



**HAL**  
open science

# GLOBAL SENSITIVITY INDICES, ESTIMATORS AND TRADEOFF BETWEEN EXPLORATIONS AND REPETITIONS FOR SOME STOCHASTIC MODELS

Gildas Mazo

► **To cite this version:**

Gildas Mazo. GLOBAL SENSITIVITY INDICES, ESTIMATORS AND TRADEOFF BETWEEN EXPLORATIONS AND REPETITIONS FOR SOME STOCHASTIC MODELS. 2021. hal-02113448v4

**HAL Id: hal-02113448**

**<https://hal.science/hal-02113448v4>**

Preprint submitted on 12 Jan 2021 (v4), 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 **GLOBAL SENSITIVITY INDICES, ESTIMATORS AND TRADEOFF**  
2 **BETWEEN EXPLORATIONS AND REPETITIONS FOR SOME**  
3 **STOCHASTIC MODELS\***

4 GILDAS MAZO†

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 such 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 averaged. These averages are seen as if they  
10 were the outputs of a deterministic model and as such can feed standard softwares. 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.  
13 The problem of finding an optimal tradeoff between the repetitions and the explorations is addressed  
14 and the sensitivity indices corresponding to the above method are formally defined. A bound on  
15 an error criterion that penalizes bad rankings of the inputs is minimized. Moreover, exploiting a  
16 representation of stochastic models obtained from minimal distributional assumptions, another kind  
17 of Sobol-like sensitivity index is considered. Estimators of both kinds are constructed and their  
18 asymptotic properties studied. The theory is illustrated on numerical experiments.

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

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

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

30 Sobol's idea [26, 27] was to notice that, if  $X_1, \dots, X_p$  are drawn independently  
31 then  $f(X)$  can be decomposed into a sum of lower-dimensional functions and that this  
32 decomposition can be used to allocate the variance of the output to the individual  
33 components of the decomposition. More precisely, we have

34 (1.1) 
$$f(X) - f_0 = f_1(X_1) + \dots + f_p(X_p)$$

35 
$$+ f_{1,2}(X_1, X_2) + \dots + f_{p-1,p}(X_{p-1}, X_p)$$

36 
$$+ \dots$$

37 
$$+ f_{1,\dots,p}(X_1, \dots, X_p),$$

39 where  $f_0 = E f(X)$ ,  $f_j(X_j) = E(f(X) - f_0 | X_j)$ ,  $j = 1, \dots, p$ , and  $f_{1,2}, \dots, f_{1,\dots,p}$  are  
40 some functions defined iteratively; see [26] and [30, p. 157] for more details. The  
41 above decomposition is sometimes called the Sobol-Hoeffding decomposition. The  
42 expectations and the covariances of the individual components in the right-hand side

---

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

†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/>).

of (1.1) are zero and hence we have the variance decomposition

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

which leads to the so-called Sobol indices

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

$j = 1, \dots, p$ ; here  $\tilde{X}_{-j}$  stands for an independent copy of  $X$  where the  $j$ th component has been replaced by that of  $X$ . Thus the Sobol index associated with the  $j$ th argument of  $f$  is defined as the proportion of the total variance associated with the lower-dimensional function that depends on the  $j$ th argument only. Sobol indices are interpreted as sensitivity measures and used to achieve various goals in uncertainty quantification [24].

If the model is nonadditive (it is said that the inputs “interact” with each other) then the Sobol indices may be inadequate. To account for interactions, the so-called total sensitivity indices [10] are often computed along with Sobol indices. The total sensitivity index associated with the  $j$ th argument of  $f$  is given by

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

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

There are numerous methods to estimate the sensitivity indices. For simplicity, we describe below Sobol’s original method to estimate  $S_j$  through Monte Carlo sampling [26]. For a review of the many other methods, see [20] or the package `sensitivity` [14] of the R software for an up-to-date list of many methods, with references. Thus, draw two independent sets of inputs  $\{X^{(i)}, i = 1, \dots, n\}$ ,  $\{\tilde{X}^{(i)} := (\tilde{X}_1, \dots, \tilde{X}_p), i = 1, \dots, n\}$  and make  $p$  more sets by combining the first two:  $\{\tilde{X}_{-j}^{(i)}, i = 1, \dots, n\}$ ,  $j = 1, \dots, p$ , where

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

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

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

This gives a simple procedure to estimate all the Sobol indices  $S_1, \dots, S_p$  with  $(p+1)n$  runs of the model. The pick-freeze estimator is asymptotically normal [6, 15]. The above formula can be improved in many ways [10, 15, 19]. Many versions of this estimator exist, the goal being always to get the most efficient estimator with the least computations. Sobol indices for multivariate, functional outputs [5, 16] or functional inputs [13] have been proposed as well.

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

86 range of possible values, assumed to come from a probability distribution depending  
 87 on the input. Examples can be found in epidemiology [2, 3, 21, 25] or ecology [28], to  
 88 name a few.

89 To do the sensitivity analysis of a stochastic model, several approaches have been  
 90 investigated. In [17], to the best of my understanding, the authors carry out the sen-  
 91 sitivity analysis of a stochastic model based on a joint metamodel. In [9], a stochastic  
 92 model is seen as a functional relation of the form  $Y(\vartheta, \omega) = f(X(\vartheta), \omega)$ , where the  $X$   
 93 is a random vector on some probability space,  $\omega$  is a point in some probability space  
 94 distinct from that on which  $X$  is defined,  $f$  is some function and  $Y(\vartheta, \omega)$  is a random  
 95 variable on the induced product probability space. The quantity  $f(X(\vartheta), \omega)$  repre-  
 96 sents the output of the stochastic model run with input  $X(\vartheta)$ ; the point  $\omega$  represents  
 97 the intrinsic randomness. The idea is then to decompose the function  $\vartheta \mapsto f(X(\vartheta), \omega)$   
 98 for each  $\omega$  and estimate the associated sensitivity indices, which depend on  $\omega$ . The  
 99 estimates are then averaged over  $\omega$  to make the final sensitivity estimates. In [1],  
 100 to the best of my understanding, the stochastic model is represented as a determi-  
 101 nistic mapping which with an input associates a probability density function. The  
 102 Sobol-Hoeffding decomposition is applied to the mapping which with an input asso-  
 103 ciates the entropy of the output evaluated at that input. Here the entropy is the  
 104 Kullback-Leibler divergence of the output density. In [31], the output of the sto-  
 105 chastic model is seen as a semiparametric statistical model—the generalized lambda  
 106 distribution—with parameters depending on the inputs. These parameters have a  
 107 polynomial chaos expansion which is estimated by maximum likelihood. Once the  
 108 law of the output conditionally on the input has been estimated, its inverse cumu-  
 109 lative distribution function is used to turn the stochastic model into a deterministic  
 110 model to which standard methods are applied. In [4], the stochastic model is seen as  
 111 a mapping that goes from the input space to a space of probability measures equipped  
 112 with the Wasserstein distance. Following [7, 8], the Wasserstein space is mapped to  
 113  $\mathbf{R}$  with some family of test functions, thus allowing for a standard Sobol-Hoeffding  
 114 decomposition which is then averaged over all possible test functions.

115 In practice, although it has not been formally defined in the literature, another  
 116 method has been used for some time [2, 21, 25, 28]. The idea is simple: at each  
 117 draw of the input  $X^{(i)}$ , one produces as many outputs  $Y^{(i,1)}, \dots, Y^{(i,m)}$  as possible,  
 118 makes the average  $m^{-1} \sum_{k=1}^m Y^{(i,k)}$  and does *as if* it were the output of some deter-  
 119 ministic model. The same is done with the inputs  $\tilde{X}_{-j}^{(i)}$  (1.2) to produce the outputs  
 120  $m^{-1} \sum_{k=1}^m Y_j^{(i,k)}$ . The obtained estimator is then the same as that in (1.3) but with  
 121  $Y^{(i)}$  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)

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

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

126 To build the estimator (1.4), the stochastic model must be run  $mn(p+1)$  times.  
 127 The number  $m$  is called the number of repetitions and the number  $n$  is called the  
 128 number of explorations. If the stochastic model is computationally intensive—that  
 129 is, each model run is time-consuming—, then the estimator is built with limited  
 130 resources. In this context, an increase of  $m$  must go along with a decrease of  $n$ , and  
 131 conversely. What is then a good balance between  $m$  and  $n$ ? How to choose  $m$  and

132  $n$  such that the estimator (1.4) will be the most efficient? This question was asked  
 133 by [28].

134 We address this problem by minimizing a bound on the missranking error. The  
 135 missranking error penalizes bad rankings of the Sobol indices associated with the  
 136 inputs. This type of error leads to an explicit solution of the induced minimization  
 137 problem and hence the “optimal” pair  $(m, n)$  can be estimated. A two-step procedure  
 138 can then be implemented to get efficient estimators. We also establish the asymptotic  
 139 normality of the estimators (1.4). The sensitivity indices to which they converge are  
 140 formally defined; they are called *the sensitivity indices of the second kind*.

141 The *sensitivity indices of the first kind* are other sensitivity indices arising from  
 142 our representation of stochastic models. We regard a stochastic model as a set of  
 143 probability measures that captures how the outputs are produced by the model. We  
 144 construct a probabilistic framework compatible with these prespecified distributions  
 145 and show that the output of the stochastic model can be seen as a function of the  
 146 inputs and some “noise”. The sensitivity indices of the first kind arise from an appli-  
 147 cation of the Sobol-Hoeffding decomposition to the above-mentioned function. The  
 148 indices of the first and of the second kinds are complementary as they offer distinct  
 149 pieces of information. Interestingly, these indices can be estimated jointly with no  
 150 additional cost, the joint estimator is asymptotically normal and the two kinds of  
 151 sensitivity indices lead to the same solution for the tradeoff problem.

152 This paper is organized as follows. Section 2 defines a class of stochastic models  
 153 and discusses the connections with [9]. Section 3 defines the indices of both kinds and  
 154 their estimators. The asymptotic properties are deferred to Section 5. Section 4 intro-  
 155 duces the tradeoff problem, gives a procedure to attack it and gives some theoretical  
 156 guarantees. Section 6 illustrates the theory on numerical simulations. A Conclusion  
 157 closes the paper.

158 **2. Representations of stochastic models.** The concept of stochastic models  
 159 is intuitive and shared by many people but there are different mathematical routes  
 160 to describe them. One is given in Section 2.1. It makes minimal distributional as-  
 161 sumptions to get to a representation in terms of random variables and establishes the  
 162 existence of a function on which the Sobol-Hoeffding decomposition can be applied.  
 163 Section 2.2 makes connections with the stochastic models of [9].

164 **2.1. Representing stochastic models from minimal distributional as-**  
 165 **sumptions.** A stochastic model is some mechanism that produces outputs at ran-  
 166 dom given some inputs. Thus, a stochastic model can be seen as family of probability  
 167 measures  $\{P_x, x \in \mathcal{X}\}$  indexed by some input space  $\mathcal{X}$ . The law  $P_x$  governs how  
 168 the outputs are produced given the input  $x$ . The stochastic experiment that consists  
 169 of drawing inputs at random according to some probability measure  $P$  and observ-  
 170 ing the corresponding outputs is then fully characterized by the family  $\{P_x\}$  and the  
 171 probability measure  $P$ . This leads us to Definition 2.1.

172 **DEFINITION 2.1.** *If  $P$  is a probability measure on some input space  $\mathcal{X}$  and  $\{P_x,$   
 173  $x \in \mathcal{X}\}$  is a family of probability measures then the pair  $(P, \{P_x\})$  is called the complete  
 174 stochastic model.*

175 From now on we assume that  $\mathcal{X}$  is a subset of  $\mathbf{R}^p$  and that  $P$  is a product measure  
 176 on the product Borel  $\sigma$ -field of  $\mathbf{R}^p$ , implying that the inputs are drawn independently.  
 177 (Note the difference between the symbols “ $P$ ” and “ $P$ ”.) We also assume that the  
 178 output produced by the stochastic model is a real number: let  $P_x$  be a measure on  $\mathbf{R}$   
 179 endowed with its Borel  $\sigma$ -field.

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

182 LEMMA 2.2. *If  $(P, \{P_x\})$  is a complete stochastic model then there exist a probabil-*  
 183 *ity space  $(\Omega, \mathcal{F}, P)$ , a random vector  $(X, Z)$  on  $\Omega$  and a real function  $f$  on  $\mathcal{X} \times \text{range } Z$*   
 184 *such that*

- 185 (i)  $f(x, Z)$  is measurable for every  $x \in \mathcal{X}$ ,
  - 186 (ii)  $P(f(x, Z) \in B) = P_x(B)$  for every  $x \in \mathcal{X}$  and every Borelian  $B$ ,
  - 187 (iii)  $P(X \in A, Z \in B) = P(A)P(Z \in B)$  for every appropriate Borelians  $A$  and  $B$ .
- 188 Moreover, if  $(X, f(X, Z))$  and  $(X', f'(X', Z'))$  are two joint vectors that satisfy the  
 189 conditions (i), (ii) and (iii) then  $(X, f(X, Z)) \stackrel{d}{=} (X', f'(X', Z'))$  where  $\stackrel{d}{=}$  means  
 190 equality in distribution.

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

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

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

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

213 In sum, we have established the existence of random variables on a common  
 214 probability space and a function  $f$  that characterize the statistical experiment that  
 215 consists of drawing inputs and observing the outputs of a stochastic model. The set  
 216 of assumptions used to represent outputs and inputs of a stochastic model is minimal:  
 217 all we need is a family  $\{P_x\}$  and a probability measure  $P$ . We remark that the above  
 218 formalism of stochastic models can be used to represent physical models as well [29].

219 **2.2. Links with the stochastic models and the sensitivity indices in [9].**  
 220 In [9], the authors consider the model  $(X'(\omega'), \varphi(X'(\omega'), \omega''))$ ,  $\omega' \in \Omega'$ ,  $\omega'' \in \Omega''$ ,  
 221 where  $(\Omega', \mathcal{F}', P')$  and  $(\Omega'', \mathcal{F}'', P'')$  are probability spaces,  $X' = (X'_1, \dots, X'_p)$  is a  
 222 random vector on  $\Omega'$  and  $\varphi$  is some function. They consider the sensitivity indices

223 
$$S_j^{\text{HAG}} = \int_{\Omega''} S_j(\omega'') P''(d\omega''),$$
  
 224

225 where

$$226 \quad S_j(\omega'') = \frac{\text{Var}[\mathbb{E}(\varphi(X', \omega'')|X'_j)]}{\text{Var}[\varphi(X', \omega'')]}; \\ 227$$

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

230 One can choose a representation in Lemma 2.2 that corresponds to the models  
231 in [9]. In particular, one can recover the sensitivity indices  $S_j^{\text{HAG}}$ ,  $j = 1, \dots, p$ . Let  
232 us illustrate this with an example. Let  $(P, \{P_x\})$  be a complete stochastic model and  
233 let  $X = (X_1, \dots, X_p)$ ,  $Z$  and  $f$  be as in Lemma 2.2. Define

$$234 \quad \tilde{S}_j^{\text{HAG}} = \mathbb{E} \left( \frac{\text{Var}(\mathbb{E}[f(X, Z)|X_j, Z]|Z)}{\text{Var}(f(X, Z)|Z)} \right), \quad j = 1, \dots, p. \\ 235$$

236 Consider the model in Example 1.1 of [9], given by

$$237 \quad (2.1) \quad \varphi(X'(\omega'), \omega'') = X_1(\omega') + X_2(\omega')\omega'',$$

239 where the law of  $X'_1$  is the uniform distribution on  $(0, 1)$ , the law of  $X'_2$  is the uniform  
240 distribution on  $(1, L + 1)$ ,  $L > 0$ , and  $P''$  is the standard normal distribution on  
241  $\Omega'' = \mathbf{R}$ . The indices in Example 1.1 of [9] are given by

$$242 \quad S_1^{\text{HAG}} = \int_{\Omega''} \frac{1}{1 + L^2\omega''} P''(d\omega'') = \int_{\mathbf{R}} \frac{1}{1 + L^2w} \exp\left(-\frac{w^2}{2}\right) \frac{1}{\sqrt{2\pi}} dw \\ 243$$

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

245 To recover these indices, let us extract the induced complete stochastic model.  
246 Set  $P((0, t_1] \times (1, t_2]) = t_1(t_2 - 1)/L$  for all  $0 < t_1 < 1$ ,  $1 < t_2 < L + 1$ ,  $L > 0$  and  
247  $P_x(-\infty, t] = \Phi((t - x_1)/x_2)$  for all  $t \in \mathbf{R}$ , where  $\Phi(t) = \int_{-\infty}^t (2\pi)^{-1/2} e^{-s^2/2} ds$  and  
248  $x = (x_1, x_2) \in \mathbf{R} \times (0, \infty)$ . Now it remains to choose a representation that fulfills the  
249 conditions in Lemma 2.2 and ensures that  $S_1^{\text{HAG}} = \tilde{S}_1^{\text{HAG}}$ . Such a representation can  
250 easily be found. For instance, take  $\Omega = (0, 1)^3$  endowed with the product Lebesgue  
251 measure and put  $Z(\omega) = \omega_3$ ,  $X_1(\omega) = F_1^{-1}(\omega_1)$  and  $X_2(\omega) = F_2(\omega_2)^{-1}$  for  $\omega =$   
252  $(\omega_1, \omega_2, \omega_3) \in \Omega$ , where  $F_1(t_1) = t_1$  for  $0 < t_1 < 1$  and  $F_2(t_2) = (t_2 - 1)/L$  for  
253  $1 < t_2 < L + 1$ . Finally take  $f(x, z) = \Phi^{-1}(z)x_2 + x_1$  for  $x_1 \in \mathbf{R}$ ,  $x_2 > 0$  and  
254  $z \in (0, 1)$ . Then the conditions of Lemma 2.2 are fulfilled by construction and the  
255 detailed calculations in Appendix A show that  $S_1^{\text{HAG}} = \tilde{S}_1^{\text{HAG}}$ .

256 In sum, the stochastic models in [9] can be expressed with the framework of  
257 Section 2.1. There is however a difference between [9] and Section 2.1. In [9], the  
258 function  $f$  is fixed. It is given as being a part of the stochastic model. In our side,  
259 we only assume that we are given a set of probability measures from which we can  
260 sample. Consequently, in our case, the indices  $\tilde{S}_j^{\text{HAG}}$  are not well-defined because  
261 they depend on the law of  $Z$ , which is not characterized by Lemma 2.2. In other  
262 words, one may find different representations such that the laws of  $(X, f(X, Z))$  will  
263 be identical but the laws of  $Z$  will differ, leading to different indices  $\tilde{S}_j^{\text{HAG}}$ .

264 **3. The sensitivity indices and their estimators.** Let  $(P, \{P_x\})$  be a complete  
265 stochastic model and let  $X, Z$  and  $f$  as in Lemma 2.2. To ensure the existence  
266 of the sensitivity indices and later to derive theoretical results for the estimators, we  
267 need to assume the following: there exists some function  $F$  with  $\mathbb{E} F(X)^8 < \infty$  such  
268 that

$$269 \quad (3.1) \quad |f(X, Z)| \leq F(X)$$

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

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

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

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

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

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

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

It is important to notice that the indices of the first kind depend on the law of  $(X, f(X, Z))$  only. Since the law of the joint vector  $(X, f(X, Z))$  fully determines the complete stochastic model, the indices of the first kind can be estimated from  $(P, \{P_x\})$  even though one may not be able to evaluate the function  $f$  nor the law of  $Z$  may be characterized. Notice that it does not make sense to define the total sensitivity indices since they depend on the law of  $Z$ .

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

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

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

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

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

On the opposite to the indices of the first kind, the total sensitivity indices of the second kind are well defined, since the function  $g$  depends on  $X$  only. For instance, the  $j$ th total sensitivity index is given by

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



312 **3.1.3. Comparison of the definitions.** The sensitivity indices of the first kind  
 313 provide more refined “first-order” information than the indices of the second kind.  
 314 Example 2 and 3 illustrate this.

315 **EXAMPLE 2.** Let  $f(X, Z) = aX_1 + cX_2\phi(Z)$ , where  $X_1, X_2, Z$  are independent  
 316 standard normal variables,  $a, c$  are real coefficients and  $\phi$  is a function such that  
 317  $E\phi(Z) = 0$ . Then

$$318 \quad S'_1 = \frac{a^2}{a^2 + c^2 E\phi(Z)^2}, S'_2 = 0, S''_1 = 1 \text{ and } S''_2 = 0.$$

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

328 Example 3 returns to the model (2.1).

329 **EXAMPLE 3.** Let  $f(X, Z) = \Phi^{-1}(Z)X_2 + X_1$  such that the law of  $X_1$  and that of  
 330  $Z$  are the uniform distribution on  $(0, 1)$ , the law of  $X_2$  is the uniform distribution on  
 331  $(1, L + 1)$ ,  $L > 0$ , and  $\Phi^{-1}$  denotes the inverse distribution function of the standard  
 332 normal distribution. The detailed calculations in Appendix A show that  $S'_2 = 0$

$$333 \quad S'_1 = \frac{1}{4(L^2 + 3(L + 1)) + 1}.$$

335 As in Example 2, the sensitivity indices of the second kind do not depend on the  
 336 coefficient  $L$ . The sensitivity indices of the first kind do depend on  $L$  but note that  
 337  $S'_1 + S'_2 \leq 1/13$ , indicating that most of the contribution to the output comes from  
 338 the intrinsic randomness or the interactions.

339 In sum, both kinds of sensitivity indices provide useful insights although neither  
 340 kind is perfect. The sensitivity indices of the second kind are good indices for doing a  
 341 sensitivity analysis of the model averaged over the intrinsic randomness but by doing  
 342 so information may be lost. The sensitivity indices of the first kind provide more  
 343 refined information into the individual contributions of the inputs but the information  
 344 is only partial because the knowledge of the interactions and the intrinsic randomness  
 345 are lacking. Remember that the sensitivity indices  $\tilde{S}_j^{\text{HAG}}$  cannot be used in our case.

346 **3.2. Construction of the estimators.** We construct estimators for the indices  
 347 in Definition 3.1 and 3.2 by Monte-Carlo simulation. The input space is “explored”  
 348  $n$  times; at each exploration two independent input vectors are drawn, combined  
 349 and passed to the stochastic model which is run  $m$  times. The integer  $n$  is called  
 350 the number of explorations and the integer  $m$  is called the number of repetitions.  
 351 The couple  $(n, m)$  is called the *design* of the Monte-Carlo sampling scheme. The  
 352 total number of calls to the stochastic model is  $mn(p + 1)$ . The details are given in  
 353 Algorithm 3.1.

354 In the algorithm above,  $\tilde{X}_{-0}^{(i)} = X^{(i)}$  by convention. By assumption, the ob-  
 355 jects  $\tilde{X}^{(i)}$ ,  $\tilde{X}_{-j}^{(i)}$  and  $Y_j^{(i,k)}$ ,  $j = 0, \dots, p$ ,  $k = 1, \dots, m$ ,  $i = 1, \dots, n$ , are ran-  
 356 dom vectors such that the sets  $\{\tilde{X}^{(i)}, \tilde{X}_{-j}^{(i)}, Y_j^{(i,k)} : j = 0, \dots, p; k = 1, \dots, m\}$ ,

---

**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**

---

357  $i = 1, \dots, n$ , are i.i.d.,  $X^{(i)}$  and  $\tilde{X}^{(i)}$  are independent and  $P(\cap_{j=0}^p \cap_{k=1}^m \{Y_j^{(i,k)} \in$   
 358  $B_j^{(k)}\} | X^{(i)}, \tilde{X}^{(i)}) = \prod_{j=0}^p \prod_{k=1}^m P(Y_j^{(i,k)} \in B_j^{(k)} | X^{(i)}, \tilde{X}^{(i)})$  for all appropriate Bore-  
 359 lians  $B_j^{(k)}$ . It is easy to see that these conditions characterize the joint law of the set  
 360  $\{\tilde{X}_{-j}^{(i)}, Y_j^{(i,k)} : j = 0, \dots, p; k = 1, \dots, m; i = 1, \dots, n\}$ , that is, the inputs and the  
 361 outputs of Algorithm 3.1.

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

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

$$369 \quad (3.4) \quad \hat{S}_{j;n,m}'' = \frac{\frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \frac{1}{m} \sum_{k'=1}^m Y_j^{(i,k')} - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2}{\frac{1}{n} \sum_{i=1}^n \left( \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2 - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2},$$

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

$$371 \quad S_j'' = \frac{\mathbb{E} g(X^{(1)}) g(\tilde{X}_{-j}^{(1)}) - (\mathbb{E} g(X^{(1)}))^2}{\mathbb{E} g(X^{(1)})^2 - (\mathbb{E} g(X^{(1)}))^2}$$

(3.5)

$$372 \quad = \frac{\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}] \mathbb{E}[f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) | \tilde{X}_{-j}^{(1)}] - (\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}])^2}{\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}]^2 - (\mathbb{E} \mathbb{E}[f(X^{(1)}, Z_0^{(1,1)}) | X^{(1)}])^2}.$$

373

374 As said in the Introduction, this estimator is used implicitly by practitioners but has  
 375 not been formally studied in the literature. A simplified version with  $m = n$  appears  
 376 in [11, 12].

377 To estimate the sensitivity indices of the first kind, we exploit a formula similar

378 to (3.5). Indeed, we have

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

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

$$387 \quad (3.7) \quad \hat{S}'_{j;n,m} = \frac{\frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \frac{1}{m} \sum_{k'=1}^m Y_j^{(i,k')} - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2}{\frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)2} - \left( \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2}.$$

388 Since the estimators for the discriminator are identical, both kinds of sensitivity  
 389 indices lead to the same estimated ranking of the inputs. All of the  $2p$  estimators  
 390 can be computed with  $mn(p+1)$  runs of the stochastic model. In (3.7) and (3.4), if  
 391  $m=1$  and if the function  $f$  does in fact not depend on  $Z$ , then the estimators reduce  
 392 to Sobol estimators [26, 27] for deterministic models.

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

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

|     |     |    |    |    |    |    |     |
|-----|-----|----|----|----|----|----|-----|
| 402 | $n$ | 50 | 25 | 10 | 5  | 2  | 1   |
|     | $m$ | 1  | 2  | 5  | 10 | 25 | 50. |

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

409 **4.1. Introducing the miss-ranking error and its bound.** To compare the  
 410 estimators, a measure of performance has to be defined. We shall consider the miss-  
 411 ranking error (MRE), defined by

$$412 \quad \text{MRE} = \mathbb{E} \sum_{j=1}^p |\hat{R}_{j;n,m} - R_j|,$$

413 where  $R_j$  is the rank of  $D_j$  among  $D_1, \dots, D_p$ , that is,  $R_j = \sum_{i=1}^p \mathbf{1}(D_i \leq D_j)$ ,  
 414 and  $\widehat{R}_{j;n,m}$  is an estimator of  $R_j$ . Recall that  $D_1, \dots, D_p$  are the upper-left terms  
 415 in (3.6) and (3.5). They determine the ranks of the sensitivity indices. Recall that  
 416 the ranks of the sensitivity indices of the first kind coincide with the ranks of the  
 417 sensitivity indices of the second kind. Thus, the MRE permits to find a unique  
 418 solution for both kinds of sensitivity indices. The MRE is small when one succeeds  
 419 in ranking the inputs from the most to the least important, a task which is called  
 420 “factors prioritization” in [24, p. 52].

421 The MRE has a bound with interesting mathematical properties. Denote by  
 422  $\text{MRE}(T, m)$  the MRE based on  $T$  number of calls and  $m$  repetitions, so that the  
 423 number of explorations is  $T/(p+1)/m$ . Remember the notation of Section 3: denote  
 424  $(X^{(1)}, \widetilde{X}^{(1)}) = \mathbf{X}$ ,  $f(X^{(1)}, Z_0^{(1,1)}) = Y_0$  and  $f(\widetilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) = Y_j$ .

425 **PROPOSITION 4.1.** *Let  $\widehat{D}_{j;n,m}$ ,  $j = 1, \dots, p$ , be the upper-left term in (3.7) or (3.4)*  
 426 *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*

$$\begin{aligned}
 427 \quad \text{MRE}(T, m) &\leq \frac{L}{nm} \left( m \sum_{j=1}^p \text{Var}(\mathbb{E}[Y_0 Y_j | \mathbf{X}]) \right. \\
 &\quad + \sum_{j=1}^p \mathbb{E}(\text{Var}[Y_0 Y_j | \mathbf{X}] - \text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}]) \\
 428 &\quad \left. + \frac{1}{m} \sum_{j=1}^p \mathbb{E}(\text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}]) \right), \\
 429 & \\
 430 &
 \end{aligned}$$

431 where

$$432 \quad L = \frac{4(p-1)}{\min_{j < j'} (|D_j - D_{j'}|^2)}.$$

434 The constant  $L$  tells us that the bound is smaller when the indices are well  
 435 separated. The bound goes to zero when the number of explorations goes to infinity.  
 436 This is true even if the number of repetitions is fixed. Most interestingly, the bound  
 437 separates  $T$  and  $m$ :

$$438 \quad (4.1) \quad \text{MRE}(T, m) \leq \frac{1}{T} v(m), \quad m \in \text{div}_p(T),$$

440 where the function  $v$  is implicitly defined through Proposition 4.1. Denote by  $m_T^\dagger$   
 441 the element  $m$  in  $\text{div}_p(T)$  that minimizes  $v(m)$ . Taking  $m = m_T^\dagger$  in (4.1), we get the  
 442 bound

$$443 \quad \text{MRE}(T, m_T^\dagger) \leq \frac{v(m_T^\dagger)}{T} \leq \frac{v(m)}{T}, \quad \text{for all } m \in \text{div}_p(T).$$

444 Thus the best guarantee coincides with  $m = m_T^\dagger$  and  $n = T/(p+1)/m_T^\dagger$  in Algo-  
 445 rithm 3.1. However  $m_T^\dagger$  is unknown.

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

451 **4.2. A two-stage procedure to estimate the sensitivity indices.** The results in Section 4.1 suggest a two-stage procedure to estimate the sensitivity indices.  
 452 The procedure is given in Algorithm 4.1. The computational budget is split into two  
 453 parts  $K$  and  $T - K$ . The first  $K$  calls to the model are used to estimate  $m_{T-K}^\dagger$ . The  
 454 last  $T - K$  calls to the model are used to estimate the sensitivity indices.  
 455

---

**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).

---

456 The estimator of  $m_{T-K}^\dagger$  is built as follows. Let  $m^*$  be the minimizer of  $v$  seen as  
 457 a function on the positive reals. Since  $v$  is convex, the minimizer is unique. It follows  
 458 from (4.1) and Proposition 4.1 that

$$459 \quad (4.2) \quad m^* := \sqrt{\frac{\sum_{j=1}^p \text{E Var}[Y_0|\mathbf{X}] \text{Var}[Y_j|\mathbf{X}]}{\sum_{j=1}^p \text{Var E}[Y_0 Y_j|\mathbf{X}]}} = \sqrt{\frac{\sum_{j=1}^p \zeta_{3,j}}{\sum_{j=1}^p \zeta_{1,j}}},$$

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

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

$$464 \quad \varphi_T(x) = \begin{cases} \lfloor x \rfloor_T & \text{if } \sqrt{\lfloor x \rfloor_T \lceil x \rceil_T} > x \geq 1 \\ \lceil x \rceil_T & \text{if } \sqrt{\lfloor x \rfloor_T \lceil x \rceil_T} \leq x \leq \frac{T}{p+1} \end{cases}$$

465 where

$$466 \quad \lfloor x^* \rfloor_T = \max\{m \in \text{div}_p(T), m \leq x\}, \quad \lceil x^* \rceil_T = \min\{m \in \text{div}_p(T), m \geq x\}.$$

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

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

472 Proposition 4.3 suggests that  $m_{T-K}^\dagger$  can be estimated by applying the function  
 473  $\varphi_{T-K}$  to an estimate of  $m^*$ . Remember that  $K = m_0 n_0 (p+1)$  and put

$$474 \quad (4.3) \quad \widehat{m}_K^* := \sqrt{\frac{\sum_{j=1}^p \widehat{\zeta}_{3,j}}{\sum_{j=1}^p \widehat{\zeta}_{1,j}}},$$

477 where

$$\begin{aligned}
 & \hat{\zeta}_{3,j} = \\
 479 \quad (4.4) \quad & \frac{1}{n_0} \sum_{i=1}^n \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})^2 \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)})^2 \\
 480 \quad (4.5) \quad & + \frac{1}{n_0} \sum_{i=1}^n \left( \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)}) \right)^2 \left( \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)}) \right)^2 \\
 481 \quad (4.6) \quad & - \frac{1}{n_0} \sum_{i=1}^n \left( \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)}) \right)^2 \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)})^2 \\
 482 \quad (4.7) \quad & - \frac{1}{n_0} \sum_{i=1}^n \frac{1}{m_0} \sum_{k_1=1}^{m_0} f(X^{(i)}, Z_0^{(i,k_1)})^2 \left( \frac{1}{m_0} \sum_{k_2=1}^{m_0} f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k_2)}) \right)^2, \\
 483
 \end{aligned}$$

484 and

$$\begin{aligned}
 & \hat{\zeta}_{1,j} = \\
 485 \quad (4.8) \quad & \frac{1}{n_0} \sum_{i=1}^n \left( \frac{1}{m_0} \sum_{k=1}^{m_0} f(X^{(i)}, Z_0^{(i,k)}) f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k)}) \right)^2 \\
 487 \quad (4.9) \quad & - \left( \frac{1}{n_0} \sum_{i=1}^n \frac{1}{m_0} \sum_{k=1}^{m_0} f(X^{(i)}, Z_0^{(i,k)}) f(\tilde{X}_{-j}^{(i)}, Z_j^{(i,k)}) \right)^2. \\
 488
 \end{aligned}$$

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

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

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

$$494 \quad \sqrt{n_0} \left( \hat{m}_K^* - \left[ m^* + \frac{C}{m_0} + \epsilon_{m_0} \right] \right) \xrightarrow{d} N(0, \sigma_{m_0}^2),$$

495 where  $C$  is some constant,  $\epsilon_{m_0} = C_1/m_0^2 + \dots + C_N/m_0^{N+1}$  for some constants  
 496  $C_1, \dots, C_N$  and  $\sigma_{m_0}^2$  is some variance depending on  $m_0$ . If  $m_0 \rightarrow \infty$  then the above  
 497 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.

498 Theorem 4.4 shows that  $\hat{m}_K^*$  is asymptotically biased. The bias is polynomial in  
 499  $1/m_0$ . Corollary 4.5 shows that letting  $m_0 \rightarrow \infty$  suffices to get the consistency of  $\hat{m}_K^*$   
 500 but to get a central limit theorem centered around  $m^*$ , it is furthermore needed that  
 501  $\sqrt{n_0}/m_0 \rightarrow 0$ .

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

$$504 \quad \sqrt{n_0}(\hat{m}_K^* - m^*) \xrightarrow{d} N(0, \lim_{m_0 \rightarrow \infty} \sigma_{m_0}^2).$$

505 To estimate  $m_{T-K}^\dagger$ , put  $\hat{m}_{T-K}^\dagger = \varphi_{T-K}(\hat{m}_K^*)$ . Proposition 4.6 states that  $\hat{m}_{T-K}^\dagger$   
 506 and  $m_{T-K}^\dagger$  are equal with probability going to one.

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

$$508 \quad P\left(\widehat{m}_{T-K}^\dagger = m_{T-K}^\dagger\right) \rightarrow 1.$$

509  
510 All the details of Algorithm 4.1 have been given.

511 **4.3. Performance.** To get some insight into the performance of the procedure  
512 given in Algorithm 4.1, we look at the performance of the sensitivity indices estimators  
513 produced in Stage 2. Since they are built with  $T - K$  calls to the model with  $\widehat{m}_{T-K}^\dagger$   
514 repetitions, they satisfy

$$515 \quad (4.10) \quad \text{MRE}(T - K, \widehat{m}_{T-K}^\dagger) \leq \frac{1}{T - K} v(\widehat{m}_{T-K}^\dagger),$$

516  
517 where the left-hand side is the conditional expectation of the MRE, given the outputs  
518 produced in Stage 1. The estimator  $\widehat{m}_{T-K}^\dagger$  is computed with  $K$  calls only.

519 It is difficult to compare the guarantee above with that which got by choosing  
520 an arbitrary number of repetitions, say  $m$ . In the later case  $K = 0$  and hence the  
521 guarantee is (4.1). The denominator in (4.10) is smaller but we expect that the  
522 numerator  $v(\widehat{m}_{T-K}^\dagger)$  will be less than  $v(m)$  for many values of  $m$ . Indeed, the numer-  
523 ator should be close to  $v(m_{T-K}^\dagger)$ . If  $T - K$  is well chosen then  $v(m_{T-K}^\dagger)$  and  $v(m_T^\dagger)$   
524 should be close and since  $v(m) \leq v(m_T^\dagger)$  for all  $m$  in  $\text{div}_p(T)$ , the numerator  $v(\widehat{m}_{T-K}^\dagger)$   
525 should be an approximate minimizer. For instance if  $K$  and  $T$  are large enough and  
526  $\text{div}_p(T - K) = \{1, 2, \dots, (T - K)/(p + 1)\}$  and  $\text{div}_p(T) = \{1, 2, \dots, T/(p + 1)\}$  hold  
527 then  $v(m_{T-K}^\dagger)$  and  $v(m_T^\dagger)$  are equal. Note that the numerator and the denominator  
528 in (4.10) cannot be good at the same time and  $K$  determines the balance.

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

$$531 \quad \frac{1}{T - K} v(\widehat{m}_{T-K}^\dagger) = \frac{1}{T} v(m_{T-K}^\dagger)(1 + o_P(1)).$$

532  
533 The bound in Theorem 4.7 is the best possible guarantee inflated by a factor not  
534 much larger than one. This result is valid if  $K$  is large but not too large with regard  
535 to  $T$ . In particular, if  $\text{div}_p(T) \cap \text{div}_p(T - K) = \text{div}_p(T - K)$  then for every fixed  
536  $m \neq m_T^\dagger$ , it holds that  $P(T^{-1}v(m_{T-K}^\dagger)(1 + o_P(1)) \leq T^{-1}v(m)) \rightarrow 1$ . In other words  
537 it is always better, in terms of obtainable guarantees, to use the procedure rather than  
538 to choose the number of repetitions arbitrarily, except for the lucky case  $m = m_T^\dagger$ .

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

544 Two frameworks are considered:

- 545 •  $n \rightarrow \infty$  and  $m$  is fixed;
- 546 •  $n \rightarrow \infty$  and  $m \rightarrow \infty$ .

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

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

$$552 \quad \sqrt{n} \left( \widehat{\mathbf{S}}''_{n,m} - \mathbf{S}'' \left[ 1 - \frac{\widehat{\mathbf{S}}'_{n,m} - \mathbf{S}'}{\frac{\text{E Var}[f(X,Z)|X]}{\text{E Var}[f(X,Z)|X] + m \text{Var E}[f(X,Z)|X]}} \right] \right) \xrightarrow{d} N(0, \Xi_m),$$

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

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

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

$$560 \quad \mathbf{S}'' \frac{\text{E Var}[f(X,Z)|X]}{\text{E Var}[f(X,Z)|X] + m \text{Var E}[f(X,Z)|X]},$$

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

566 COROLLARY 5.2. Assume (3.1) holds. Let  $n \rightarrow \infty$ . If  $m \rightarrow \infty$  such that  $\sqrt{n}/m \rightarrow$   
567 0 then

$$568 \quad \sqrt{n} \left( \widehat{\mathbf{S}}''_{n,m} - \mathbf{S}'' \right) \xrightarrow{d} N(0, \Xi_{22}),$$

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

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

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

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

582 **6.1. Comparison of Monte-Carlo designs.** The effect of the number of rep-  
583 etitions on the sensitivity indices estimators and the effect of the calibration in the  
584 two-stage procedure are examined in two kinds of experiments: the “direct” experi-  
585 ments and the “calibration” experiments.

586 In the direct experiments, the sensitivity indices are estimated directly with the  
587 given number of repetitions. Increasing numbers of repetitions  $m$  are tested. (Since  
588 the budget is fixed, this goes with decreasing numbers of explorations.) For each  $m$ ,  
589 the mean squared errors (MSEs), given by  $\text{E} \sum_{j=1}^p (\widehat{S}'_{j;n,m} - S'_j)^2$  and  $\text{E} \sum_{j=1}^p (\widehat{S}''_{j;n,m} -$   
590  $S''_j)^2$ , are estimated with replications. They are also split into the sum of the squared



591 biases and the sum of the variances to get further insight about the behavior of the  
 592 estimators. The MREs are estimated as well. A normalized version is considered:  
 593 it is the MRE divided by the number of variables. For models with two inputs, the  
 594 normalized MRE is interpreted directly as the probability that the two inputs are  
 595 ranked incorrectly.

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

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

606 **6.1.1. High noise context.** The coefficients are  $\beta = 1.2$  and  $\sigma = 4$ . The  
 607 sensitivity indices are  $S'_1 = 0.05$ ,  $S'_2 = 0.08$ ,  $S''_1 = 0.41$  and  $S''_2 = 0.59$ . The real  
 608  $m^*$  is about 5.8. The total budget is  $T = 3 \times 500 = 1500$  and hence  $\text{div}_2(1500) =$   
 609  $\{1, 2, 4, 5, 10, 20, 25, 50, 100, 125, 250, 500\}$ . The integer  $m_{1500}^\dagger$  is equal to  $\varphi_{1500}(m^*) =$   
 610 5. Since the budget is kept fixed, the numbers of explorations are, respectively,  
 611 500, 250, 125, 100, 50, 25, 20, 10, 5, 4, 2, 1. The number of replications is 1500.

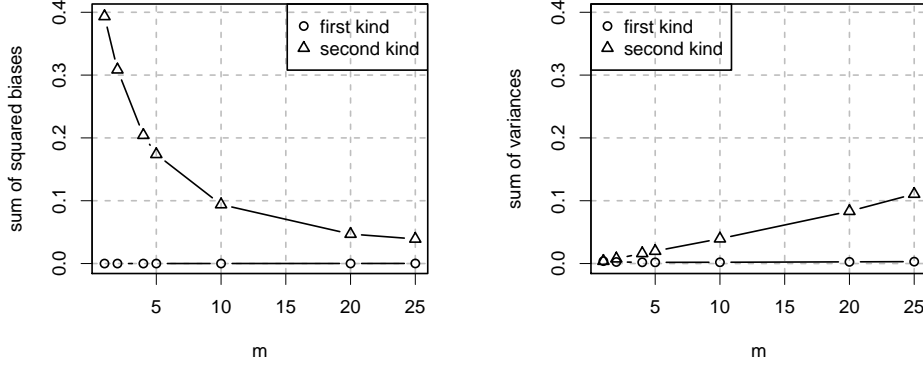
612 The results of the direct experiment are given in Figure 1 for  $m = 1, 2, 4, 5, 10,$   
 613 20, 25. The MSE of first kind does not vary with the number of repetitions and is  
 614 much lower than the MSE of second kind, see (c). The estimators of the second kind  
 615 are highly biased for small numbers of repetitions (a) and they have a higher variance  
 616 for larger numbers of repetitions (b). The fact that the bias is high for small numbers  
 617 of repetitions agrees with the theory, according to which the bias should vanish as  $m$   
 618 goes to infinity. Overall, the sensitivity indices of the second kind seem to be much  
 619 harder to estimate than the indices of the first kind, the estimators of which have a  
 620 negligible bias and a very small variance whatever the number of repetitions.

621 According to Figure 1(c), the normalized MRE curve has a banana shape with a  
 622 minimum of about slightly less than 30% reached around  $m \in \{5, 10\}$  and endpoints  
 623 with a value of about 35%. A value of 30% means that the probability of ranking  
 624 the inputs correctly is about 70%. The region of observed optimal performance  $m \in$   
 625  $\{5, 10\}$  coincides with  $m_{1500}^\dagger = 5$ , the point at which the bound is minimal.

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

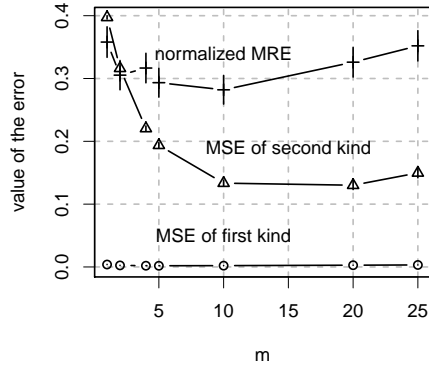
633 **6.1.2. Low noise context.** The coefficients are  $\beta = 1.2$  and  $\sigma = 0.9$ . The  
 634 sensitivity indices are  $S'_1 = 0.31$ ,  $S'_2 = 0.44$ ,  $S''_1 = 0.41$  and  $S''_2 = 0.59$ . The real  
 635  $m^*$  is about 0.30 and hence the integer  $m_{1500}^\dagger$  is equal to 1. As expected, these  
 636 numbers are smaller than the ones found in the high noise context. The total budget  
 637 is  $T = 3 \times 500 = 1500$ . The number of replications is 500.

638 The results for the direct experiment are given in Figure 2. The MSE of first  
 639 kind increases with the number of repetitions, see (c): this is due to the increase



(a) Squared bias

(b) Variance



(c) Errors

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

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

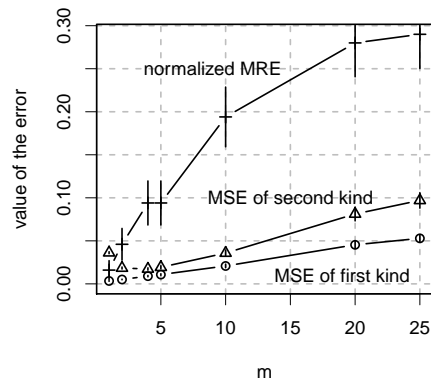
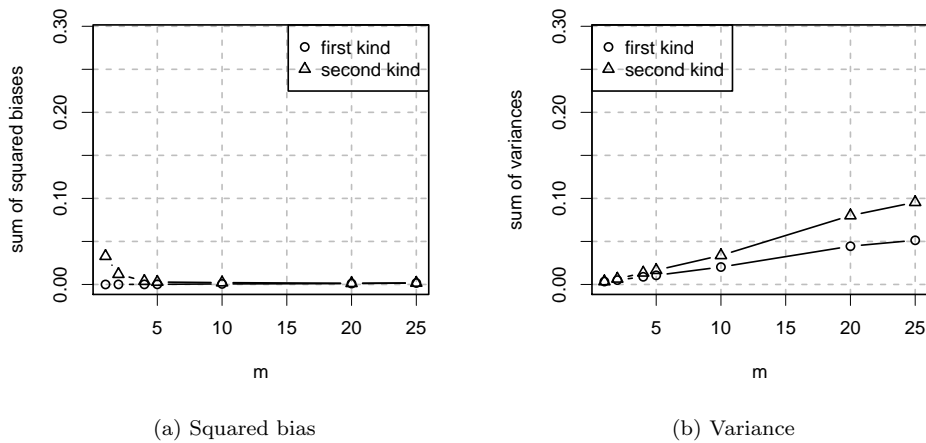


Fig. 2: Sum of squared biases (a), sum of variances (b) and errors (c) of the sensitivity indices estimators for the linear model in the low noise context. Confidence intervals of level 95% are added in (c).

640 of the variance (b), while the bias is negligible (a). As in the high noise context,  
 641 the estimators of the second kind have a decreasing bias and an increasing variance,  
 642 although the decrease of the bias is of much less magnitude. This agrees with the  
 643 theory, where we have seen that, for the sensitivity indices of the second kind, the  
 644 biases of the estimators are small when the noise of the model is low.

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

649 The results of the calibration experiment for the normalized MRE is given in

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

650 Table 2. The best performance is reached at the bottom left of the table with numbers  
 651 that correspond to the optimal performance in Figure 2 (c). Moreover, notice that a  
 652 large spectrum of calibration parameters ( $K, m_0$ ) yield low errors.

653 **6.2. Asymptotic behavior of the sensitivity indices estimators.** To illus-  
 654 trate the asymptotic behavior of the sensitivity indices estimators, Sobol's g-function,  
 655 a benchmark in sensitivity analysis [22, 18], is considered. Sobol's g-function is given  
 656 by

$$657 \quad g(U_1, \dots, U_{p+1}) = \prod_{j=1}^{p+1} \frac{|4U_j - 2| + a_j}{1 + a_j},$$

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

$$662 \quad S_j^{(a_1, \dots, a_{p+1})} = \frac{1}{3(1 + a_j)^2} \left( -1 + \prod_{j=1}^{p+1} \frac{(4/3 + a_j^2 + 2a_j)}{(1 + a_j)^2} \right)^{-1}.$$

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

$$667 \quad (6.1) \quad Y = f(X_1, \dots, X_p, Z) = g(X_1, \dots, X_{i-1}, Z, X_i, \dots, X_p).$$

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

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

$$673 \quad S'_j = \begin{cases} S_j^{(a_1, \dots, a_{p+1})} & \text{if } 1 \leq j \leq i-1 \\ S_{j+1}^{(a_1, \dots, a_{p+1})} & \text{if } i \leq j \leq p \end{cases}$$

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

$$676 \quad b_{ij} = \begin{cases} a_j & \text{if } 1 \leq j \leq i-1, \\ a_{j+1} & \text{if } i \leq j \leq p. \end{cases}$$

678 For each kind of Sobol index, we produced 500 estimates of the  $p$  Sobol indices  
 679 and computed the values of the mean squared error (MSE) by averaging over the  
 680 500 replications and summing over the  $p$  indices. We tested  $n = 100, 500, 2500$  and  
 681  $m = 1, 10, 100$ .

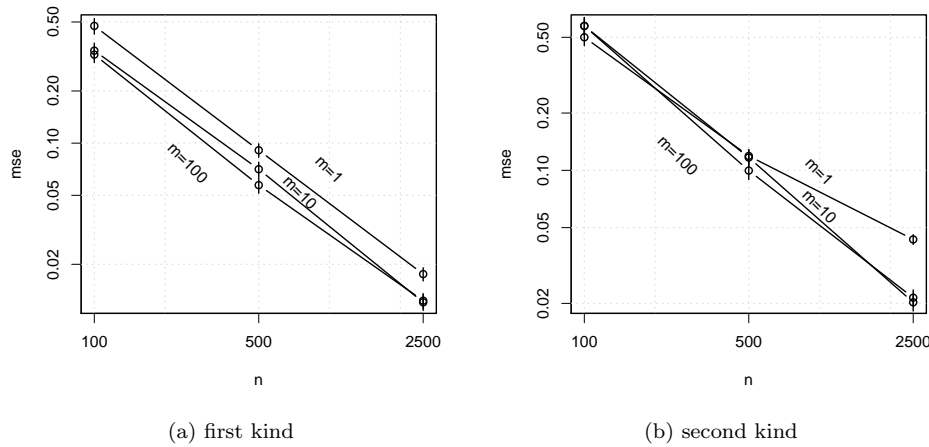


Fig. 3: MSEs for the Sobol index estimators of the first and second kind (logarithmic scale).

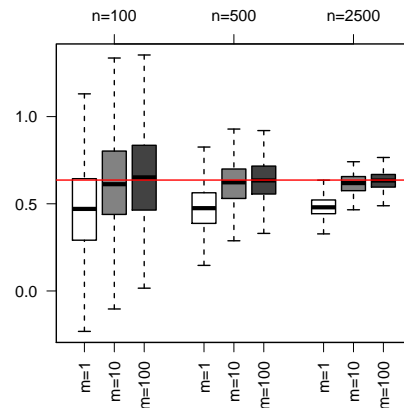


Fig. 4: Boxplots of the estimates for the Sobol index of the second kind associated with  $X_1$ . The red horizontal line is the truth.

682 The MSEs are shown in Figure 3. Let us look at 3a. As  $n$  increases, the decrease  
 683 is linear for each  $m$ . This indicates that the MSEs go to zero at a polynomial rate,  
 684 even if  $m$  is fixed (look at the line  $m = 1$ ). This agrees with the theoretical results

685 of Section 5. The picture is different for the estimator of Sobol indices of the second  
 686 kind. In 3b, the curve for  $m = 1$  is not a straight line, indicating that the MSE may  
 687 not go to zero. Indeed, the MSE for  $m$  fixed is not expected to go to zero because  
 688 of the bias depending on  $m$ . To make the MSE go to zero, one has to force  $m$  go to  
 689 infinity.

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

694 **7. Conclusion.** The practical method that consists of repeating the stochastic  
 695 model at each exploration of the input space was analysed in the context of global  
 696 sensitivity analysis. To address the problem of finding an optimal tradeoff between the  
 697 number of repetitions and the number of explorations, the MRE was considered. It  
 698 was shown that the MRE is less than some bound, the global minimum of which can be  
 699 found explicitly and used in a two-stage procedure to construct asymptotically efficient  
 700 estimators. To do this analysis, we had to formally define the sensitivity indices to  
 701 which the estimator converges, which led us to consider a definition of stochastic  
 702 models arising from minimal distributional assumptions. A link was made with the  
 703 model in [9]. Our representation of stochastic models allowed to define another kind  
 704 of sensitivity indices, called the indices of the first kind, for which estimators were  
 705 constructed. During our asymptotic analysis, it was discovered that the estimators  
 706 for the indices of the second kind are asymptotically biased, while the estimators  
 707 for the indices of the first kind are not. To test the theory, simulation experiments  
 708 were conducted, where the bias of the sensitivity estimator of the second kind was  
 709 confirmed. Optimal compromises between repetitions and explorations have been  
 710 identified and compared with the output of the two-stage procedure for different  
 711 values of the tuning parameters.

712 This work opens many research directions. First, the sensitivity estimators of the  
 713 two stages could be aggregated to build estimators with a lower variance. Second,  
 714 other methods might be developed to optimize the Monte-Carlo sampling scheme. For  
 715 instance the MSE might be approximated or asymptotic variance-covariance matrices  
 716 might be minimized. Third, multilevel Monte-Carlo sampling schemes might be con-  
 717 sidered to alleviate the bias issue. Fourth, a finite-sample analysis could be conducted  
 718 to get insight into the tradeoff  $K$  is subjected to. Fifth, since the bias is known, it  
 719 could be estimated to build bias-corrected sensitivity indices estimators. Sixth, the  
 720 problem of choosing a number of calls with many divisors must be addressed. It may  
 721 be worth to call the model a bit less if this permits to have a better set  $\text{div}_p(T)$ . Sev-  
 722 enth, the connection between our representation of stochastic models and that of [9]  
 723 could be investigated further.

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

725 **A.1. Calculations for  $\tilde{S}_1^{\text{HAG}}$ .** We have

726 
$$\tilde{S}_1^{\text{HAG}} = \mathbb{E} \left( \frac{\text{Var}(\mathbb{E}[f(X, Z)|X_j, Z]|Z)}{\text{Var}(f(X, Z)|Z)} \right) = \int_{\Omega} \frac{\text{Var}(\mathbb{E}[f(X, Z)|X_j, Z]|Z)}{\text{Var}(f(X, Z)|Z)} dP.$$

727  
728

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

$$731 \quad \widetilde{S}_1^{\text{HAG}} = \int_{(0,1)} \frac{\text{Var}(\mathbb{E}[f(X, z)|X_2, Z = z]|Z = z)}{\text{Var}(f(X, z)|Z = z)} dz = \int_{(0,1)} \frac{\text{Var}(\mathbb{E}[f(X, z)|X_1])}{\text{Var}(f(X, z))} dz.$$

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

$$734 \quad \begin{aligned} \text{Var}(f(X, z)) &= \text{Var}(\Phi^{-1}(z)X_2 + X_1) = \Phi^{-1}(z)^2 \text{Var}(X_2) + \text{Var}(X_1) \\ &= \Phi^{-1}(z)^2 \frac{L^2}{12} + \frac{1}{12}, \end{aligned}$$

737 and

$$738 \quad \begin{aligned} \text{Var}(\mathbb{E}[f(X, z)|X_1]) &= \text{Var}(\mathbb{E}[\Phi^{-1}(z)X_2 + X_1|X_1]) \\ &= \text{Var}(\Phi^{-1}(z) \mathbb{E}[X_2|X_1] + \mathbb{E}[X_1|X_1]) \\ &= \text{Var}(\Phi^{-1}(z) \mathbb{E}[X_2] + X_1) \\ &= \text{Var}(X_1) \\ &= \frac{1}{12} \end{aligned}$$

744 and hence

$$745 \quad \widetilde{S}_1^{\text{HAG}} = \int_{(0,1)} \frac{1}{\Phi^{-1}(z)^2 L^2 + 1} dz = \int_{-\infty}^{\infty} \frac{1}{z^2 L + 1} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz.$$

747 **A.2. Calculations for  $S'_1$  in Example 3.** The sensitivity index of the first  
748 kind associated with the first input is given by

$$749 \quad S'_1 = \frac{\text{Var}[\mathbb{E}(X_1 + X_2 \Phi^{-1}(Z)|X_1)]}{\text{Var}[X_1 + X_2 \Phi^{-1}(Z)]}.$$

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

$$754 \quad \begin{aligned} \text{Var}[X_2 \Phi^{-1}(Z)] &= \text{Var}[\mathbb{E}(X_2 \Phi^{-1}(Z)|Z)] + \mathbb{E}(\text{Var}[X_2 \Phi^{-1}(Z)|Z]) \\ &= \text{Var} \left[ \Phi^{-1}(Z) \left( \frac{L}{2} + 1 \right) \right] + \int_0^1 \Phi^{-1}(z)^2 \text{Var}[X_2] dz \\ &= \left( \frac{L}{2} + 1 \right)^2 + \frac{L^2}{12}, \end{aligned}$$

758 so that

$$759 \quad S'_1 = \frac{1/12}{1/12 + (L/2 + 1)^2 + L^2/12} = \frac{1}{4(L^2 + 3(L + 1)) + 1}.$$

761 **Appendix B. Proofs.**

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

$$770 \quad P(f(x, Z) \leq t)$$

$$771 \quad = P(Z \leq F_x(t)) = \lambda^{\otimes_{j=1}^{p+1}}\{\omega \in \Omega : \omega_{p+1} \leq F_x(t)\} = \lambda(0, F_x(t)] = F_x(t).$$

773 Finally, by the same token,

$$774 \quad P(X_1 \leq t_1, \dots, X_p \leq t_p, Z \leq t_{p+1})$$

$$775 \quad = P\{\omega : \omega_1 \leq F_1(t_1), \dots, \omega_p \leq F_p(t_p), \omega_{p+1} \leq t_{p+1}\} = t_{p+1} \prod_{j=1}^p F_j(t_j).$$

777 The proof is complete.

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

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

$$782 \quad P(\widehat{D}_i - \widehat{D}_j \geq 0) \leq \frac{\text{Var } \widehat{D}_i + \text{Var } \widehat{D}_j}{\frac{1}{2}|D_i - D_j|^2}.$$

784 *Proof. We have*

$$785 \quad P(\widehat{D}_i - \widehat{D}_j \geq 0) \leq P(|\widehat{D}_i - D_i| + |\widehat{D}_j - D_j| \geq D_j - D_i)$$

$$786 \quad \leq P(|\widehat{D}_i - D_i|^2 + |\widehat{D}_j - D_j|^2 \geq \frac{1}{2}|D_j - D_i|^2)$$

788

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

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

$$791 \quad \sum_{i=1}^p \mathbf{E}|\widehat{R}_i - R_i| \leq \sum_{i=1}^p \sum_{j=1}^p \mathbf{E}|\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i) - \mathbf{1}(D_j \leq D_i)|$$

$$792 \quad \leq \sum_{i=1}^p \sum_{j \neq i} \frac{\text{Var } \widehat{D}_i + \text{Var } \widehat{D}_j}{\frac{1}{2}|D_i - D_j|^2}$$

$$793 \quad \leq \frac{4(p-1)}{\min_{j < j'} |D_j - D_{j'}|^2} \sum_{i=1}^p \text{Var } \widehat{D}_i,$$

794

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

$$796 \quad \mathbf{E}|\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i) - \mathbf{1}(D_j \leq D_i)| = \begin{cases} \mathbf{E}|\mathbf{1}(\widehat{D}_j > \widehat{D}_i)| & \text{if } j < i, \\ 0 & \text{if } j = i, \\ \mathbf{E}|\mathbf{1}(\widehat{D}_j \leq \widehat{D}_i)| & \text{if } j > i. \end{cases}$$

797



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

$$\begin{aligned} 800 \quad \text{Var } \widehat{D}_j &= \frac{1}{n} \{ \text{Var } E[Y_0 Y_j | \mathbf{X}] + \frac{1}{m} (E \text{Var}[Y_0 Y_j | \mathbf{X}] - \text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}]) \\ 801 &\quad + \frac{1}{m^2} E \text{Var}[Y_0 | \mathbf{X}] \text{Var}[Y_j | \mathbf{X}] \}. \end{aligned}$$

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

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

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

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

$$812 \quad \varphi_{T-K}(m^*) = \begin{cases} \llcorner m^* \lrcorner_{T-K} & \text{if } \sqrt{\llcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}} > m^* \\ 813 \quad \lrcorner m^* \lrcorner_{T-K} & \text{if } \sqrt{\llcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}} \leq m^*. \end{cases}$$

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

$$817 \quad v(\llcorner m^* \lrcorner_{T-K}) < v(\lrcorner m^* \lrcorner_{T-K}) \text{ iff } \sqrt{\llcorner m^* \lrcorner_{T-K} \lrcorner m^* \lrcorner_{T-K}} > \frac{\zeta_3}{\zeta_1} = m^*.$$

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

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

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

$$826 \quad \widehat{m}_K^* = \sqrt{\frac{\sum_{j=1}^p \widehat{\zeta}_{3,j}}{\sum_{j=1}^p \widehat{\zeta}_{1,j}}} = \sqrt{\frac{\sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.4)} + \xi_{j;m,i}^{(4.5)} - \xi_{j;m,i}^{(4.6)} - \xi_{j;m,i}^{(4.7)}}{\sum_{j=1}^p \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.8)} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.9)} \right)^2}},$$

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

$$\begin{aligned} 830 \quad \bar{\xi} &= \frac{1}{n} \sum_{i=1}^n \xi_{m,i}, \\ 831 \quad \xi_{m,i} &= (\xi_{1;m,i}^\top, \dots, \xi_{p;m,i}^\top)^\top, \quad i = 1, \dots, n, \\ 832 \quad \xi_{j;m,i} &= (\xi_{j;m,i}^{(4.4)}, \dots, \xi_{j;m,i}^{(4.9)})^\top, \quad j = 1, \dots, p, \quad i = 1, \dots, n. \end{aligned}$$

834 Let  $s$  be the function defined by

$$835 \quad s(\mathbf{x}) = \sqrt{\frac{\sum_{j=1}^p x_j^{(4.4)} + x_j^{(4.5)} - x_j^{(4.6)} - x_j^{(4.7)}}{\sum_{j=1}^p x_j^{(4.8)} - x_j^{(4.9)2}},$$

836

837 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  
 838 notation we have  $\widehat{m}_K^* = s(\bar{\boldsymbol{\xi}})$ . Moreover, elementary calculations show that

$$839 \quad (\text{B.1}) \quad \mathbf{E} \boldsymbol{\xi}_{m,1} = \boldsymbol{\theta} + \sum_{\nu=1}^4 \frac{\mathbf{C}_\nu}{m^\nu},$$

840

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

$$842 \quad \boldsymbol{\theta}_j = \mathbf{E} \begin{pmatrix} Y_0^{(1,1)2} Y_j^{(1,1)2} \\ Y_0^{(1,1)} Y_0^{(1,2)} Y_j^{(1,1)} Y_j^{(1,2)} \\ Y_0^{(1,1)} Y_0^{(1,2)} Y_j^{(1,1)2} \\ Y_j^{(1,1)} Y_j^{(1,2)} Y_0^{(1,1)2} \\ Y_0^{(1,1)} Y_0^{(1,2)} Y_j^{(1,1)} Y_j^{(1,2)} \\ Y_j^{(1,1)} Y_0^{(1,1)} \end{pmatrix}.$$

843

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

$$845 \quad \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbf{E} \boldsymbol{\xi}_{m,1})^\top \dot{s}(\mathbf{E} \boldsymbol{\xi}_{m,1}) + \frac{1}{2}(\bar{\boldsymbol{\xi}} - \mathbf{E} \boldsymbol{\xi}_{m,1})^\top \ddot{s}_{n,m}(\bar{\boldsymbol{\xi}} - \mathbf{E} \boldsymbol{\xi}_{m,1})$$

846 (B.2)  $= \sqrt{n}(s(\bar{\boldsymbol{\xi}}) - s(\mathbf{E} \boldsymbol{\xi}_{m,1}))$

$$847 \quad = \sqrt{n}(s(\bar{\boldsymbol{\xi}}) - s(\boldsymbol{\theta}) - (\mathbf{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \dot{s}(\boldsymbol{\theta}) - \frac{1}{2}(\mathbf{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \ddot{s}_m(\mathbf{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})),$$

848

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

$$853 \quad \epsilon_m = \frac{1}{2}(\mathbf{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta})^\top \ddot{s}_m(\mathbf{E} \boldsymbol{\xi}_{m,1} - \boldsymbol{\theta}) + \sum_{\nu=2}^4 \frac{C_\nu}{m^\nu},$$

854

855 it follows from (B.2) that

$$857 \quad (\text{B.3}) \quad \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbf{E} \boldsymbol{\xi}_{m,1})^\top \dot{s}(\mathbf{E} \boldsymbol{\xi}_{m,1}) + \frac{1}{2}(\bar{\boldsymbol{\xi}} - \mathbf{E} \boldsymbol{\xi}_{m,1})^\top \ddot{s}_{n,m}(\bar{\boldsymbol{\xi}} - \mathbf{E} \boldsymbol{\xi}_{m,1})$$

$$858 \quad = \sqrt{n}(\widehat{m}_K^* - m^* - \frac{C_1}{m} - \epsilon_m).$$

859

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

$$861 \quad \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbf{E} \boldsymbol{\xi}_{m,1}) \rightarrow N(0, \Sigma_m),$$

862

863 for some variance-covariance matrix  $\Sigma_m$  of size  $6p \times 6p$ . Moreover, the second term in  
 864 the left-hand side of (B.3) is  $o_P(1)$  by Cauchy-Schwartz's inequality and the continuity

865 of the second derivatives of  $s$ . The first term goes to  $N(0, \dot{s}(\mathbf{E}\boldsymbol{\xi}_{m,1})^\top \Sigma_m \dot{s}(\mathbf{E}\boldsymbol{\xi}_{m,1}))$   
 866 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  $C = C_1$ .

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

$$868 \quad \sqrt{n}(\bar{\boldsymbol{\xi}} - \mathbf{E}\boldsymbol{\xi}_{m,1}) \rightarrow N(0, \lim_{m \rightarrow \infty} \Sigma_m).$$

870 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.  
 871 The proof is complete.

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

$$875 \quad \text{div}_p(T-K) \setminus \left\{ 1, \frac{T-K}{p+1} \right\}.$$

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

$$878 \quad \mathcal{D}_{T-K} \subset \{ \sqrt{ij} : i \text{ and } j \text{ are two consecutive integers} \} = \mathcal{E}.$$

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

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

$$887 \quad P \left( \left| \frac{\frac{1}{T-K} v(\hat{m}_{T-K}^\dagger) - \frac{1}{T} v(m_{T-K}^\dagger)}{\frac{1}{T} v(m_{T-K}^\dagger)} > \varepsilon \right| \right)$$

$$888 \quad \leq P \left( \left| \frac{T}{T-K} (\hat{m}_{T-K}^\dagger - m_{T-K}^\dagger) v'(\tilde{m}) + \frac{K}{T-K} v(m_{T-K}^\dagger) \right| > v(m_{T-K}^\dagger) \varepsilon \right),$$

890 where  $\tilde{m}$  denotes a real between  $\hat{m}_{T-K}^\dagger$  and  $m_{T-K}^\dagger$ . A decomposition of the probability  
 891 above according to whether  $\hat{m}_{T-K}^\dagger - m_{T-K}^\dagger \neq 0$  or  $\hat{m}_{T-K}^\dagger - m_{T-K}^\dagger = 0$  yields the  
 892 bound

$$893 \quad P \left( \hat{m}_{T-K}^\dagger - m_{T-K}^\dagger \neq 0 \right) + P \left( \frac{K}{T-K} > \varepsilon \right).$$

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

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

$$899 \quad \hat{S}_{j;n,m}' = \frac{\frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{\text{UL}} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}} \right)^2}{\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{LL}} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}} \right)^2}, \quad j = 1, \dots, p,$$

901 and

$$902 \quad \hat{S}_{j;n,m}'' = \frac{\frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{\text{UL}} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}} \right)^2}{\frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{LL}} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{\text{UR}} \right)^2}, \quad j = 1, \dots, p,$$

903

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

$$906 \quad \mathbf{E} \boldsymbol{\xi}_{m,1} = \begin{pmatrix} \mathbf{E} \mathbf{E}[f(X, Z)|X] \mathbf{E}[f(\tilde{X}_{-1}, Z)|\tilde{X}_{-1}] \\ \vdots \\ \mathbf{E} \mathbf{E}[f(X, Z)|X] \mathbf{E}[f(\tilde{X}_{-p}, Z)|\tilde{X}_{-p}] \\ \mathbf{E} f(X, Z) \\ \mathbf{E} f(X, Z)^2 \\ \mathbf{E} \mathbf{E}[f(X, Z)|X]^2 + \frac{\mathbf{E} \text{Var}[f(X, Z)|X]}{m} \end{pmatrix}.$$

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

$$908 \quad s(x_1, \dots, x_p, x_{p+1}, x_{p+2}, x_{p+3}) \\ 909 \\ 910 \quad = \left( \frac{x_1 - x_{p+1}^2}{x_{p+2} - x_{p+1}^2}, \dots, \frac{x_p - x_{p+1}^2}{x_{p+2} - x_{p+1}^2}, \frac{x_1 - x_{p+1}^2}{x_{p+3} - x_{p+1}^2}, \dots, \frac{x_p - x_{p+1}^2}{x_{p+3} - x_{p+1}^2} \right).$$

912 Clearly, we have

$$913 \quad s \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} \right) = \begin{pmatrix} \hat{\mathbf{S}}'_{n,m} \\ \hat{\mathbf{S}}''_{n,m} \end{pmatrix}$$

914 and

$$915 \quad s(\mathbf{E} \boldsymbol{\xi}_{m,1}) = \begin{pmatrix} \mathbf{S}' \\ \mathbf{S}'' \left[ 1 - \frac{\mathbf{S}'}{\mathbf{E} \text{Var}[f(X, Z)|X] + m \text{Var} \mathbf{E}[f(X, Z)|X]} \right] \end{pmatrix}.$$

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

$$917 \quad \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} - \mathbf{E} \boldsymbol{\xi}_{m,1} \right) \xrightarrow{d} N(0, \Sigma_m),$$

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

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

$$922 \quad \sqrt{n} \left( s \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} \right) - s(\mathbf{E} \boldsymbol{\xi}_{m,1}) \right) \\ 923 \quad = \sqrt{n} \left( \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} - \mathbf{E} \boldsymbol{\xi}_{m,1} \right) \dot{s}_m \right. \\ 924 \quad \left. + \frac{1}{2} \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} - \mathbf{E} \boldsymbol{\xi}_{m,1} \right)^\top \ddot{s}_{n,m} \left( \frac{1}{n} \sum_{i=1}^n \boldsymbol{\xi}_{m,i} - \mathbf{E} \boldsymbol{\xi}_{m,1} \right) \right), \\ 925$$

926 where  $\dot{s}_m$  is the gradient of  $s$  at  $\mathbf{E} \boldsymbol{\xi}_{m,1}$  and  $\ddot{s}_{n,m}$  is the Hessian matrix of  $s$  at a  
 927 point between  $n^{-1} \sum_i \boldsymbol{\xi}_{m,i}$  and  $\mathbf{E} \boldsymbol{\xi}_{m,1}$ . Since that point goes to a constant and  $s$  has

928 continuous second derivatives, it holds that  $\ddot{s}_{n,m}$  goes to a constant as well. So does  
 929  $\dot{s}_m$  and the claim follows by Slutsky's lemma.

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

$$932 \quad (C.1) \quad \frac{1}{n} \sum_{i=1}^n \xi_{m,i},$$

933 with

$$934 \quad (C.2) \quad \xi_{m,i} = \prod_{l=1}^L \frac{1}{m} \sum_{k=1}^m \prod_{j=0}^p Y_j^{(i,k) b_{j;l}},$$

936 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  
 937  $b_{j;l}$ ,  $j = 0, \dots, p$ ,  $l = 1, \dots, L$ , are nonnegative coefficients. The coefficients are  
 938 arranged in a matrix  $(b_{j;l})$  with  $L$  rows and  $p+1$  columns, where  $b_{j;l}$  is the element in  
 939 the  $l$ th row and  $(j+1)$ th column. This way, all estimators of the form (C.1) and (C.2),  
 940 or, equivalently, all summands (C.2), can be represented by a matrix. We sometimes  
 941 write  $\xi_{m,i} \simeq (b_{j;l})$ , where  $(b_{j;l})$  is the matrix of size  $L \times (p+1)$  with coefficients  $b_{j;l}$ ,  
 942  $j = 0, \dots, p$ ,  $l = 1, \dots, L$ .

943 **C.1. Examples.** The estimator

$$944 \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \frac{1}{m} \sum_{k'=1}^m Y_j^{(i,k')}$$

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

$$947 \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{pmatrix},$$

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

$$950 \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)}, \quad \frac{1}{n} \sum_{i=1}^n \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)2},$$

$$951 \quad \frac{1}{n} \sum_{i=1}^n \left( \frac{1}{m} \sum_{k=1}^m Y_0^{(i,k)} \right)^2$$

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

$$954 \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

$$955 \quad \begin{pmatrix} 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 & 0 & 0 & \cdots & 0 \end{pmatrix},$$

956 respectively.

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

$$960 \quad \widehat{\zeta}_{3,j} = \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.4)} + \xi_{j;m,i}^{(4.5)} - \xi_{j;m,i}^{(4.6)} - \xi_{j;m,i}^{(4.7)} \quad \text{and,}$$

$$961 \quad \widehat{\zeta}_{1,j} = \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.8)} - \left( \frac{1}{n} \sum_{i=1}^n \xi_{j;m,i}^{(4.9)} \right)^2,$$

962 where

$$963 \quad \begin{aligned} 964 \quad & \xi_{j;m,i}^{(4.4)}, & \xi_{j;m,i}^{(4.5)} \\ 965 \quad & \xi_{j;m,i}^{(4.6)}, & \xi_{j;m,i}^{(4.7)} \\ 966 \quad & \xi_{j;m,i}^{(4.8)}, & \xi_{j;m,i}^{(4.9)} \end{aligned}$$

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

$$968 \quad \begin{pmatrix} 2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \end{pmatrix},$$

$$969 \quad \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 2 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

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

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

$$972 \quad \begin{pmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

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

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

$$976 \quad \xi_{1;m,i}^{\text{UL}} \simeq \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \end{pmatrix}, \dots, \xi_{p;m,i}^{\text{UL}} \simeq \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 1 \end{pmatrix}$$

977 for the upper left (UL) terms,

$$978 \quad \xi_{m,i}^{\text{UR}} \simeq \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

979 for the upper right (UR) term,

$$980 \quad \xi_{m,i}^{\text{LL}} \simeq \begin{pmatrix} 2 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \end{pmatrix}$$

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

$$982 \quad \xi_{m,i}^{\text{LL}} \simeq \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 0 & \dots & 0 \end{pmatrix}$$

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

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

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

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

$$993 \begin{pmatrix} b_{0;1} & \cdots & b_{p;1} \\ \vdots & & \vdots \\ b_{0;L} & \cdots & b_{p;L} \\ b'_{0;1} & \cdots & b'_{p;1} \\ \vdots & & \vdots \\ b'_{0;L} & \cdots & b'_{p;L} \end{pmatrix}.$$

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

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

$$997 |\xi| \leq \left( \bigvee_{j=0}^p F_j(\mathbf{X}^{(i)}) \right)^{\sum_{j=0}^p \sum_{l=1}^L b_{j;l}},$$

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

1000 *Proof.* The proof of (i) is trivial. Let us prove (ii). We have

$$\begin{aligned} 1001 \mathbb{E}\xi &= \frac{1}{m^L} \sum_{(k_1, \dots, k_L) \in \{1, \dots, m\}^L} \mathbb{E} \prod_{l=1}^L \prod_{j=0}^p Y_j^{(1, k_l) b_{j;l}} \\ 1002 &= \frac{1}{m^L} \sum_{(k_1, \dots, k_L) \in \{1, \dots, m\}^L} \mathbb{E} \mathbb{E} \left( \prod_{l=1}^L \prod_{j=0}^p Y_j^{(1, k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right) \\ 1003 \text{ (C.3)} &= \frac{1}{m^L} \sum_{(k_1, \dots, k_L) \in \{1, \dots, m\}^L} \mathbb{E} \prod_{j=0}^p \mathbb{E} \left( \prod_{l=1}^L Y_j^{(1, k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right). \\ 1004 \end{aligned}$$

1005 Since (i)  $\mathbf{X}^{(1)}$  and  $\{\mathbf{Z}^{(1, k)}, k = 1, \dots, m\}$  are independent and (ii) the law of

$$1006 (\mathbf{Z}^{(1, k_1)}, \dots, \mathbf{Z}^{(1, k_L)})$$

1007 is invariant through any permutation of distinct  $k_1, \dots, k_L$ , all the inner expectations  
 1008 in (C.3) are equal to some others. For if  $k_1, \dots, k_L$  are distinct then

$$1009 \mathbb{E} \left( \prod_{l=1}^L Y_j^{(1, k_l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right) = \mathbb{E} \left( \prod_{l=1}^L Y_j^{(1, l) b_{j;l}} \middle| \mathbf{X}^{(1)} \right)$$

1010

1011 for all  $j = 0, \dots, p$ . The number of inner expectations equal to the one above is  
 1012  $m(m-1) \cdots (m-L+1)$ , a polynomial in  $m$  with degree  $L$ . If some components of  
 1013 the tuple  $(k_1, \dots, k_L)$  are equal, then we can always write

$$1014 \quad \mathbb{E} \left( \prod_{l=1}^L Y_j^{(1, k_l) b_{jl}} \middle| \mathbf{X}^{(1)} \right) = \mathbb{E} \left( \prod_{l=1}^{L'} Y_j^{(1, l) \beta_{j;l}} \middle| \mathbf{X}^{(1)} \right)$$

1016 for some  $L' \leq L$  and coefficients  $\beta_{jl}$ . It is easy to see that the number of inner expect-  
 1017 ations equal to the one above is a polynomial in  $m$  with degree at most  $L$ . (Looking  
 1018 at examples helps to see this; see e.g. the proof of Lemma D.2 in Appendix D.)  
 1019 Therefore, the sum in (C.3) is also a polynomial in  $m$  with degree at most  $L$  and the  
 1020 claim follows ( $\mathbb{E} \xi$  can be zero). To prove (iii), simply remember that, by assumption,  
 1021  $|Y^{(1, k)}| \leq F(X^{(1)})$  and  $|Y_j^{(1, k)}| \leq F(\tilde{X}_{-j}^{(1)})$  for all  $k$  and all  $j$ .  $\square$

1022 Two frameworks are considered:

- 1023 •  $n \rightarrow \infty$  and  $m$  is fixed;
- 1024 •  $n \rightarrow \infty$  and  $m \rightarrow \infty$ .

1025 In the second framework  $m_n$  is a sequence indexed by  $n$  that goes to infinity as  $n$  goes  
 1026 to infinity.

1027 LEMMA C.2. Let  $\xi_{m,i}^{(I)}$ ,  $I = 1, \dots, N$ , be elements of  $\mathcal{E}_{m,i}(L)$  with coefficients  
 1028  $(b_{j;l}^{(I)})$ . Assume

$$1029 \quad \mathbb{E} F(X^{(1)})^2 \sum_{j=0}^p \sum_{l=1}^L b_{j;l}^{(I)} < \infty$$

1030 for all  $I = 1, \dots, N$ . Let  $n \rightarrow \infty$ . If  $m$  is fixed then

$$1031 \quad \sqrt{n} \left[ \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{(1)} - \mathbb{E} \xi_{m,1}^{(1)}, \dots, \frac{1}{n} \sum_{i=1}^n \xi_{m,i}^{(N)} - \mathbb{E} \xi_{m,1}^{(N)} \right]^\top \xrightarrow{d} N(0, \Sigma_m),$$

1033 where  $\Sigma_m$  is the variance-covariance matrix of  $\boldsymbol{\xi}_{m,i} = (\xi_{m,i}^{(1)}, \dots, \xi_{m,i}^{(N)})^\top$ . If  $m \rightarrow$   
 1034  $\infty$  then  $\lim_{m \rightarrow \infty} \Sigma_m$  exists elementwise and the above display with  $\Sigma_m$  replaced by  
 1035  $\lim_{m \rightarrow \infty} \Sigma_m$  is true.

1036 *Proof.* Let  $m$  be fixed. By Proposition C.1 (i),  $\xi_{m,i}^{(I)2}$ ,  $I = 1, \dots, N$ , belongs to  
 1037  $\mathcal{E}_{m,i}(2L)$  and has coefficients

$$1038 \quad \xi_{m,i}^{(I)2} \simeq \begin{pmatrix} b_{0;1}^{(I)} & \cdots & b_{p;1}^{(I)} \\ \vdots & & \vdots \\ b_{0;L}^{(I)} & \cdots & b_{p;L}^{(I)} \\ b_{0;1}^{(I)} & \cdots & b_{p;1}^{(I)} \\ \vdots & & \vdots \\ b_{0;L}^{(I)} & \cdots & b_{p;L}^{(I)} \end{pmatrix}.$$

1040 Thus, denoting  $\sum_{j=0}^p \sum_{l=1}^L b_{j;l}^{(I)}$  by  $\beta$ , Proposition C.1 (iii) yields

$$1041 \quad (C.4) \quad \xi_{m,i}^{(I)2} \leq \bigvee_{j=0}^p F_j(\mathbf{X}^{(i)})^{2\beta}$$

1042



1043 and hence

$$1044 \quad \mathbb{E} \xi_{m,i}^{(I)2} \leq \mathbb{E} \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{2\beta} \leq (p+1) \mathbb{E} \left( F(X^{(1)}) \right)^{2\beta} < \infty.$$

1045

1046 Therefore we can apply the central limit theorem to finish the proof for  $m$  fixed.

1047 Let  $m \rightarrow \infty$ . According to Lindeberg-Feller's central limit theorem (see e.g. [30]),  
1048 it suffices to show

1049 (i) for all  $\epsilon > 0$ ,

$$1050 \quad \sum_{i=1}^n \mathbb{E} \left\| \frac{1}{\sqrt{n}} \boldsymbol{\xi}_{m,i} \right\|^2 \mathbf{1} \left\{ \left\| \frac{1}{\sqrt{n}} \boldsymbol{\xi}_{m,i} \right\| > \epsilon \right\} \rightarrow 0,$$

1051

1052 and

1053 (ii) the limit  $\sum_{i=1}^n \text{Cov}(\boldsymbol{\xi}_{m,i}/\sqrt{n})$  exists and is finite.

1054 Let us show (i). Denoting  $\mathbf{X} = (X^{(1)}, \tilde{X}^{(1)})$ , we have

$$1055 \quad \sum_{i=1}^n \mathbb{E} \left\| \frac{\boldsymbol{\xi}_{m,i}}{\sqrt{n}} \right\|^2 \mathbf{1} \{ \|\boldsymbol{\xi}_{m,i}\| > \sqrt{n}\epsilon \} = \mathbb{E} \|\boldsymbol{\xi}_{m,1}\|^2 \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \}$$

$$1056 \quad = \mathbb{E} \sum_{I=1}^N \xi_{m,1}^{(I)2} \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \}$$

$$1057 \quad = \sum_{I=1}^N \mathbb{E} \left[ \mathbb{E} \left( \xi_{m,1}^{(I)2} \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \} \mid \mathbf{X} \right) \right].$$

1058

1059 By (C.4), we have

$$1060 \quad \mathbb{E} \left( \xi_{m,1}^{(I)2} \mathbf{1} \{ \|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \} \mid \mathbf{X} \right) \leq \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{2\beta} P(\|\boldsymbol{\xi}_{m,1}\| > \sqrt{n}\epsilon \mid \mathbf{X})$$

$$1061 \quad \leq \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{2\beta} \frac{\sum_{I=1}^N \mathbb{E} \left( \xi_{m,1}^{(I)2} \mid \mathbf{X} \right)}{n\epsilon^2}$$

$$1062 \quad \leq \frac{N \bigvee_{j=0}^p F_j(\mathbf{X}^{(1)})^{4\beta}}{n\epsilon^2},$$

1063

1064 where the last inequality holds by using (C.4) once more. The upper bound goes to  
1065 zero and is dominated by an integrable function. Thus, we can apply the dominated  
1066 convergence theorem to complete the proof.

1067 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  
1068 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  
1069  $\mathbb{E} \xi_{m,1}^{(I)2} < \infty$ ,  $I = 1, \dots, N$ , and hence  $\mathbb{E} \xi_{m,1}^{(I)} \xi_{m,1}^{(J)} \leq \mathbb{E} \xi_{m,1}^{(I)2}/2 + \xi_{m,1}^{(J)2}/2 < \infty$ . Therefore  
1070 the limit of  $\text{Cov} \boldsymbol{\xi}_{m,1}$  exists and is finite. The proof is complete.  $\square$

1071 **Appendix D. Explicit moment calculations.** Explicit moment calculations  
1072 are given for the summands in the proof of Theorem 5.1. In this section,  $\mathbb{E} f(X, Z)$   
1073 and  $\mathbb{E} \mathbb{E}[f(X, Z) \mid X]^2$  are denoted by  $\mu$  and  $D$ , respectively. Recall that the upper-left  
1074 term in (3.6) and (3.5) is denoted by  $D_j$ . The moments are given in Lemma D.1  
1075 and Lemma D.2. The variances and covariances are given in Lemma D.3. Let  $\mathbf{X} =$

1076  $(X^{(1)}, \tilde{X}^{(1)})$ . Whenever there is a superscript  $\mathbf{X}$  added to the expectation symbol  $\mathbb{E}$   
 1077 or the variance symbol  $\text{Var}$ , this means that these operators are to be understood  
 1078 conditionally on  $\mathbf{X}$ . An integral with respect to  $P(d\mathbf{x})$  means that we integrate with  
 1079 respect to the law of  $\mathbf{X}$ .

1080 LEMMA D.1 (Moments of order 1). *The moments of order 1 are given by*

$$\begin{aligned} 1081 \quad & \mathbb{E} \xi_{j;m1}^{UL} = D_j, \\ 1082 \quad & \mathbb{E} \xi_{m1}^{UR} = \mu, \\ 1083 \quad & \mathbb{E} \xi_{m1}^{LL} = \frac{1}{m} \mathbb{E} \text{Var}^X f(X^{(1)}, Z^{(1,1)}) + D. \\ 1084 \end{aligned}$$

1085

1086 *Proof.* One has

$$\begin{aligned} 1087 \quad \mathbb{E} \xi_{j;m1}^{UL} &= \frac{1}{m^2} \sum_{k,k'} \mathbb{E} f(X^{(1)}, Z^{(1,k)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,k')}) \\ 1088 \quad &= \frac{1}{m^2} \sum_{k,k'} \int \mathbb{E} f(x, Z^{(1,k)}) f(\tilde{x}_{-j}, Z_j^{(1,k')}) P(d\mathbf{x}) \\ 1089 \quad &= \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\ 1090 \quad &= D_j, \end{aligned}$$

1092 where the integral is taken with respect to the law of  $\mathbf{x} = (x, \tilde{x})$ , and,

$$\begin{aligned} 1093 \quad \mathbb{E} \xi_{m1}^{LL} &= \frac{1}{m^2} \sum_{k,k'} \mathbb{E} f(X^{(1)}, Z^{(1,k)}) f(X^{(1)}, Z^{(1,k')}) \\ 1094 \quad &= \frac{1}{m} \mathbb{E} \text{Var}^X f(X, Z) + \mathbb{E}(\mathbb{E}^X f(X, Z))^2 \\ 1095 \quad &= \frac{1}{m} \mathbb{E} \text{Var}^X f(X, Z) + D. \\ 1096 \end{aligned}$$

1097 The proof for  $\xi_{m1}^{UR}$  is similar. □

1098 LEMMA D.2 (Moments of order 2). *The moments of order 2 are given by*

$$\begin{aligned}
1099 \quad \mathbb{E} \xi_{j;m_1}^{(UL)2} &= \text{Var} \mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) + D_j^2 \\
1100 \quad &+ \frac{1}{m} [\mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1101 \quad &- \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})] \\
1102 \quad &+ \frac{1}{m^2} \mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}), \\
1103 \quad \mathbb{E} \xi_{m_1}^{(UR)2} &= \frac{1}{m} \mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) + \mathbb{E} (\mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}))^2, \\
1104 \quad \mathbb{E} \xi_{m_1}^{\mu(LL)2} &= \frac{m(m-1)(m-2)(m-3)}{m^4} \\
1105 \quad &\mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) f(X^{(1)}, Z^{(1,4)}) \\
1106 \quad &+ \frac{\binom{4}{2} m(m-1)(m-2)}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) \\
1107 \quad &+ \frac{\binom{4}{3} m(m-1)}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^3 f(X^{(1)}, Z^{(1,2)}) \\
1108 \quad &+ \frac{m}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^4 \\
1109 \quad &+ \frac{\binom{4}{2} m(m-1)/2}{m^4} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)})^2 \\
1110 \quad & \\
1111 \quad &
\end{aligned}$$

1112 *Proof.* Let us first deal with  $\xi_{j;m_1}^{\text{UL}}$ . We have

$$\begin{aligned}
1113 \quad \mathbb{E} \xi_{j;m_1}^{(\text{UL})2} &= \frac{1}{m^4} \sum_{k_1, k_2, k_3, k_4} \mathbb{E} f(X^{(1)}, Z^{(1, k_1)}) f(X^{(1)}, Z^{(1, k_2)}) \\
1114 \quad & \\
1115 \quad & f(\tilde{X}_{-j}^{(1)}, Z_j^{(1, k_3)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1, k_4)}) \\
1116 \quad &
\end{aligned}$$

1117 where, in the sum, the indices run over  $1, \dots, m$ . We split the sum into four parts.  
1118 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  
1119 part, all the terms are equal to

$$\begin{aligned}
1120 \quad (\text{term 1}) \quad & \mathbb{E} \left( \mathbb{E}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \right)^2. \\
1121 \quad &
\end{aligned}$$

1122 The second part contains the  $m^2(m-1)$  terms that satisfy  $k_1 \neq k_2$  and  $k_3 = k_4$  and  
1123 that are equal to

$$\begin{aligned}
1124 \quad (\text{term 2}) \quad & \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})^2. \\
1125 \quad &
\end{aligned}$$

1126 The third part contains the  $m^2(m-1)$  terms that satisfy  $k_1 = k_2$  and  $k_3 \neq k_4$  and  
1127 that are equal to

$$\begin{aligned}
1128 \quad (\text{term 3}) \quad & \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,2)}). \\
1129 \quad &
\end{aligned}$$

1130 Finally, the fourth part contains the  $m^2$  terms that satisfy  $k_1 = k_2$  and  $k_3 = k_4$  and  
1131 are equal to

$$\begin{aligned}
1132 \quad (\text{term 4}) \quad & \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})^2. \\
1133 \quad &
\end{aligned}$$

1134 (One can see that the number of terms is  $m^4$ .) Thus,

$$\begin{aligned}
 1135 \quad \mathbb{E} \xi_{m1}^{(\text{UL})2} &= (\text{term 1}) \\
 1136 \quad &+ \frac{1}{m} [(\text{term 2}) + (\text{term 3}) - 2(\text{term 1})] \\
 1137 \quad &+ \frac{1}{m^2} [(\text{term 1}) - (\text{term 2}) - (\text{term 3}) + (\text{term 4})]. \\
 1138
 \end{aligned}$$

1139 Furthermore,  $[(\text{term 1}) - (\text{term 2}) - (\text{term 3}) + (\text{term 4})]$  is equal to

$$\begin{aligned}
 1140 \quad &\int \left( \mathbb{E}^{\mathbf{X}} f(x, Z) f(\tilde{x}_{-j}, Z_j) \right)^2 \\
 1141 \quad &- \mathbb{E}^{\mathbf{X}} f(x, Z^{(1,1)}) f(x, Z^{(1,2)}) f(\tilde{x}_{-j}, Z_j^{(1,1)})^2 \\
 1142 \quad &- \mathbb{E}^{\mathbf{X}} f(x, Z^{(1,1)})^2 f(\tilde{x}_{-j}, Z_j^{(1,1)}) f(\tilde{x}_{-j}, Z_j^{(1,2)}) \\
 1143 \quad &+ \mathbb{E}^{\mathbf{X}} f(x, Z^{(1,1)})^2 f(\tilde{x}_{-j}, Z_j^{(1,1)})^2 dP(\mathbf{x}) \\
 1144 \quad &= \int \left( \mathbb{E}^{\mathbf{X}} f(x, Z) \right)^2 \left( \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j) \right)^2 \\
 1145 \quad &- \left( \mathbb{E}^{\mathbf{X}} f(x, Z) \right)^2 \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)^2 \\
 1146 \quad &- \mathbb{E}^{\mathbf{X}} f(x, Z)^2 \left( \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j) \right)^2 \\
 1147 \quad &+ \mathbb{E}^{\mathbf{X}} f(x, Z)^2 \mathbb{E}^{\mathbf{X}} f(\tilde{x}_{-j}, Z_j)^2 dP(\mathbf{x}) \\
 1148 \quad &= \int \text{Var}^{\mathbf{X}} f(X, Z) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}, Z_j) dP(\mathbf{x}). \\
 1149
 \end{aligned}$$

1150 Likewise, we find that  $[(\text{term 2}) + (\text{term 3}) - 2(\text{term 1})]$  is equal to

$$1151 \quad \mathbb{E} \text{Var}^{\mathbf{X}} f(X, Z) f(\tilde{X}_{-j}, Z_j) - \text{Var}^{\mathbf{X}} f(X, Z) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}, Z_j),$$

1153 and term 1 is  $\text{Var} \mathbb{E}^{\mathbf{X}} f(X, Z) f(\tilde{X}_{-j}, \tilde{Z}) + D_j^2$ .

1154 We now deal with  $\xi_{m1}^{\text{LL}}$ . We have

$$\begin{aligned}
 1155 \quad \mathbb{E} \xi_{m1}^{\text{LL}2} &= \frac{1}{m^4} \sum_{k_1, k_2, k_3, k_4} \mathbb{E} f(X^{(1)}, Z^{(1, k_1)}) f(X^{(1)}, Z^{(1, k_2)}) \\
 1156 \quad &f(X^{(1)}, Z^{(1, k_3)}) f(X^{(1)}, Z^{(1, k_4)}). \\
 1157
 \end{aligned}$$

1159 The sum is split into five parts. The first part consists of the  $m(m-1)(m-2)(m-3)$   
 1160 terms with different indices; those terms are equal to

$$1161 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}) f(X^{(1)}, Z^{(1,4)}).$$

1162 The second part consists of the  $\binom{4}{2} m(m-1)(m-2)$  terms with exactly two equal  
 1163 indices; those terms are equal to

$$1164 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) f(X^{(1)}, Z^{(1,3)}).$$

1165 The third part consists of the  $\binom{4}{3} m(m-1)$  terms with exactly three equal indices;  
 1166 those terms are equal to

$$1167 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^3 f(X^{(1)}, Z^{(1,2)}).$$

1168 The fourth part consists of the  $m$  terms with exactly four equal indices; those terms  
1169 are equal to

$$1170 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^4.$$

1171 The fifth and last part consists of the  $\binom{4}{2}m(m-1)/2$  terms with exactly two pairs of  
1172 equal indices; those terms are equal to

$$1173 \quad \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)})^2.$$

1174 (One can check that the total number of terms is  $m^4$ .) □

1175 LEMMA D.3 (Variances and covariances).

$$\begin{aligned}
1176 \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)}) \\
1177 & + \frac{1}{m} [\mathbb{E} \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1178 & - \text{Var}^{\mathbf{X}} f(X^{(1)}, Z^{(1,1)}) \text{Var}^{\mathbf{X}} f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)})] \\
1179 & + \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)}), \\
1180 \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)}) \\
1181 & + \frac{1}{m} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) - D_j \mu \\
1182 \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)}) \\
1183 \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 \\
1184 \quad (iv) \quad & \text{Var } \xi_{m1}^{UR} = \frac{1}{m} \text{Var } f(X, Z) \\
1185 \quad (v) \quad & \text{Cov}(\xi_{m1}^{UR}, f(X, Z)^2) = \frac{1}{m} f(X, Z)^3 \\
1186 & + \frac{m-1}{m} \mathbb{E} f(X^{(1)}, Z^{(1,1)})^2 f(X^{(1)}, Z^{(1,2)}) - \mu \kappa \\
1187 \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)}) \\
1188 & + \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)}) \\
1189 & + \frac{m(m-1)(m-2)}{m^3} \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(X^{(1)}, Z^{(1,2)}) \\
1190 & f(X^{(1)}, Z^{(1,3)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1191 & - \mathbb{E} f(X^{(1)}, Z^{(1,1)}) f(\tilde{X}_{-j}^{(1)}, Z_j^{(1,1)}) \\
1192 & \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\} \\
1193 & \\
1194 &
\end{aligned}$$

1195 *Proof.* The proof follows from direct calculations. □

### 1196 Appendix E. Calculations for the linear model.

1197 LEMMA E.1. Suppose that  $f(X, Z) = \beta_0 + \beta_{p+1}Z + \sum_{j=1}^p \beta_j X_j$  where  $X =$   
1198  $(X_1, \dots, X_p), Z_k, \tilde{Z}_{ik}$  are independent,  $\mathbb{E} X_j = \mathbb{E} Z = 0$ ,  $\mathbb{E} X_j^2 = \mathbb{E} Z^2 = 1$ ,  $\mathbb{E} X_j^3 = 0$ ,

1199  $\mathbb{E} X_j^4 = 3$ . Then the squared optimal number of repetitions is given by

$$1200 \quad (m_i^*)^2 = \frac{\beta_{p+1}^4}{(\beta_0 + \beta_i)^2 - 2\beta_0^4 + (\sum_{j=0}^p \beta_j^2)^2}$$

1201 and the discriminator (the upper-left term in (3.6) and (3.5)) is

$$1202 \quad \beta_0^2 + \beta_i^2.$$

1203

1204 *Proof.* We have

$$1205 \quad m_i^* = \frac{A_i + B_i + C_i + D_i}{E_i},$$

1206 with

$$\begin{aligned} 1207 \quad A_i &= \mathbb{E} f(X, Z_1)^2 f(\tilde{X}_{-i}, \tilde{Z}_{i1})^2 \\ 1208 \quad B_i &= \mathbb{E} f(X, Z_1) f(\tilde{X}_{-i}, \tilde{Z}_{i1}) f(X, Z_2) f(\tilde{X}_{-i}, \tilde{Z}_{i2}) \\ 1209 \quad C_i &= -\mathbb{E} f(X, Z_1)^2 f(\tilde{X}_{-i}, \tilde{Z}_{i1}) f(\tilde{X}_{-i}, \tilde{Z}_{i2}) \\ 1210 \quad D_i &= -\mathbb{E} f(\tilde{X}_{-i}, \tilde{Z}_{i1})^2 f(X, Z_1) f(X, Z_2) \\ 1211 \quad E_i &= B - [\mathbb{E} f(X, Z_1) f(\tilde{X}_{-i}, \tilde{Z}_{i1})]^2 \end{aligned}$$

1213 where  $X = (X_1, \dots, X_p)$ ,  $Z_k, \tilde{Z}_{ik}$  are independent,  $\mathbb{E} X_j = \mathbb{E} Z = 0$ ,  $\mathbb{E} X_j^2 = \mathbb{E} Z^2 = 1$ ,  
1214  $\mathbb{E} X_j^3 = 0$ ,  $\mathbb{E} X_j^4 = 3$ . We deal with the case

$$1215 \quad f(X, Z) = \beta_0 + \beta_{p+1} Z + \sum_{j=1}^p \beta_j X_j.$$

1216 We calculate the terms one by one as follows. We have

$$\begin{aligned} 1217 \quad A_j &= \mathbb{E} \left( \beta_0 + \sum_{j=1}^p \beta_j X_j \right)^2 \left( \beta_0 + \beta_i X_i + \sum_{j:1 \leq j \neq i} \beta_j \tilde{X}_j \right)^2 \\ 1218 \quad &+ \left( \beta_0 + \sum_{j=1}^p \beta_j X_j \right)^2 \beta_{p+1}^2 \tilde{Z}_{i1}^2 + \beta_{p+1}^4 Z_1^2 \tilde{Z}_{i1}^2 \\ 1219 \quad &+ \beta_{p+1}^2 Z_1^2 \left( \beta_0 + \beta_i X_i + \sum_{j:1 \leq j \neq i} \beta_j \tilde{X}_j \right)^2 \\ 1220 \quad &= A_{j1} + A_{j2} + A_{j3}, \end{aligned}$$

1222 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-  
1223 what tedious calculations yield

$$1224 \quad \mathbb{E}(A1) = \beta_0^4 + 3\beta_i^4 + 6\beta_0^2 \beta_i^2 + 2(\beta_0^2 + \beta_i^2) \sum_{j:1 \leq j \neq i} \beta_j^2 + \left( \sum_{j:1 \leq j \neq i} \beta_j^2 \right)^2.$$

1225

1226 Similar calculations show that  $B_j = A_{j1}$ ,  $C_j = -A_{j1} - A_{j3}$ ,  $D_j = -A_{j1} - A_{j3}$ ,  
 1227  $E_j = A_{j1} - (\beta_0^2 + \beta_i^2)^2$ . Thus,

$$1228 \quad (m_i^*)^2 = \frac{\beta_{p+1}^4}{(\beta_0 + \beta_i)^2 - 2\beta_0^4 + (\sum_{j=0}^p \beta_j^2)^2}. \quad \square$$

1229 **Acknowledgments.** The author thanks two anonymous referees who made use-  
 1230 ful suggestions. The author thanks Hervé Monod and Elisabeta Vergu for stimulating  
 1231 discussions from which this work arose and Bertrand Iooss for the references [9, 11].

## 1232 REFERENCES

- 1233 [1] S. AZZI, B. SUDRET, AND J. WIART, *Sensitivity analysis for stochastic simulators using differ-*  
 1234 *ential entropy*, International Journal for Uncertainty Quantification, 10 (2020).  
 1235 [2] A. COURCOUL, H. MONOD, M. NIELEN, D. KLINCKENBERG, L. HOGERWERF, F. BEAUDEAU, AND  
 1236 E. VERGU, *Modelling the effect of heterogeneity of shedding on the within herd coxiella*  
 1237 *burnetii spread and identification of key parameters by sensitivity analysis*, Journal of  
 1238 Theoretical Biology, 284 (2011), pp. 130–141.  
 1239 [3] D. J. DALEY AND J. GANI, *Epidemic Modelling*, Cambridge, 1999.  
 1240 [4] J.-C. FORT, T. KLEIN, AND A. LAGNOUX, *Global sensitivity analysis and Wasserstein spaces*,  
 1241 arXiv preprint arXiv:2007.12378, (2020).  
 1242 [5] F. GAMBOA, A. JANON, T. KLEIN, AND A. LAGNOUX, *Sensitivity analysis for multidimensional*  
 1243 *and functional outputs*, Electron. J. Stat., 8 (2014), pp. 575–603, [https://doi.org/10.1214/](https://doi.org/10.1214/14-EJS895)  
 1244 [14-EJS895](https://doi.org/10.1214/14-EJS895).  
 1245 [6] F. GAMBOA, A. JANON, T. KLEIN, A. LAGNOUX, AND C. PRIEUR, *Statistical inference for Sobol*  
 1246 *pick-freeze Monte Carlo method*, Statistics, 50 (2016), pp. 881–902.  
 1247 [7] F. GAMBOA, T. KLEIN, AND A. LAGNOUX, *Sensitivity analysis based on cramér–von mises*  
 1248 *distance*, SIAM/ASA Journal on Uncertainty Quantification, 6 (2018), pp. 522–548.  
 1249 [8] F. GAMBOA, T. KLEIN, A. LAGNOUX, AND L. MORENO, *Sensitivity analysis in general metric*  
 1250 *spaces*, arXiv preprint arXiv:2002.04465, (2020).  
 1251 [9] J. L. HART, A. ALEXANDERIAN, AND P. A. GREMAUD, *Efficient computation of Sobol’indices*  
 1252 *for stochastic models*, SIAM Journal on Scientific Computing, 39 (2017), pp. A1514–A1530.  
 1253 [10] T. HOMMA AND A. SALTELLI, *Importance measures in global sensitivity analysis of non-*  
 1254 *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)  
 1255 [science/article/pii/0951832096000026](http://www.sciencedirect.com/science/article/pii/0951832096000026).  
 1256 [11] B. IOOSS, L. L. GRATIET, A. LAGNOUX, AND T. KLEIN, *Sensitivity analysis for stochastic*  
 1257 *computer codes: Theory and estimation methods*, tech. report, EDF R&D, 2014.  
 1258 [12] B. IOOSS, T. KLEIN, AND A. LAGNOUX, *Sobol’ sensitivity analysis for stochastic numerical*  
 1259 *codes*, in 8th International Conference Sensitivity Analysis of Model Output, 2016, pp. 48–  
 1260 49.  
 1261 [13] B. IOOSS AND M. RIBATET, *Global sensitivity analysis of computer models with functional in-*  
 1262 *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)  
 1263 <https://doi.org/10.1016/j.ress.2008.09.010>, [http://www.sciencedirect.com/science/article/](http://www.sciencedirect.com/science/article/pii/S0951832008002299)  
 1264 [pii/S0951832008002299](http://www.sciencedirect.com/science/article/pii/S0951832008002299).  
 1265 [14] B. IOOSS, S. D. VEIGA, A. JANON, G. PUJOL, WITH CONTRIBUTIONS FROM BAPTISTE BROTO,  
 1266 K. BOUMHAOUT, T. DELAGE, R. E. AMRI, J. FRUTH, L. RENT GILQUIN, J. GUILLAUME,  
 1267 L. L. GRATIET, P. LEMAITRE, A. MARREL, A. MEY-NAOUI, B. L. NELSON, F. IPPO  
 1268 MONARI, R. OOMEN, O. RAKOVEC, B. RAMOS, O. ROUSTANT, E. HYE SONG, J. STAUM,  
 1269 R. SUEUR, T. TOUATI, AND F. WEBER, *sensitivity*, 2020, [https://cran.r-project.org/](https://cran.r-project.org/package=sensitivity)  
 1270 [package=sensitivity](https://cran.r-project.org/package=sensitivity). R package version 1.23.0.  
 1271 [15] A. JANON, T. KLEIN, A. LAGNOUX, M. NODET, AND C. PRIEUR, *Asymptotic normality and*  
 1272 *efficiency of two Sobol index estimators*, ESAIM: Probability and Statistics, 18 (2014),  
 1273 pp. 342–364, <https://doi.org/10.1051/ps/2013040>.  
 1274 [16] M. LAMBONI, H. MONOD, AND D. MAKOWSKI, *Multivariate sensitivity analysis to measure*  
 1275 *global contribution of input factors in dynamic models*, Reliability Engineering and System  
 1276 Safety, 96 (2010), pp. 450–459.  
 1277 [17] A. MARREL, B. IOOSS, S. DA VEIGA, AND M. RIBATET, *Global sensitivity analysis of stochastic*  
 1278 *computer models with joint metamodells*, Statistics and Computing, 22 (2012), pp. 833–847,  
 1279 <https://doi.org/10.1007/s11222-011-9274-8>, <https://doi.org/10.1007/s11222-011-9274-8>.  
 1280

- 1281 [18] A. MARREL, B. IOOSS, B. LAURENT, AND O. ROUSTANT, *Calculations of sobol indices for*  
1282 *the gaussian process metamodel*, Reliability Engineering & System Safety, 94 (2009),  
1283 pp. 742 – 751, <https://doi.org/https://doi.org/10.1016/j.res.2008.07.008>, <http://www.sciencedirect.com/science/article/pii/S0951832008001981>.
- 1284 [19] H. MONOD, C. NAUD, AND D. MAKOWSKI, *Working with dynamic crop models: Evaluation,*  
1285 *analysis, parameterization, and applications*, in Uncertainty and sensitivity analysis for  
1286 crop models, Elsevier, 2006, pp. 55–100.
- 1287 [20] C. PRIEUR AND S. TARANTOLA, *Variance-based sensitivity analysis: theory and estimation*  
1288 *algorithms*, in Handbook of Uncertainty Quantification, Springer, 2015, pp. 1–23.
- 1289 [21] L. RIMBAUD, C. BRUCHOU, S. DALLOT, D. R. J. PLEYDELL, E. JACQUOT, S. SOUBEYRAND,  
1290 AND G. THÉBAUD, *Using sensitivity analysis to identify key factors for the propaga-*  
1291 *tion of a plant epidemic*, Open Science, 5 (2018), <https://doi.org/10.1098/rsos.171435>,  
1292 <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)  
1293 [rsos.royalsocietypublishing.org/content/5/1/171435.full.pdf](https://arxiv.org/abs/http://rsos.royalsocietypublishing.org/content/5/1/171435.full.pdf).
- 1294 [22] A. SALTELLI AND I. M. SOBOL', *About the use of rank transformation in sensitivity analysis*  
1295 *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/science/article/pii/0951832095000992>.
- 1296 [23] A. SALTELLI, S. TARANTOLA, AND F. CAMPOLONGO, *Sensitivity analysis as an ingredient of*  
1297 *modeling*, Statistical Science, 15 (2000), pp. 377–395.
- 1300 [24] A. SALTELLI, S. TARANTOLA, F. CAMPOLONGO, AND M. RATTO, *Sensitivity analysis in practice*,  
1301 Wiley, 2004.
- 1302 [25] J. F. SAVALL, C. BIDOT, M. LEBLANC-MARIDOR, C. BELLOC, AND S. TOUZEAU, *Modelling*  
1303 *salmonella transmission among pigs from farm to slaughterhouse: interplay between man-*  
1304 *agement variability and epidemiological uncertainty*, International Journal of Food Micro-  
1305 biology, 229 (2016), pp. 33–43.
- 1306 [26] I. M. SOBOL, *Sensitivity estimates for nonlinear mathematical models*, Mathematical modelling  
1307 and computational experiments, 1 (1993), pp. 407–414.
- 1308 [27] I. M. SOBOL, *Global sensitivity indices for nonlinear mathematical models and their Monte*  
1309 *Carlo estimates*, Mathematics and computers in simulation, 55 (2001), pp. 271–280.
- 1310 [28] M. SPENCE, *Statistical issues in ecological simulation models*, PhD thesis, University of  
1311 Sheffield, 2015. <http://etheses.whiterose.ac.uk/10517>.
- 1312 [29] L. SZEWCZYK, J.-L. GRIMAUD, AND I. COJAN, *Experimental evidence for bifurcation angles*  
1313 *control on abandoned channel fill geometry*, Earth Surface Dynamics, 8 (2020), pp. 275–288,  
1314 <https://doi.org/10.5194/esurf-8-275-2020>, <https://hal.archives-ouvertes.fr/hal-02562670>.
- 1315 [30] A. W. VAN DER VAART, *Asymptotic Statistics*, Cambridge University Press, 1998.
- 1316 [31] X. ZHU AND B. SUDRET, *Global sensitivity analysis for stochastic simulators based on general-*  
1317 *ized lambda surrogate models*, arXiv preprint arXiv:2005.01309, (2020).
- 1318