



**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-02113448v5

**HAL Id: hal-02113448**

**<https://hal.science/hal-02113448v5>**

Preprint submitted on 7 May 2021 (v5), 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 a method that exploits the availability of softwares for deterministic models. For each  
9 input, the stochastic model is repeated and the outputs are averaged. These averages are seen as if  
10 they came from a deterministic model and hence Sobol's method can be used. However, in the context  
11 of limited computational resources, one must ensure that the number of repetitions of the stochastic  
12 model multiplied by the number of explorations of the input space is less than a fixed threshold. The  
13 problem of finding an optimal tradeoff between the repetitions and the explorations is addressed. A  
14 bound on an error criterion that penalizes bad rankings of the inputs' sensitivities is minimized. The  
15 estimator induced by the empirical method described above is shown to be asymptotically biased if  
16 the number of repetitions goes to infinity too slowly. A functional relationship between the output,  
17 the input and some random noise is derived from minimal distributional assumptions, which leads  
18 to a new sensitivity index with better statistical properties. The theory is illustrated on numerical  
19 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

35 (1.1) 
$$f(X) - f_0 = f_1(X_1) + \dots + f_p(X_p)$$

36 
$$+ f_{1,2}(X_1, X_2) + \dots + f_{p-1,p}(X_{p-1}, X_p)$$

37 
$$+ \dots$$

38 
$$+ f_{1,\dots,p}(X_1, \dots, X_p),$$

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

---

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

<sup>†</sup>MaIAGE, 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 of the individual components in the right-hand side of (1.1) are zero and hence we  
45 have the variance decomposition

$$46 \quad \text{Var } f(X) = \text{Var } f_1(X_1) + \cdots + \text{Var } f_p(X_p) + \cdots ,$$

48 which leads to the so-called Sobol indices

$$49 \quad S_j = \frac{\text{Var } f_j(X_j)}{\text{Var } f(X)} = \frac{\text{Var } \mathbb{E}(f(X)|X_j)}{\text{Var } f(X)} = \frac{\mathbb{E} f(X)f(\tilde{X}_{-j}) - (\mathbb{E} f(X))^2}{\mathbb{E} f(X)^2 - (\mathbb{E} f(X))^2},$$

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

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

$$61 \quad S_{T_j} = 1 - \frac{\text{Var } \mathbb{E}(f(X)|X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p)}{\text{Var } f(X)}.$$

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

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

$$72 \quad (1.2) \quad \tilde{X}_{-j}^{(i)} := (\tilde{X}_1^{(i)}, \dots, \tilde{X}_{j-1}^{(i)}, X_j^{(i)}, \tilde{X}_{j+1}^{(i)}, \dots, \tilde{X}_p^{(i)}).$$

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

$$77 \quad (1.3) \quad \hat{S}_j = \frac{\frac{1}{n} \sum_{i=1}^n Y^{(i)} Y_j^{(i)} - \left(\frac{1}{n} \sum_{i=1}^n Y^{(i)}\right)^2}{\frac{1}{n} \sum_{i=1}^n Y^{(i)2} - \left(\frac{1}{n} \sum_{i=1}^n Y^{(i)}\right)^2}.$$

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

85 The big difference between a deterministic model and a stochastic model is that  
86 the stochastic model is not a function anymore. To a particular value of the input there

87 does not correspond any particular value for the output. Instead, there corresponds a  
 88 range of possible values, assumed to come from a probability distribution depending  
 89 on the input. Examples can be found in epidemiology [2, 3, 24, 28] or ecology [31], to  
 90 name a few.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

226 **2.2. Links with the stochastic models and the sensitivity indices in [10].**  
 227 In [10], the authors consider the model  $(X'(\omega'), \varphi(X'(\omega'), \omega''))$ ,  $\omega' \in \Omega'$ ,  $\omega'' \in \Omega''$ ,

228 where  $(\Omega', \mathcal{F}', P')$  and  $(\Omega'', \mathcal{F}'', P'')$  are probability spaces,  $X' = (X'_1, \dots, X'_p)$  is a  
 229 random vector on  $\Omega'$  and  $\varphi$  is some function. They consider the sensitivity indices

$$230 \quad S_j^{\text{HAG}} = \int_{\Omega''} S_j(\omega'') P''(d\omega''),$$

232 where

$$233 \quad S_j(\omega'') = \frac{\text{Var}[\mathbb{E}(\varphi(X', \omega'') | X'_j)]}{\text{Var}[\varphi(X', \omega'')]},$$

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

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

$$241 \quad \tilde{S}_j^{\text{HAG}} = \mathbb{E} \left( \frac{\text{Var}(\mathbb{E}[f(X, Z) | X_j, Z] | Z)}{\text{Var}(f(X, Z) | Z)} \right), \quad j = 1, \dots, p.$$

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

$$244 \quad (2.1) \quad \varphi(X'(\omega'), \omega'') = X_1(\omega') + X_2(\omega')\omega'',$$

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

$$249 \quad S_1^{\text{HAG}} = \int_{\Omega''} \frac{1}{1 + L^2\omega''} P''(d\omega'') = \int_{\mathbb{R}} \frac{1}{1 + L^2w} \exp\left(-\frac{w^2}{2}\right) \frac{1}{\sqrt{2\pi}} dw$$

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

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

254 **EXAMPLE 2.** *Let us first extract the induced complete stochastic model. Set  $P^*((0, t_1] \times$   
 255  $(1, t_2]) = t_1(t_2 - 1)/L$  for all  $0 < t_1 < 1$ ,  $1 < t_2 < L + 1$ ,  $L > 0$  and  $Q_x(-\infty, t] =$   
 256  $\Phi((t - x_1)/x_2)$  for all  $t \in \mathbb{R}$ , where  $\Phi(t) = \int_{-\infty}^t (2\pi)^{-1/2} e^{-s^2/2} ds$  and  $x = (x_1, x_2) \in$   
 257  $\mathbb{R} \times (0, \infty)$ . Now it remains to choose a representation that fulfills the conditions in  
 258 Lemma 2.2 and ensures that  $S_1^{\text{HAG}} = \tilde{S}_1^{\text{HAG}}$ . Such a representation can easily be  
 259 found. For instance, take  $\Omega = (0, 1)^3$  endowed with the product Lebesgue measure and  
 260 put  $Z(\omega) = \omega_3$ ,  $X_1(\omega) = F_1^{-1}(\omega_1)$  and  $X_2(\omega) = F_2(\omega_2)^{-1}$  for  $\omega = (\omega_1, \omega_2, \omega_3) \in \Omega$ ,  
 261 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  
 262 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  
 263 of Lemma 2.2 are fulfilled by construction and the detailed calculations in Appendix A  
 264 show that  $S_1^{\text{HAG}} = \tilde{S}_1^{\text{HAG}}$ .*

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



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

$$278 \quad (3.1) \quad |f(X, Z)| \leq F(X)$$

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

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

288 **3.1.1. Indices of the first kind.** Applying the Sobol-Hoeffding decomposition  
 289 to  $f$  yields

$$290 \quad (3.2) \quad f(X, Z) - f_0 = f_1(X_1) + \cdots + f_p(X_p) + f_{p+1}(Z) + \cdots,$$

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

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

$$298 \quad S'_j = \frac{\text{Var } E(f(X, Z) | X_j)}{\text{Var } f(X, Z)}, \quad j = 1, \dots, p.$$

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

303 **3.1.2. Indices of the second kind.** Let  $g(X) := E[f(X, Z) | X]$  be the condi-  
 304 tional expectation of the output of the stochastic model given the input. The object  
 305  $g$  is a function and the Sobol-Hoeffding decomposition can be applied to it, yielding

$$306 \quad g(X) - g_0 = g_1(X_1) + \cdots + g_p(X_p) + \cdots,$$

308 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  
 309 interaction terms. Since the components of  $X$  are independent, we have  $\text{Var } g(X) =$   
 310  $\text{Var } g_1(X_1) + \cdots + \text{Var } g_p(X_p) + \cdots$ , leading to the indices in Definition 3.2.

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

$$313 \quad S''_j = \frac{\text{Var } E(g(X) | X_j)}{\text{Var } g(X)} = \frac{\text{Var } E(E[f(X, Z) | X] | X_j)}{\text{Var } E[f(X, Z) | X]}, \quad j = 1, \dots, p.$$



314 The total sensitivity indices, defined by

$$315 \quad (3.3) \quad S''_{Tj} = 1 - \frac{\text{Var E}(g(X)|X_1, \dots, X_{j-1}, X_{j+1}, \dots, X_p)}{\text{Var } g(X)},$$

316

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

318 **3.1.3. Comparison of the definitions.** The sensitivity indices of the first kind  
 319 provide more refined “first-order” information than the indices of the second kind.  
 320 Example 3 and 4 illustrate this.

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

$$324 \quad S'_1 = \frac{a^2}{a^2 + c^2 E\phi(Z)^2}, S'_2 = 0, S''_1 = 1 \text{ and } S''_2 = 0.$$

325

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

334 Example 4 returns to the model (2.1).

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

$$340 \quad S'_1 = \frac{1}{4(L^2 + 3(L + 1)) + 1}.$$

341

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

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

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

358 The couple  $(n, m)$  is called the *design* of the Monte-Carlo sampling scheme. The  
 359 total number of calls to the stochastic model is  $mn(p + 1)$ . The details are given in  
 360 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**

---

361 In the algorithm above,  $\tilde{X}_{-0}^{(i)} = X^{(i)}$  by convention. By assumption, the objects  
 362  $\tilde{X}^{(i)}$ ,  $\tilde{X}_{-j}^{(i)}$  and  $Y_j^{(i,k)}$ ,  $j = 0, \dots, p$ ,  $k = 1, \dots, m$ ,  $i = 1, \dots, n$ , are random vectors  
 363 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  
 364 i.i.d.,  $X^{(i)}$  and  $\tilde{X}^{(i)}$  are independent and  $P(\cap_{j=0}^p \cap_{k=1}^m \{Y_j^{(i,k)} \in B_j^{(k)}\} | X^{(i)}, \tilde{X}^{(i)}) =$   
 365  $\prod_{j=0}^p \prod_{k=1}^m P(Y_j^{(i,k)} \in B_j^{(k)} | X^{(i)}, \tilde{X}^{(i)})$  for all Borel sets  $B_j^{(k)} \in \mathcal{B}$ . It is easy to see that  
 366 these conditions characterize the joint law of the set  $\{\tilde{X}_{-j}^{(i)}, Y_j^{(i,k)} : j = 0, \dots, p; k =$   
 367  $1, \dots, m; i = 1, \dots, n\}$ , that is, the inputs and the outputs of Algorithm 3.1.

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

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

$$375 \quad (3.4) \quad \hat{S}_{j;n,m}'' = \frac{\frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \frac{1}{m} \sum_{k'=1}^m Y_j^{(i,k')} - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2}{\frac{1}{n} \sum_{i=1}^n \left( \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2 - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2},$$

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

$$377 \quad S_j'' = \frac{\mathbb{E} g(X^{(1)}) g(\tilde{X}_{-j}^{(1)}) - (\mathbb{E} g(X^{(1)}))^2}{\mathbb{E} g(X^{(1)})^2 - (\mathbb{E} g(X^{(1)}))^2}$$

$$(3.5)$$

$$378 \quad = \frac{\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}] \mathbb{E}[f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) | \tilde{X}_{-j}^{(1)}] - (\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}])^2}{\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}]^2 - (\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}])^2}.$$

379

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

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

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

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

$$393 \quad (3.7) \quad \hat{S}'_{j;n,m} = \frac{\frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \frac{1}{m} \sum_{k'=1}^m Y_j^{(i,k')} - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2}{\frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)^2} - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2}.$$

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

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

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

408	$n$	50	25	10	5	2	1
	$m$	1	2	5	10	25	50.

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

415 **4.1. Introducing the miss-ranking error and its bound.** To compare the  
 416 estimators, a measure of performance has to be defined. We shall consider the miss-  
 417 ranking error (MRE), defined by

$$418 \quad \text{MRE} = \mathbb{E} \sum_{j=1}^p |\hat{R}_{j;n,m} - R_j|,$$

419 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  
 420  $\widehat{R}_{j;n,m}$  is an estimator of  $R_j$ . Recall that  $D_1, \dots, D_p$  are the upper-left terms in (3.6)  
 421 and (3.5). They determine the ranks of the sensitivity indices. Recall that the ranks of  
 422 the sensitivity indices of the first kind coincide with the ranks of the sensitivity indices  
 423 of the second kind. Thus, the MRE permits to find a unique solution for both kinds  
 424 of sensitivity indices. The MRE is small when one succeeds in ranking the inputs  
 425 from the most to the least important, a task which is called ‘‘factors prioritization’’  
 426 in [27, p. 52].

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

431 **PROPOSITION 4.1.** *Let  $\widehat{D}_{j;n,m}$ ,  $j = 1, \dots, p$ , be the upper-left term in (3.7) or (3.4) █*  
 432 *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*

$$433 \quad \text{MRE}(T, m) \leq \frac{L}{nm} \left( m \sum_{j=1}^p \text{Var}(\mathbb{E}[Y_0 Y_j | \mathbf{X}]) \right. \\
 434 \quad \quad \quad \left. + \sum_{j=1}^p \mathbb{E}(\text{Var}[Y_0 Y_j | \mathbf{X}] - \text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}]) \right. \\
 435 \quad \quad \quad \left. + \frac{1}{m} \sum_{j=1}^p \mathbb{E}(\text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}]) \right),$$

436 where

$$437 \quad L = \frac{4(p-1)}{\min_{j < j'} (|D_j - D_{j'}|^2)}.$$

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

$$443 \quad (4.1) \quad \text{MRE}(T, m) \leq \frac{1}{T} v(m), \quad m \in \text{div}_p(T),$$

444 where

$$445 \quad v(m) = L(p+1) \left( m \sum_{j=1}^p \text{Var}(\mathbb{E}[Y_0 Y_j | \mathbf{X}]) + \sum_{j=1}^p \mathbb{E}(\text{Var}[Y_0 Y_j | \mathbf{X}] - \text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}]) \right. \\
 446 \quad \quad \quad \left. + \frac{1}{m} \sum_{j=1}^p \mathbb{E}(\text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}]) \right).$$

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

$$449 \quad \text{MRE}(T, m_T^\dagger) \leq \frac{v(m_T^\dagger)}{T} \leq \frac{v(m)}{T}, \quad \text{for all } m \in \text{div}_p(T).$$

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

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

463 **4.2. A two-stage procedure to estimate the sensitivity indices.** The re-  
 464 sults in Section 4.1 suggest a two-stage procedure to estimate the sensitivity indices.  
 465 The procedure is given in Algorithm 4.1. The computational budget is split into two  
 466 parts  $K$  and  $T - K$ . Denote by  $m_{T-K}^\dagger$  the element  $m \in \text{div}_p(T - K)$  that minimizes  
 467 the function  $v(m)$ . The first  $K$  calls to the model are used to estimate  $m_{T-K}^\dagger$ . The  
 468 last  $T - K$  calls to the model are used to estimate the sensitivity indices.

---

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

---

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

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

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

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

---

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

$$472 \quad (4.2) \quad m^* := \sqrt{\frac{\sum_{j=1}^p \text{E Var}[Y_0|\mathbf{X}] \text{Var}[Y_j|\mathbf{X}]}{\sum_{j=1}^p \text{Var E}[Y_0 Y_j|\mathbf{X}]}} = \sqrt{\frac{\sum_{j=1}^p \zeta_{3,j}}{\sum_{j=1}^p \zeta_{1,j}}},$$

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

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

$$477 \quad \varphi_T(x) = \begin{cases} \lfloor x \rfloor_T & \text{if } \sqrt{\lfloor x \rfloor_T \lceil x \rceil_T} > x \geq 1 \\ \lceil x \rceil_T & \text{if } \sqrt{\lfloor x \rfloor_T \lceil x \rceil_T} \leq x \leq \frac{T}{p+1} \end{cases}$$

478 where

$$480 \quad \lfloor x \rfloor_T = \max\{m \in \text{div}_p(T), m \leq x\}, \quad \lceil x \rceil_T = \min\{m \in \text{div}_p(T), m \geq x\}.$$

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

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

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

$$491 \quad (4.3) \quad \widehat{m}_K^* := \sqrt{\frac{\sum_{j=1}^p \widehat{\zeta}_{3,j}}{\sum_{j=1}^p \widehat{\zeta}_{1,j}}},$$

492 where

$$493 \quad \widehat{\zeta}_{3,j} =$$

$$494 \quad (4.4) \quad \frac{1}{n_0} \sum_{i=1}^n \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})^2 \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\widetilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)})^2$$

$$495 \quad (4.5) \quad + \frac{1}{n_0} \sum_{i=1}^n \left( \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)}) \right)^2 \left( \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\widetilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)}) \right)^2$$

$$496 \quad (4.6) \quad - \frac{1}{n_0} \sum_{i=1}^n \left( \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)}) \right)^2 \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\widetilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)})^2$$

$$497 \quad (4.7) \quad - \frac{1}{n_0} \sum_{i=1}^n \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})^2 \left( \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\widetilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)}) \right)^2,$$

498 and

$$500 \quad \widehat{\zeta}_{1,j} =$$

$$501 \quad (4.8) \quad \frac{1}{n_0} \sum_{i=1}^n \left( \frac{1}{m_0} \sum_{k=1}^{m_0} f(X^{(i)}, Z_0^{(i,k)}) f(\widetilde{X}_{-j}^{(i)}, Z_j^{(i,k)}) \right)^2$$

$$502 \quad (4.9) \quad - \left( \frac{1}{n_0} \sum_{i=1}^n \frac{1}{m_0} \sum_{k=1}^{m_0} f(X^{(i)}, Z_0^{(i,k)}) f(\widetilde{X}_{-j}^{(i)}, Z_j^{(i,k)}) \right)^2.$$

504 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  
 505  $\widehat{m}_K^* = 0$ .

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

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

$$509 \quad \sqrt{n_0} \left( \widehat{m}_K^* - \left[ m^* + \frac{C}{m_0} + \epsilon_{m_0} \right] \right) \xrightarrow{d} \mathcal{N}(0, \sigma_{m_0}^2),$$

510 where  $C$  is some constant,  $\epsilon_{m_0} = C_1/m_0^2 + \dots + C_N/m_0^{N+1}$  for some constants  
 511  $C_1, \dots, C_N$  and  $\sigma_{m_0}^2$  is some variance depending on  $m_0$ . If  $m_0 \rightarrow \infty$  then the above  
 512 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.

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

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

$$519 \quad \sqrt{n_0}(\widehat{m}_K^* - m^*) \xrightarrow{d} \mathcal{N}(0, \lim_{m_0 \rightarrow \infty} \sigma_{m_0}^2).$$

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

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

$$524 \quad P\left(\widehat{m}_{T-K}^\dagger = m_{T-K}^\dagger\right) \rightarrow 1.$$

525 All the details of Algorithm 4.1 have been given.

527 **4.3. Performance.** To get some insight into the performance of the procedure  
 528 given in Algorithm 4.1, we look at the performance of the sensitivity indices estimators  
 529 produced in Stage 2. Since they are built with  $T - K$  calls to the model with  $\widehat{m}_{T-K}^\dagger$   
 530 repetitions, equation (4.1) yields

$$531 \quad (4.10) \quad \text{MRE}(T - K, \widehat{m}_{T-K}^\dagger) \leq \frac{1}{T - K} v(\widehat{m}_{T-K}^\dagger),$$

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

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

542 THEOREM 4.7. Assume that the conditions of Proposition 4.6 are fulfilled. Sup-  
 543 pose furthermore that  $K \rightarrow \infty$  such that  $K/T \rightarrow 0$ . Then

$$544 \quad \frac{1}{T - K} v(\widehat{m}_{T-K}^\dagger) = \frac{1}{T} v(m_{T-K}^\dagger)(1 + o_P(1)).$$

545 Theorem 4.7 says that, roughly speaking, our guarantee is  $v(m_{T-K}^\dagger)/T$ , which  
 546 may or may not be better than the best obtainable guarantee  $v(m_T^\dagger)/T$ , depending  
 547 on the potential difference between  $m_{T-K}^\dagger$  and  $m_T^\dagger$ . This problem stems from the  
 548 fact that the sets  $\text{div}_p(T - K)$  and  $\text{div}_p(T)$  may be different even if  $T$  and  $T - K$   
 549 are of the same order of magnitude. For instance, to ensure that our guarantee is the  
 550 best possible, we may choose  $T$  and  $K$  large enough and such that  $\text{div}_p(T - K) =$   
 551  $\{1, 2, \dots, (T - K)/(p + 1)\}$  and  $\text{div}_p(T) = \{1, 2, \dots, T/(p + 1)\}$ . Then  $v(m_{T-K}^\dagger)$  and  
 552  $v(m_T^\dagger)$  will be equal. Corollary 4.8 below puts this formally.

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

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



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

565 Two frameworks are considered:

- 566 •  $n \rightarrow \infty$  and  $m$  is fixed;
- 567 •  $n \rightarrow \infty$  and  $m \rightarrow \infty$ .

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

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

$$573 \quad \sqrt{n} \left( \widehat{\mathbf{S}}''_{n,m} - \mathbf{S}'' \left[ 1 - \frac{\widehat{\mathbf{S}}'_{n,m} - \mathbf{S}'}{\frac{\mathbb{E} \text{Var}[f(X,Z)|X]}{\mathbb{E} \text{Var}[f(X,Z)|X] + m \text{Var} \mathbb{E}[f(X,Z)|X]}} \right] \right) \xrightarrow{d} \mathcal{N}(0, \Xi_m),$$

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

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

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

$$581 \quad \mathbf{S}'' \frac{\mathbb{E} \text{Var}[f(X,Z)|X]}{\mathbb{E} \text{Var}[f(X,Z)|X] + m \text{Var} \mathbb{E}[f(X,Z)|X]},$$

582  
 583 is null whenever  $f$  actually does not depend on  $Z$ , and large whenever the stochastic  
 584 model is highly stochastic. As **Theorem 5.1** shows, the bias is still present even if  $m$   
 585 goes to infinity. **Corollary 5.2** shows that  $m$  must go to infinity fast enough to avoid  
 586 the estimator to be tightly concentrated around the wrong target.

587 **COROLLARY 5.2.** *Assume (3.1) holds. Let  $n \rightarrow \infty$ . If  $m \rightarrow \infty$  such that  $\sqrt{n}/m \rightarrow$   
 588  $0$  then*

$$589 \quad \sqrt{n} \left( \widehat{\mathbf{S}}''_{n,m} - \mathbf{S}'' \right) \xrightarrow{d} \mathcal{N}(0, \Xi_{22}),$$

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

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

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

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

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

603 **6.1. Comparison of Monte-Carlo designs.** The effect of the number of rep-  
 604 etitions on the sensitivity indices estimators and the effect of the calibration in the  
 605 two-stage procedure are examined in two kinds of experiments: the “direct” experi-  
 606 ments and the “calibration” experiments.

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

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

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

627 **6.1.1. High noise context.** The coefficients are  $\beta = 1.2$  and  $\sigma = 4$ . The  
 628 sensitivity indices are  $S'_1 = 0.05$ ,  $S'_2 = 0.08$ ,  $S''_1 = 0.41$  and  $S''_2 = 0.59$ . The real  
 629  $m^*$  is about 5.8. The total budget is  $T = 3 \times 500 = 1500$  and hence  $\text{div}_2(1500) =$   
 630  $\{1, 2, 4, 5, 10, 20, 25, 50, 100, 125, 250, 500\}$ . The integer  $m_{1500}^\dagger$  is equal to  $\varphi_{1500}(m^*) =$   
 631 5. Since the budget is kept fixed, the numbers of explorations are, respectively,  
 632 500, 250, 125, 100, 50, 25, 20, 10, 5, 4, 2, 1. The number of replications is 1500.

633 The results of the direct experiment are given in Figure 1 for  $m = 1, 2, 4, 5, 10,$   
 634 20, 25. The MSE of first kind does not vary with the number of repetitions and is  
 635 much lower than the MSE of second kind, see (c). The estimators of the second kind  
 636 are highly biased for small numbers of repetitions (a) and they have a higher variance  
 637 for larger numbers of repetitions (b). The fact that the bias is high for small numbers  
 638 of repetitions agrees with the theory, according to which the bias should vanish as  $m$   
 639 goes to infinity. Overall, the sensitivity indices of the second kind seem to be much  
 640 harder to estimate than the indices of the first kind, the estimators of which have a  
 641 negligible bias and a very small variance whatever the number of repetitions.

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

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

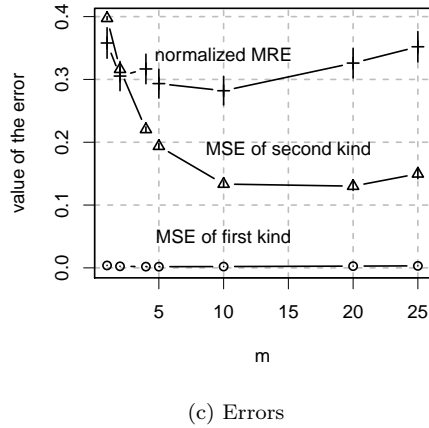
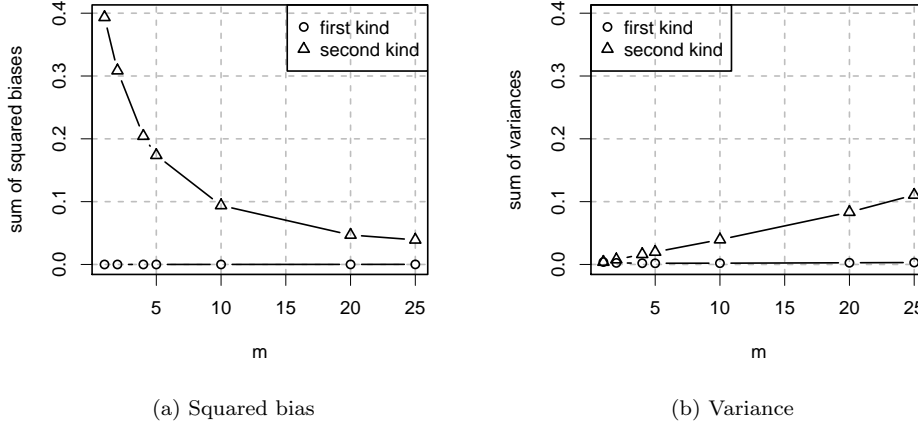


Fig. 1: Sum of squared biases (a), sum of variances (b) and errors (c) of the sensitivity indices estimators for the linear model in the high noise setting. Confidence intervals of level 95% are added in (c).

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

654 **6.1.2. Low noise context.** The coefficients are  $\beta = 1.2$  and  $\sigma = 0.9$ . The  
 655 sensitivity indices are  $S'_1 = 0.31$ ,  $S'_2 = 0.44$ ,  $S''_1 = 0.41$  and  $S''_2 = 0.59$ . The real  
 656  $m^*$  is about 0.30 and hence the integer  $m_{1500}^\dagger$  is equal to 1. As expected, these  
 657 numbers are smaller than the ones found in the high noise context. The total budget

$K/3$	$m_0$				$n_0$			
	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  $m_0 = 2, 5, 10, 20, 40, 100$ .

$K/3$	$m_0$				$n_0$			
	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  $m_0 = 2, 5, 10, 20, 40, 100$ .

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

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

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

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

674 **6.2. Asymptotic behavior of the sensitivity indices estimators.** To illus-  
675 trate the asymptotic behavior of the sensitivity indices estimators, Sobol's g-function,  
676 a benchmark in sensitivity analysis [25, 20], is considered. Sobol's g-function is given  
677 by

$$678 \quad g(U_1, \dots, U_{p+1}) = \prod_{j=1}^{p+1} \frac{|4U_j - 2| + a_j}{1 + a_j},$$

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

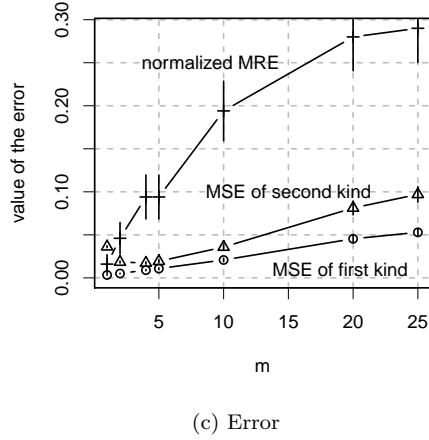
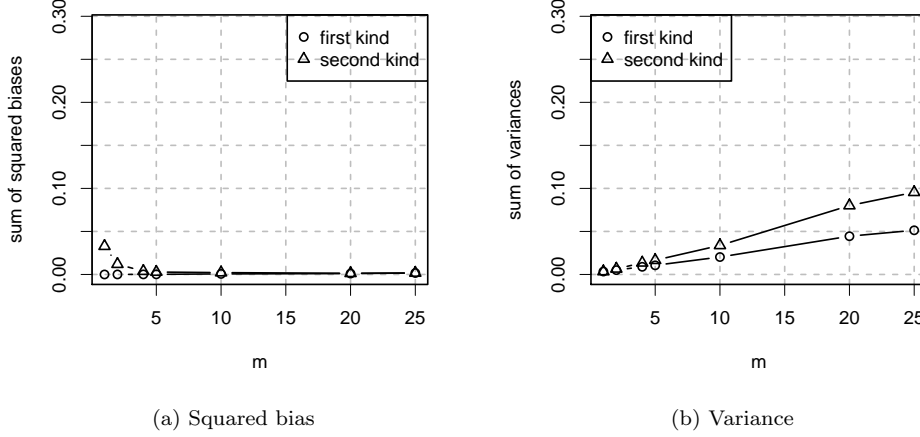


Fig. 2: Sum of squared biases (a), sum of variances (b) and errors (c) of the sensitivity indices estimators for the linear model in the low noise context. Confidence intervals of level 95% are added in (c).

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

$$683 \quad S_j^{(a_1, \dots, a_{p+1})} = \frac{1}{3(1+a_j)^2} \left( -1 + \prod_{j=1}^{p+1} \frac{(4/3 + a_j^2 + 2a_j)}{(1+a_j)^2} \right)^{-1} .$$

684

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

$$688 \quad (6.1) \quad Y = f(X_1, \dots, X_p, Z) = g(X_1, \dots, X_{i-1}, Z, X_i, \dots, X_p).$$

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

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

$$694 \quad S'_j = \begin{cases} S_j^{(a_1, \dots, a_{p+1})} & \text{if } 1 \leq j \leq i - 1 \\ S_{j+1}^{(a_1, \dots, a_{p+1})} & \text{if } i \leq j \leq p \end{cases}$$

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

$$697 \quad b_{ij} = \begin{cases} a_j & \text{if } 1 \leq j \leq i - 1, \\ a_{j+1} & \text{if } i \leq j \leq p. \end{cases}$$

699 For each kind of Sobol index, we produced 500 estimates of the  $p$  Sobol indices  
 700 and computed the values of the mean squared error (MSE) by averaging over the  
 701 500 replications and summing over the  $p$  indices. We tested  $n = 100, 500, 2500$  and  
 702  $m = 1, 10, 100$ .

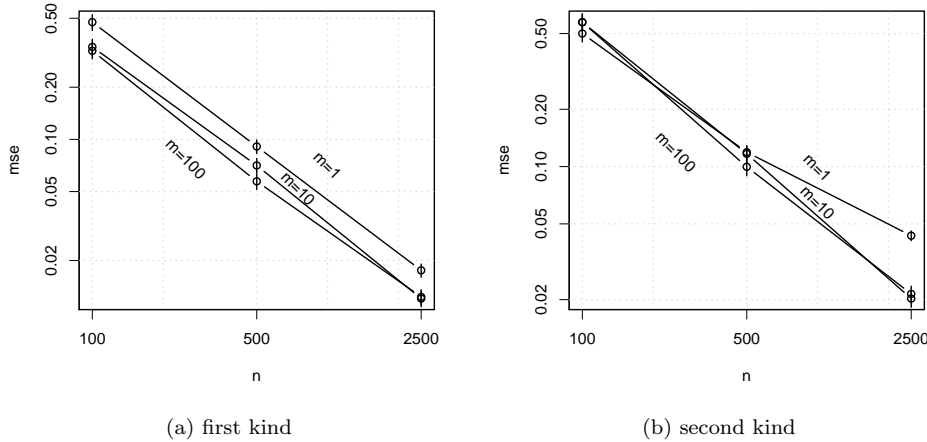


Fig. 3: MSEs for the Sobol index estimators of the first and second kind (logarithmic scale).

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

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

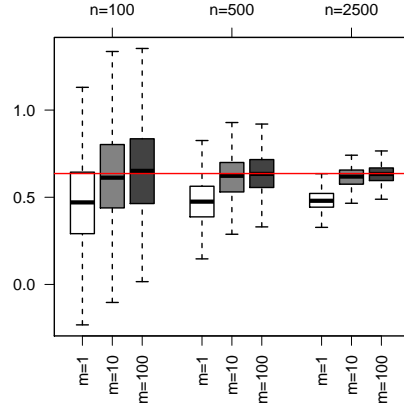


Fig. 4: Boxplots of the estimates for the Sobol index of the second kind associated with  $X_1$ . The red horizontal line is the truth.

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

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



744 enth, the connection between our representation of stochastic models and that of [10]  
745 could be investigated further.

746 **Appendix A. Calculations of some sensitivity indices.**

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

$$748 \quad \tilde{S}_1^{\text{HAG}} = \mathbb{E} \left( \frac{\text{Var}(\mathbb{E}[f(X, Z)|X_j, Z]|Z)}{\text{Var}(f(X, Z)|Z)} \right) = \int_{\Omega} \frac{\text{Var}(\mathbb{E}[f(X, Z)|X_j, Z]|Z)}{\text{Var}(f(X, Z)|Z)} dP.$$

749

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

$$753 \quad \tilde{S}_1^{\text{HAG}} = \int_{(0,1)} \frac{\text{Var}(\mathbb{E}[f(X, z)|X_j, Z = z]|Z = z)}{\text{Var}(f(X, z)|Z = z)} dz = \int_{(0,1)} \frac{\text{Var}(\mathbb{E}[f(X, z)|X_1])}{\text{Var}(f(X, z))} dz.$$

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

$$756 \quad \begin{aligned} \text{Var}(f(X, z)) &= \text{Var}(\Phi^{-1}(z)X_2 + X_1) = \Phi^{-1}(z)^2 \text{Var}(X_2) + \text{Var}(X_1) \\ &= \Phi^{-1}(z)^2 \frac{L^2}{12} + \frac{1}{12}, \end{aligned}$$

757

758

759 and

$$760 \quad \begin{aligned} \text{Var}(\mathbb{E}[f(X, z)|X_1]) &= \text{Var}(\mathbb{E}[\Phi^{-1}(z)X_2 + X_1|X_1]) \\ &= \text{Var}(\Phi^{-1}(z) \mathbb{E}[X_2|X_1] + \mathbb{E}[X_1|X_1]) \\ &= \text{Var}(\Phi^{-1}(z) \mathbb{E}[X_2] + X_1) \\ &= \text{Var}(X_1) \\ &= \frac{1}{12} \end{aligned}$$

764

765

766 and hence

$$767 \quad \tilde{S}_1^{\text{HAG}} = \int_{(0,1)} \frac{1}{\Phi^{-1}(z)^2 L^2 + 1} dz = \int_{-\infty}^{\infty} \frac{1}{z^2 L + 1} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

768

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

$$771 \quad S'_1 = \frac{\text{Var}[\mathbb{E}(X_1 + X_2 \Phi^{-1}(Z)|X_1)]}{\text{Var}[X_1 + X_2 \Phi^{-1}(Z)]}.$$

772

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

$$776 \quad \begin{aligned} \text{Var}[X_2 \Phi^{-1}(Z)] &= \text{Var}[\mathbb{E}(X_2 \Phi^{-1}(Z)|Z)] + \mathbb{E}(\text{Var}[X_2 \Phi^{-1}(Z)|Z]) \\ &= \text{Var} \left[ \Phi^{-1}(Z) \left( \frac{L}{2} + 1 \right) \right] + \int_0^1 \Phi^{-1}(z)^2 \text{Var}[X_2] dz \\ &= \left( \frac{L}{2} + 1 \right)^2 + \frac{L^2}{12}, \end{aligned}$$

778

779

780 so that

$$781 \quad S'_1 = \frac{1/12}{1/12 + (L/2 + 1)^2 + L^2/12} = \frac{1}{4(L^2 + 3(L + 1)) + 1}.$$

783 **Appendix B. Proofs.**

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

$$792 \quad P(f(x, Z) \leq t)$$

$$793 \quad = P(Z \leq F_x(t)) = \lambda^{\otimes_{j=1}^{p+1}}\{\omega \in \Omega : \omega_{p+1} \leq F_x(t)\} = \lambda(0, F_x(t)] = F_x(t).$$

795 Finally, by the same token,

$$796 \quad P(X_1 \leq t_1, \dots, X_p \leq t_p, Z \leq t_{p+1})$$

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

799 The proof is complete.

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

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

$$804 \quad P(\widehat{D}_i - \widehat{D}_j \geq 0) \leq \frac{\text{Var } \widehat{D}_i + \text{Var } \widehat{D}_j}{\frac{1}{2}|D_i - D_j|^2}.$$

806 *Proof. We have*

$$807 \quad P(\widehat{D}_i - \widehat{D}_j \geq 0) \leq P(|\widehat{D}_i - D_i| + |\widehat{D}_j - D_j| \geq D_j - D_i)$$

$$808 \quad \leq P(|\widehat{D}_i - D_i|^2 + |\widehat{D}_j - D_j|^2 \geq \frac{1}{2}|D_j - D_i|^2)$$

809

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

$$813 \quad \sum_{i=1}^p \mathbb{E}|\widehat{R}_i - R_i| \leq \sum_{i=1}^p \sum_{j=1}^p \mathbb{E}|\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i) - \mathbf{1}(D_j \leq D_i)|$$

$$814 \quad \leq \sum_{i=1}^p \sum_{j \neq i} \frac{\text{Var } \widehat{D}_i + \text{Var } \widehat{D}_j}{\frac{1}{2}|D_i - D_j|^2}$$

$$815 \quad \leq \frac{4(p-1)}{\min_{j < j'} |D_j - D_{j'}|^2} \sum_{i=1}^p \text{Var } \widehat{D}_i,$$

816

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

$$818 \quad \mathbb{E}|\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i) - \mathbf{1}(D_j \leq D_i)| = \begin{cases} \mathbb{E}|\mathbf{1}(\widehat{D}_j > \widehat{D}_i)| & \text{if } j < i, \\ 0 & \text{if } j = i, \\ \mathbb{E}|\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i)| & \text{if } j > i. \end{cases}$$

819

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

$$822 \quad \text{Var } \widehat{D}_j = \frac{1}{n} \{ \text{Var } \mathbb{E}[Y_0 Y_j | \mathbf{X}] + \frac{1}{m} (\mathbb{E} \text{Var}[Y_0 Y_j | \mathbf{X}] - \text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}]) \\ 823 \quad + \frac{1}{m^2} \mathbb{E} \text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}] \}.$$

824

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

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

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

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

$$834 \quad \varphi_{T-K}(m^*) = \begin{cases} \lrcorner m^* \lrcorner_{T-K} & \text{if } \sqrt{\lrcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}} > m^* \\ \lrcorner m^* \lrcorner_{T-K} & \text{if } \sqrt{\lrcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}} \leq m^*. \end{cases}$$

835

836 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  
837  $m_{T-K}^\dagger = \lrcorner m^* \lrcorner_{T-K} = \varphi_{T-K}(m^*)$ . Otherwise, since  $v(x) = \zeta_1 x + \zeta_2 + \zeta_3/x$ ,  $x > 0$ ,  
838 for some constants  $\zeta_1, \zeta_2$  and  $\zeta_3$  such that  $\zeta_3/\zeta_1 = m^*$ , we have

$$839 \quad v(\lrcorner m^* \lrcorner_{T-K}) < v(\lrcorner m^* \lrcorner_{T-K}) \text{ iff } \sqrt{\lrcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}} > \frac{\zeta_3}{\zeta_1} = m^*.$$

840

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

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

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

$$848 \quad \widehat{m}_K^* = \sqrt{\frac{\sum_{j=1}^p \widehat{\zeta}_{3,j}}{\sum_{j=1}^p \widehat{\zeta}_{1,j}}} = \sqrt{\frac{\sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.4)} + \xi_{j;m,i}^{(4.5)} - \xi_{j;m,i}^{(4.6)} - \xi_{j;m,i}^{(4.7)}}{\sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.8)} - \left(\frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.9)}\right)^2}},$$

849

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

$$\begin{aligned} 852 \quad \bar{\boldsymbol{\xi}} &= \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i}, \\ 853 \quad \boldsymbol{\xi}_{m,i} &= (\boldsymbol{\xi}_{1;m,i}^\top, \dots, \boldsymbol{\xi}_{p;m,i}^\top)^\top, \quad i = 1, \dots, n, \\ 854 \quad \boldsymbol{\xi}_{j;m,i} &= (\xi_{j;m,i}^{(4.4)}, \dots, \xi_{j;m,i}^{(4.9)})^\top, \quad j = 1, \dots, p, \quad i = 1, \dots, n. \end{aligned}$$

856 Let  $s$  be the function defined by

$$857 \quad s(\mathbf{x}) = \sqrt{\frac{\sum_{j=1}^p x_j^{(4.4)} + x_j^{(4.5)} - x_j^{(4.6)} - x_j^{(4.7)}}{\sum_{j=1}^p x_j^{(4.8)} - x_j^{(4.9)2}}},$$

859 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  
860 notation we have  $\hat{m}_K^* = s(\bar{\boldsymbol{\xi}})$ . Moreover, elementary calculations show that

$$861 \quad (\text{B.1}) \quad \mathbb{E} \boldsymbol{\xi}_{m,1} = \boldsymbol{\theta} + \sum_{\nu=1}^4 \frac{\mathbf{C}_\nu}{m^\nu},$$

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

$$864 \quad \boldsymbol{\theta}_j = \mathbb{E} \begin{pmatrix} Y_0^{(1,1)2} Y_j^{(1,1)2} \\ Y_0^{(1,1)} Y_0^{(1,2)} Y_j^{(1,1)} Y_j^{(1,2)} \\ Y_0^{(1,1)} Y_0^{(1,2)} Y_j^{(1,1)2} \\ Y_j^{(1,1)} Y_j^{(1,2)} Y_0^{(1,1)2} \\ Y_0^{(1,1)} Y_0^{(1,2)} Y_j^{(1,1)} Y_j^{(1,2)} \\ Y_j^{(1,1)} Y_0^{(1,1)} \end{pmatrix}.$$

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

$$\begin{aligned} 867 \quad & \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1})^\top \dot{s}(\mathbb{E} \boldsymbol{\xi}_{m,1}) + \frac{1}{2}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1})^\top \ddot{s}_{n,m}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1}) \\ 868 \quad (\text{B.2}) \quad & = \sqrt{n}(s(\bar{\boldsymbol{\xi}}) - s(\mathbb{E} \boldsymbol{\xi}_{m,1})) \\ 869 \quad & = \sqrt{n}(s(\bar{\boldsymbol{\xi}}) - s(\boldsymbol{\theta}) - (\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \dot{s}(\boldsymbol{\theta}) - \frac{1}{2}(\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \ddot{s}_m(\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})), \end{aligned}$$

871 where  $\dot{s}$  is the gradient of  $s$ ,  $\ddot{s}_{n,m}$  is the Hessian matrix of  $s$  at a point between  $\bar{\boldsymbol{\xi}}$   
872 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  
873 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  
874 constants  $C_\nu$ . Putting

$$875 \quad \epsilon_m = \frac{1}{2}(\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \ddot{s}_m(\mathbb{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta}) + \sum_{\nu=2}^4 \frac{C_\nu}{m^\nu},$$

877 it follows from (B.2) that

$$\begin{aligned} 879 \quad (\text{B.3}) \quad & \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1})^\top \dot{s}(\mathbb{E} \boldsymbol{\xi}_{m,1}) + \frac{1}{2}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1})^\top \ddot{s}_{n,m}(\bar{\boldsymbol{\xi}} - \mathbb{E} \boldsymbol{\xi}_{m,1}) \\ 880 \quad & = \sqrt{n}(\hat{m}_K^* - m^* - \frac{C_1}{m} - \epsilon_m). \end{aligned}$$

881

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

$$883 \quad \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbf{E} \boldsymbol{\xi}_{m,1}) \rightarrow \mathcal{N}(0, \Sigma_m),$$

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

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

$$890 \quad \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbf{E} \boldsymbol{\xi}_{m,1}) \rightarrow \mathcal{N}(0, \lim_{m \rightarrow \infty} \Sigma_m).$$

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

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

$$897 \quad \operatorname{div}_p(T-K) \setminus \left\{ 1, \frac{T-K}{p+1} \right\}.$$

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

$$900 \quad \mathcal{D}_{T-K} \subset \{\sqrt{ij} : i \text{ and } j \text{ are two consecutive integers}\} = \mathcal{E}.$$

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

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

$$909 \quad P \left( \left| \frac{\frac{1}{T-K} v(\widehat{m}_{T-K}^\dagger) - \frac{1}{T} v(m_{T-K}^\dagger)}{\frac{1}{T} v(m_{T-K}^\dagger)} > \varepsilon \right| \right)$$

$$910 \quad \leq P \left( \left| \frac{T}{T-K} (\widehat{m}_{T-K}^\dagger - m_{T-K}^\dagger) v'(\tilde{m}) + \frac{K}{T-K} v(m_{T-K}^\dagger) \right| > v(m_{T-K}^\dagger) \varepsilon \right),$$

912 where  $\tilde{m}$  denotes a real between  $\widehat{m}_{T-K}^\dagger$  and  $m_{T-K}^\dagger$ . A decomposition of the probability  
 913 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  
 914 bound

$$915 \quad P \left( \widehat{m}_{T-K}^\dagger - m_{T-K}^\dagger \neq 0 \right) + P \left( \frac{K}{T-K} > \varepsilon \right).$$

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

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

$$921 \quad \widehat{S}'_{j;n,m} = \frac{\frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{\text{UL}} - \left(\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}}\right)^2}{\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{LL}} - \left(\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}}\right)^2}, \quad j = 1, \dots, p,$$

923 and

$$924 \quad \widehat{S}''_{j;n,m} = \frac{\frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{\text{UL}} - \left(\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}}\right)^2}{\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{LL}} - \left(\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}}\right)^2}, \quad j = 1, \dots, p,$$

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

$$928 \quad \mathbb{E} \boldsymbol{\xi}_{m,1} = \begin{pmatrix} \mathbb{E} \mathbb{E}[f(X, Z)|X] \mathbb{E}[f(\widetilde{X}_{-1}, Z)|\widetilde{X}_{-1}] \\ \vdots \\ \mathbb{E} \mathbb{E}[f(X, Z)|X] \mathbb{E}[f(\widetilde{X}_{-p}, Z)|\widetilde{X}_{-p}] \\ \mathbb{E} f(X, Z) \\ \mathbb{E} f(X, Z)^2 \\ \mathbb{E} \mathbb{E}[f(X, Z)|X]^2 + \frac{\mathbb{E} \text{Var}[f(X, Z)|X]}{m} \end{pmatrix}.$$

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

$$930 \quad s(x_1, \dots, x_p, x_{p+1}, x_{p+2}, x_{p+3}) \\ 931 \quad = \left( \frac{x_1 - x_{p+1}^2}{x_{p+2} - x_{p+1}^2}, \dots, \frac{x_p - x_{p+1}^2}{x_{p+2} - x_{p+1}^2}, \frac{x_1 - x_{p+1}^2}{x_{p+3} - x_{p+1}^2}, \dots, \frac{x_p - x_{p+1}^2}{x_{p+3} - x_{p+1}^2} \right).$$

934 Clearly, we have

$$935 \quad s\left(\frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i}\right) = \begin{pmatrix} \widehat{\mathbf{S}}'_{n,m} \\ \widehat{\mathbf{S}}''_{n,m} \end{pmatrix}$$

936 and

$$937 \quad s(\mathbb{E} \boldsymbol{\xi}_{m,1}) = \begin{pmatrix} \mathbf{S}' \\ \mathbf{S}'' \left[ 1 - \frac{\mathbb{E} \text{Var}[f(X, Z)|X]}{\mathbb{E} \text{Var}[f(X, Z)|X] + m \text{Var} \mathbb{E}[f(X, Z)|X]} \right] \end{pmatrix}.$$

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

$$939 \quad \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} - \mathbb{E} \boldsymbol{\xi}_{m,1} \right) \xrightarrow{d} \mathcal{N}(0, \Sigma_m),$$

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

942 If  $m \rightarrow \infty$ , Lemma C.2 still holds with the variance-covariance matrix replaced

943 by its limit. Taylor's expansion yields

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

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

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

$$954 \quad (\text{C.1}) \quad \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i},$$

955 with

$$956 \quad (\text{C.2}) \quad \boldsymbol{\xi}_{m,i} = \prod_{l=1}^L \frac{1}{m} \sum_{k=1}^m \prod_{j=0}^p Y_j^{(i,k) b_{j;l}},$$

958 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  
959  $b_{j;l}$ ,  $j = 0, \dots, p$ ,  $l = 1, \dots, L$ , are nonnegative coefficients. The coefficients are  
960 arranged in a matrix  $(b_{j;l})$  with  $L$  rows and  $p+1$  columns, where  $b_{j;l}$  is the element in  
961 the  $l$ th row and  $(j+1)$ th column. This way, all estimators of the form (C.1) and (C.2),  
962 or, equivalently, all summands (C.2), can be represented by a matrix. We sometimes  
963 write  $\boldsymbol{\xi}_{m,i} \simeq (b_{j;l})$ , where  $(b_{j;l})$  is the matrix of size  $L \times (p+1)$  with coefficients  $b_{j;l}$ ,  
964  $j = 0, \dots, p$ ,  $l = 1, \dots, L$ .

965 **C.1. Examples.** The estimator

$$966 \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \frac{1}{m} \sum_{k'=1}^m Y_j^{(i,k')}$$

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

$$969 \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix},$$

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

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



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

$$976 \quad \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

$$977 \quad \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

978 respectively.

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

$$982 \quad \widehat{\zeta}_{3,j} = \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.4)} + \xi_{j;m,i}^{(4.5)} - \xi_{j;m,i}^{(4.6)} - \xi_{j;m,i}^{(4.7)}, \quad \text{and,}$$

$$983 \quad \widehat{\zeta}_{1,j} = \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.8)} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.9)} \right)^2,$$

984 where

$$985 \quad \begin{matrix} \xi_{j;m,i}^{(4.4)} & \xi_{j;m,i}^{(4.5)} \\ \xi_{j;m,i}^{(4.6)} & \xi_{j;m,i}^{(4.7)} \\ \xi_{j;m,i}^{(4.8)} & \xi_{j;m,i}^{(4.9)} \end{matrix}$$

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

$$991 \quad \begin{pmatrix} 2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$

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

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

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

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

$$997 \quad \xi_{1;m,i}^{\text{UL}} \simeq \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \end{pmatrix}, \dots, \xi_{p;m,i}^{\text{UL}} \simeq \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

998 for the upper left (UL) terms,

$$1000 \quad \xi_{m,i}^{\text{UR}} \simeq \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

1001 for the upper right (UR) term,

$$1002 \quad \xi_{m,i}^{\prime LL} \simeq \begin{pmatrix} 2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

1003 for the lower left (LL) term of  $\widehat{S}'_{j;n,m}$  and

$$1004 \quad \xi_{m,i}^{\prime\prime LL} \simeq \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

1005 for the lower left (LL) term of  $\widehat{S}''_{j;n,m}$ .

1006 **C.2. A central limit theorem.** For each  $n$ , the random variables  $\xi_{m,1}, \dots, \xi_{m,n}$   
 1007 are independent and identically distributed. Denote by  $\mathcal{E}_{m,i}(L)$  the set of all sum-  
 1008 mands (C.2). In other words,  $\mathcal{E}_{m,i}(L)$  is the set of all nonnegative matrices of size  
 1009  $L \times (p+1)$ . This set has useful properties, gathered in Proposition C.1 for subsequent  
 1010 use.

1011 **PROPOSITION C.1.** *Let  $\xi$  be an element of  $\mathcal{E}_{m,i}(L)$  with coefficients  $(b_{j;l})$ . The*  
 1012 *following statements are true.*

1013 (i) *If  $\xi'$  is an element of  $\mathcal{E}_{m,i}(L)$  with coefficients  $(b'_{j;l})$  then  $\xi\xi'$  is an element of*  
 1014  *$\mathcal{E}_{m,i}(2L)$  with coefficients*

$$1015 \quad \begin{pmatrix} b_{0;1} & \cdots & b_{p;1} \\ \vdots & & \vdots \\ b_{0;L} & \cdots & b_{p;L} \\ b'_{0;1} & \cdots & b'_{p;1} \\ \vdots & & \vdots \\ b'_{0;L} & \cdots & b'_{p;L} \end{pmatrix}.$$

1016 (ii) *The limit of  $\mathbb{E}\xi$  exists as  $m \rightarrow \infty$ .*

1017 (iii) *If there exists some function  $F$  such that  $|f(x, z)| \leq F(x)$  for all  $x$  and  $z$  in the*  
 1018 *domain of definition of  $f$  then*

$$1020 \quad |\xi| \leq \left( \bigvee_{j=0}^p F_j(\mathbf{X}^{(i)}) \right)^{\sum_{j=0}^p \sum_{l=1}^L b_{j;l}},$$

1021 where  $F_j(\mathbf{X}^{(i)})$  is  $F(X^{(i)})$  if  $j = 0$  and  $F(\widetilde{X}_{-j}^{(i)})$  if  $j \geq 1$ .

1022 *Proof.* The proof of (i) is trivial. Let us prove (ii). We have

$$1023 \quad \mathbb{E}\xi = \frac{1}{m^L} \sum_{(k_1, \dots, k_L) \in \{1, \dots, m\}^L} \mathbb{E} \prod_{l=1}^L \prod_{j=0}^p Y_j^{(1, k_l) b_{j;l}}$$

$$1024 \quad = \frac{1}{m^L} \sum_{(k_1, \dots, k_L) \in \{1, \dots, m\}^L} \mathbb{E} \mathbb{E} \left( \prod_{l=1}^L \prod_{j=0}^p Y_j^{(1, k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right)$$

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

1026

1027 Since (i)  $\mathbf{X}^{(1)}$  and  $\{\mathbf{Z}^{(1,k)}, k = 1, \dots, m\}$  are independent and (ii) the law of

1028 
$$(\mathbf{Z}^{(1,k_1)}, \dots, \mathbf{Z}^{(1,k_L)})$$

1029 is invariant through any permutation of distinct  $k_1, \dots, k_L$ , all the inner expectations  
1030 in (C.3) are equal to some others. For if  $k_1, \dots, k_L$  are distinct then

1031 
$$\mathbb{E} \left( \prod_{l=1}^L Y_j^{(1,k_l)b_{j;l}} \middle| \mathbf{X}^{(1)} \right) = \mathbb{E} \left( \prod_{l=1}^L Y_j^{(1,l)b_{j;l}} \middle| \mathbf{X}^{(1)} \right)$$

1032  
1033 for all  $j = 0, \dots, p$ . The number of inner expectations equal to the one above is  
1034  $m(m-1) \cdots (m-L+1)$ , a polynomial in  $m$  with degree  $L$ . If some components of  
1035 the tuple  $(k_1, \dots, k_L)$  are equal, then we can always write

1036 
$$\mathbb{E} \left( \prod_{l=1}^L Y_j^{(1,k_l)b_{j;l}} \middle| \mathbf{X}^{(1)} \right) = \mathbb{E} \left( \prod_{l=1}^{L'} Y_j^{(1,l)\beta_{j;l}} \middle| \mathbf{X}^{(1)} \right)$$

1037  
1038 for some  $L' \leq L$  and coefficients  $\beta_{j;l}$ . It is easy to see that the number of inner expec-  
1039 tations equal to the one above is a polynomial in  $m$  with degree at most  $L$ . (Looking  
1040 at examples helps to see this; see e.g. the proof of Lemma D.2 in Appendix D.)  
1041 Therefore, the sum in (C.3) is also a polynomial in  $m$  with degree at most  $L$  and the  
1042 claim follows ( $\mathbb{E} \xi$  can be zero). To prove (iii), simply remember that, by assumption,  
1043  $|Y^{(1,k)}| \leq F(X^{(1)})$  and  $|Y_j^{(1,k)}| \leq F(\tilde{X}_{-j}^{(1)})$  for all  $k$  and all  $j$ .  $\square$

1044 Two frameworks are considered:

- 1045 •  $n \rightarrow \infty$  and  $m$  is fixed;
- 1046 •  $n \rightarrow \infty$  and  $m \rightarrow \infty$ .

1047 In the second framework  $m_n$  is a sequence indexed by  $n$  that goes to infinity as  $n$  goes  
1048 to infinity.

1049 LEMMA C.2. Let  $\xi_{m,i}^{(I)}$ ,  $I = 1, \dots, N$ , be elements of  $\mathcal{E}_{m,i}(L)$  with coefficients  
1050  $(b_{j;l}^{(I)})$ . Assume

1051 
$$\mathbb{E} F(X^{(1)})^2 \sum_{j=0}^p \sum_{l=1}^L b_{j;l}^{(I)} < \infty$$

1052 for all  $I = 1, \dots, N$ . Let  $n \rightarrow \infty$ . If  $m$  is fixed then

1053 
$$\sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{(1)} - \mathbb{E} \xi_{m,1}^{(1)}, \dots, \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{(N)} - \mathbb{E} \xi_{m,1}^{(N)} \right] \xrightarrow{d} \mathcal{N}(0, \Sigma_m),$$

1054  
1055 where  $\Sigma_m$  is the variance-covariance matrix of  $\boldsymbol{\xi}_{m,i} = (\xi_{m,i}^{(1)}, \dots, \xi_{m,i}^{(N)})^\top$ . If  $m \rightarrow$   
1056  $\infty$  then  $\lim_{m \rightarrow \infty} \Sigma_m$  exists elementwise and the above display with  $\Sigma_m$  replaced by  
1057  $\lim_{m \rightarrow \infty} \Sigma_m$  is true.

1058 *Proof.* Let  $m$  be fixed. By Proposition C.1 (i),  $\xi_{m,i}^{(I)2}$ ,  $I = 1, \dots, N$ , belongs to  
1059  $\mathcal{E}_{m,i}(2L)$  and has coefficients

1060 
$$\xi_{m,i}^{(I)2} \simeq \begin{pmatrix} b_{0;1}^{(I)} & \cdots & b_{p;1}^{(I)} \\ \vdots & & \vdots \\ b_{0;L}^{(I)} & \cdots & b_{p;L}^{(I)} \\ b_{0;1}^{(I)} & \cdots & b_{p;1}^{(I)} \\ \vdots & & \vdots \\ b_{0;L}^{(I)} & \cdots & b_{p;L}^{(I)} \end{pmatrix}.$$

1061

1062 Thus, denoting  $\sum_{j=0}^p \sum_{l=1}^L b_{j;l}^{(I)}$  by  $\beta$ , Proposition C.1 (iii) yields

$$1063 \quad (\text{C.4}) \quad \xi_{m,i}^{(I)2} \leq \bigvee_{j=0}^p F_j(\mathbf{X}^{(i)})^{2\beta}$$

1064

1065 and hence

$$1066 \quad \mathbb{E} \xi_{m,i}^{(I)2} \leq \mathbb{E} \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{2\beta} \leq (p+1) \mathbb{E} \left( F(X^{(1)}) \right)^{2\beta} < \infty.$$

1067

1068 Therefore we can apply the central limit theorem to finish the proof for  $m$  fixed.

1069 Let  $m \rightarrow \infty$ . According to Lindeberg-Feller's central limit theorem (see e.g. [33]),  
1070 it suffices to show

1071 (i) for all  $\epsilon > 0$ ,

$$1072 \quad \sum_{i=1}^n \mathbb{E} \left\| \frac{1}{\sqrt{n}} \boldsymbol{\xi}_{m,i} \right\|^2 \mathbf{1} \left\{ \left\| \frac{1}{\sqrt{n}} \boldsymbol{\xi}_{m,i} \right\| > \epsilon \right\} \rightarrow 0,$$

1073

1074 and

1075 (ii) the limit  $\sum_{i=1}^n \text{Cov}(\boldsymbol{\xi}_{m,i}/\sqrt{n})$  exists and is finite.

1076 Let us show (i). Denoting  $\mathbf{X} = (X^{(1)}, \tilde{X}^{(1)})$ , we have

$$1077 \quad \sum_{i=1}^n \mathbb{E} \left\| \frac{\boldsymbol{\xi}_{m,i}}{\sqrt{n}} \right\|^2 \mathbf{1} \{ \|\boldsymbol{\xi}_{m,i}\| > \sqrt{n}\epsilon \} = \mathbb{E} \|\boldsymbol{\xi}_{m,1}\|^2 \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \}$$

$$1078 \quad = \mathbb{E} \sum_{I=1}^N \xi_{m,1}^{(I)2} \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \}$$

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

1080

1081 By (C.4), we have

$$1082 \quad \mathbb{E} \left( \xi_{m,1}^{(I)2} \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \} \mid \mathbf{X} \right) \leq \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{2\beta} P(\|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \mid \mathbf{X})$$

$$1083 \quad \leq \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{2\beta} \frac{\sum_{I=1}^N \mathbb{E} \left( \xi_{m,1}^{(I)2} \mid \mathbf{X} \right)}{n\epsilon^2}$$

$$1084 \quad \leq \frac{N \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{4\beta}}{n\epsilon^2},$$

1085

1086 where the last inequality holds by using (C.4) once more. The upper bound goes to  
1087 zero and is dominated by an integrable function. Thus, we can apply the dominated  
1088 convergence theorem to complete the proof.

1089 Let us show that (ii) holds. We have  $\sum_{i=1}^n \text{Cov}(\boldsymbol{\xi}_{m,i}/\sqrt{n}) = \text{Cov}(\boldsymbol{\xi}_{m,1})$ . The  
1090 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  
1091  $\mathbb{E} \xi_{m,1}^{(I)2} < \infty$ ,  $I = 1, \dots, N$ , 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  
1092 the limit of  $\text{Cov} \boldsymbol{\xi}_{m,1}$  exists and is finite. The proof is complete.  $\square$

1093 **Appendix D. Explicit moment calculations.** Explicit moment calculations  
 1094 are given for the summands in the proof of Theorem 5.1. In this section,  $E f(X, Z)$   
 1095 and  $E E[f(X, Z)|X]^2$  are denoted by  $\mu$  and  $D$ , respectively. Recall that the upper-left  
 1096 term in (3.6) and (3.5) is denoted by  $D_j$ . The moments are given in Lemma D.1  
 1097 and Lemma D.2. The variances and covariances are given in Lemma D.3. Let  $\mathbf{X} =$   
 1098  $(X^{(1)}, \tilde{X}^{(1)})$ . Whenever there is a superscript  $\mathbf{X}$  added to the expectation symbol  $E$   
 1099 or the variance symbol  $\text{Var}$ , this means that these operators are to be understood  
 1100 conditionally on  $\mathbf{X}$ . An integral with respect to  $\mathbf{P}^*(d\mathbf{x})$  means that we integrate with  
 1101 respect to the law of  $\mathbf{X}$ .

1102 LEMMA D.1 (Moments of order 1). *The moments of order 1 are given by*

$$\begin{aligned} 1103 \quad & E \xi_{j;m1}^{UL} = D_j, \\ 1104 \quad & E \xi_{m1}^{UR} = \mu, \\ 1105 \quad & E \xi_{m1}^{\prime LL} = \frac{1}{m} E \text{Var}^X f(X^{(1)}, Z^{(1,1)}) + D. \\ 1106 \end{aligned}$$

1107

1108 *Proof.* One has

$$\begin{aligned} 1109 \quad E \xi_{j;m1}^{UL} &= \frac{1}{m^2} \sum_{k,k'} E f(X^{(1)}, Z^{(1,k)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,k')}) \\ 1110 \quad &= \frac{1}{m^2} \sum_{k,k'} \int E f(x, Z^{(1,k)}) f(\tilde{x}_{-j}, Z_j^{(1,k')}) \mathbf{P}^*(d\mathbf{x}) \\ 1111 \quad &= E f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\ 1112 \quad &= D_j, \end{aligned}$$

1114 where the integral is taken with respect to the law of  $\mathbf{x} = (x, \tilde{x})$ , and,

$$\begin{aligned} 1115 \quad E \xi_{m1}^{\prime LL} &= \frac{1}{m^2} \sum_{k,k'} E f(X^{(1)}, Z^{(1,k)}) f(X^{(1)}, Z^{(1,k')}) \\ 1116 \quad &= \frac{1}{m} E \text{Var}^X f(X, Z) + E(E^X f(X, Z))^2 \\ 1117 \quad &= \frac{1}{m} E \text{Var}^X f(X, Z) + D. \\ 1118 \end{aligned}$$

1119 The proof for  $\xi_{m1}^{UR}$  is similar. □

1120 LEMMA D.2 (Moments of order 2). *The moments of order 2 are given by*

$$\begin{aligned}
1121 \quad \mathbb{E} \xi_{j;m_1}^{(UL)2} &= \text{Var} \mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) + D_j^2 \\
1122 \quad &+ \frac{1}{m} [\mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1123 \quad &- \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})] \\
1124 \quad &+ \frac{1}{m^2} \mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}), \\
1125 \quad \mathbb{E} \xi_{m_1}^{(UR)2} &= \frac{1}{m} \mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) + \mathbb{E} (\mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}))^2, \\
1126 \quad \mathbb{E} \xi_{m_1}^{\mu(LL)2} &= \frac{m(m-1)(m-2)(m-3)}{m^4} \\
1127 \quad &\mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) f(X^{(1)}, Z^{(1,4)}) \\
1128 \quad &+ \frac{\binom{4}{2} m(m-1)(m-2)}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) \\
1129 \quad &+ \frac{\binom{4}{3} m(m-1)}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^3 f(X^{(1)}, Z^{(1,2)}) \\
1130 \quad &+ \frac{m}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^4 \\
1131 \quad &+ \frac{\binom{4}{2} m(m-1)/2}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)})^2 \\
1132 \quad & \\
1133 \quad &
\end{aligned}$$

1134 *Proof.* Let us first deal with  $\xi_{j;m_1}^{\text{UL}}$ . We have

$$\begin{aligned}
1135 \quad \mathbb{E} \xi_{j;m_1}^{(\text{UL})2} &= \frac{1}{m^4} \sum_{k_1, k_2, k_3, k_4} \mathbb{E} f(X^{(1)}, Z^{(1, k_1)}) f(X^{(1)}, Z^{(1, k_2)}) \\
1136 \quad & \\
1137 \quad & f(\tilde{X}_{-j}^{(1)}, Z_j^{(1, k_3)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1, k_4)}) \\
1138 \quad &
\end{aligned}$$

1139 where, in the sum, the indices run over  $1, \dots, m$ . We split the sum into four parts.  
1140 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  
1141 part, all the terms are equal to

$$\begin{aligned}
1142 \quad (\text{term 1}) \quad & \mathbb{E} \left( \mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \right)^2. \\
1143 \quad &
\end{aligned}$$

1144 The second part contains the  $m^2(m-1)$  terms that satisfy  $k_1 \neq k_2$  and  $k_3 = k_4$  and  
1145 that are equal to

$$\begin{aligned}
1146 \quad (\text{term 2}) \quad & \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})^2. \\
1147 \quad &
\end{aligned}$$

1148 The third part contains the  $m^2(m-1)$  terms that satisfy  $k_1 = k_2$  and  $k_3 \neq k_4$  and  
1149 that are equal to

$$\begin{aligned}
1150 \quad (\text{term 3}) \quad & \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,2)}). \\
1151 \quad &
\end{aligned}$$

1152 Finally, the fourth part contains the  $m^2$  terms that satisfy  $k_1 = k_2$  and  $k_3 = k_4$  and  
1153 are equal to

$$\begin{aligned}
1154 \quad (\text{term 4}) \quad & \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})^2. \\
1155 \quad &
\end{aligned}$$

1156 (One can see that the number of terms is  $m^4$ .) Thus,

$$\begin{aligned}
1157 \quad \mathbb{E} \xi_{m1}^{(\text{UL})2} &= (\text{term 1}) \\
1158 \quad &+ \frac{1}{m} [(\text{term 2}) + (\text{term 3}) - 2(\text{term 1})] \\
1159 \quad &+ \frac{1}{m^2} [(\text{term 1}) - (\text{term 2}) - (\text{term 3}) + (\text{term 4})]. \\
1160
\end{aligned}$$

1161 Furthermore,  $[(\text{term 1}) - (\text{term 2}) - (\text{term 3}) + (\text{term 4})]$  is equal to

$$\begin{aligned}
1162 \quad &\int \left( \mathbb{E}^{\mathbf{X}} f(x, Z) f(\tilde{x}_{-j}, Z_j) \right)^2 \\
1163 \quad &- \mathbb{E}^{\mathbf{X}} f(x, Z^{(1,1)}) f(x, Z^{(1,2)}) f(\tilde{x}_{-j}, Z_j^{(1,1)})^2 \\
1164 \quad &- \mathbb{E}^{\mathbf{X}} f(x, Z^{(1,1)})^2 f(\tilde{x}_{-j}, Z_j^{(1,1)}) f(\tilde{x}_{-j}, Z_j^{(1,2)}) \\
1165 \quad &+ \mathbb{E}^{\mathbf{X}} f(x, Z^{(1,1)})^2 f(\tilde{x}_{-j}, Z_j^{(1,1)})^2 d\mathbf{P}^*(\mathbf{x}) \\
1166 \quad &= \int \left( \mathbb{E}^{\mathbf{X}} f(x, Z) \right)^2 \left( \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j) \right)^2 \\
1167 \quad &- \left( \mathbb{E}^{\mathbf{X}} f(x, Z) \right)^2 \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)^2 \\
1168 \quad &- \mathbb{E}^{\mathbf{X}} f(x, Z)^2 \left( \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j) \right)^2 \\
1169 \quad &+ \mathbb{E}^{\mathbf{X}} f(x, Z)^2 \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)^2 d\mathbf{P}^*(\mathbf{x}) \\
1170 \quad &= \int \text{Var}^{\mathbf{X}} f(X, Z) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}, Z_j) d\mathbf{P}^*(\mathbf{x}). \\
1171
\end{aligned}$$

1172 Likewise, we find that  $[(\text{term 2}) + (\text{term 3}) - 2(\text{term 1})]$  is equal to

$$1173 \quad \mathbb{E} \text{Var}^{\mathbf{X}} f(X, Z) f(\tilde{X}_{-j}, Z_j) - \text{Var}^{\mathbf{X}} f(X, Z) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}, Z_j),$$

1175 and term 1 is  $\text{Var} \mathbb{E}^{\mathbf{X}} f(X, Z) f(\tilde{X}_{-j}, \tilde{Z}) + D_j^2$ .

1176 We now deal with  $\xi_{m1}^{\prime\prime\text{LL}}$ . We have

$$\begin{aligned}
1178 \quad \mathbb{E} \xi_{m1}^{\prime\prime\text{LL}2} &= \frac{1}{m^4} \sum_{k_1, k_2, k_3, k_4} \mathbb{E} f(X^{(1)}, Z^{(1, k_1)}) f(X^{(1)}, Z^{(1, k_2)}) \\
1179 \quad &f(X^{(1)}, Z^{(1, k_3)}) f(X^{(1)}, Z^{(1, k_4)}).
\end{aligned}$$

1181 The sum is split into five parts. The first part consists of the  $m(m-1)(m-2)(m-3)$   
1182 terms with different indices; those terms are equal to

$$1183 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) f(X^{(1)}, Z^{(1,4)}).$$

1184 The second part consists of the  $\binom{4}{2} m(m-1)(m-2)$  terms with exactly two equal  
1185 indices; those terms are equal to

$$1186 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}).$$

1187 The third part consists of the  $\binom{4}{3} m(m-1)$  terms with exactly three equal indices;  
1188 those terms are equal to

$$1189 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^3 f(X^{(1)}, Z^{(1,2)}).$$

1190 The fourth part consists of the  $m$  terms with exactly four equal indices; those terms  
1191 are equal to

$$1192 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^4.$$

1193 The fifth and last part consists of the  $\binom{4}{2}m(m-1)/2$  terms with exactly two pairs of  
1194 equal indices; those terms are equal to

$$1195 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)})^2.$$

1196 (One can check that the total number of terms is  $m^4$ .) □

1197 **LEMMA D.3** (Variances and covariances).

$$\begin{aligned}
1198 \quad (i) \quad & \text{Var } \xi_{m1}^{UL} = \text{Var } \mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1199 \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)}) \\
1200 \quad & - \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})] \\
1201 \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)}), \\
1202 \quad (ii) \quad & \text{Cov}(\xi_{m1}^{UL}, \xi_{m1}^{UR}) = \frac{m-1}{m} \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1203 \quad & + \frac{1}{m} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) - D_j \mu \\
1204 \quad (iii) \quad & \text{Cov}(\xi_{m1}^{UL}, f(X, Z)^2) = \frac{1}{m} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^3 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1205 \quad (iii) \quad & + \frac{m-1}{m} \mathbb{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 \\
1206 \quad (iv) \quad & \text{Var } \xi_{m1}^{UR} = \frac{1}{m} \text{Var } f(X, Z) \\
1207 \quad (v) \quad & \text{Cov}(\xi_{m1}^{UR}, f(X, Z)^2) = \frac{1}{m} f(X, Z)^3 \\
1208 \quad & + \frac{m-1}{m} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) - \mu \kappa \\
1209 \quad (vi) \quad & \text{Cov}(\xi_{m_n1}^{UL}, \xi_{m_n1}^{LL}) = \frac{m}{m^3} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^3 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1210 \quad & + \frac{3m(m-1)}{m^3} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1211 \quad & + \frac{m(m-1)(m-2)}{m^3} \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) \\
1212 \quad & f(X^{(1)}, Z^{(1,3)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1213 \quad & - \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1214 \quad & \left\{ \frac{1}{m} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 + \frac{m-1}{m} \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) \right\} \\
1215 \quad & \\
1216 \quad &
\end{aligned}$$

1217 *Proof.* The proof follows from direct calculations. □

## 1218 **Appendix E. Calculations for the linear model.**

1219 **LEMMA E.1.** Suppose that  $f(X, Z) = \beta_0 + \beta_{p+1}Z + \sum_{j=1}^p \beta_j X_j$  where  $X =$   
1220  $(X_1, \dots, X_p), Z_k, \tilde{Z}_{ik}$  are independent,  $\mathbb{E} X_j = \mathbb{E} Z = 0$ ,  $\mathbb{E} X_j^2 = \mathbb{E} Z^2 = 1$ ,  $\mathbb{E} X_j^3 = 0$ ,



1221  $\mathbb{E} X_j^4 = 3$ . Then the squared optimal number of repetitions is given by

$$1222 \quad (m_i^*)^2 = \frac{\beta_{p+1}^4}{(\beta_0 + \beta_i)^2 - 2\beta_0^4 + (\sum_{j=0}^p \beta_j^2)^2}$$

1223 and the discriminator (the upper-left term in (3.6) and (3.5)) is

$$1224 \quad \beta_0^2 + \beta_i^2.$$

1225

1226 *Proof.* We have

$$1227 \quad m_i^* = \frac{A_i + B_i + C_i + D_i}{E_i},$$

1228 with

$$\begin{aligned} 1229 \quad A_i &= \mathbb{E} f(X, Z_1)^2 f(\tilde{X}_{-i}, \tilde{Z}_{i1})^2 \\ 1230 \quad B_i &= \mathbb{E} f(X, Z_1) f(\tilde{X}_{-i}, \tilde{Z}_{i1}) f(X, Z_2) f(\tilde{X}_{-i}, \tilde{Z}_{i2}) \\ 1231 \quad C_i &= -\mathbb{E} f(X, Z_1)^2 f(\tilde{X}_{-i}, \tilde{Z}_{i1}) f(\tilde{X}_{-i}, \tilde{Z}_{i2}) \\ 1232 \quad D_i &= -\mathbb{E} f(\tilde{X}_{-i}, \tilde{Z}_{i1})^2 f(X, Z_1) f(X, Z_2) \\ 1233 \quad E_i &= B - [\mathbb{E} f(X, Z_1) f(\tilde{X}_{-i}, \tilde{Z}_{i1})]^2 \end{aligned}$$

1235 where  $X = (X_1, \dots, X_p)$ ,  $Z_k, \tilde{Z}_{ik}$  are independent,  $\mathbb{E} X_j = \mathbb{E} Z = 0$ ,  $\mathbb{E} X_j^2 = \mathbb{E} Z^2 = 1$ ,  
1236  $\mathbb{E} X_j^3 = 0$ ,  $\mathbb{E} X_j^4 = 3$ . We deal with the case

$$1237 \quad f(X, Z) = \beta_0 + \beta_{p+1} Z + \sum_{j=1}^p \beta_j X_j.$$

1238 We calculate the terms one by one as follows. We have

$$\begin{aligned} 1239 \quad A_j &= \mathbb{E} \left( \beta_0 + \sum_{j=1}^p \beta_j X_j \right)^2 \left( \beta_0 + \beta_i X_i + \sum_{j:1 \leq j \neq i} \beta_j \tilde{X}_j \right)^2 \\ 1240 \quad &+ \left( \beta_0 + \sum_{j=1}^p \beta_j X_j \right)^2 \beta_{p+1}^2 \tilde{Z}_{i1}^2 + \beta_{p+1}^4 Z_1^2 \tilde{Z}_{i1}^2 \\ 1241 \quad &+ \beta_{p+1}^2 Z_1^2 \left( \beta_0 + \beta_i X_i + \sum_{j:1 \leq j \neq i} \beta_j \tilde{X}_j \right)^2 \\ 1243 \quad &= A_{j1} + A_{j2} + A_{j3}, \end{aligned}$$

1244 where  $\mathbb{E}(A2) = \beta_{p+1}^4 + \beta_{p+1}^2 \sum_{j=0}^p \beta_j^2$ ,  $\mathbb{E}(A3) = \beta_{p+1}^2 \sum_{j=0}^p \beta_j^2$ . Elementary but some-  
1245 what tedious calculations yield

$$1246 \quad \mathbb{E}(A1) = \beta_0^4 + 3\beta_i^4 + 6\beta_0^2 \beta_i^2 + 2(\beta_0^2 + \beta_i^2) \sum_{j:1 \leq j \neq i} \beta_j^2 + \left( \sum_{j:1 \leq j \neq i} \beta_j^2 \right)^2.$$

1247

1248 Similar calculations show that  $B_j = A_{j1}$ ,  $C_j = -A_{j1} - A_{j3}$ ,  $D_j = -A_{j1} - A_{j3}$ ,  
 1249  $E_j = A_{j1} - (\beta_0^2 + \beta_i^2)^2$ . Thus,

$$1250 \quad (m_i^*)^2 = \frac{\beta_{p+1}^4}{(\beta_0 + \beta_i)^2 - 2\beta_0^4 + (\sum_{j=0}^p \beta_j^2)^2}. \quad \square$$

1251 **Acknowledgments.** The author thanks two anonymous referees who made use-  
 1252 ful suggestions. The author thanks Hervé Monod and Elisabeta Vergu for stimulating  
 1253 discussions from which this work arose and Bertrand Iooss for the references [10, 13].

#### 1254 REFERENCES

- 1255 [1] S. AZZI, B. SUDRET, AND J. WIART, *Sensitivity analysis for stochastic simulators using differ-*  
 1256 *ential entropy*, International Journal for Uncertainty Quantification, 10 (2020).  
 1257 [2] A. COURCOUL, H. MONOD, M. NIELEN, D. KLINKENBERG, L. HOGERWERF, F. BEAUDEAU, AND  
 1258 E. VERGU, *Modelling the effect of heterogeneity of shedding on the within herd coxiella*  
 1259 *burnetii spread and identification of key parameters by sensitivity analysis*, Journal of  
 1260 Theoretical Biology, 284 (2011), pp. 130–141.  
 1261 [3] D. J. DALEY AND J. GANI, *Epidemic Modelling*, Cambridge, 1999.  
 1262 [4] P. ETORÉ, C. PRIEUR, D. K. PHAM, AND L. LI, *Global sensitivity analysis for models described*  
 1263 *by stochastic differential equations*, Methodology and Computing in Applied Probability,  
 1264 (2019), pp. 1–29.  
 1265 [5] J.-C. FORT, T. KLEIN, AND A. LAGNOUX, *Global sensitivity analysis and Wasserstein spaces*,  
 1266 arXiv preprint arXiv:2007.12378, (2020).  
 1267 [6] F. GAMBOA, A. JANON, T. KLEIN, AND A. LAGNOUX, *Sensitivity analysis for multidimensional*  
 1268 *and functional outputs*, Electron. J. Stat., 8 (2014), pp. 575–603, [https://doi.org/10.1214/](https://doi.org/10.1214/14-EJS895)  
 1269 [14-EJS895](https://doi.org/10.1214/14-EJS895).  
 1270 [7] F. GAMBOA, A. JANON, T. KLEIN, A. LAGNOUX, AND C. PRIEUR, *Statistical inference for Sobol*  
 1271 *pick-freeze Monte Carlo method*, Statistics, 50 (2016), pp. 881–902.  
 1272 [8] F. GAMBOA, T. KLEIN, AND A. LAGNOUX, *Sensitivity analysis based on cramér–von mises*  
 1273 *distance*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 522–548.  
 1274 [9] F. GAMBOA, T. KLEIN, A. LAGNOUX, AND L. MORENO, *Sensitivity analysis in general metric*  
 1275 *spaces*, arXiv preprint arXiv:2002.04465, (2020).  
 1276 [10] J. L. HART, A. ALEXANDERIAN, AND P. A. GREMAUD, *Efficient computation of Sobol’indices*  
 1277 *for stochastic models*, SIAM Journal on Scientific Computing, 39 (2017), pp. A1514–A1530.  
 1278 [11] W. HOEFFDING, *A class of statistics with asymptotically normal distribution*, The Annals of  
 1279 Mathematical Statistics, 19 (1948), pp. 293–325.  
 1280 [12] T. HOMMA AND A. SALTELLI, *Importance measures in global sensitivity analysis of non-*  
 1281 *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)  
 1282 [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.ress.2008.09.010)  
 1291 <https://doi.org/10.1016/j.ress.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 [22] M. NAVARRO JIMENEZ, O. P. LE MAÎTRE, AND O. M. KNIO, *Global sensitivity analysis in*  
1316 *stochastic simulators of uncertain reaction networks*, The Journal of Chemical Physics, 145  
1317 (2016), p. 244106, <https://doi.org/10.1063/1.4971797>, <https://doi.org/10.1063/1.4971797>,  
1318 <https://arxiv.org/abs/https://doi.org/10.1063/1.4971797>.
- 1319 [23] C. PRIEUR AND S. TARANTOLA, *Variance-based sensitivity analysis: theory and estimation*  
1320 *algorithms*, in Handbook of Uncertainty Quantification, Springer, 2015, pp. 1–23.
- 1321 [24] L. RIMBAUD, C. BRUCHOU, S. DALLOT, D. R. J. PLEYDELL, E. JACQUOT, S. SOUBEYRAND,  
1322 AND G. THÉBAUD, *Using sensitivity analysis to identify key factors for the propaga-*  
1323 *tion of a plant epidemic*, Open Science, 5 (2018), <https://doi.org/10.1098/rsos.171435>,  
1324 <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)  
1325 [rsos.royalsocietypublishing.org/content/5/1/171435.full.pdf](https://arxiv.org/abs/http://rsos.royalsocietypublishing.org/content/5/1/171435.full.pdf).
- 1326 [25] A. SALTELLI AND I. M. SOBOL', *About the use of rank transformation in sensitivity analysis*  
1327 *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)  
1328 [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).