR AND S. TARANTOLA, *Variance-based sensitivity analysis: theory and estimation*  
1321 *algorithms*, in Handbook of Uncertainty Quantification, Springer, 2015, pp. 1–23.
- 1322 [24] L. RIMBAUD, C. BRUCHOU, S. DALLOT, D. R. J. PLEYDELL, E. JACQUOT, S. SOUBEYRAND,  
1323 AND G. THÉBAUD, *Using sensitivity analysis to identify key factors for the propaga-*  
1324 *tion of a plant epidemic*, Open Science, 5 (2018), <https://doi.org/10.1098/rsos.171435>,  
1325 <http://rsos.royalsocietypublishing.org/content/5/1/171435>, [https://arxiv.org/abs/http://](https://arxiv.org/abs/http://rsos.royalsocietypublishing.org/content/5/1/171435.full.pdf)  
1326 [rsos.royalsocietypublishing.org/content/5/1/171435.full.pdf](http://rsos.royalsocietypublishing.org/content/5/1/171435.full.pdf).
- 1327 [25] A. SALTELLI AND I. M. SOBOL', *About the use of rank transformation in sensitivity analysis*  
1328 *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](https://doi.org/https://doi.org/10.1016/0951-8320(95)00099-2), [http://www.sciencedirect.com/](http://www.sciencedirect.com/science/article/pii/0951832095000992)  
1329 [science/article/pii/0951832095000992](http://www.sciencedirect.com/science/article/pii/0951832095000992).
- 1330 [26] A. SALTELLI, S. TARANTOLA, AND F. CAMPOLONGO, *Sensitivity analysis as an ingredient of*  
1331 *modeling*, Statistical Science, 15 (2000), pp. 377–395.
- 1332 [27] A. SALTELLI, S. TARANTOLA, F. CAMPOLONGO, AND M. RATTO, *Sensitivity analysis in practice*,  
1333 Wiley, 2004.
- 1334 [28] J. F. SAVALL, C. BIDOT, M. LEBLANC-MARIDOR, C. BELLOC, AND S. TOUZEAU, *Modelling*  
1335 *salmonella transmission among pigs from farm to slaughterhouse: interplay between man-*  
1336 *agement variability and epidemiological uncertainty*, International Journal of Food Micro-  
1337 biology, 229 (2016), pp. 33–43.
- 1338 [29] I. M. SOBOL, *Sensitivity estimates for nonlinear mathematical models*, Mathematical modelling  
1339 and computational experiments, 1 (1993), pp. 407–414.
- 1340 [30] I. M. SOBOL, *Global sensitivity indices for nonlinear mathematical models and their Monte*  
1341 *Carlo estimates*, Mathematics and computers in simulation, 55 (2001), pp. 271–280.
- 1342 [31] M. SPENCE, *Statistical issues in ecological simulation models*, PhD thesis, University of  
1343 Sheffield, 2015. <http://etheses.whiterose.ac.uk/10517>.
- 1344 [32] L. SZEWCZYK, J.-L. GRIMAUD, AND I. COJAN, *Experimental evidence for bifurcation angles*  
1345 *control on abandoned channel fill geometry*, Earth Surface Dynamics, 8 (2020), pp. 275–288,  
1346 <https://doi.org/10.5194/esurf-8-275-2020>, <https://hal.archives-ouvertes.fr/hal-02562670>.
- 1347 [33] A. W. VAN DER VAART, *Asymptotic Statistics*, Cambridge University Press, 1998.
- 1348 [34] X. ZHU AND B. SUDRET, *Global sensitivity analysis for stochastic simulators based on general-*  
1349 *ized lambda surrogate models*, arXiv preprint arXiv:2005.01309, (2020).
- 1350