

1 Qualitative GSA measures for functions with correlated input parameters

This chapter explains the computation of the EE-based screening measures that are applied to the model in Keane and Wolpin (1994). First, I outline two sampling schemes that are tailored to the computation of EEs. Second, I present a simplified version of the approach to extend EEs to models with correlated parameters by Ge and Menendez (2017). Third, I develop an improvement to these measures that sustains the EE's fundamental derivative characteristic. The section concludes with the analysis of a test function. This analysis shows, first, my ability to replicate the results in Ge and Menendez (2017), and second, validates my improvement.

1.1 Sampling Schemes

The two presented sampling schemes are the trajectory and the radial design. Although the schemes are tailored to the computation of EEs, positional differences between them cause differences in their post-processing to compute the measures.

According to several experiments with common test functions by Campolongo et al. (2011), the best design is the radial design (Saltelli (2002)) and the most commonly used is the trajectory design (Morris (1991)). Both designs are comprised by a $(k + 1) \times k$ -dimensional matrix. The elements are generated in $[0, 1]$. Afterwards, they can potentially be transformed to the distributions of choice. The columns represent the different input parameters and each row is a complete input parameter vector. To compute the aggregate qualitative measures, a set of multiple matrices, or sample subsets, of input parameters has to be generated.

A matrix in radial design is generated the following way: Draw a vector of length $2k$ from a quasi-random sequence. The first row, or parameter vector, is the first half of the sequence. Then, copy the first row to the remaining k rows. For each row k' of the remaining $2, \dots, k + 1$ rows, replace the k' -th element by the k' -th element of the second half of the vector. This generates a matrix of the following form:

$$\mathbf{R}_{(k+1) \times k} = \begin{pmatrix} a_1 & a_2 & \dots & a_k \\ \mathbf{b}_1 & a_2 & \dots & a_k \\ a_1 & \mathbf{b}_2 & \dots & a_k \\ \vdots & \vdots & \ddots & \vdots \\ a_1 & a_2 & \dots & \mathbf{b}_k \end{pmatrix} \quad (1)$$

Note here, that each column consists only of the respective first row element, except of in one row. From this matrix, one EE can be obtained for each parameter X_i . This is

achieved by using the $(i + 1)$ -th row as function argument for the minuend and the first row as the subtrahend in the EE formula in Equation (??). Then, $\Delta^{(i,j)} = b_i^{(j)} - a_i^{(j)}$. The asterisk is an index for all elements of a vector.

$$d_i = \frac{Y(\mathbf{a}_{\sim i}, b_i) - Y(\mathbf{a})}{b_i - a_i} = \frac{Y(\mathbf{R}_{i+1,*}) - Y(\mathbf{R}_{1,*})}{b_i - a_i}. \quad (2)$$

If the number of radial subsamples is high, the draws from the quasi-random sequence lead to a fast coverage of the input space (compared to a random sequence). However, a considerable share of steps will be large, at maximum $1 - \epsilon$ in a sample space of $[0, 1]$. This amplifies the aforementioned problem of EE-based measures with non-linear functions. The quasi-random sequence considered here is the Sobol' sequence. It is comparably successful in the dense coverage of the unit hypercube, but also conceptually more involved. Therefore, its presentation is beyond the scope of this work. As the first elements of each Sobol' sequence, the direction numbers, are equal, I draw the sequence at once for all sets of radial matrices.

Next, I present the trajectory design. As we will see, it can lead to a relatively representative coverage for a very small number of subsamples but also to repetitions of similar draws. I skip the equations that generate a trajectory and present the method verbally instead. There are different forms of trajectories. I focus on the common version presented in Morris (1991) that generates equiprobable elements. The first step is to decide the number p of equidistant grid points in interval $[0, 1]$. Then, the first row of the trajectory is composed of the lower half value of these grid points. Now, fix $\Delta = p/[2(p - 1)]$. This function implies, that adding Δ to the lowest point in the lowest half results in the lowest point of the upper half of the grid points, and so on. It also implies that 0.5 is the lower bound of Δ . The rest of the rows is constructed, first, by copying the row one above and, second, by adding Δ to the i -th element of the $i + 1$ -th row. The created matrix scheme is depicted below.

$$\mathbf{T}_{(k+1) \times k} = \begin{pmatrix} a_1 & a_2 & \dots & a_k \\ \mathbf{b}_1 & a_2 & \dots & a_k \\ \mathbf{b}_1 & \mathbf{b}_2 & \dots & a_k \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_k \end{pmatrix} \quad (3)$$

In contrary to the radial scheme, each b_i is copied to the subsequent row. Therefore, the EEs have to be determined by comparing each row with the row above instead of with the first row. Importantly, two random transformations are common. These are, first, randomly switching rows, and second, randomly interchanging the i -th column with the $(k - i)$ -th column and then reversing the column. The first transformation is skipped

as it does not add additional coverage and because we need the stairs-shaped design to facilitate later transformations which account for correlations between input parameters. The second transformation is adapted because it is important to also have negative steps and because it sustains the stairs shape. Yet, this implies that Δ is also parameter- and trajectory-specific. Let f and h be additional indices representing the input parameters. The derivative formula is adapted to the trajectory design as follows:¹

$$d_i = \frac{Y(\mathbf{b}_{f \leq i}, \mathbf{a}_{h > i}) - Y(\mathbf{b}_{f < i}, \mathbf{a}_{h \geq i})}{b_i - a_i} = \frac{Y(\mathbf{T}_{i+1,*}) - Y(\mathbf{T}_{i,*})}{b_i - a_i}. \quad (4)$$

The trajectory design involves first, a fixed grid, and second and more importantly, a fixed step Δ , s.t. $\{\Delta\} = \{\pm\Delta\}$. This implies less step variety and less space coverage vis-à-vis the radial design for larger number of draws.

To improve the the sample space coverage by the trajectory design, Campolongo et al. (2007) develop a post-selection approach based on distances. The approach creates enormous costs for more than a small number of trajectories. This problem is effectively mitigated by Ge and Menendez (2014). The following describes the main ideas of both contributions.

The objective of both contributions is to select k trajectories from a set of N matrices. Campolongo et al. (2007) assign a distance to each pair of trajectories in the start set. To each set of pair combinations in every possible subset of k trajectories, Campolongo et al. assign an aggregate distance based on the pair distances. Then, the optimized trajectory set is the subset with the highest aggregate distance.

This is computationally very costly, because each aggregate distance is a sum of a binomial number of pair distances². To decrease the computation time, Ge and Menendez (2014) propose two improvements. First, in each iteration i , they select only $N(i) - 1$ matrices from a set containing $N(i)$ trajectories until the set size has decreased to k . Second, they compute the pair distances in each iteration based on the aggregate distances and the pair distances from the first set. Due to numerical imprecisions, their improvement does not always result in the same set as obtained from Campolongo et al. (2007). However, the sets are usually very similar in terms of the aggregate distance. This thesis only uses the first step in Ge and Menendez (2014) to post-select the trajectory set because the second step does not provide any gain.³

So far, we have only considered draws in $[0,1]$. For uncorrelated input parameters from arbitrary distributions with well-defined CDF, Φ , one would simply evaluate each element (potentially including the addition of the step) by the inverse CDF, or quantile function, Φ^{-1} , of the respective parameter. One intuition is, that Φ maps the sample space to $[0,1]$.

¹In contrary to most authors, I also denote the step as a subtraction instead of Δ when referring to the trajectory design. This provides additional clarity.

²For example, $\binom{30}{15} = 155117520$.

³This refers only to my implementation.

Hence Φ^{-1} can be used to transform random draws in $[0,1]$ to the sample space of the arbitrary distribution. This is a basic example of so-called inverse transform sampling which we will recall in the next section.

1.2 The approach for correlated input parameters in Ge and Menendez (2017)

This section describes the incomplete approach by Ge and Menendez (2017) to extend the EE-based measures to input parameters that are correlated. Their main achievement is to outline a transformation of samples in radial and trajectory design that incorporates the correlation between the input parameters. This implies, that the trajectory and radial samples cannot be written as in Equation (2) and Equation (4). The reason is that the correlations of parameter X_i , to which step Δ^i is added, imply that all other parameters $\mathbf{X}_{\sim i}$ in the same row with non-zero correlation with X_i are changed as well. Therefore, the rows cannot be denoted and compared as easily by a 's and b 's as in Equation (2) and (4). Transforming these matrices allows to re-define the EE-based measures accordingly, such that they sustain the main properties of the ordinary measures for uncorrelated parameters. The property is being a function of the mean derivative. Yet, Ge and Menendez (2017) do not fully develop these measures. I explain how their measures lead to arbitrary results for correlated input parameters. This section covers their approach in a simplified form, focussing on normally distributed input parameters, and presents their measures.

The next paragraph deals with developing a recipe for transforming draws $\mathbf{u} = \{u_1, u_2, \dots, u_k\}$ from $[0, 1]$ for an input parameter vector to draws $\mathbf{x} = \{x_1, x_2, \dots, x_k\}$ from an arbitrary joint normal distribution. We will do this in three steps.

For this purpose, let Σ be a non-singular variance-covariance matrix and let $\boldsymbol{\mu}$ be the mean vector. The k -variate normal distribution is denoted by $\mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$.

Creating potentially correlated draws \mathbf{x} from $\mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$ is simple. Following Gentle (2006)⁴, this can be achieved the following way: Draw a k -dimensional row vector of i.i.d standard normal deviates from the univariate $N(0, 1)$ distribution, such that $\mathbf{z} = \{z_1, z_2, \dots, z_k\}$, and compute the Cholesky decomposition of Σ , such that $\Sigma = \mathbf{T}^T \mathbf{T}$. The lower triangular matrix is denoted by \mathbf{T}^T . Then apply the operation in Equation (10) to obtain the correlated deviates from $\mathcal{N}_k(\boldsymbol{\mu}, \Sigma)$.

$$\mathbf{x} = \boldsymbol{\mu} + \mathbf{T}^T \mathbf{z}^T \quad (5)$$

Intuition for the underlying mechanics is provided in Appendix A.

The next step is to understand that we can split the operation in Equation (10) into

⁴See p. 197

two subsequent operations. The separated first part allows to potentially map correlated standard normal deviates to other distributions than the normal one. For this, let $\boldsymbol{\sigma}$ be the vector of standard deviations and let \mathbf{R}_k be the correlation matrix of \mathbf{x} .

The first operation is to transform the standard deviates \mathbf{z} to correlated standard deviates \mathbf{z}_c by using $\mathbf{z}_c = \mathbf{Q}^T \mathbf{z}^T$. In this equation, \mathbf{Q}^T is the lower matrix from the Cholesky decomposition $\mathbf{R}_k = \mathbf{Q}^T \mathbf{Q}$. This is equivalent to the above approach in Gentle (2006) for the specific case of the multivariate standard normal distribution $\mathcal{N}_k(0, R_k)$. This is true because for multivariate standard normal deviates, the correlation matrix is equal to the covariance matrix.

The second operation is to scale the correlated standard normal deviates: $\mathbf{z} = \mathbf{z}_c(\mathbf{i})\boldsymbol{\sigma}(\mathbf{i}) + \boldsymbol{\mu}$, where the i s indicate an element-wise multiplication. This equation is specific to normally distributed parameters.

The last step to construct the final approach is to recall the inverse transform sampling method. Therewith, we can transform the input parameter draws \mathbf{u} to uncorrelated standard normal draws \mathbf{z} . Then we will continue with the two operations in the above paragraph. The transformation from \mathbf{u} to \mathbf{z} is denoted by $F^{-1}(\Phi^c)$, where the c in Φ^c stands for the introduced correlations. This transformation is summarized by the following three steps:

$$\left. \begin{array}{l} \text{Step 1: } \mathbf{z} = \Phi^{-1}(\mathbf{u}) \\ \text{Step 2: } \mathbf{z}_c = \mathbf{Q}^T \mathbf{z}^T \\ \text{Step 3: } \mathbf{x} = \boldsymbol{\mu} + \mathbf{z}_c(\mathbf{i})\boldsymbol{\sigma}(\mathbf{i}) \end{array} \right\} F^{-1}(\Phi^c) \stackrel{\text{def}}{=} \mathcal{T}_2$$

To map u to different sample spaces, Step 3 can be substituted. For instance, this could be achieved by applying Φ^u , and the inverse CDF of the desired distribution to z_c .⁵

The one most important point to understand is that the transformation comprised by the three steps is not unique for correlated input parameters. Rather, the transformation changes with the order of parameters in vector \mathbf{u} ⁶. This can be seen from the lower triangular matrix \mathbf{Q}^T . To prepare the next equation, let $\mathbf{R}_k = (\rho_{ij})_{ij=1}^k$ and sub-matrix $\mathbf{R}_h = (\rho_{ij})_{ij=1}^h$, $h \leq k$. Also let $\boldsymbol{\rho}_i^{*,j} = (\rho_{1,j}, \rho_{2,j}, \dots, \rho_{i-1,j})$ for $j \geq i$ with the following abbreviation $\boldsymbol{\rho}_i \stackrel{\text{def}}{=} \boldsymbol{\rho}_i^{*,i}$. Following Madar (2015), the lower matrix can be written as

⁵The procedure described in the three steps above is equivalent to an inverse Rosenblatt transformation and a linear inverse Nataf transformation for parameters in normal sample space and connects to Gaussian copulas. For the first two transformations, see Lemaire (2013), p 78 - 113. These concepts can be used to transform deviates in $[0,1]$ to the sample space of arbitrary distributions by using the properties sketched above under varying assumptions.

⁶This point is more obvious in the formal definition of the Rosenblatt transformation.

$$\mathbf{Q}^T = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ \rho_{1,2} & \sqrt{1 - \rho_{1,2}^2} & 0 & \dots & 0 \\ \rho_{1,3} & \frac{\rho_{2,3} - \rho_{1,2}\rho_{1,3}}{\sqrt{1 - \rho_{1,2}^2}} & \sqrt{1 - \rho_3 \mathbf{R}_2^{-1} \rho_3^T} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{1,k} & \frac{\rho_{2,k} - \rho_{1,2}\rho_{1,k}}{\sqrt{1 - \rho_{1,2}^2}} & \frac{\rho_{3,k} - \rho_3^{*,k} \mathbf{R}_2^{-1} \rho_3^T}{\sqrt{1 - \rho_3 \mathbf{R}_2^{-1} \rho_3^T}} & \dots & \sqrt{1 - \rho_k \mathbf{R}_2^{-1} \rho_k^T} \end{pmatrix}. \quad (6)$$

Equation (6), together with Step 2, implies that the order of the uncorrelated standard normal deviates \mathbf{z} constitutes a hierarchy amongst the correlated deviates \mathbf{z}_c in the following manner: The first parameter is not subject to any correlations, the second parameter is subject to the correlation with the first parameter, the third parameter is subject to the correlations with the parameters before, etc. Therefore, if parameters are correlated, typically $\mathbf{Q}^T \mathbf{z}^T \neq \mathbf{Q}^T (\mathbf{z}')^T$ and $F^{-1}(\Phi)(\mathbf{u}) \neq F^{-1}(\Phi)(\mathbf{u}')$, where \mathbf{z}' and \mathbf{u}' denote \mathbf{z} and \mathbf{u} in different orders.

Coming back to the EE-based measures for correlated inputs, Ge and Menendez (2017) attempt to construct two variations of d_i . They name one EE the independent Elementary Effects, d_i^{ind} , and the other one the full Elementary Effects, d_i^{full} . For each of these two EEs, they derive the aggregate measures μ , μ^* and σ . The difference between the two EEs is that d_i^{ind} excludes and d_i^{full} includes the effect of the correlations from adding step Δ^i to X_i on the other parameters $\mathbf{X}_{\sim i}$. Knowing d_i^{ind} is important because the correlations can decrease d_i^{full} (close to 0). However, if the independent-EE-based measures are not close to zero, X_i is still important. Additionally, fixing this parameter can potentially change d_i^{full} for $\mathbf{X}_{\sim i}$.

For the trajectory design, the formula for the full Elementary Effect, is given by

$$d_i^{full,T} = \frac{f(\mathcal{T}(\mathbf{T}_{i+1,*}; i-1)) - f(\mathcal{T}(\mathbf{T}_{i,*}; i))}{\Delta}. \quad (7)$$

In Equation (12), $\mathcal{T}(\cdot; i) \stackrel{\text{def}}{=} \mathcal{T}_3(\mathcal{T}_2(\mathcal{T}_1(\cdot; i)); i)$. $\mathcal{T}_1(\cdot; i)$ orders the parameters, or row elements, to establish the right correlation hierarchy. \mathcal{T}_2 , or $F^{-1}(\Phi^c)$, correlates the draws in $[0, 1]$ and transforms them to the sample space. $\mathcal{T}_3(\cdot; i)$ reverses the element order back to the start, to be able to apply the subtraction in the numerator of the EEs row-by-row. Index i in $\mathcal{T}_1(\cdot; i)$ and $\mathcal{T}_3(\cdot; i)$ stands for the number of first row elements that is cut and moved to the back of the row in the same order. Applying $\mathcal{T}(\mathbf{T}_{i+1,*}; i-1)$ and $\mathcal{T}(\mathbf{T}_{i,*}; i)$ to all rows i of trajectory \mathbf{T} as in Equation (12) gives the following two transformed trajectories:

$$\mathcal{T}_1(\mathbf{T}_{\mathbf{i}+1,*}; i-1) = \begin{pmatrix} a_k & a_1 & \dots & \dots & a_{k-1} \\ \mathbf{b}_1 & a_2 & \dots & \dots & a_k \\ \mathbf{b}_2 & a_3 & \dots & \dots & \mathbf{b}_1 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_k & \mathbf{b}_1 & \dots & \dots & \mathbf{b}_{k-1} \end{pmatrix} \quad (8)$$

$$\mathcal{T}_1(\mathbf{T}_{\mathbf{i},*}; i-1) = \begin{pmatrix} a_1 & a_2 & \dots & \dots & a_k \\ a_2 & \dots & \dots & a_k & \mathbf{b}_1 \\ a_3 & \dots & \dots & \mathbf{b}_1 & \mathbf{b}_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_1 & \mathbf{b}_2 & \dots & \dots & \mathbf{b}_k \end{pmatrix} \quad (9)$$

Two points can be seen from Equation (8) and (9). First, $\mathcal{T}_1(\mathbf{T}_{\mathbf{i}+1,*}; i-1)$ and $\mathcal{T}_1(\mathbf{T}_{\mathbf{i},*}; i)$, i.e. the $(i+1)$ -th row in Eq. (8) and the (i) -th row in Eq. (9), only differ in the i -th element. The difference is $b_i - a_i$. Thus, these two rows can be used to compute the EEs like in the uncorrelated case in Equation (??). However, in this order, the parameters are in the wrong positions to be directly handed over to the function, as the i -th parameter is always in front. The second point is that in $\mathcal{T}_1(\mathbf{T}_{\mathbf{i}+1,*}; i-1)$, b_i is in front of the i -th row. This order prepares the establishing of the right correlation hierarchy by \mathcal{T}_2 , such that the Δ in $a_i + \Delta$ is included to transform all other elements representing $X_{\sim i}$. Importantly, to perform \mathcal{T}_2 , mean vector \mathbf{x} and covariance matrix $\mathbf{\Sigma}$ and its transformed representatives have always to be re-ordered according to each row. Then, \mathcal{T}_3 restores the original row order and d_i^{full} can comfortably be computed by comparing function evaluations of row $i+1$ in $\mathcal{T}(\mathbf{T}_{\mathbf{i}+1,*}; i-1)$ with function evaluations of row i in $\mathcal{T}(\mathbf{T}_{\mathbf{i},*}; i-1)$. Now, the two transformed trajectories only differ in the i -th element in each row i .

The formula for the independent Elementary Effect for the trajectory samples is given by

$$d_i^{ind,T} = \frac{f(\mathcal{T}(\mathbf{T}_{\mathbf{i}+1,*}; i)) - f(\mathcal{T}(\mathbf{T}_{\mathbf{i},*}; i))}{\Delta}. \quad (10)$$

Note that $\mathcal{T}(\mathbf{T}_{\mathbf{i}+1,*}; i)$ equals $\mathcal{T}(\mathbf{T}_{\mathbf{i},*}; i-1)$ which is the function argument in the subtrahend in Equation (7). Therefore this transformation can be skipped for the trajectory design and the transformed trajectory in Equation (10) can be recycled. Here, the X_i that takes the step in the denominator of the Elementary Effect is moved to the back such that the step does not affect the other input parameters $\mathbf{X}_{\sim i}$ through the correlations. The left trajectory is constructed such that for each row i , a_i does take the step that b_i took in the aforementioned minuend trajectory. The argument of the subtrahend in the denominator in d_i^{ind} is given by the rows in

$$\mathcal{T}_1(\mathbf{T}_{\mathbf{i},*}; i-1) = \begin{pmatrix} a_2 & a_3 & \dots & \dots & a_1 \\ a_3 & \dots & a_k & \mathbf{b}_1 & a_2 \\ a_4 & \dots & \mathbf{b}_1 & \mathbf{b}_2 & a_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{b}_2 & \mathbf{b}_3 & \dots & \dots & \mathbf{b}_1 \end{pmatrix} \quad (11)$$

The transformation for the samples in radial design work equally except that the minuend samples are composed of the first row copied to each row below because the steps in the radial design are always taken from the draws in the first row. To account for the correlation hierarchy and the transformation to the sample space performed by \mathcal{T}_2 , it is important to also reorder the minuend trajectories. In fact, it is enough to compute one minuend trajectory if the order is regarded correctly. The formulae of the Elementary Effects for the radial design are given by Equation (17) and (18).

$$d_i^{full,R} = \frac{f(\mathcal{T}(\mathbf{R}_{\mathbf{i}+1,*}; i-1)) - f(\mathcal{T}(\mathbf{R}_{\mathbf{1},*}; i))}{b_i - a_i}. \quad (12)$$

$$d_i^{ind,R} = \frac{f(\mathcal{T}(\mathbf{R}_{\mathbf{i}+1,*}; i)) - f(\mathcal{T}(\mathbf{R}_{\mathbf{1},*}; i))}{b_i - a_i}. \quad (13)$$

The trajectory scheme requires $N(3k+1)$ and the radial scheme $N3k$ function evaluations to compute both EEs. The first row of the minuend and the last row of the subtrahend trajectories can be skipped. If one transformed trajectory contains information for two EEs, then all rows are used for the trajectory design. The radial scheme requires one row evaluation less because the subsamples that contain the repeated first rows include a pair of identical rows.

In the next section, I present the shortcomings of the EEs by Ge and Menendez (2017). Furthermore, I develop improved measures that finally correspond to the actual EE definition for uncorrelated parameters.

1.3 Drawbacks in Ge and Menendez (2017) and Corrected Elementary Effects.

The drawback in the EE definitions by Ge and Menendez (2017) is that Δ is transformed multiple times in the numerator but not in the denominator. Therefore, these measures are not Elementary Effects in the sense of a derivative. To understand the intuition, it is easier to view Δ as $b - a$. The transformation in the numerator is performed by applying

$F^{-1}(\Phi^c)$ to $w_i^j = a_i^j + \Delta^{(i,j)}$. The implications of this approach is twofold. For the following explanation, I refer to the normal sample space. The first implication is that function f is evaluated at arguments that are non-linearly transformed by Step 1 and Step 2. Then, the difference in the numerator is divided by the difference between the changed input parameter and the base input parameter - in unit space. This makes the results hard to interpret. It also increases the influence of more extreme draws in $[0, 1]$ because Φ^{-1} is very sensitive to these. Therefore, it will take a larger sample for the aggregate EE measures in Ge and Menendez (2017) to converge. Additionally, these problems are amplified if there are large differences between the inputs' standard deviation through the subsequent multiplicative scaling. The large sensitivity to more extreme draws implies also a higher sensitivity to larger differences in $b - a$. Therefore, the results will differ in their level depending on the sampling scheme that is used. The largest drawback, however, is that $b_i - a_i$ in the denominator of d_i^{ind} does not account for the transformation of the $b_i - a_i$ in the nominator by the establishing of correlations in Step 2. This transformation decreases $b_i - a_i$ proportional to the degree of correlations of the respective input parameter as can be seen by the last row of the lower Cholesky matrix in Eq. (6). Hence, d_i^{ind} is inflated by the input parameters' correlations even though this measure is introduced as an independent effect. Because the full EE has to be interpreted with regards to the independent Effect, this problem spills over to d_i^{full} . For these reasons, I argue that these measures can not be used for screening. I refine the measures in Ge and Menendez (2017) by scaling the Δ in the denominator according to the nominator. I refer to these measures as the correlated and the uncorrelated Elementary Effects, d_i^c and d_i^u . They are given below for arbitrary input distributions and for samples in trajectory and radial design. For this purpose, let $Q_{k,*k-1}^T$ be the last row except of the last element of the lower triangular Cholesky matrix of the correlation matrix in the respective order of $\mathcal{T}_1(\mathbf{T}_{i+1,*}; i-1)$ and $\mathcal{T}_1(\mathbf{T}_{i-1,*}; i-1)$. Let $Q_{k,k}^T$ be the last element of the same matrix. Let Φ^u be the transformation from $[0, 1]$ to standard normal space without regarding correlations in Step 3. Let F^{-1} be the transformation that maps standard normal deviates to an arbitrary sample space like in Step 1. Index $*k-1$ represents all elements except of the last one of a vector of length k . Index j is used to indicate an element-wise multiplication.

$$d_i^{c,T} = \frac{f(\mathcal{T}(\mathbf{T}_{i+1,*}; i-1)) - f(\mathcal{T}(\mathbf{T}_{i-1,*}; i))}{F^{-1}(\Phi^u(b_i)) - F^{-1}(\Phi^u(a_i))} \quad (14)$$

$$d_i^{u,T} = \frac{f(\mathcal{T}(\mathbf{T}_{i+1,*}; i)) - f(\mathcal{T}(\mathbf{T}_{i,*}; i))}{F^{-1}(Q_{k,*k-1}^T(j)T_{i+1,*k-1}^T(j) + Q_{k,k}^T\Phi^u(b_i)) - F^{-1}(Q_{k,*k-1}^T(j)T_{i,*k-1}^T(j) + Q_{k,k}^T\Phi^u(a_i))} \quad (15)$$

$$d_i^{c,R} = \frac{f(\mathcal{T}(\mathbf{R}_{i+1,*}; i-1)) - f(\mathcal{T}(\mathbf{R}_{1,*}; i-1))}{F^{-1}(\Phi^u(b_i)) - F^{-1}(\Phi^u(a_i))} \quad (16)$$

$$d_i^{u,R} = \frac{f(\mathcal{T}(\mathbf{R}_{i+1,*}; i)) - f(\mathcal{T}(\mathbf{R}_{1,*}; i))}{F^{-1}(Q_{k,*k-1}^T(j)R_{i+1,*k-1}^T(j) + Q_{k,k}^T\Phi^u(b_i)) - F^{-1}(Q_{k,*k-1}^T(j)R_{i,*k-1}^T(j) + Q_{k,k}^T\Phi^u(a_i))}. \quad (17)$$

In Equation (15) and (17), $Q_{k,*k-1}^T(j)R_{i,*k-1}^T(j)$ refers to the transformed trajectory of the copied first rows, $\mathcal{T}(\mathbf{R}_{i+1,*}; i)$, as previously described.

The denominator of the correlated Elementary Effect, d_i^c , consists of Φ and F^{-1} . These functions account for the transformation from uniform to arbitrary sample space. In d_i^c , the denominator must not account for a correlating transformation of a_i and b_i because $Q_{1,*}^T = (1, 0, 0, \dots, 0)$. Therefore, b_i and a_i are multiplied by one in the correlation step, Step 2, of the standard normal deviates.

The last point is not true for the uncorrelated Elementary Effect $d_i^{c,*}$. One accounts for the correlation step by multiplying $\Phi^u(b_i)$ by $Q_{k,*k-1}^T(j)T_{i+1,*k-1}^T(j)/\Phi^u(b_i) + Q_{k,k}^T$. The respective operation is also done for $\Phi^u(a_i)$. One can also write the denominator equally but less explicitly for all four EEs as $\mathcal{T}(b_i) - \mathcal{T}(a_i)$ ⁷. Not accounting for Q_t like Ge and Menendez (2017) leads to arbitrarily decreased independent Elementary Effects for input parameters with higher correlations. As d_i^{full} and d_i^{ind} are interpreted jointly, both measures are in fact useless.

For $X_1, \dots, X_k \sim \mathcal{N}_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the denominator of $d_i^{u,*}$ simplifies drastically to

$$\begin{aligned} & (\mu_i + \sigma_i(Q_{k,*k-1}^T(j)T_{i+1,*k-1}^T(j) + Q_{k,k}^T\Phi^u(b_i)) \\ & - (\mu_i + \sigma_i(Q_{k,*k-1}^T(j)T_{i+1,*k-1}^T(j) + Q_{k,k}^T\Phi^u(a_i)) \\ & = \sigma_i Q_{k,k}^T(\Phi^u(b_i) - \Phi^u(a_i)). \end{aligned} \quad (18)$$

The next section presents results for a test function to demonstrate a replication of Ge and Menendez (2017) and to validate this thesis' measures.

⁷From an implementation perspective, this is also the simplest and most general way. It only requires to transfer the right elements from the final, transformed samples to the denominator.

1.4 Replication and Validation

Let $g(X_1, \dots, X_k) = \sum_{i=1}^k c_i X_i$ be an arbitrary linear function. Let $\rho_{i,j}$ be the linear correlation between X_i and X_j . Then, for all $i \in 1, \dots, k$, I expect:⁸

$$d_i^{u,*} = c_i \quad (19)$$

$$d_i^{c,*} = \sum_{j=1}^k \rho_{i,j} c_j. \quad (20)$$

Both equations state that, conceptually, the result does not depend on the sampling scheme.

Equation (19) shows that the uncorrelated EE of each input equals the derivative of $f(\mathbf{X})$ with respect to X_i . This corresponds to the concept of the original EE as a derivative. As the function is linear, the derivative is equal for all values in the sample space of each X_i . This implies $\mu = \mu^* = d_i$ and $\sigma = 0$ for both sampling schemes and both EE versions.

Equation (20) states that $d_i^{c,*}$ also includes the effects of the other input parameters that are correlated with X_i . Then, the function's linearity implies that these effects are included by adding the respective coefficients multiplied by the correlation with X_i to the uncorrelated effect $d_i^{u,*}$.

Table 1 and Table 2 show results for a simple linear test function with 3 input parameters, where $X_i \sim \mathcal{N}(0, 1)$ for $i = \{1, 2, 3\}$, and with correlations $\rho_{1,2} = 0.9$, $\rho_{1,3} = 0.4$, $\rho_{2,3} = 0.01$. The first columns in both tables show the aggregate EE-based measures by Ge and Menendez (2017). The second (and third) columns show my replications for these measures. Table 1 consists of two replications because the trajectory design contains two numerical decisions. The details are explained in the following. The last columns in both tables depict the results for the measures developed in this thesis. Ge and Menendez generate 240 trajectories and 260 radial samples to compute their results. My figures base on 10,000 samples for each scheme.

⁸These results correspond to the intuition provided by the example in Saltelli et al. (2008), p. 123.

Table 1. Replication and validation - trajectory design

Measure	GM'17	Repl. $\mu^{*\dagger}$	Repl. σ^\ddagger	S'20
$\mu^{*,ind}$	1.20	1.36	0.83	1.00
	1.30	1.48	0.91	1.00
	3.20	3.11	1.94	1.00
σ^{ind}	0.55	0.00	0.56	0.00
	0.60	0.00	0.62	0.00
	1.30	0.00	1.32	0.00
$\mu^{*,full}$	14.90	16.20	9.97	2.30
	12.50	13.45	8.31	1.91
	10.00	9.93	6.18	1.41
σ^{full}	6.50	0.00	6.74	0.00
	5.50	0.00	5.63	0.00
	4.00	0.00	4.20	0.00

 $\dagger 0^{num} = 0.00001$ and $l = 4$. $\ddagger 0^{num} = 0.00000001$ and $l = 24$.**Table 2.** Replication and validation - radial design

Measure	GM'17	Replication	S'20
$\mu^{*,ind}$	0.60	0.57	1.00
	0.75	0.85	1.00
	1.50	1.31	1.00
σ^{ind}	0.20	0.10	0.00
	0.30	0.41	0.00
	0.85	0.22	0.00
$\mu^{*,full}$	7.50	6.84	2.30
	6.80	7.77	1.91
	4.75	4.19	1.41
σ^{full}	2.90	1.15	0.00
	2.65	3.68	0.00
	2.50	0.70	0.00

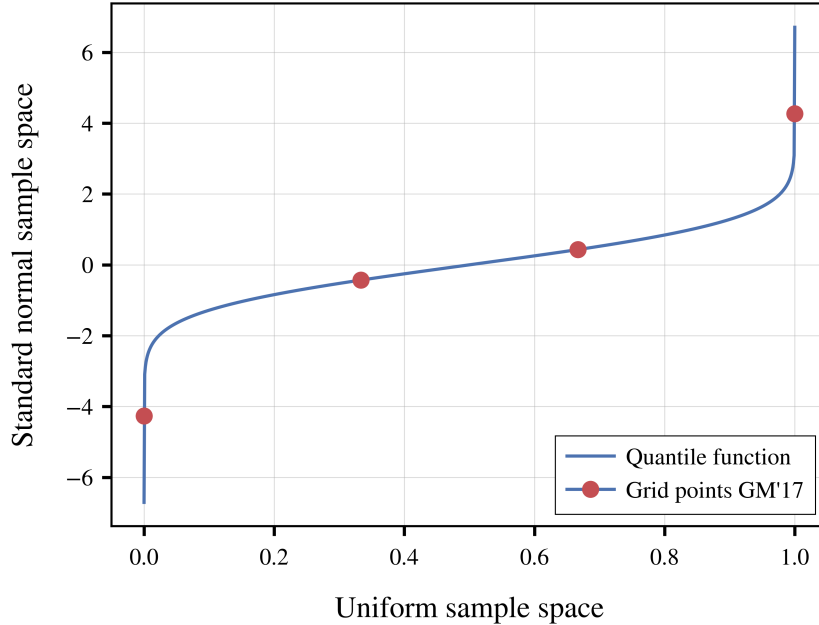
From the first columns, we can observe three things about the measures in Ge and Menendez (2017). First the measures are not related to derivatives, as the non-linear transformation $F(\Phi^c(\cdot))$ is only applied to the EE numerator. This transformation leads to two additional observations.

The second result is that the measures differ between trajectory and radial design. This is because the non-linearity introduces variation to these measure according to the values of the draws in $[0, 1]$. This variation is due to two reasons. The first reason are the general differences between both sampling schemes.

The last observation results from the fact that the trajectory design, in contrary to the radial design, includes two numeric parameters. These are, first, the approximation of 0 and 1 to prevent the inverse CDF from returning infinite values. It is enough to approximate 0 by 0^{num} . The lower 0^{num} , the more extreme are the transformed draws for 0 and 1. Therefore, lower 0^{num} also increase σ . The second approximation is the number of levels, or grid points, l in $[0, 1]$. The more grid points, the less imprecise are the results. The authors indicate that they use $0^{num} = 0.00001$ and l for the trajectory design.⁹

My replications for the trajectory design are depicted in the second column of Table 1. The mean and mean absolute EEs, $\mu^{*,ind}$ and $\mu^{*,full}$ come reasonably close. However, the standard deviations, $\sigma^{*,ind}$ and $\sigma^{*,full}$ equal 0. The reason is that four grid points re-linearise the transformed numerator in the EEs. This argument is illustrated by Figure 1.

Figure 1. Grid points in standard normal sample space for trajectory design with $l = 4$



The grid points with four levels are $\{0^{num}, 1/3, 2/3, 1 - 0^{num}\}$ and $\Delta = 2/3$. The quantile function is non-linear but point-symmetric to $(0.5, 0)$. Thus, $\Phi(\frac{1/3 - (1 - 0^{num})}{2/3}) = \Phi(\frac{0^{num} - 2/3}{2/3})$.

⁹I obtained this information from an unofficial email from Qiao Ge.

Hence, for each single X_i the Elementary Effects are constant the transformed Δ -specific derivation is equal at each of the two base grid points.

Therefore, I used $0^{num} = 0.00000001$ and $l = 24$ to increase the variation amongst the EEs by Ge and Menendez (2017) in order to achieve similar results.

The replication in Table 2 for the radial design is relatively close for the mean Effects. However, there are considerable differences for the standard deviations. In my view, this results, first, from different start values from which the Sobol' sequence is generated, and second, from a too low number of samples in Ge and Menendez (2017). This indicates that the appropriate number of samples for correlated parameters has to be higher than what is common for functions with uncorrelated input parameters.

The last column in both Table 1 and Table 2 depicts the aggregate measures based on the improved EEs. These are equal to what is anticipated at the beginning of this section. Therefore, the correlated and uncorrelated Elementary Effects developed in this thesis can be regarded as a valid extension of Elementary Effects for models with correlated input parameters.

To facilitate the analysis in the results section, I describe the link from EE-measures to Sobol' indices by means of another linear test function h . However, let us consider the case without any correlations between the inputs. Additionally, let $c_i = \{3, 2, 1\}$ and $\sigma_{X_i}^2 = \{1, 4, 9\}$. The following results are derived from Saltelli et al. (2008)¹⁰. Let us first compute the Sobol' indices. As g does not include any interactions, $S_i^T = S_i$. Additionally, we have $\text{Var}(Y) = \sum_{i=1}^k c_i^2 \sigma_{X_i}^2$ and $\text{Var}_{\mathbf{X}_{\sim i}}(\mathbb{E}_{X_{\sim i}}[Y|\mathbf{X}_{\sim i}]) = c_i^2 \sigma_{X_i}^2$. Table 3 compares three different sensitivity measures. These are the total Sobol' indices, S_i^T , the mean absolute EE, γ_i^* , and the *squared* sigma-normalized mean absolute EE, $(\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y})^2$.

Table 3. Comparison of sensitivity measures for a linear function

Parameters	S_i^T	γ_i^*	$(\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y})^2$
X_1	9	3	9
X_2	8	2	8
X_3	9	1	9

In context of screening, S_i^T is the objective measure that we would like to predict approximately. We observe that γ_i^* ranks the parameters incorrectly. The reason, is that γ_i^* is only a measure of the influence of X_i on Y and not of the influence of the variation and the level of X_i on the variation of Y . We also see that $(\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y})^2$ is a exact predictor for S_i^T as it does not only generate the correct ranking but also the right effect size. Of course, this result is specific to a linear function without interactions and correlation. However, it

¹⁰See p.15-25.

underlines the point that γ_i^* alone is not sufficient for screening. One approach would be to additionally consider the EE variation, σ_i . However, analysing two measures at once is difficult for models with a large number of input parameters. Table 3 indicates that $\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y}$ and also $\mu_i^* \frac{\sigma_{X_i}}{\sigma_Y}$ can be an appropriate alternative. The actual derivative version of this measure is also recommended by guidelines of the Intergovernmental Panel for Climate Change (IPCC (1999, 2000)).

The next chapter presents the occupational choice model in Keane and Wolpin (1994).

References

- Campolongo, F., J. Cariboni, and A. Saltelli (2007). An effective screening design for sensitivity analysis of large models. *Environmental modelling & software* 22(10), 1509–1518.
- Campolongo, F., A. Saltelli, and J. Cariboni (2011). From screening to quantitative sensitivity analysis. a unified approach. *Computer Physics Communications* 182(4), 978–988.
- Ge, Q. and M. Menendez (2014). An efficient sensitivity analysis approach for computationally expensive microscopic traffic simulation models. *International Journal of Transportation* 2(2), 49–64.
- Ge, Q. and M. Menendez (2017). Extending morris method for qualitative global sensitivity analysis of models with dependent inputs. *Reliability Engineering & System Safety* 100(162), 28–39.
- Gentle, J. E. (2006). *Random number generation and Monte Carlo methods*. Springer Science & Business Media.
- IPCC (1999). Ipcc expert meetings on good practice guidance and uncertainty management in national greenhouse gas inventories. Background papers.
- Keane, M. P. and K. I. Wolpin (1994). The solution and estimation of discrete choice dynamic programming models by simulation and interpolation: Monte Carlo evidence. *Review of Economics and Statistics* 76(4), 648–672.
- Lemaire, M. (2013). *Structural reliability*. John Wiley & Sons.
- Madar, V. (2015). Direct formulation to cholesky decomposition of a general nonsingular correlation matrix. *Statistics & probability letters* 103, 142–147.
- Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics* 33(2), 161–174.
- Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer physics communications* 145(2), 280–297.
- Saltelli, A., M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola (2008). *Global Sensitivity Analysis: The Primer*. John Wiley & Sons.