Local Polynomial Estimation for Sensitivity Analysis on Models With Correlated Inputs

Sebastien DA VEIGA
Institut Francais du Petrole
92500 Rueil-Malmaison
France
(sebastien.da-veiga@ifp.fr)

Francois WAHL
Institut Francais du Petrole
69360 Solaize
France
(francois.wahl@ifp.fr)

Fabrice GAMBOA
Institut de Mathematiques
Universite de Toulouse III
31000 Toulouse
France
(Gambo@CICT.fr)

Sensitivity indexes when the inputs of a model are not independent are derived from local polynomial techniques. Two original estimators based on local polynomial smoothers are proposed. Both have good theoretical properties, which are illustrated through analytical examples. Comparison with the Bayesian approach developed by Oakley and O’Hagan (2004) is also performed. The two proposed estimators are used to carry out a sensitivity analysis on two real case models with correlated parameters.

KEY WORDS: Conditional moments estimation; Global sensitivity index; Nonparametric regression.

Achieving better knowledge of refining processes usually calls for a kinetic model in which the inputs are the detailed description of both the components introduced in the unit and the operating conditions and the outputs are the components produced in the yield. Such a model is based on the choice of a reaction mechanism depending on various parameters (e.g., kinetic constants). But the complexity of the mechanism, the variability in the behaviors of catalysts when used and the sparsity of observations and experiments imply that most often these parameters cannot be inferred from theoretical considerations and need to be estimated through practical experiments. As a result, there is uncertainty in parameter values that propagates through to model predictions. This can be highly problematic in real situations. It is then essential to quantify this uncertainty and to study its influence on model outputs through uncertainty and sensitivity analysis.

Over the last several decades, much effort in mathematical analysis of physical processes has focused on modeling and reasoning with uncertainty and sensitivity. Model calibration and validation are examples of areas in which sensitivity and uncertainty analysis have become essential investigative scientific tools. Roughly speaking, uncertainty analysis refers to the inherent variations of a model (e.g., a modeled physical process) and is helpful in determining the relationship between the variability and probability distribution of input parameters and the variability and probability distribution of outputs. In contrast, sensitivity analysis investigates the impact on outputs due to variation in inputs; it ascertains how much a model depends on each or some of its inputs.

Several mathematical and computer-assisted methods have been developed to carry out global sensitivity analysis (see Saltelli, Chan, and Scott 2000 for a wide-ranging and thorough review). Among these methods, a particularly popular class is that comprising “variance-based” methods, which we describe in detail later. Consider a mathematical model given by

\[ Y = \eta(X), \]

where \( \eta: \mathbb{R}^d \to \mathbb{R} \) is the modeling function, \( Y \in \mathbb{R} \) represents the output or prediction of the model, and \( X = (X_1, \ldots, X_d) \) is the \( d \)-dimensional real vector of the input factors or parameters. The vector of input parameters encompasses a set of random variables, implying that output is also a random variable. With variance-based methods, we aim to explain the variance of \( Y \) through variations in \( X_i \), \( i = 1, \ldots, d \). \( \text{Var}(Y) \) can be expressed as follows:

\[ \text{Var}(Y) = \text{Var}(\mathbb{E}(Y|X_i)) + \mathbb{E}([\text{Var}(Y|X_i)]) \]

for all \( i = 1, \ldots, d \). \( \mathbb{E}(Y|X_i) \) and \( \text{Var}(Y|X_i) \) are the conditional expectation and variance of \( Y \) given \( X_i \), respectively. The influence of \( X_i \) on the variance of \( Y \) depends on how well \( \mathbb{E}(Y|X_i) \) fits \( Y \) and can be measured by the first-order sensitivity index

\[ S_i = \frac{\text{Var}(\mathbb{E}(Y|X_i))}{\text{Var}(Y)} \]

\( S_i \) is also called correlation ratio. Sensitivity indexes of higher order can be introduced to account for input interactions. For example, the second-order sensitivity index for \( X_i \) and \( X_j \) is defined as

\[ S_{ij} = \frac{\text{Var}(\mathbb{E}(Y|X_i,X_j)) - \mathbb{E}(\text{Var}(Y|X_i)) - \mathbb{E}(\text{Var}(Y|X_j))}{\text{Var}(Y)} \]

Such formulations can be extended to other orders (see Saltelli, Chan, and Scott 2000 for details).
In the case of independent inputs, the Sobol (Sobol’ 1993) and FAST (Cukier et al. 1973) techniques are the most popular for estimating the $S_i$ indexes. Although powerful and computationally efficient, these methods can be used only when inputs are independent. But this is not the case in many practical problems, such as the kinetic models considered here. Three original methods were introduced by Ratto, Tarantola, and Saltelli (2001), Jacques, Lavergne, and Devictor (2004), and Oakley and O’Hagan (2004) to handle correlated parameters. The first of these calculates sensitivity indexes using replicated Latin hypercube sampling. This approach requires a large number of model evaluations to reach an acceptable accuracy. The second method extends the definition of sensitivity indexes by taking into account blocks of correlations among the inputs. This method is useless when many input factors are correlated, however. The third approach approximates the function $f$ in model (1) by a so-called “kriging” response surface (Santner, Williams, and Notz 2003); thus analytical expressions can be obtained for the sensitivity indexes. Although appealing and accurate, these analytical expressions involve multidimensional integrals that are tractable only when the conditional densities of the input factors are known and easy to integrate. If they are not, then the multidimensional integrals must be numerically approximated, which is very time-consuming.

Consequently, we propose a novel and intermediate approach for estimating sensitivity indexes. This is an intermediate technique because it involves a sample from the joint density of inputs and output like the method of Ratto, Tarantola, and Saltelli (2001) and a nonparametric regression model like the method of Oakley and O’Hagan (2004). But our approach does not require as many model evaluations as the method of Ratto, Tarantola, and Saltelli (2001) and does not require any approximation of multidimensional integrals like the method of Oakley and O’Hagan (2004) in the general case.

In this article we consider a new method based on local polynomial approximations for conditional moments to deal with correlated inputs. (See Fan and Gijbels 1996 and Fan et al. 1996 on conditional expectation and Fan and Yao 1998 and Ruppert et al. 1997 on the conditional variance.) Given the form of the sensitivity indexes, local polynomial regression can be used to estimate them. This approach not only yields sensitivity indexes through a simple black-box procedure, but also achieves good precision.

The article is organized as follows. In Section 1 we review the methods of Ratto, Tarantola, and Saltelli (2001), Jacques, Lavergne, and Devictor (2004), and Oakley and O’Hagan (2004) and discuss their advantages and drawbacks. In Section 2 we develop two new estimators for sensitivity indexes that rely on local polynomial methods. In Section 3 we investigate several analytical examples to illustrate the theoretical behavior of our method. We also apply our method to a practical example based on a kinetic problem.

1. MODELS WITH CORRELATED INPUTS

When inputs are independent, Sobol showed, based on an orthogonal decomposition of the function $f$ (Sobol’ 1993), that the sum of the sensitivity indexes of all orders is equal to 1. Sensitivity indexes arise naturally from this functional ANOVA decomposition. Nevertheless, when inputs are correlated, this property no longer holds, because such a decomposition cannot be done without taking into account the joint distribution of the inputs. If one decides to estimate sensitivity indexes assuming that inputs are independent while they are not, results and consequently interpretation can be inappropriate (see the first example in Sec. 3.1). Nonetheless, we can still consider the initial ANOVA decomposition and work with the original sensitivity indexes without ignoring correlation. Then, when quantifying the first-order sensitivity index of a particular input factor, some part of the sensitivity of all the other input factors correlated with it is taken into account as well. As a result, the same information is considered several times. Interpretation of sensitivity indexes for correlated inputs becomes problematic. However, mathematically speaking, regardless of whether or not the input factors are independent, the first-order sensitivity index still points out which factor (if fixed) will reduce the variance of the output to the greatest degree. If the practitioner wants to conduct a “factors prioritization” ([Saltelli et al. 2004]; i.e., identifying the factor that must be fixed to most reduce the uncertainty in the output), then first-order sensitivity indexes remain the key measure of importance (see Saltelli et al. 2004).

Considering that this is a usual goal for practitioners, it is of interest to be able to compute first-order sensitivity indexes when inputs are correlated.

Beyond this interpretation problem, correlation also makes the FAST and Sobol computation methods unusable, because they have been designed for independent inputs. To overcome these difficulties, we can first build “new” sensitivity indexes that generalize the original ones and match their properties, thus allowing interpretation. This is the idea behind the multidimensional sensitivity analysis of Jacques (Jacques, Lavergne, and Devictor 2004), which is detailed in the following section. Ratto, Tarantola, and Saltelli (2001) attempted to continue working with the original sensitivity indexes and to compute them as described in Section 3.1, even if they did not explain how to go through interpretation. These authors generated replicated Latin hypercube samples to approximate conditional densities. Finally, Oakley and O’Hagan (2004) suggested approximating the function $f$ in model (1) by a kriging response surface, which leads to analytical expressions of sensitivity indexes through multidimensional integrals.

1.1 Multidimensional Sensitivity Analysis

To define multidimensional sensitivity indexes, Jacques, Lavergne, and Devictor (2004) suggest to split $X$ into $p$ vectors $U_j, j = 1, \ldots, p$, each of size $k_j$ such that $U_j$ is independent from $U_l$ for $1 \leq j, l \leq p, j \neq l$:

$$X = (X_1, \ldots, X_d)$$

$$= (X_{1,1}, X_{1,2}, \ldots, X_{1,k_1}), \ldots, (X_{p,1}, X_{p,2}, \ldots, X_{p, k_p}).$$

where $k_1 + k_2 + \cdots + k_p = d$. For example, let us consider $X = (X_1, X_2, X_3)$. We assume that $X_1$ does not depend on $X_2$ and $X_3$, while $X_2$ and $X_3$ are correlated. Then, we set $U_1 = X_1$ and $U_2 = (X_2, X_3)$. 

TECHNOMETRICS, NOVEMBER 2009, VOL. 51, NO. 4
Therefore, Jacques, Lavergne, and Devictor (2004) built first-order multidimensional sensitivity indexes using the vectors $U_j$:

$$S_j = \frac{\text{Var}(E(Y|U_j))}{\text{Var}(Y)} = \frac{\text{Var}(E(Y|X_{i_1+k_2+...+k_{n-1}+1,X_{i_1+k_2+...+k_n}))}{\text{Var}(Y)}$$

for $j = 1, \ldots, p$. Note that for independent inputs, the foregoing expression reduces to classical sensitivity analysis. Finally, it is possible to derive these indexes from Monte Carlo estimations. Nevertheless, this method is unsuitable when all the inputs are correlated. In such a case, the factors cannot be split into disjoint groups, thus making the method useless for interpretation. When there are many correlated inputs, the problem essentially remains, as only a small number of multidimensional indexes would be defined. Moreover, it is still impossible to identify which inputs in an influential group are the most important ones. We will illustrate this phenomenon in the second example of Section 3.1.

1.2 Correlation Ratios With Known Conditional Density Functions

The estimator introduced by Ratto, Tarantola, and Saltelli (2001) was first discussed in McKay (1996) and is based on samples from the conditional density functions of $Y$ given $X_i$, $i = 1, \ldots, d$.

Let $(X_i)^{n}_{j=1,} \ldots, n$ be an iid sample of size $n$ from the distribution of vector $X_i = (X_i^{1}, \ldots, X_i^{r})$ is then an iid sample of size $n$ from the distribution of the input factor $X_i$. For each realization $X_i$ of this sample, let $(Y_{i}^{jk})_{k=1,\ldots,r}$ be an iid sample of size $r$ from the conditional density function of $Y$ given $X_i = X_i^{j}$. We define the sample means as follows:

$$\bar{Y}_i^j = \frac{1}{r} \sum_{k=1}^{r} Y_{i}^{jk}, \quad \bar{Y}_i = \frac{1}{n} \sum_{j=1}^{n} \bar{Y}_i^j.$$ \(\bar{Y}_i^j\) and \(\frac{1}{r} \sum_{k=1}^{r} (Y_{i}^{jk} - \bar{Y}_i^j)^2\) are the estimates of the conditional expectation $E(Y|X_i = X_i^{j})$ and the conditional variance $\text{Var}(Y|X_i = X_i^{j})$, respectively, while \(\bar{Y}_i\) is the estimate of the expectation $E(Y)$.

Referring to these moment estimators, the numerator of the first-order sensitivity index $S_i$, $\text{Var}(E(Y|X_i))$, can be derived from the empirical estimator

$$SSB = \frac{1}{n} \sum_{j=1}^{n} (\bar{Y}_i - \bar{Y}_i)^2.$$ \(SSB\) and \(\frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{r} (Y_{i}^{jk} - \bar{Y}_i)^2\) are the estimates of the conditional expectation $E(Y|X_i = X_i^{j})$ and the conditional variance $\text{Var}(Y|X_i = X_i^{j})$, respectively, while \(\bar{Y}_i\) is the estimate of the expectation $E(Y)$.

Similarly, the denominator of $S_i$, $\text{Var}(Y)$, is estimated by

$$SST = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{r} \sum_{k=1}^{r} (Y_{i}^{jk} - \bar{Y}_i)^2.$$ \(SST\) and \(\frac{1}{n} \sum_{j=1}^{n} \sum_{k=1}^{r} (Y_{i}^{jk} - \bar{Y}_i)^2\) are the estimates of the conditional expectation $E(Y|X_i = X_i^{j})$ and the conditional variance $\text{Var}(Y|X_i = X_i^{j})$, respectively, while \(\bar{Y}_i\) is the estimate of the expectation $E(Y)$.

The estimator of the first-order sensitivity index $S_i$ of the input factor $X_i$, $i = 1, \ldots, d$, is then defined as

$$S_i = \frac{SSB}{SST}.$$ \(S_i\) and \(\frac{SSB}{SST}\) are the estimates of the conditional expectation $E(Y|X_i = X_i^{j})$ and the conditional variance $\text{Var}(Y|X_i = X_i^{j})$, respectively, while \(\bar{Y}_i\) is the estimate of the expectation $E(Y)$.

To compute these indexes and generate the required samples, Ratto used two different methods: pure Monte-Carlo sampling and a singly replicated Latin hypercube ($r$-LHS) sample. But these two methods call for a huge number of model evaluations to achieve good precision, and thus can be used only for cases in which model runs have very low computational cost.

1.3 Bayesian Sensitivity Analysis

The idea of Oakley and O'Hagan (2004) is to treat the function $\eta(\cdot)$ as an unknown smooth function and formulate its prior distribution. In other words, it is modeled as the realization of a Gaussian random field with mean 0 and covariance functions. Then, given a set of values $y_{1} = \eta(x_{1})$, we can derive the posterior distribution of $\eta(\cdot)$ from classical Bayesian considerations. The prior distribution of $\eta(x)$ is a Gaussian stationary field,

$$\eta(x) \equiv h(x)^{T}\beta + Z(x)$$

conditionally on $\beta$ and $\sigma^2$, where $h(\cdot)$ is a vector of $q$ known regression functions and $Z(\mathbf{x})$ is a Gaussian stationary random field with mean 0 and covariance function $\sigma^2 c(\mathbf{x}, \mathbf{x}')$. Vector $h(\cdot)$ and correlation function $c(\cdot, \cdot)$ are selected so that they incorporate information about how the output responds to the inputs and about the degree of smoothness required for the output, respectively. (See Santner, Williams, and Notz 2003 and to Kennedy and O'Hagan 2001 for a detailed discussion on this topic.) The second-stage prior concerns the conjugate prior form for $\beta$ and $\sigma^2$, which is chosen to be a normal inverse gamma distribution.

Now, assuming that we observe a set $y$ of $n$ values of $y_{1} = \eta(x_{1})$, we can show that the posterior distribution of $\eta(\cdot)$ given these data have a Student distribution (see Oakley and O'Hagan 2004 for details).

Using this posterior distribution, sensitivity indexes can be computed analytically through multidimensional integrals involving functions of the observations and the conditional distributions of the input factors only. The main advantage of this Bayesian approach is that the model is only evaluated to calculate the foregoing quantities, that is, to “fit” the response surface. Once this is done, estimating the sensitivity indexes involves just the conditional distributions of the input factors. When the number of model runs is fixed, this method clearly reduces the standard errors in the estimated sensitivity indexes obtained by Monte Carlo methods such as that of Sobol (when the input factors are independent) and can still be used for correlated input factors.

But the multidimensional integrals leading to the computation of the sensitivity indexes, when not analytically tractable, must be estimated. Here we focus on one of the estimators proposed by Oakley and O’Hagan (2004). We keep the authors’ notations and denote by $E^{*}$ the expectations defined with respect to the posterior distribution of $\eta(\cdot)$. The numerator of the first-order sensitivity index of $Y$ with respect to $X_1$ is written as

$$E^{*}(\text{Var}(E(Y|X_1))) = E^{*}(E(E(Y|X_1)^2)) - E^{*}(E(Y)^2).$$

One of the quantities involved in the computation of $E^{*}(E(E(Y|X_1)^2))$ is

$$U_1 = \int_{\mathbb{R}} \int_{\mathbb{R}^d-1} \int_{\mathbb{R}} c(\mathbf{x}, \mathbf{x}'') dF_{-11}(\mathbf{x} - 1|\mathbf{x}_1) dF_{-11}(\mathbf{x}' - 1|\mathbf{x}_1) dF_{1}(\mathbf{x}_1),$$
where \( F_{-1|1} \) is the marginal distribution of \( X_{-1} \) (subvector of \( X \) containing all elements except \( X_1 \)) given \( X_1 \), \( F_1 \) is the marginal distribution of \( X_1 \), and \( \mathbf{x}^* \) denotes the vector with elements made up of \( x_1 \) and \( X_{-1}^* \) in the same way as \( \mathbf{x} \) is composed of \( x_1 \) and \( X_{-1} \). If the conditional distribution \( F_{-1|1} \) is not analytically known, then it must be estimated from a sample from the joint distribution \( F \). Many methods to this have been developed, including kernel techniques; however, in high dimension, data are very sparsely distributed, making an accurate approximation of conditional distributions difficult. This is the so-called “curse of dimensionality.” For instance, the best possible MSE rate with kernel techniques is \( n^{-d/(4+d)} \), which becomes worse as \( d \) grows larger.

Moreover, even if we can get a good approximation for \( F_{-1|1} \), we still must solve the problem of evaluating the multidimensional integrals. Indeed, the dimensionality of these integrals can reach \( 2d - 1 \) as in the expression of \( U_1 \) earlier. Because these integrals cannot in general be split into unidimensional integrals, approximating them with sufficient accuracy has no obvious mathematical solution. Deterministic schemes cannot reasonably be envisioned, and with Monte Carlo or quasi Monte Carlo sampling (Owen 2005), thousands (or millions) of draws are required to get reasonable accuracy.

A public domain software based on this analysis has been developed by Oakley and O’Hagan (2004) (http://www.ctcd.group.shef.ac.uk/gem.html). This software does not allow for handling dependent inputs, however. Numerical examples are reported in Section 3.

## 2. NEW ESTIMATION METHODOLOGY

We propose to estimate the conditional moments \( \mathbb{E}(Y|X_i = X_i^0) \) and \( \text{Var}(Y|X_i = X_i^0) \) with an intermediate method between the methods of Ratto, Tarantola, and Saltelli (2001) and Oakley and O’Hagan (2004). We first use a sample \((X_i, Y_i)\) to estimate the conditional moments with nonparametric tools (provided that they are smooth functions of the input factors). We then derive sensitivity indexes from another sample of the input factors. We propose using local polynomial regression to approximate the conditional moments. These expressions clearly lead to two alternatives to estimate \( S \).

### 2.1 Formulation of the Estimators

Let \((X_i, Y_i)_{i=1,...,n}\) be a two-dimensional iid sample of a real random vector, \((X, Y)\). Assuming that \(X\) and \(Y\) are square-integrable, we may write an heteroscedastic regression model of \(Y_i\) on \(X_i\) as

\[
Y_i = m(X_i) + \sigma(X_i)\epsilon_i, \quad i = 1, \ldots, n.
\]

\(m(x) = \mathbb{E}(Y|X = x)\) and \(\sigma^2(x) = \text{Var}(Y|X = x)\) are the conditional mean and variance. Errors \(\epsilon_1, \ldots, \epsilon_n\) are independent random variables satisfying \(\mathbb{E}(\epsilon_i|X_i) = 0\) and \(\text{Var}(\epsilon_i|X_i) = 1\). Usually \(\epsilon_i\) and \(X_i\) are assumed to be independent, although this is not the case in our work. Results for correlated errors have been recently reported (see, e.g., Vilar-Fernández, and Francisco-Fernández 2002 for the autoregressive case). Local polynomial fitting consists in locally approximating the regression function \(m\) by a \(p\)-th order polynomial

\[
m(z) \approx \sum_{j=0}^{p} \beta_j (z-x)^j
\]

for \(z\) in a neighborhood of \(x\). This polynomial is then fitted to the observations \((X_i, Y_i)\) by solving the weighted least squares problem

\[
\min_{\beta} \sum_{i=1}^{n} \left( Y_i - \sum_{j=0}^{p} \beta_j (X_i-x)^j \right)^2 K_1 \left( \frac{X_i-x}{h_1} \right).
\]

\(K_1(\cdot)\) is a kernel function and \(h_1\) is a smoothing parameter (or bandwidth). In this case, if \(\hat{\beta}(x) = (\hat{\beta}_0(x), \ldots, \hat{\beta}_p(x))^T\) is the minimizer of (3), we can write

\[
\hat{m}(x) = \hat{\beta}_0(x).
\]

The \(v\)-th derivative of \(m(x)\) is estimated via the relation

\[
\hat{\beta}_v(x) = \frac{\hat{m}^{(v)}(x)}{v!}
\]

(see Fan and Gijbels 1996 for more detail). As discussed later, the smoothing parameter \(h_1\) is chosen to balance bias and variance of the estimator. Finally, the particular case \(p = 0\) (constant fit) reduces to the well-known Nadaraya–Watson estimator, \(\hat{m}_{NW}(x)\), of the conditional expectation, given explicitly by

\[
\hat{m}_{NW}(x) = \sum_{i=1}^{n} Y_i K((X_i-x)/h) / \sum_{i=1}^{n} K((X_i-x)/h)
\]

(Wand and Jones 1994).

Estimating the conditional variance is less straightforward. If the regression function \(m\) were known, then the problem of estimating \(\sigma^2(\cdot)\) would be considered a local polynomial regression of \(\epsilon_i^2\) on \(X_i\) with \(\epsilon_i^2 = (Y_i - m(X_i))^2\), as \(\mathbb{E}(\epsilon_i^2|X_i) = \sigma^2(x)\) with \(\epsilon_i^2 = (Y - m(X))^2\). But in practice, \(m\) is unknown. An intuitive approach involves substituting \(m(\cdot)\) by its estimate \(\hat{m}(\cdot)\) and to get the the residual-based estimator \(\hat{\sigma}^2(\cdot)\) by solving the associated weighted least squares problem,

\[
\min_{\gamma} \sum_{i=1}^{n} \left( \epsilon_i^2 - \sum_{j=1}^{q} \gamma_j (X_i-x)^j \right)^2 K_2 \left( \frac{X_i-x}{h_2} \right)
\]

TECHNOMETRICS, NOVEMBER 2009, VOL. 51, NO. 4
Here \( \hat{\sigma}^2(x) = \hat{\gamma}_0(x) \), where \( \hat{\gamma}(x) = (\hat{\gamma}_0(x), \ldots, \hat{\gamma}_p(x)) \) is the minimizer of (4). As before, the smoothing parameter \( h_2 \) must be chosen to balance the bias and variance of the estimator (see Fan and Yao 1998).

Going back to the equalities (2), we still must estimate the quantities \( \text{Var}(\hat{E}(Y|X)) \) and \( \text{Var}(\hat{V}(Y|X)) \) by using the local polynomial estimators for the conditional moments defined earlier. To do this, we assume that we have another i.i.d sample \( (\tilde{X}_j)_{j=1,\ldots,n'} \) with the same distribution as \( X \).

If the functions \( m(\cdot) \) and \( \sigma^2(\cdot) \) are known, we could estimate \( \text{Var}(\hat{E}(Y|X)) = \text{Var}(m(X)) \) and \( \text{Var}(\hat{V}(Y|X)) = \text{E}(\sigma^2(\cdot)) \) with the classical empirical moments

\[
\frac{1}{n'-1} \sum_{j=1}^{n'} (\hat{m}(\tilde{X}_j) - \bar{m})^2 \quad \text{and} \quad \frac{1}{n'} \sum_{j=1}^{n'} \hat{\sigma}^2(\tilde{X}_j),
\]

where \( \bar{m} = \frac{1}{n'} \sum_{j=1}^{n'} \hat{m}(\tilde{X}_j) \). Because \( m(\cdot) \) and \( \sigma^2(\cdot) \) are unknown, the main idea is to replace them by their local polynomial estimators, which leads us to consider

\[
\hat{T}_1 = \frac{1}{n'-1} \sum_{j=1}^{n'} (\hat{m}(\tilde{X}_j) - \bar{m})^2 \quad \text{and} \quad \hat{T}_2 = \frac{1}{n'} \sum_{j=1}^{n'} \hat{\sigma}^2(\tilde{X}_j),
\]

where \( \bar{m} = \frac{1}{n'} \sum_{j=1}^{n'} \hat{m}(\tilde{X}_j) \) and \( \hat{m}(\cdot) \) and \( \hat{\sigma}^2(\cdot) \) are the local polynomial estimators of \( m(\cdot) \) and \( \sigma^2(\cdot) \) introduced earlier. It is important to note that we need two samples, the first one, \((X_i, Y_i)_{i=1,\ldots,n'}\), to compute \( \hat{m}(\cdot) \) and \( \hat{\sigma}^2(\cdot) \) and the second one, \((\tilde{X}_j)_{j=1,\ldots,n'}\), to finally compute the empirical estimators \( \hat{T}_1 \) and \( \hat{T}_2 \).

### 2.2 Bandwidth and Orders Selection

The selection of the smoothing parameters \( h_1 \) and \( h_2 \) and, to a lesser extent, of the polynomials orders \( p \) and \( q \) can be crucial to reduce the mean squared error (MSE) of the estimators \( \hat{T}_1 \) and \( \hat{T}_2 \). Classically, the MSE consists of a bias term plus a variance term and so is minimized by finding a compromise between bias and variance.

Concerning this choice, the reader is referred to the work of Fan et al. (1996), Fan and Yao (1998), and Ruppert (1997). Most of the methods suggested by these authors rely on asymptotic arguments and their efficiency for finite-sample cases is not clear. In practice, cross-validation methods can be used for finite-sample cases (Jones, Marron, and Sheather 1996); however, in the examples of Section 3 we prefer to use the empirical-bias bandwidth selector (EBBS) of Ruppert, which appears to be efficient on simulated data. The EBBS involves estimating the MSE empirically instead of using asymptotic expressions. Selecting the polynomials orders is more subjective. Concerning the conditional expectation, Fan and Gijbels (1996) recommend using a \( v+1 \)th- or \( v+3 \)th-order polynomial to estimate the \( v \)th derivative of \( m(x) \). Their prescription relies on theoretical considerations about the asymptotic bias of \( \hat{m}(x) \) on the boundary. Thus, we take \( p = 1 \) or \( p = 3 \) to estimate the \( 0 \)th derivative \( m(x) \). On the other hand, Ruppert, Wand, and Carroll (2003) suggest that this conclusion should be balanced by simulation studies and stress that \( p = 2 \) often outperforms \( p = 1 \) and \( p = 3 \). The only point of agreement is that local linear regression (\( p = 1 \)) is usually superior to kernel regression (Nadaraya–Watson estimator obtained with \( p = 0 \)). This is the reason why we consider and study only local linear regression for \( m(x) \) in the theoretical and practical discussions that follow. For the conditional variance, the choice is at least as difficult, because we have to choose \( p \) and \( q \) simultaneously. To reiterate, there is no universal answer: Fan and Yao (1998) recommend setting \( p = 1 \), \( q = 1 \), whereas Ruppert et al. (1997) suggested using \( p = 2 \), \( q = 1 \) or \( p = 3 \), \( q = 1 \). Based on our simulations, we concluded that choosing \( p = 1 \), \( q = 1 \) is adequate and satisfactory in terms of precision. Thus, we consider only the \( p = 1 \), \( q = 1 \) case for both theoretical and practical developments.

### 2.3 Theoretical Properties of the Estimators

The properties of \( \hat{T}_1 \) and \( \hat{T}_2 \) depend strongly on the asymptotic results on the bias and variance of the local linear estimators \( \hat{m}(\cdot) \) and \( \hat{\sigma}^2(\cdot) \). We report only two main results, all assumptions (A0)–(A4) and (C0), with proofs given in the Appendix. \( \text{E}_X \) and \( \text{Var}_X \) represent the conditional expectation and variance given the predictors \( X = (X_1, \ldots, X_n) \). The expression \( op(\phi(h)) \) is equal to \( \phi(h)op(1) \) for a given function \( \phi \). Here \( op(1) \) is the standard notation for a sequence of random variables that converges to 0 in probability.

**Theorem 1.** Under assumptions (A0)–(A4) and (C0), the estimator \( \hat{T}_1 \) is asymptotically unbiased. More precisely,

\[
\text{E}_X(\hat{T}_1) = \text{Var}(\hat{E}(Y|X)) + M_1 h_1^2 + \frac{M_2}{nh_1} + op(h_1^2),
\]

where \( M_1 \) and \( M_2 \) are constants given in the Appendix.

**Remark 1.** Calculating the variance of this estimator would be of interest, but it requires the expressions of the third and fourth moments of the local linear estimator, \( \hat{m}(\cdot) \) (see the Appendix). This problem is not trivial. To the best of our knowledge, it has not been addressed in the literature. It should motivate research in the near future. However, in practical cases, the variance can be estimated through bootstrap methods, for example, that of (Efron and Tibshirani 1994).

**Theorem 2.** Under assumptions (B0)–(B4) and (C0), the estimator \( \hat{T}_2 \) is consistent. More precisely,

\[
\text{E}_X(\hat{T}_2) = \text{E}(\text{Var}(Y|X)) + V_1 h_2^2 + op(h_1^2 + h_2^2)
\]

and

\[
\text{Var}_X(\hat{T}_2) = \frac{1}{n'} \left[ \text{E}(\text{Var}(Y|X)) + V_2 h_2^2 + V_3 h_1^2 + \frac{V_4}{nh_2} + op\left(h_1^2 + h_2^2 + \frac{1}{\sqrt{nh_2}}\right) \right],
\]

where \( V_1, V_2, V_3, \) and \( V_4 \) are constants given in the Appendix.
2.4 Application to Sensitivity Analysis

We return to model (1), where $X$ is multidimensional. The goal is to get an estimate of $S_i$ for $i = 1, \ldots, d$ by using one of the two estimators $\hat{T}_1$ and $\hat{T}_2$. We need two samples to compute each of them, that is, a sample $(X^k, Y^k)_{k=1,\ldots,n}$ to estimate $\hat{m}(\cdot)$ and $\hat{\sigma}^2(\cdot)$ and a sample $(X_l^i)_{l=1,\ldots,n^l}$ to get $\hat{T}_1$ and $\hat{T}_2$, where $(X^k_i)_{k=1,\ldots,n}$ and $(X_l^i)_{l=1,\ldots,n^l}$ are samples from the joint distribution of the $d$-dimensional input factors $X = (X_i)_{i=1,\ldots,d}$ and $(Y^k)_{k=1,\ldots,n}$, a sample of the output $Y$. Note that the model is run just for the first sample, not for the second sample. Three situations can arise:

1. Sampling from the joint distribution of $X$ has low computational cost, and running the model to compute $(Y^k)_{k=1,\ldots,n}$ is cheap. In this very ideal case, the two samples $(X^k, Y^k)_{k=1,\ldots,n}$ and $(X_l^i)_{l=1,\ldots,n^l}$ can be generated independently and be as large as required.

2. Sampling from the joint distribution of $X$ has low computational cost, unlike model evaluations. In this case (as also pointed out in Oakley and O'Hagan 2004), a moderate-sized sample, $(X^k, Y^k)_{k=1,\ldots,n}$, is used to fit the conditional moments. Then, to compute $\hat{T}_1$ and $\hat{T}_2$, we use a sample $(X_l^i)_{l=1,\ldots,n^l}$ of large size.

3. Sampling from the joint distribution of $X$ has high computational cost. In practice this is the case when, for example, the input factors are derived from a procedure based on experimental data. We then have an initial sample, $(X^i)_{i=1,\ldots,N}$, of limited size $N$ that must be used for the two steps of the estimation. The first idea is to split this sample in two parts, one part used for $(X^k, Y^k)_{k=1,\ldots,n}$ and the other part used for $(X_l^i)_{l=1,\ldots,n^l}$. The drawback of this method is clear for very small $N$. Another way to face the problem is to use the well-known leave-one-out procedure, which gives a better approximation than data splitting.

It also would be possible to use the sample of size $N$ to estimate the conditional moments and estimate the marginal densities of each input using a nonparametric density estimator, for instance. One could then use these density estimates to get a sample $(X_l^i)_{l=1,\ldots,n^l}$. The clear disadvantage of this procedure is that it may bias the final estimators. Simulation runs not reported here for lack of space show that such a procedure yields less-efficient estimates, probably because of the large bias produced by nonparametric methods.

This last situation obviously leads to less-accurate approximations of first-order sensitivity indexes. However, in general, literature and results about sensitivity analysis assume that, if not analytically known, the joint distribution of the input factors can at least be generated at a low computational cost. This is the reason why here we describe only the procedure for estimating first-order sensitivity indexes in case 1 or 2. We now assume that we have two samples, $(X^k)_{k=1,\ldots,n}$ and $(X_l^i)_{l=1,\ldots,n^l}$, given by one of the methods described earlier.

The estimation procedure for $S_i = \frac{\text{Var}(E(Y|X_i))}{\text{Var}(Y)}$ is as follows:

**Step 1.** Compute the output sample $(Y^k)_{k=1,\ldots,n}$ by running the model at $(X^k)_{k=1,\ldots,n}$.

**Step 2.** Compute $\hat{\sigma}^2_Y$, the classical unbiased estimator of the variance $\text{Var}(Y)$,

$$\hat{\sigma}^2_Y = \frac{1}{n-1} \sum_{k=1}^{n} (Y^k - \bar{Y})^2.$$

**Step 3.** Use the sample $(X^k, Y^k)_{k=1,\ldots,n}$ to obtain $\hat{m}(\bar{X}_l^i)$ for $l = 1, \ldots, n^l$ and $\hat{m}(X^k_i)$ for $k = 1, \ldots, n$ using the smoothing parameter $h_1$ given by EBBS.

**Step 4.** Compute squared residuals $\hat{r}_k = (Y^k - \hat{m}(X^k_i))^2$ for $k = 1, \ldots, n$ and apply the smoothing parameter $h_2$ obtained by EBBS to compute $\hat{\sigma}^2(X_l^i)$ for $l = 1, \ldots, n^l$.

**Step 5.** Compute $\hat{T}_1$ with $\hat{m}(\bar{X}_l^i)$ for $l = 1, \ldots, n^l$ from step 3 and compute $\hat{T}_2$ with $\hat{\sigma}^2(X_l^i)$ for $l = 1, \ldots, n^l$ from step 4.

**Step 6.** The estimates of $S_i$ are then

$$\hat{S}_i(1) = \frac{\hat{T}_1}{\hat{\sigma}^2_Y} \quad \text{and} \quad \hat{S}_i(2) = 1 - \frac{\hat{T}_2}{\hat{\sigma}^2_Y}, \quad (5)$$

To obtain all first-order sensitivity indexes, repeat the same procedure from step 3 to step 6 for $i = 1, \ldots, d$.

**Remark 2.** Given the theoretical properties of $\hat{T}_1$ and $\hat{T}_2$ and, more precisely, their nonparametric convergence rate, we also can expect a nonparametric convergence rate for $\hat{S}_i(1)$ and $\hat{S}_i(2)$.

**Remark 3.** In practice, in our simulations, we took values of about 100 and 2,000 for $n$ and $n^l$, and the estimates of sensitivity indexes were sufficiently accurate.

3. EXAMPLES

In all of the following examples, we use the two estimators $\hat{S}_i(1)$ and $\hat{S}_i(2)$ defined in (5). As stated in Section 2.2, the conditional expectation is estimated here with local linear regression ($p = 1$) and the conditional variance with $p = 1$ and $q = 1$, with the bandwidths selected by the estimated-bias method of Ruppert (1997).

3.1 Analytical Examples

In this section we report two comparison studies to investigate the capabilities of our two estimators from a numerical standpoint. The first model was chosen to underline their precision in correlated cases when FAST and Sobol methods are no longer efficient and when the approach of Jacques for multidimensional sensitivity analysis is limited. We also show that interpreting sensitivity indexes obtained while neglecting correlation between variables can be misleading. The second model is an example illustrating the performance of our estimators with respect to the method of Oakley and O’Hagan in a two-dimensional setting.

TECHNOMETRICS, NOVEMBER 2009, VOL. 51, NO. 4
3.1.1 Linear Model With Correlated Inputs. In the first analytical example, we study the model

\[ Y = X_1 + X_2 + X_3, \]

where \((X_1, X_2, X_3)\) is a three-dimensional normal vector with mean \(0\) and covariance matrix

\[ \Gamma = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & \rho \sigma \\ 0 & \rho \sigma & \sigma^2 \end{bmatrix}, \]

\(\rho\) is the correlation of \(X_2\) and \(X_3\) and \(\sigma > 0\) is the standard deviation of \(X_3\). The first-order sensitivity indexes can be evaluated analytically:

\[ S_1 = \frac{1}{2 + \sigma^2 + 2 \rho \sigma}, \]
\[ S_2 = \frac{(1 + \rho \sigma)^2}{2 + \sigma^2 + 2 \rho \sigma}, \]
\[ S_3 = \frac{(\sigma + \rho)^2}{2 + \sigma^2 + 2 \rho \sigma}. \]

The first crucial remark in this case is that we must account for correlations to estimate sensitivity indexes if we want to conduct a serious investigation of this model. As an example, let us consider the case where \(\sigma = 1.2\) and \(\rho = -0.8\). We then have

\[ S_1 = 0.6579, \quad S_2 = 0.0011, \quad \text{and} \quad S_3 = 0.1053, \]

which means that \(X_1\) should be the input to be fixed to reach the largest variance reduction on \(Y\). If we had neglected correlation and derived the sensitivity indexes from the FAST method, (i.e., working with a three-dimensional normal vector with mean \(0\) and covariance matrix \(\Gamma\) with \(\rho = 0\)), then we would have obtained

\[ S_1^0 = 0.2907, \quad S_2^0 = 0.2907, \quad \text{and} \quad S_3^0 = 0.4186. \]

Here \(S_0^0\) represents sensitivity indexes with \(\rho = 0\). These results indicate that \(X_3\) should be fixed to provide the greatest reduction of the variance of \(Y\), which is inconsistent with the calculations shown earlier. This simple example stresses how misleading it can be to neglect correlations between inputs.

Otherwise, applying the approach of Jacques to \(X_1\) and the couple \((X_2, X_3)\), we also get the expression of the first-order multidimensional sensitivity index

\[ S_{[2,3]} = \frac{1 + \sigma^2 + 2 \rho \sigma}{2 + \sigma^2 + 2 \rho \sigma}. \]

Choosing \(\rho = -0.2\) and \(\sigma = 0.4\), we have

\[ S_1 = S_{[2,3]} = 0.5, \quad S_2 = 0.4232, \quad \text{and} \quad S_3 = 0.02. \]

If we interpret these indexes as suggested by Jacques’s multidimensional sensitivity analysis, then the only obvious conclusion is that the couple \((X_2, X_3)\) is as important as \(X_1\), because \(S_{[2,3]} = S_1\). But the high value of \(S_{[2,3]}\) comes from \(X_2\), as shown by the exact calculations given earlier. This implies that the information about \(S_{[2,3]}\) alone is not sufficient.

With our method, we can estimate all of the first-order sensitivity indexes. An average of their values over 100 simulations with \(n = 50\) and \(n' = 1,000\) is as follows:

\[ \hat{S}_1 = 0.4895, \quad \hat{S}_2 = 0.4250, \quad \hat{S}_3 = 0.0234, \]
\[ \hat{S}_1 = 0.5081, \quad \hat{S}_2 = 0.4368, \quad \hat{S}_3 = 0.0361. \]

Figure 1 shows boxplots corresponding to the distribution of the sensitivity indexes for these 100 simulations with the estimator \(\hat{T}_2\). Because of the mathematical complexity for computing the variance of \(\hat{T}_1\), we are not able to recommend one estimator over another from a theoretical standpoint. However, in practice, we observed that the variance of \(\hat{T}_2\) is at least comparable to the variance of \(\hat{T}_1\), and sometimes lower. Nevertheless, the computation of \(\hat{T}_2\) is more difficult, as illustrated in Section 2.4.

The \(S_2\) and \(S_3\) sensitivity indexes provided by our method, even if they account for correlations, confirm the expected result that all of the variability comes from \(X_2\), not from \(X_3\). This simple example illustrates the limitation of Jacques’s multidimensional approach.

3.1.2 Nonlinear Model. In this second analytical example, we consider the model

\[ Y = 0.2 \exp(X_1 - 3) + 2.2|X_2| + 1.3X_3^2 - 2X_2^2 \\
- 0.5X_3^4 - 0.5X_1^4 + 2.5X_1^2 + 0.7X_3^3 \\
+ \frac{3}{(8X_1 - 2)^2 + (5X_2 - 3)^2 + 1} + \sin(5X_1)\cos(3X_1^2). \]

Here \(X_1\) and \(X_2\) are independent random variables distributed uniformly on \([-1, 1]\) (see Figure 2).

In this case, the sensitivity indexes are

\[ S_1 = 0.9375 \quad \text{and} \quad S_2 = 0.0625. \]

We consider a \(6 \times 6\) regular grid on \([0, 1]^2\) and use it to estimate first the posterior distribution in the method of Oakley and O’Hagan (OOH) and then the conditional moments in our method. We then calculate analytically the multidimensional integrals in the Bayesian approach while using a sample of size...
5,000 to compute \( \hat{S}_i^{(2)} \) for \( i = 1, 2 \). The multidimensional integrals in OOH also could have been estimated numerically without significant loss of precision. The Bayesian approach leads to
\[
\hat{S}_1^{(1)} = 0.9038 \quad \text{and} \quad \hat{S}_2^{(1)} = 0.0961,
\]
while the local polynomial technique gives
\[
\hat{S}_1^{(2)} = 0.9127 \quad \text{and} \quad \hat{S}_2^{(2)} = 0.0452.
\]
The results obtained with the two methods are comparable; however, for this simple case, the multidimensional integrals were computed analytically, which usually is not feasible in a nonindependent setting.

### 3.1.3 Comparison
The third example, known as the Rastigrin test, function deals with an analytical formula in which the number of entries can be arbitrarily large,
\[
Y = 10d + \sum_{k=1}^{d} (X_k^2 - \cos(2\pi X_k))^2.
\]
The number of entries can be increased to show how dimension affects both the OOH method and the method studied here. Due to the cosine, this function shows a lot of maxima and minima. In this example, sensitivity indexes are derived from the two methods, for input space dimension \( d = 2, 5, \) and 10. The OOH method is run with the public domain software GEM-SA (Kennedy 2005). The size of the observed samples is \( n = 400 \) for all dimensions and methods. The variables \( X_k \) are independently and uniformly drawn over the range \([-3, 3]\). Theoretically, in such conditions, the sensitivity indexes for all of the variables should be \( 1/d \). For \( d = 2 \), we find very similar results, close to the theoretical values: 0.4779 for \( X_1 \) and 0.4880 for \( X_2 \) with OOH and 0.4593 and 0.4754 with our method. Figure 3 displays the results obtained when \( d = 5 \) and \( d = 10 \). It can be seen that OOH method systematically underestimates the sensitivity indexes. Because of the small number of observations in dimension \( d = 10 \), the results show great variability. In this case, using a PC with 1 GB of memory and a 4 kHz processor, the OOH method requires 1 hour of computation to get the final estimates, while our method provides the results in only 30 seconds. In addition, the new method works with the 2 order marginal distributions and thus always stays in dimension 2, while OOH must cope with the full dimensionality. No comparison has been made with correlated entries, because this feature is not available in GEM-SA.

### 3.2 Practical Example From Chemistry: Isomerization of Normal Butane
Here we report the results obtained for a real example in kinetic modeling. We show that all of the parameters are influential for at least one setting. The isomerization of normal butane (i.e., molecules with four carbon atoms) is a chemical process aimed at transforming normal butane \((nC_4)\) into isobutane \((iC_4)\) to obtain a higher octane number, favored by \( iC \). A simplified reaction mechanism has been used, as follows:
\[
\begin{align*}
nC_4 & \leftrightarrow iC_4, \quad (1) \\
2iC_4 & \rightarrow C_3 + C_5, \quad (3) \\
nC_4 + iC_4 & \rightarrow C_3 + C_5. \quad (4)
\end{align*}
\]
Reaction (1) is the main reversible reaction that converts the normal butane into isobutane. Reactions (3) and (4) are secondary and irreversible reactions that produce propane \((C_3)\) and a lump of normal and isopentane \((C_5)\), paraffins with three and five carbon atoms. The model associated to this process can be written as
\[
Y = \eta(c, \theta),
\]
where
\[
- \ Y \ is the three-dimensional result vector \ (mole fractions of components nC_4, iC_4, C_3, and C_5; \ note that their sum is 1) \\
- \ c \ is the vector describing the operating conditions \ (e.g., pressure, temperature) and the mole fraction of the input components \ (nC_4 and iC_4, called the feed) \\
\]
- $\theta = (\theta^j)_{j=1,...,8}$ is the eight-dimensional random vector of the parameters of the reactions (e.g., preexponential factors, activation energies, adsorption constants)
- $f$ is the function modeling the chemical reactor, in which the reaction occurs. It is evaluated through the resolution of a system of ordinary differential equations, which is solved on a numerical basis.

The first step involves obtaining the distribution of $\theta$, which is unknown. It can be reasonably approximated based on experience and knowledge of chemical engineers. Classically, we assume that $\theta$ has a multivariate Gaussian distribution with mean 0 (once parameters are centered). As for the correlation matrix, it is built with experts and with the help of bootstrap simulations. We generate a sample of size $n = 5,000$ from this distribution to compute sensitivity indexes.

Our objective is to estimate, for given operating conditions and feed vector $c$, the sensitivity indexes of the outputs with respect to the input factors in $\theta$ [i.e., $S_j^c = \text{Var}(\mathbb{E}(Y_j|\theta))/\text{Var}(Y_j)$] for $j = 1, \ldots, 3$ and $i = 1, \ldots, 8$. We want to identify the factor on which we have to focus to reduce the uncertainty, by carrying out new experiments. This factor should be chosen to reduce as much as possible the uncertainty of the outputs.

We consider two particular vectors, $c_1$ and $c_2$, representing the same operating conditions, but different feeds ($c_1$: nC4 = 1 and iC4 = 0, $c_2$: iC4 = 1 and nC4 = 0). For each vector $c_i$, $i = 1, 2$, we draw a sample of size $n$ from $Y$ by Monte Carlo simulations, that is, by computing $Y_j = \eta(c_i, \theta)$ for $j = 1, \ldots, n$. Thus we get a sample of $(Y, \theta)$ for each particular $c_1$ and $c_2$. Figure 4 shows the estimates of the sensitivity indexes of the third output, C3 + C5, derived from estimator $\hat{T}_1$. The filled bars correspond to $c_1$, and the empty bars correspond to $c_2$.

The estimates provided by the $\hat{T}_2$ estimator are similar. These results highlight the behavior of the C3 + C5 output when feed changes. When we use only nC4 in the feed ($c_1$), the production of C3 + C5 is linked mainly to the production of iC4 by reaction (1). This is confirmed by the importance of parameters 1 and 6 in Figure 4, which are the parameters involved in reaction (1). When the feed contains only iC4 ($c_2$), the first reaction is no longer dominating for the production of C3 + C5, which is then mainly linked to reaction (3). Parameters 4 and 2, which are the most important in Figure 4 for $c_2$, are connected to reaction (3). Thus we can conclude that the results confirm the expected behavior for the C3 + C5 output.

Obviously, we could study the sensitivity indexes for the other outputs and for other operating conditions. Such a study was carried out and showed that the most influential parameters depend on the operating conditions and the feed. This study also underscored that each parameter of the model has an influence on at least one output for at least one operating condition. In this case, the sensitivity indexes estimates emphasize the fact that all parameters are potentially important. This conclusion comforts the experts on the adequacy of their model. To the best of our knowledge, there is no sensitivity index capable of aggregating together such a diversity of responses and settings.

4. DISCUSSION AND CONCLUSION

The estimation method proposed in this article is efficient for carrying out sensitivity analysis by computing first-order sensitivity indexes when the inputs are not independent. The use of local polynomial estimators is the key point of the estimation procedure. This guarantees some interesting theoretical properties and ensures good qualities for the estimators that we have introduced. Beyond these theoretical results, practical examples also indicate good precision for a rather low computation time. Obviously, higher precision calls for more CPU time, and the user has the option to adapt the estimators by fixing some hyperparameter values, such as polynomial orders.

The main advantage of our estimators is that they only make the assumption that the marginals are smooth and then require fewer model runs than classical sampling methods. Our method is based on the same philosophy as the Bayesian approach of Oakley and O’Hagan (2004), because it uses model runs to fit a response surface under smoothness assumptions. We avoid its numerical integration issue in high dimension, however.

Moreover, our approach is appealing for practitioners in the sense that they can see it as a black-box routine, because each step of the procedure is data-driven once the user has given the two samples required for the estimation. Finally, we believe that a practitioner willing to carry out a sensitivity analysis should combine different approaches to get the most accurate result, for example, computing the indexes with the method that we introduce here and the method of Oakley and O’Hagan, if possible.

Future work also will be dedicated to the building of multioutput sensitivity indexes through multivariate nonparametric regression techniques. Interested readers may contact the authors for a copy of the MATLAB® code for the algorithm.

APPENDIX: PROOFS OF THEOREMS

A.1 Assumptions

Here we list all of the assumptions used in the development of our proofs. Note that the bandwidths $h_1$ and $h_2$ are, by definition, positive real numbers.
(A0) As \( n \to \infty \), \( h_1 \to 0 \) and \( nh_1 \to \infty \).

(A1) The kernel \( K(\cdot) \) is a bounded symmetric and continuous density function with finite seventh moment.

(A2) \( f_X(x) > 0 \) and \( \tilde{f}_X(\cdot) \) is bounded in a neighborhood of \( x \), where \( f_X(\cdot) \) denotes the marginal density function of \( X \).

(A3) \( \hat{\sigma}^2(\cdot) \) exists and is continuous in a neighborhood of \( x \).

(A4) \( \sigma^2(\cdot) \) has a bounded third derivative in a neighborhood of \( x \) and \( \tilde{m}(\cdot) \neq 0 \).

(B0) As \( n \to \infty \), \( h_i \to 0 \) and \( \lim inf nh_i^4 > 0 \) for \( i = 1, 2 \).

(B1) The kernel \( K(\cdot) \) is a symmetric density function with a bounded support in \( \mathbb{R} \). Furthermore, \( |K(x_i) - K(x_2)| \leq c|x_i - x_2| \) for \( x_1, x_2 \in \mathbb{R} \).

(B2) The marginal density function \( f_X(\cdot) \) satisfies \( f_X(x) > 0 \) and \( |f_X(x_1) - f_X(x_2)| \leq c|x_i - x_2| \) for \( x_1, x_2 \in \mathbb{R} \).

(B3) \( \mathbb{E}(Y^4) < \infty \).

(B4) \( \sigma^2(\cdot) > 0 \), and the function \( \mathbb{E}(Y^k|X = \cdot) \) is continuous at \( x \) for \( k = 3, 4 \). Furthermore, \( \tilde{m}(\cdot) \) and \( \hat{\sigma}^2(\cdot) \) are uniformly continuous on an open set containing the point \( x \).

(C0) \( f_X(\cdot) \) has compact support \([a, b]\).

Assumptions (A0) and (B0) are standard assumptions in kernel estimation theory. Some classical considerations on MSE or mean integrated squared error (MISE) lead to theoretical optimal constant bandwidths of order \( n^{-1/5} \).

Assumptions (A1) and (B1) are directly satisfied by commonly used kernel functions. We note that they require a kernel with bounded support, but this is only a technical assumption for brevity of proofs. The Gaussian kernel can be used, for example.

The assumption \( f_X(x) > 0 \) in (A2) and (B2) simply ensures that the experimental design is sufficiently rich. The fact that (A2) also requires \( \tilde{f}_X(\cdot) \) to be bounded in a neighborhood of \( x \) is natural. The Lipschitz condition on \( f \) in (B2) is directly satisfied if \( f \) is sufficiently regular with \( \tilde{f}_X(\cdot) \) also compact support.

Assumptions (A3), (A4), (B3), and (B4) are natural and ensure sufficient regularity to the conditional moments.

Assumption (C0) is made to make the presentation easier. It can be relaxed by means of the conventional truncation techniques used in real cases (Mack and Silverman 1982). Nevertheless, in practice, the input factors considered in sensitivity analysis are bounded and have densities with compact support.

### A.2 Proof of Theorem 1

This theorem is a direct consequence of the asymptotic behavior of the bias and variance in local linear regression.

Under assumptions (A0)–(A4), Fan et al. (1996) established that for a given kernel \( K(\cdot) \),

\[
\mathbb{E}_X(\hat{m}(x)) = m(x) + \frac{1}{2} \mu_2 \tilde{m}(x) h_1^2 + o_P(h_1^2), \tag{A.1}
\]

and

\[
\text{Var}_X(\hat{m}(x)) = \frac{\nu_0 \sigma^2(x)}{f_X(x) nh_1} + o_P(h_1^2), \tag{A.2}
\]

where \( \mu_k = \int u^K K(u) \, du \) and \( v_k = \int u^K K^2(u) \, du \). Now, because the estimator \( \hat{T}_1 \) is

\[
\hat{T}_1 = \frac{1}{n' - 1} \sum_{j=1}^{n'} (\hat{m}(\tilde{X}_j) - \hat{m})^2,
\]

we can write

\[
\hat{T}_1 = \frac{1}{n' - 1} \sum_{j=1}^{n'} (Z_j - \bar{Z})^2,
\]

where \( (Z_j)_{j=1,...,n'} := (\hat{m}(\tilde{X}_j))_{j=1,...,n'} \) and \( \bar{Z} = 1/n' \sum_{j=1}^{n'} Z_j \). By conditioning on the predictors \( X \), the sample \((Z_j|X)_{j=1,...,n'}\) is an iid sample distributed as \( Z_1|X \), and then the conditional bias of \( \hat{T}_1 \) can be obtained through the classical formula for the empirical estimator of the variance:

\[
\mathbb{E}_X(\hat{T}_1) = \text{Var}_X(Z_1) = \mathbb{E}_X(Z_1^2) - \mathbb{E}_X(Z_1)^2.
\]

Note that we can also compute its variance,

\[
\text{Var}_X(\hat{T}_1) = \frac{1}{n'} \left( \mathbb{E}_X((Z_1 - \mathbb{E}_X(Z_1))^2) - \frac{n' - 3}{n' - 1} (\text{Var}_X(Z_1))^2 \right),
\]

even though we do not use this result here (see Remark 1).

Because \( X \) is independent of \( X \) and \( Y \), we write

\[
\mathbb{E}_X(Z_1^2) = \int \mathbb{E}_X(\hat{m}(x)^2 f_X(x) \, dx
\]

\[
= \int \{\text{Var}_X(\hat{m}(x)) + \mathbb{E}_X(\hat{m}(x)^2)\} f_X(x) \, dx.
\]

Considering assumptions (A3), (A4), and (C0), using (A.1) and (A.2) in a similar way as for the standard MISE evaluation, we get

\[
\mathbb{E}_X(Z_1^2) = \int m(x)^2 f_X(x) \, dx + \frac{\nu_0}{nh_1} \int \sigma^2(x) \, dx
\]

\[
+ \frac{\mu_2 h_1^2}{2} \int m(x) \tilde{m}(x) f_X(x) \, dx + o_P(h_1^2),
\]

and by the same arguments, we also have

\[
\mathbb{E}_X(Z_1) = \int m(x) f_X(x) \, dx + \frac{1}{2} \mu_2 h_1 \int \tilde{m}(x) f_X(x) \, dx + o_P(h_1^2),
\]

which finally leads to

\[
\mathbb{E}_X(\hat{T}_1) = \mathbb{E}_X(Z_1^2) - \mathbb{E}_X(Z_1)^2
\]

\[
= \text{Var}(\mathbb{E}(Y|X))
\]

\[
+ \mu_2 h_1^2 \left[ \int m(x) \tilde{m}(x) f_X(x) \, dx \right]
\]

\[
- \left( \int m(x) f_X(x) \, dx \right) \left( \int \tilde{m}(x) f_X(x) \, dx \right)
\]

\[
+ \frac{\nu_0}{nh_1} \int \sigma^2(x) \, dx + o_P(h_1^2)
\]

\[
= \text{Var}(\mathbb{E}(Y|X)) + M_1 h_1^2 + \frac{M_2}{nh_1} + o_P(h_1^2),
\]
where
\[
M_1 = \mu_2 \left[ \int m(x)\tilde{m}(x)f_X(x)\,dx \right. \\
\left. - \left( \int m(x)f_X(x)\,dx \right) \left( \int \tilde{m}(x)f_X(x)\,dx \right) \right]
\]
and
\[
M_2 = \nu_0 \int \sigma^2(x)\,dx.
\]

A.3 Proof of Theorem 2

Similarly, we first recall the asymptotic results for the residual-based estimator of the conditional variance. Under assumptions (B0)–(B4), Fan and Yao (1998) showed that
\[
\mathbb{E}(\hat{\sigma}^2(x)) = \sigma^2(x) + \frac{1}{2} \mu_2 \hat{\sigma}^2(x)h_2^2 + o_P(h_1^2 + h_2^2)
\]
and
\[
\text{Var}(\hat{\sigma}^2(x)) = \frac{\nu_0 \sigma^4(x)\lambda^2(x)}{f_X(x)nh_2} + o_P\left( \frac{1}{\sqrt{nh_2}} \right),
\]
where \(\lambda^2(x) = \mathbb{E}((\epsilon - 1)^2|X = x)\) and \(\mu_2\) and \(\nu_0\) are as defined earlier. The estimator \(\hat{T}_2\) can be written as
\[
\hat{T}_2 = \frac{1}{n'} \sum_{j=1}^{n'} U_j,
\]
where \((U_j)_{j=1, \ldots, n'} := (\hat{\sigma}^2(X_j))_{j=1, \ldots, n'}\). As in the proof of Theorem 1, we then get the conditional bias and variance of \(\hat{T}_2\):
\[
\mathbb{E}(\hat{T}_2) = \mathbb{E}(U_1)
\]
and
\[
\text{Var}(\hat{T}_2) = \frac{1}{n'} \text{Var}(U_1).
\]

As \(\tilde{X}\) is independent of \(X\) and \(Y\), we have
\[
\mathbb{E}(U_1) = \int \mathbb{E}(\hat{\sigma}^2(x)f_X(x))\,dx.
\]

Considering assumptions (B4) and (C0) as in the proof of Theorem 1, we then get
\[
\mathbb{E}(\hat{T}_2) = \mathbb{E}(\text{Var}(Y|X)) + \frac{1}{2} \mu_2 h_2^2 \int \hat{\sigma}^2(x)f_X(x)\,dx \\
+ o_P(h_1^2 + h_2^2) \\
= \mathbb{E}(\text{Var}(Y|X)) + V_1 h_2^2 + o_P(h_1^2 + h_2^2),
\]
where
\[
V_1 = \frac{1}{2} \mu_2 \int \hat{\sigma}^2(x)f_X(x)\,dx,
\]
and, using the same arguments,
\[
\text{Var}(\hat{T}_2) = \frac{1}{n} \left\{ \mathbb{E}(\text{Var}(Y|X)^2) \\
+ \mu_2 h_2^2 \int \sigma^2(x)\hat{\sigma}^2(x)f_X(x)\,dx \\
- \mu_2 h_1^2 \int \hat{\sigma}^2(x)f_X(x)\,dx \left( \int \sigma^2(x)f_X(x)\,dx \right) \\
+ \frac{\nu_0}{nh_2} \int \sigma^4(x)\lambda^2(x)\,dx \\
+ o_P\left( h_1^2 + h_2^2 + \frac{1}{\sqrt{nh_2}} \right) \right\} \\
= \frac{1}{n} \left\{ \mathbb{E}(\text{Var}(Y|X)^2) + V_2 h_2^2 + V_3 h_1^2 + \frac{V_4}{nh_2} \\
+ o_P\left( h_1^2 + h_2^2 + \frac{1}{\sqrt{nh_2}} \right) \right\}.
\]

where
\[
V_2 = \mu_2 \int \sigma^2(x)\hat{\sigma}^2(x)f_X(x)\,dx,
\]
\[
V_3 = -\mu_2 \left( \int \hat{\sigma}^2(x)f_X(x)\,dx \right) \left( \int \sigma^2(x)f_X(x)\,dx \right),
\]
\[
V_4 = \nu_0 \int \sigma^4(x)\lambda^2(x)\,dx.
\]

ACKNOWLEDGMENTS

The authors thank Professor Anestis Antoniadis for very helpful discussions and the referees and associate editor for their very useful suggestions.

[Received June 2008. Revised April 2009.]

REFERENCES


