Total Effects with Constrained Features

Emanuele Borgonovo, Elmar Plischke, Clémentine Prieur

Total Effects with Constrained Features

Emanuele Borgonovo* Elmar Plischke† Clémentine Prieur‡

May 16, 2023

Recent studies have emphasized the connection between machine learning feature importance measures and total order sensitivity indices (total effects, henceforth). Feature correlations and the need to avoid unrestricted permutations make the estimation of these indices challenging. Additionally, there is no established theory or approach for non-Cartesian domains. We propose four alternative strategies for computing total effects that account for both dependent and constrained features. Our first approach involves a generalized winding stairs design combined with the Knothe-Rosenblatt transformation. Our second approach is a U-statistic estimator that combines the Jansen intuition with a weighting factor. The U-statistic framework allows the derivation of a central limit theorem for this estimator. However, this design is computationally intensive. Then, our third approach uses derangements to significantly reduce computational burden. We prove consistency and central limit theorems for these estimators as well. Our fourth approach is based on a nearest-neighbour intuition and it further reduces computational burden. We test these estimators through a series of increasingly complex computational experiments with features constrained on compact and connected domains (circle, simplex), non-compact and non-connected domains (Sierpinski gaskets), and conclude with an application to a realistic simulator.

Keywords: Feature Importance; Constrained Features; Winding Stairs; U-Statistics

1 Introduction

Determining feature importance is a crucial task in machine learning and statistical investigations. In machine learning, it is an integral part of post-hoc explainability

*Università Commerciale Luigi Bocconi and Bocconi Institute for Data Science and Analytics, Milano 20138, Italy

†Technische Universität Clausthal, 38678 Clausthal-Zellerfeld, Germany

‡Univ. Grenoble Alpes, CNRS, Inria, Grenoble INP, LJK, 38000 Grenoble, France

[41, 15], where it helps us understand the degree to which a model relies on the available features. This understanding has two final objectives [19, 2]. The first objective is dimensionality reduction, which involves screening out features that do not contribute to the model's predictions. The second objective is identifying the features that are most important for further modeling efforts or data collection by domain experts.

Over the years, several feature importance measures have been developed to perform this task. On the one hand, in machine learning, a particularly important family is represented by Breiman's permutation importance measures [4]. Breiman originally defines them based on the notion of mean decrease accuracy [2]. The intuition is as follows: A given machine learning model (e.g., a random forest) is fitted to a feature-target dataset, yielding a given predictive accuracy. The values of a specific feature are then permuted to break its relation to the target. The predictive accuracy is then reassessed for this perturbed dataset. The difference between the new (possibly degraded) and the original accuracies provides us with an indication about the importance of the feature. However, [2, Proposition 2] show that there is no consensus on the exact mathematical formulation of the mean decrease accuracy and they prove that alternative software implementations yield different values. On the other hand, in statistics, a central role is played by measures of statistical association. Several indicators have been developed over the years: From the original Pearson linear correlation coefficient [47], to the new correlation coefficient of [8]. In this family, a significant role is played by the so-called total order sensitivity effects [25]— total effects for short.

Total effects are defined as the difference between the variance of the target and the portion that remains after all features have been fixed with the exception of the feature of interest. [25] show that, when features are independent, the total effect of a given feature equals the sum of all terms in the ANOVA expansion associated with that feature. Also, we can calculate total effects as the expectation of the squared difference of the values of the model output in two points that differ only in the value of the feature of interest. The new point can be obtained by a simple permutation of the values of the features in the dataset. This formulation is known as the Jansen's estimation method [30, 31] and has inspired the so-called pick-and-freeze designs. In the class of pick-and-freeze estimators the one proposed by [29] is proven to be asymptotically efficient.

[22] show that even in the case of dependent features, total effects retain an interpretation from a relative error perspective. Under a squared loss function, a total index is the expected loss increase for approximating the input-output mapping with a function that does not contain the terms associated with the feature of interest. Also, [2] show that total effects are closely related to Breiman's mean decrease accuracy. In particular, [2, Proposition 2] show that the different software implementations do not converge to the total effects but to a quantity whose bias increases with dependence, and is potentially amplified by interactions. They propose corrections so that the calculation of Breiman's mean decrease accuracy indeed converges to a total effect in the case the machine learning model is a random forest.

The presence of statistical dependence complicates the calculation of total effects (we refer to [9, Ch. 5] for a thorough account). First, the interpretation in terms of the correspondence with the sum of terms in the ANOVA decomposition is lost. Second,

also the possibility to use a Jansen-type estimator is not straightforward. In fact, while under independence the new points can be obtained with unrestricted permutations, the presence of dependences challenges such procedure. The problem is similar to the one signalled by [26] in the machine learning literature: unrestricted permutations may make the new points fall in regions that are far from where the data lie, forcing the machine learning model to extrapolate. These difficulties are compounded when features are not only dependent but also constrained on non-Cartesian supports. Constraints arise in applications when physical or business reasons require features to be located in certain regions. Here, an unrestricted permutation could lead to a feature that falls outside a given constraint, making the evaluation of the model not only at risk of extrapolation, but also meaningless. Furthermore, if constraints give rise to disconnected feature domains, they make the functional ANOVA expansion ill-defined [46]. [34] propose a numerical approach based on the combination of rejection sampling and quadrature for the calculation of variance-based indices, with focus on numerical aspects. However, a statistical analysis of possible estimators with constrained inputs is missing.

Our goal is to address the estimation of total effects with both dependent and constrained features, considering numerical as well as theoretical aspects. We proceed as follows. First, we extend the estimator of [31] to the case of dependent inputs. We show that it is still possible to estimate total effects under input dependence using a Jansen-like approach if the new value of the feature is obtained under conditional independence. We then propose a generalized winding stairs design based on the Knothe-Rosenblatt transform that can be used in association with a vast family of input dependencies. However, while this design conceptually pushes the boundaries of available methods for dependent inputs, it becomes impractical when inputs are constrained.

We then introduce a new estimator of total indices by applying a weighting factor (called density quotient) to the extended Jansen's estimator. We show that the density quotient can be reinterpreted as a block-copula density, that vanishes when inputs are outside the constraints and that becomes unity when inputs are independent. We then formulate a U-statistic version of the estimator and obtain a central limit theorem. However, the new U-statistic estimator turns out to be computationally expensive as it requires the evaluation of the model at n^2 points. We then propose an alternative estimator based on a single permutation that reduces computational burden. We consider first the simplest estimator with the permutation given by a one-shift in the coordinate of interest and prove a central limit theorem of this estimator. We then extend the result for general derangements in the coordinate of interest. To further abate computational burden, we also introduce a nearest-neighbour estimator that makes the estimation cost independent of the number of features.

We derive analytical expressions for the estimators in the case of linear models and Gaussian inputs. We then challenge the estimators on test cases of increasing complexity, starting with Cartesian domains with dependent features, and moving to connected non-Cartesian domains (e.g., circle and simplex) to disconnected non-Cartesian domains (such as non-overlapping triangles and Sierpinski gaskets), and conclude with application to a realistic simulator, the flood example of [10].

2 Total Effects: Bridging Old and New

In this section, we review total effects from a fresh perspective. We discuss the covariance representation of total effects and establish a link with an early result by Fréchet. We underline the role of conditional independence in the estimation of total effects with independent as well as dependent inputs. The analysis allows us to propose a new estimator for total effects with dependent inputs based on winding stairs and the Knothe-Rosenblatt transformation. In Section 2.2, we highlight the roles of conditional independence for such a representation. In Section 2.3, we propose a new estimator based on winding stairs and on the Knothe-Rosenblatt transformation for the case in which input dependence can be expressed via a Gaussian copula.

2.1 A Fréchet Perspective

Let us consider a reference probability space $(\Omega, \mathcal{B}(\Omega), \mathrm{pr})$, where $\mathcal{B}(\Omega)$ is a Borel σ-algebra. Let also X, Y be random variables on $(\Omega, \mathcal{B}(\Omega), \mathrm{pr})$, with supports $\mathcal{X}$, $\mathcal{Y}$. We let $X = (X_1, X_2, \ldots, X_d)$ be a d-dimensional random vector in $\mathbb{R}^d$, so that $\mathcal{X} \subseteq \mathbb{R}^d$ and consider a univariate Y, with $\mathcal{Y} \subseteq \mathbb{R}$. For the moment, we make the further assumption that the support of X is Cartesian, that is $\mathcal{X} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_d$, where $\mathcal{X}_i$ is the support of X_i, $i = 1, 2, \ldots, d$. We denote the cumulative distribution function and probability density functions of X and Y by $F_X(x)$, $f_X(x)$ and $F_Y(y)$, $f_Y(y)$, respectively. For notation simplicity, we regard X and Y as continuous in the remainder. We suppose that Y has finite second moment, $\mathbb{V}[Y] < \infty$. Let $u \subseteq [d] = \{1, \ldots, d\}$, e.g., $u = \{i_1, i_2, \ldots, i_k\}$ with $k \leq d$. Let x_u correspond to the $|u|$-dimensional vector whose components are indexed by u, and x_{-u} the $(d - |u|)$-dimensional vector whose components are indexed by the complement of u, $-u = [d] \setminus u$. For a single factor i, we have $u = i$ and we write the all-but-one set $[d] \setminus \{i\}$ as $-i$. The total effect of X_u is defined as [25]

$$\tau_u = \mathbb{E}[\mathbb{V}[Y|X_{-u}]] = \mathbb{V}[Y] - \mathbb{V}[\mathbb{E}[Y|X_{-u}]]. \tag{1}$$

The literature has also introduced the normalized total effect as $T_u = \tau_u / \mathbb{V}[Y]$. Using an argument of [16], we find the following useful equalities.

Lemma 1. *Let Y' be a replicate of Y conditionally independent on X_{-u}, i.e., Y and Y' have same distribution and satisfy $\mathrm{pr}(Y \cdot Y'|X_{-u}) = \mathrm{pr}(Y|X_{-u})\,\mathrm{pr}(Y'|X_u)$. Then,*

$$\tau_u = \mathbb{V}[Y] - \mathrm{cov}(Y, Y') = \tfrac{1}{2}\,\mathbb{E}\left[(Y - Y')^2\right]. \tag{2}$$

Proof. [16] shows that

$$\begin{aligned}\mathbb{V}[\mathbb{E}[Y|X_{-u}]] &= \mathbb{E}[\mathbb{E}[Y|X_{-u}]^2] - \mathbb{E}[Y]^2 = \mathbb{E}[\mathbb{E}[Y \cdot \mathbb{E}[Y|X_{-u}] - \mathbb{E}[Y]^2|X_{-u}]] \\ &= \mathrm{cov}(Y, \mathbb{E}[Y|X_{-u}]).\end{aligned}$$

Then, we obtain the covariance representation of τ_u:

$$\tau_u = \mathbb{E}[\mathbb{V}[Y|X_{-u}]] = \mathbb{V}[Y] - \mathrm{cov}(Y, \mathbb{E}[Y|X_{-u}]). \tag{3}$$

Now, let Y' be an independent replicate of Y conditionally on X_{-u}. As a consequence of conditional independence, $\mathbb{E}[Y \cdot Y'|X_{-u}] = \mathbb{E}[Y|X_{-u}]\,\mathbb{E}[Y'|X_{-u}]$. Hence

$$\begin{aligned}\operatorname{cov}(Y, Y') &= \mathbb{E}[\mathbb{E}[Y \cdot Y' - \mathbb{E}[Y]^2|X_{-u}]] = \mathbb{E}[\mathbb{E}[Y|X_{-u}] \cdot \mathbb{E}[Y'|X_{-u}]] - \mathbb{E}[Y]^2 \\ &= \mathbb{V}[\mathbb{E}[Y|X_{-u}]],\end{aligned}$$

so that

$$\tau_u = \mathbb{V}[Y] - \operatorname{cov}(Y, Y') = \tfrac{1}{2}\left(\mathbb{V}[Y] + \mathbb{V}[Y']\right) - \operatorname{cov}(Y, Y') = \tfrac{1}{2}\,\mathbb{E}\left[(Y - Y')^2\right]$$

where $\mathbb{V}[Y] = \mathbb{V}[Y']$ as both are identically distributed. □

The first equality in (2) substitutes the possibly high-dimensional nonlinear regression $\mathbb{E}[Y|X_{-u}]$ in (1) with a covariance operation. When replacing the regression surface by Y', the error term $Y' - \mathbb{E}[Y|X_{-u}]$ is uncorrelated to Y, because $\operatorname{cov}(Y, Y' - \mathbb{E}[Y|X_{-u}]) = 0$. The second equality can be interpreted in terms of Jansen's equality for total effects [30, 31] of which it is a generalization, as it does not require feature independence.
In simulation and machine learning Y is often a function of X, $Y = g(X)$, $g : \mathcal{X} \to \mathbb{R}$. Suppose that g is square integrable, and that it can be decomposed as

$$g(x) = g_0 + \sum_{u \in 2^{[d]}, u \neq \emptyset} g_u(X_u), \tag{4}$$

with $2^{[d]}$ the power set of $[d]$, $g_0 = \mathbb{E}[Y]$, and $g_u(x_u) = \mathbb{E}[Y|X_u = x_u] - \sum_{v \subset u} g_v(x_v)$. Under input independence, we can expand $\mathbb{V}[Y]$ via the well-known functional ANOVA decomposition [14, 55, 43, 58]

$$\mathbb{V}[Y] = \sum_{u \in 2^{[d]}, u \neq \emptyset} \mathbb{V}[g_u(X_u)] \tag{5}$$

with $\mathbb{V}[g_u(X_u)]$ the variance of $g_u(X_u)$ in (4). In [25, 54], the total effect of input X_j is defined as the sum of all terms in the right hand side of (5) that contain index j and it is shown that

$$\tau_u = \mathbb{E}[\mathbb{V}[Y|X_{-u}]] = \mathbb{V}[Y] - \mathbb{V}[\mathbb{E}[Y|X_{-u}]] = \sum_{v \in 2^{[d]}, v \cap u \neq \emptyset} \mathbb{V}[g_v(X_v)]. \tag{6}$$

However, this identity does not hold if features are statistically dependent. Under dependence, τ_u remains defined as in (1) and enjoys an interpretation in terms of the L^2 approximation error, as established in [22]. In an argument similar to [51], [22] consider that the space L^2 can be decomposed into a direct sum $L^2(\mathcal{X}) = M_{-u} \oplus M_{-u}^{\perp}$ where M_{-u} contains all L^2 functions which solely depend on x_{-u} and $M_{-u}^{\perp}$ is its orthogonal complement. In general, for dependent features, $M_{-u}^{\perp} \neq M_u$. Then we can write $g(x) = g_0 + g_{-u}(x_{-u}) + g_{-u}^{\perp}(x)$ with $g_{-u} \in M_{-u}$ and $g_{-u}^{\perp} \in M_{-u}^{\perp}$. If we ask the question of how accurately $g(x) - g_0$ can be approximated without the features in x_u, then the answer is

$$\|g - g_0 - g_{-u}\|_{L^2}^2 = \left\|g_{-u}^{\perp}\right\|_{L^2}^2 = \|g - g_0\|_{L^2}^2 - \|g_{-u}\|_{L^2}^2\,. \tag{7}$$

If we consider the L^2 norm weighted with the density of X, then we regain (1) from (7), as then $g_{-u}(x_{-u}) = \mathbb{E}[g(X) - g_0|X_{-u} = x_{-u}]$.
Conditional independence plays a central role in deriving (2): it is this property that enables one to replace $\mathbb{E}[Y|X_u]$ by Y' in (2). We show that it also plays a central role in estimating τ_u under input dependence via winding stairs and pick-and-freeze designs.

2.2 Conditional Independence and Total Effect Estimation

The gold standard for obtaining estimates for total effects *under input independence* is the Sobol' method, i.e., a pick-and-freeze design paired with Jansen's estimator [31]. Let X, X' be d-dimensional input vectors. For $u \subseteq [d]$, we use the notation $X'_u : X_{-u}$ to denote the d-dimensional vector whose components indexed by u are taken from X' and whose components indexed by $-u$ are taken from X. Now, let X'_u be a replicate of X_u, independent of X_u conditionally on X_{-u}. Then, $(X'_u : X_{-u})$ and X are identically distributed and $Y = g(X)$ and $Y' = g(X'_u : X_{-u})$ are identically distributed and conditionally independent given X_{-u}. The second equality in Equation (2) can then be rewritten as

$$\tau_u = \tfrac{1}{2}\,\mathbb{E}\left[\left(g(X) - g(X'_u : X_{-u})\right)^2\right]. \tag{8}$$

By Lemma 1, Equation (8) is true even under feature dependence. However, independence makes the design of an estimator for τ_u straightforward. One generates two independent samples of size n from the input distribution. Let us denote them by $X^A = (X^{A,i})_{i=1,\ldots,n}$ and $X^B = (X^{B,i})_{i=1,\ldots,n}$. The columns of these sample matrix blocks are recombined, copying input realizations for factor(s) $j \in u$ from the second sample (B) into the first sample (A) to form pick-and-freeze input sample blocks $X^B_u : X^A_{-u}$. The model is then evaluated to obtain the output samples $Y^A = g(X^A)$ and $Y^{BA}_u = g(X^B_u : X^A_{-u})$. Combining them via Jansen's equality, we obtain the estimator

$$\widehat{\tau}^{\mathrm{PF}}_u = \frac{1}{2n}\sum_{i=1}^{n}\left(g(X^{A,i}) - g(X^{B,i}_u : X^{A,i}_{-u})\right)^2 = \frac{1}{2n}\left(Y^{A,i} - Y^{BA,i}_u\right)^2. \tag{9}$$

After the introduction of this design in [55], [25], works such as [53, 56, 18, 17, 50] have developed it further refining several aspects. Most of these works rely on the independence assumption, while we remove it in the remainder of this section.
Let Z be a general m-dimensional random vector that takes its values on a Cartesian product space. Again for notation simplicity, we assume that the joint probability distribution of Z has a density function $h(z)$ with respect to Lebesgue measure. For u, v two disjoint subsets of $[m] = \{1, 2, \ldots, m\}$, we denote the disjoint union of u and v by $u + v$.

Definition 2. Let u, v, w be pairwise disjoint index sets in $[m]$. Then Z_u and Z_v are conditionally independent given Z_w ($Z_u \perp\!\!\!\perp Z_v|Z_w$) if for all $z_u \in \mathcal{Z}_u$, $z_v \in \mathcal{Z}_v$, $z_w \in \mathcal{Z}_w$, the density h satisfies

$$h_{u+v|w}(z_{u+v}|z_w) = h_{u|w}(z_u|z_w) \cdot h_{v|w}(z_v|z_w). \tag{10}$$

Multiplying (10) by $h_w(z_w)$ we find

$$h_{u+v+w}(z_{u+v+w}) = h_{u|w}(z_u|z_w) \cdot h_{v|w}(z_v|z_w)h_w(z_w) = \begin{cases} h_{v|w}(z_v|z_w) \cdot h_{u+w}(z_{u+w}), \\ h_{u|w}(z_u|z_w) \cdot h_{v+w}(z_{v+w}). \end{cases} \tag{11}$$

Thus, under conditional independence the joint density factors into the product of two terms that one can symmetrically write as per the second identity in (11). As a direct consequence for the Jansen's estimator, when considering the joint distribution of X_u, X'_u and X_{-u}, we obtain the following result.

Proposition 3. *Let X' be a replicate of X, conditionally independent given X_{-u}. Letting $Y = g(X_u : X_{-u})$ and $Y' = g(X'_u : X_{-u})$, then Y is a replicate of Y' conditionally independent given X_{-u}. Written in density terms, we find two interchangeable representations of the total effect,*

$$\begin{aligned} \tau_u &= \frac{1}{2} \int_{\mathbb{R}^{d+|u|}} \left(g(x'_u : x_{-u}) - g(x_u : x_{-u})\right)^2 f_{u|-u}(x_u|x_{-u}) f(x'_u : x_{-u}) dx'_u dx_u dx_{-u} \\ &= \frac{1}{2} \int_{\mathbb{R}^{d+|u|}} \left(g(x_u : x_{-u}) - g(x'_u : x_{-u})\right)^2 f_{u|-u}(x'_u|x_{-u}) f(x_u : x_{-u}) dx'_u dx_u dx_{-u}. \end{aligned} \tag{12}$$

The second term in (12) is the numerator of Equation (2.11) in [35, p. 939]. Equation (12) is an essential ingredient for the estimation of total Sobol' indices under feature dependence. In the next section, we exploit Equation (12) to create a generalized design. In Section 3, we use it for the definition of total effect estimators in the presence of constrained (i.e., non-Cartesian) input domains.

2.3 Winding stairs for Dependent Inputs with Gaussian Copula

We propose a new estimation strategy that combines Proposition 3 with the Knothe-Rosenblatt transformation [33, 52]. This transformation is discussed in association with the estimation of variance-based sensitivity indices with dependent features in the works of [38, 39, 36]. The intuition is to move from the dependent features X to a set of independent features U uniformly distributed in the unit hypercube. Then, the U features are independent and one can apply the theory and algorithms of the functional ANOVA expansion under independence. Formally, the Knothe-Rosenblatt transformation is

$$U_1 = F_1(X_1), \qquad U_2 = F_{2|1}(X_2|X_1), \quad \dots \quad U_d = F_{d|1,\dots d-1}(X_d|X_1, \dots X_{d-1}). \tag{13}$$

Hence after applying the transformation, one can calculate global sensitivity indices on the independent coordinates in U. However, after transformation the physical meaning of the original features may be lost and it might be difficult to transfer the results back to the original scale. Furthermore, the ranking is dependent on the order with which the features enter the transformation.
We propose an intuition to use the Knothe-Rosenblatt transformation for the calculation of total indices that avoids the rank dependence on the feature ordering and allows us to remain within the original feature space. The key is to combine these two facts. The

first is that, by Proposition 3, the total effects of individual features, τ_j, are associated with the conditional density $f_{j|-j}$. The second is that this coincides with the density of the last term in (13). Then, if this last term is available from the transformation, we can simply draw realizations of an independent standard uniform random variable and apply the inverse transformation $t_j : u \mapsto x_j = F^{-1}_{j|-j}(u|X_{-j} = x_{-j})$. This transformation is, indeed, the inverse of the last term in (13) (up to a re-oredering of input variables): we have an X_j which is conditional on all the remaining features.

We exploit this fact to introduce a generalized winding stairs total effect estimator for the case of dependent features. The term winding stairs originates with [30]. We refer to [6], [45] and [21] for further reviews. Assume that $X^{(0)}$ is a random copy of X, and U is a random vector of d independent standard uniformly distributed random variables, independent of $X^{(0)}$. In the classical winding stairs design, under independence, the j^{th} column in the feature sample matrix is replaced by an independent copy of X_j. Under dependence, using Lemma 1, we can cyclically replace the jth entry in the input vector with a conditionally independent one. An appropriate way to obtain this conditionally independent sample is a Knothe-Rosenblatt transformation of the form: In the j^{th} step, the j^{th} component of $X^{(0)}$ is altered via

$$X_\ell^{(j)} = \begin{cases} t_j(U_j|X_{-j}^{(j-1)}) = F^{-1}_{j|-j}\left(U_j \middle| X_{-j}^{(j-1)}\right), & \text{for } \ell = j, \\ X_\ell^{(j-1)}, & \text{otherwise.} \end{cases} \tag{14}$$

for $\ell, j = 1, 2, \ldots, d$. When sampling, we obtain blocks of the type $X^{(j)} = (X_j^{(j)} : X_{-j}^{(j-1)})$ for $j = 1, \ldots, d$. By construction, $Y^{(j)} = g(X^{(j)})$ is a replicate of $Y^{(j-1)}$ conditionally independent given X_{-j}. From Lemma 1, one obtains a winding stairs total effect formula

$$\tau_j^{\text{WS}} = \frac{1}{2}\,\mathbb{E}\left[\left(Y^{(j)} - Y^{(j-1)}\right)^2\right], \qquad j \in [d]. \tag{15}$$

The associated estimator is a variant of (9). As observed in [21], such an estimator is a sample average, so that the sample variance can be used to approximate the empirical variance.

If we model input dependence via Gaussian copulas, the transformations are linear in standard normal coordinates. For this, let Ψ_j, $j = 1, \ldots, d$ be transformations from the marginal into the standard normal distribution and let Z_j be a standard normal random variable independent of X. Then there exist linear combinations such that the random vectors $[\Psi_1(X_1) \ldots \Psi_d(X_d)]$ and

$$\left[\Psi_1(X_1) \quad \ldots \quad \gamma_j^{(j)} Z_j + \textstyle\sum_{\ell \neq i} \gamma_\ell^{(j)} \Psi_\ell(X_\ell) \quad \ldots \quad \Psi_d(X_d)\right] \tag{16}$$

are identically $\mathcal{N}(0, \Sigma)$ distributed and conditionally independent given X_{-j}. These linear combinations can be extracted from a Cholesky decomposition of a reordered covariance matrix where the j^{th} row/column is moved to the last position (we refer to the proof of Theorem 14 for the computation of the coefficients $\gamma_\ell^{(j)}$ in the linear

combination in (16)). Hence (14) specializes to

$$X_j^{(j)} = \Psi_j^{-1}\left(\gamma_j^{(j)} Z_j + \sum_{\ell \neq j} \gamma_\ell^{(j)} \Psi_\ell\left(X_\ell^{(j-1)}\right)\right) = t_j(\Phi^{-1}(Z_j)|X_{-j}^{(j-1)}), \tag{17}$$

where Φ is the standard normal cumulative distribution function. From a numerical viewpoint, the computational cost associated with the winding stairs approach is $n(d+1)$ model evaluations. This cost is explained as follows: one samples n random values of X and then considers one-at-a-time variations in each input for each of the n values.
This design generalizes a winding stairs approach to dependence structures that include the broad family of Gaussian copulas. However, for cases in which the Knothe-Rosenblatt transformation is not available, the generalized winding stairs design becomes impractical. This happens as soon as features do not leave on a support which is Cartesian. We then introduce alternative approaches that allow to generalize Lemma 1 to more complex dependence structures in the next sections.

3 Total Effects under Feature Dependence via Reweighting

Our purpose in this section is to introduce an estimation strategy that allows us to relax the traditional condition of a Cartesian domain, that is, we allow for $\mathcal{X} \neq \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_d$. We assume that the features are distributed with a joint density f such that $f(x) > 0$ if $x \in \mathcal{X}$ and $f(x) = 0$ if $x \notin \mathcal{X}$. In order to use an estimation strategy with a classical pick-and-freeze design, we start with the following definition.

Definition 4. Let X' be an independent copy of X. We call the function

$$\iota_u(X', X) = \frac{f(X'_u : X_{-u})}{f_u(X'_u) f_{-u}(X_{-u})} \tag{18}$$

the density quotient of X on $\mathcal{X}$ for the feature list u.

By Proposition 3, the density quotient in (18) satisfies

$$\iota_u(X', X) = \frac{f_{u|-u}(X'_u|X_{-u})}{f_u(X'_u)} = \frac{f_{-u|u}(X_{-u}|X'_u)}{f_{-u}(X_{-u})} = \iota_{-u}(X, X').$$

To illustrate, for a Cartesian domain and independent inputs, we have $\iota_u(X', X) = 1$. Also, we have a compact expression for the case in which the dependence among two features can be expressed via a Gaussian copula.

Example 5. Under a bivariate Gaussian copula, the density quotient for pairwise dependence can be obtained in a compact form as follows. Let X_i and X_j be two random variables with a rank-correlation of ϱ. Setting $u_i = F_i(x_i)$ and $u_j = F_j(x_j)$, [32, Section 4.3.1] derives the density of the bivariate Gaussian copula as

$$\iota_i(u_i, u_j; \varrho) = \frac{\phi(\Phi^{-1}(u_i), \Phi^{-1}(u_j); \varrho)}{\phi(\Phi^{-1}(u_i))\phi(\Phi^{-1}(u_j))} = \frac{1}{\sqrt{1-\varrho^2}} \exp\left(\frac{-\varrho}{2(1-\varrho^2)}\left(\varrho(z_i^2 + z_j^2) - 2z_i z_j\right)\right), \tag{19}$$

where one uses the transformation $z_i = \Phi^{-1}(u_i) = \Phi^{-1}(F_i(x_i))$. The right hand side in (19) is the density quotient for a bivariate Gaussian copula.

Proposition 6. *Let X and X' be i.i.d. random vectors. The following equality holds:*

$$\tau_u = \mathbb{E}\left[\iota_u(X', X)\left(g(X_u : X_{-u}) - g(X'_u : X_{-u})\right)^2\right]. \tag{20}$$

Given n independent copies X^i of X, $i = 1, 2, \ldots, n$, then an unbiased estimator of τ_u is

$$\widehat{\tau}^U_{u,n} = \frac{1}{2n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \frac{f(X^j_u : X^i_{-u})}{f_u(X^j_u) f_{-u}(X^i_{-u})} \left(g(X^i_u : X^i_{-u}) - g(X^j_u : X^i_{-u})\right)^2. \tag{21}$$

Proof. Consider an independent copy X' of X. Projections onto index subsets u and $-u$ keep independence intact, i.e., $X'_u = P_u(X')$ and $X_{-u} = P_{-u}(X)$ are independent. The random vector obtained by glueing these two vectors together therefore has a density $f_u \cdot f_{-u}$, breaking the inter-block dependence. Hence generally for a measurable function $h : \mathbb{R}^d \to \mathbb{R}$, we have

$$\mathbb{E}\left[h(X'_u : X_{-u})\right] = \iint h(x'_u : x_{-u}) f_{-u}(x_{-u}) f_u(x'_u) dx_{-u} dx'_u.$$

Now, in order to compare the expectation of $h(X)$ with X possibly being dependent, we split the argument X into two arguments via projections onto subdimensions indexed by u and $-u$, so that we may write the joint density as product of marginal and conditional density,

$$\mathbb{E}\left[h(X_u : X_{-u})\right] = \iint h(x_u : x_{-u}) f_{u|-u}(x_u|x_{-u}) f_{-u}(x_{-u}) dx_u dx_{-u}.$$

Considering all three terms in a function $h_2 : \mathbb{R}^{|u|} \times \mathbb{R}^{d-|u|} \times \mathbb{R}^{|u|} \to \mathbb{R}$ and taking its expectation then leads to

$$\mathbb{E}\left[h_2(X_u, X_{-u}, X'_u)\right] = \iiint h_2(x_u, x_{-u}, x'_u) f_{u|-u}(x_u|x_{-u}) f_{-u}(x_{-u}) f_u(x'_u) dx_u dx_{-u} dx'_u.$$

Let us now consider the weighted Jansen's estimator,

$$h_2(X_u, X_{-u}, X'_u) = \frac{1}{2} \frac{f_{u,-u}(X'_u : X_{-u})}{f_u(X'_u) f_{-u}(X_{-u})} \left(g(X_u : X_{-u}) - g(X'_u : X_{-u})\right)^2.$$

Then by (12),

$$\begin{aligned}
\mathbb{E}[h_2(X_u, X_{-u}, X'_u)] =& \frac{1}{2} \iiint \frac{f_{u,-u}(x'_u : x_{-u})}{f_u(x'_u) f_{-u}(x_{-u})} \left(g(x_u : x_{-u}) - g(x'_u : x_{-u})\right)^2 \cdot \\
& f_{u|-u}(x_u|x_{-u}) f_{-u}(x_{-u}) f_u(x'_u) d(x_u, x_{-u}, x'_u) \\
=& \frac{1}{2} \iiint f_{u|-u}(x'_u|x_{-u}) \left(g(x_u : x_{-u}) - g(x'_u : x_{-u})\right)^2 \cdot \\
& f_{u|-u}(x_u|x_{-u}) f_{-u}(x_{-u}) d(x_u, x_{-u}, x'_u) = \tau_u.
\end{aligned} \tag{22}$$

The last equality follows from (12). By definition, a sample consists of realizing n independent copies of X. Hence two different copies of X, X^i and X^j, $i, j = 1, \ldots, n$, $i \neq j$, are independent. Then, an estimator of τ_u is

$$\widehat{\tau}^U_{u,n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} h_2(X^i_u, X^i_{-u}, X^j_u), \tag{23}$$

which yields (21). □

Equation (21) combines a brute-force double-loop sample design together with a Jansen's estimator modified by a weight from importance sampling. First, it allows for a pick-and-freeze sample design. However, a pick-and-freeze sample only provides us with a product of marginal distributions. The density quotient in Proposition 6 introduces a correction factor that allows one to switch from the product of the marginals back to the (correct) joint distribution. Moreover, the density quotient allows us to consider non-Cartesian input domains, as it vanishes for points outside the region where the inputs are defined. If features are independent we regain Jansen's classical estimator, because the density quotient is identically equal to one.

Lemma 7. *The mix-and-reweight estimator* (21) *of Proposition 6 is a U-statistic of order* 2.

Proof. Let us define the random vector $W = (W_1, W_2) = (X_u, X_{-u})$ (the random vector X split according to the index set u). Let W^i, $i = 1, \ldots, n$, be identical copies of W. We then write the estimator as follows:

$$\widehat{\tau}^U_{u,n} = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} \Phi(W^i, W^j) = \binom{n}{2}^{-1} \sum_{i=1}^{n} \sum_{j>i} \phi^s(W^i, W^j) \tag{24}$$

with

$$\Phi^s(W^i, W^j) = \tfrac{1}{2} \left(\Phi(W^i, W^j) + \Phi(W^j, W^i) \right), \tag{25}$$

$$\Phi(W^i, W^j) = \tfrac{1}{2} i_u(W^i_1, W^j_2) \left(g\left(W^j_1 : W^j_2\right) - g\left(W^i_1 : W^j_2\right) \right)^2. \tag{26}$$

Hence, $\widehat{\tau}^U_{u,n}$ defines a U-statistic of order 2 for $\tau_u = \mathbb{E}[\Phi(W^i, W^j)] = \mathbb{E}[\widehat{\tau}^U_{u,n}]$. □

We call the design associated with Proposition 6 mix-and-reweight approach. It is associated with a computational cost of $dn^2 - nd + n$ evaluations (Table 1, third row).
From the general theory of U-statistics, and drawing from the classical findings of [24] it is possible to obtain the variance and derive a central limit theorem for the mix-and-reweight estimator.

Lemma 8. *Let* $\delta_1 = \mathbb{V}[\mathbb{E}[\Phi^s(W_1, W_2)|W_1]]$ *and* $\delta_2 = \mathbb{V}[\Phi^s(W_1, W_2)]$ *with* Φ^s *defined by* (25) *and* (26). *Assume that* $\mathbb{V}[\delta_2] < +\infty$. *Then*

$$\mathbb{V}[\widehat{\tau}^U_{u,n}] = \frac{2}{n(n-1)} \left(2(n-2)\delta_1 + \delta_2\right) = \frac{4}{n}\delta_1 + O(n^{-2}). \tag{27}$$

If $\delta_1 \neq 0$ *then the U statistic is non-degenerate and* $\sqrt{n}(\widehat{\tau}^U_{u,n} - \tau_u) \to \mathcal{N}(0, 4\,\delta_1)$.

Proof. The proof of (27) follows directly from (24) using the computations in [3, Section 1.2.1]. The asymptotic normality is a consequence of [3, Theorem 1.1]. □

The asymptotic variance in Lemma 8 can be estimated with the plug-in Jackknife estimator of [23] (see also [3, p. 106]), defined as

$$\widehat{\mathbb{V}}[\widehat{\tau}^U_{u,n}] = \frac{4}{n}\frac{n-1}{(n-2)^2}\sum\nolimits_{i=1}^{n}\left(\frac{1}{n-1}\sum\nolimits_{j\neq i}\Phi^s(X^i_u, X^j_{-u}) - \widehat{\tau}^U_{u,n}\right)^2.$$

As an alternative, a bootstrap distribution of the estimator $\widehat{\tau}^U_{u,n}$ can be derived from the sample of $\Phi^s(X^i_u, X^j_{-u})$.

We find earlier accounts on the use of reweighting techniques in sensitivity analysis. Let us mention the estimation of first-order and total Sobol' indices in [57] in which the already available sample of simulations is reweighted with a sample weighting scheme based on importance sampling, but with the roles of the target distribution and the sampling distribution reversed. Also in [1, Section 5.4], the authors discuss reweighting and rejection techniques to measure the potential impact of small changes in the input probability distribution on the output mean. In [34], a rejection technique to handle non-Cartesian input domains is implemented.

4 Derangement and Shift Estimators

The mix-and-reweight estimator of Lemma 8 possesses the clear theoretical advantages associated with the notion of U-statistics. However, the associated estimation cost may turn into a notable disadvantage in practical applications. To reduce the cost for estimating τ_u under input constraints, we propose two new estimators based on derangements and shifts. The intuition here is to compare a given realization to the next one (or via a random pick) instead of comparing it against all other realizations. This makes the costs drop from being quadratic in the sample size to being linear.

Having fixed a sample block of size n, we introduce the cyclic shift-by-one of $\{1,\dots,n\}$ defined by $s_n(i) = i+1$ for $i < n$, and $s_n(n) = 1$. We also define the acyclic shift-by-one by $s^a_{n-1}(i) = i+1$ for $i < n-1$, and $s^a_{n-1}(n-1) = n$. Here, $s^a_{n-1}(\cdot)$ is a fixpoint-free map from $\{1,\dots,n-1\}$ to $\{2,\dots,n\}$. We have the following result.

Theorem 9. *Let X be a sample of size n subject to the joint density f. Then the shift-and-reweight total effect estimator for factor j defined as*

$$\widehat{\tau}^S_{j,n} = \frac{1}{2n}\sum_{i=1}^{n}\iota_j(X^{s_n(i)}, X^i)\left(g(X^i_j : X^i_{-j}) - g(X^{s_n(i)}_j : X^i_{-j})\right)^2 \tag{28}$$

is unbiased. Assume that $\sigma^2_j = \mathbb{E}[V^1_j V^1_j] + 2\,\mathbb{E}[V^1_j V^2_j] < +\infty$ with

$$V^i_j = \frac{1}{2}\iota_j(X^{s_n(i)}, X^i)\left(g(X^i_j : X^i_{-j}) - g(X^{i+1}_j : X^i_{-j})\right)^2.$$

Then $\widehat{\tau}^S_{j,n}$ defined by (28) satisfies the following central limit theorem:

$$\sqrt{n}\left(\widehat{\tau}^S_{j,n} - \tau_j\right) \xrightarrow[n\to+\infty]{} \mathcal{N}(0, \sigma^2_j). \tag{29}$$

Proof. First, the ith realization and the $s_n(i)$th one in the input sample are independent. We thus deduce from (22) that for any $1 \leq i \leq n$,

$$\mathbb{E}\left[\frac{1}{2}\iota_j(X^{s_n(i)}, X^i)\left(g(X_j^i : X_{-j}^i) - g(X_j^{s_n(i)} : X_{-j}^i)\right)^2\right] = \tau_j$$

thus the estimator $\hat{\tau}_{j,n}$ is unbiased. To prove the central limit theorem, we first decompose $\widehat{\tau}_{j,n}^S$ as:

$$\widehat{\tau}_{j,n}^S = \frac{n-1}{n}\tilde{\tau}_{j,n-1} + \frac{1}{2n}\frac{f_{j,-j}(X_j^1 : X_{-j}^n)}{f_j(X_j^1)f_{-j}(X_{-j}^n)}\left(g(X_j^n : X_{-j}^n) - g(X_j^1 : X_{-j}^n)\right)^2$$

where $\tilde{\tau}_{j,n-1}$ stands for the estimator built from Formula (28) on $(X_1, \ldots, X_{n-1})$ and by replacing the cyclic shift-by-one s_n by the acyclic one s_{n-1}^a. Then, $\tilde{\tau}_{j,n-1} = \frac{1}{n-1}\sum_{i=1}^{n-1} V_j^i$ and the sequence $\left(V_j^i\right)_{i\geq n}$ is stationary and 1-dependent. Thus, the limit $\sqrt{n}(\tilde{\tau}_{j,n-1} - \tau_j) \to \mathcal{N}(0, \sigma_j^2)$ follows from [13, Theorem 5]. Finally, noting that

$$\mathbb{E}\left[\left|\sqrt{n}\frac{1}{2n}\frac{f_{j,-j}(X_j^1 : X_{-j}^n)}{f_j(X_j^1)f_{-j}(X_{-j}^n)}\left(g(X_j^n : X_{-j}^n) - g(X_j^1 : X_{-j}^n)\right)^2\right|\right] = \frac{\tau_j}{\sqrt{n}} \xrightarrow[n\to+\infty]{} 0$$

and applying Slutsky's Theorem we get (29). □

Corollary 10. *Let X be a sample of size n subject to the joint density f. Define the shift-and-reweight normalized total effect estimator for factor j by $\widehat{T}_j^S = S_{Y,n}^{-2}\widehat{\tau}_{j,n}^S$ with $S_{Y,n}^2$ the empirical variance associated to n-sample $Y = g(X)$. Assume $\mathbb{V}[Y] < +\infty$ and $\sigma_j^2 < +\infty$, with σ_j^2 defined in Theorem 9. Then we have:*

$$\sqrt{n}\left(\widehat{T}_{j,n}^S - T_j\right) \xrightarrow[n\to+\infty]{} \mathcal{N}(0, \sigma_{\text{norm},j}^2)$$

where $\sigma_{\text{norm},j}^2 = \rho_j^T\Sigma_j\rho_j$ with $\rho_j = \frac{1}{\mathbb{V}[Y]}[1, 2T_j\,\mathbb{E}[Y], -T_j]^T$ and

$$\Sigma_j = \begin{pmatrix} \mathbb{V}[V_j^1] & \text{cov}(V_j^1, Y^1) & \text{cov}(V_j^1, (Y^1)^2) \\ \text{cov}(V_j^1, Y^1) & \mathbb{V}(Y^1) & \text{cov}(Y^1, (Y^1)^2) \\ \text{cov}(V_j^1, (Y^1)^2) & \text{cov}(Y^1, (Y^1)^2) & \mathbb{V}[(Y^1)^2] \end{pmatrix}.$$

Proof. The result follows from Theorem 9 and by applying the Delta method (see, e.g., [59]). First, by mimicking the proof of Theorem 9, it is possible to prove that for any α, β, γ, a central limit theorem holds true for $\alpha\widehat{\tau}_{j,n}^S + \beta\frac{1}{n}\sum_{i=1}^n g(X^i) + \gamma\frac{1}{n}\sum_{i=1}^n g^2(X^i)$. It yields a central limit theorem for $(U_{1,n}, U_{2,n}, U_{3,n}) = \left(\widehat{\tau}_{j,n}^S, \frac{1}{n}\sum_{i=1}^n g(X^i), \frac{1}{n}\sum_{i=1}^n g^2(X^i)\right)$, namely

$$\sqrt{n}\left((U_{1,n}, U_{2,n}, U_{3,n})^T - \theta_j\right) \xrightarrow[n\to+\infty]{} \mathcal{N}(0, \Sigma_j)$$

with $\theta_j = (\tau_j, \mathbb{E}[Y], \mathbb{E}[Y^2])$ and

$$\Sigma_j = \begin{pmatrix} \mathbb{V}[V_j^1] & \text{cov}(V_j^1, Y^1) & \text{cov}(V_j^1, (Y^1)^2) \\ \text{cov}(V_j^1, Y^1) & \mathbb{V}(Y^1) & \text{cov}(Y^1, (Y^1)^2) \\ \text{cov}(V_j^1, (Y^1)^2) & \text{cov}(Y^1, (Y^1)^2) & \mathbb{V}[(Y^1)^2] \end{pmatrix}.$$

Then we can prove the central limit theorem for T_j, using the Delta method on $\psi(x,y,z) = \frac{x}{z-y^2}$ and $\theta_j = (\tau_j, \mathbb{E}[Y], \mathbb{E}[Y^2])$. More precisely, let ρ_j denote the gradient of ψ at θ_j. We have $\rho_j = \nabla\psi(\theta_j) = \frac{1}{\mathbb{V}[Y]}\left[1, 2T_j\,\mathbb{E}[Y], -T_j\right]^T$. Thus

$$\sqrt{n}\left(\hat{T}^S_{j,n} - T_j\right) \xrightarrow[n\to+\infty]{} \mathcal{N}(0, \rho_j^T \Sigma_j \rho_j)\cdot$$

It concludes the proof of Corollary 10. □

It is possible to generalize Theorem 9 by dealing with more general permutations than the cyclic shift-by-one, as stated in Theorem 11 below.

Theorem 11. *Let X be a sample of size n subject to the joint density f. Let $(\pi_n)_{n\geq 1}$ be a sequence of derangements (fixpoint-free permutations) of $\{1,\ldots,n\}$. Then the derange-and-reweight total effect estimator for factor j defined as*

$$\widehat{\tau}^D_{j,n} = \frac{1}{2n}\sum_{i=1}^{n} \iota_j(X^{\pi_n(i)}, X^i)\left(g(X^i_j : X^i_{-j}) - g(X^{\pi_n(i)}_j : X^i_{-j})\right)^2 \tag{30}$$

is unbiased. Suppose that there exists $\delta > 0$ such that $\mathbb{E}\left[|V^1_j|^{2+\delta}\right] < +\infty$ with V^1_j defined as in Theorem 9 and $\lim_{n\to+\infty} m_n^{1+\delta} n^{-\delta/2} \to 0$, with m_n the number of cycles of π_n. Then we have the following central limit theorem:

$$\sqrt{n}\left(\widehat{\tau}^D_{j,n} - \tau_j\right) \xrightarrow[n\to+\infty]{} \mathcal{N}(0, \sigma_j^2) \tag{31}$$

where $\sigma_j^2 = \mathbb{V}[V^1_j] + 2\,\mathrm{cov}(V^1_j, V^2_j)$ with

$$V^i_j = \frac{1}{2}\frac{f_{j,-j}(X^{i+1}_j : X^i_{-j})}{f_j(X^{i+1}_j) f_{-j}(X^i_{-j})}\left(g(X^i_j : X^i_{-j}) - g(X^{i+1}_j : X^i_{-j})\right)^2 \cdot$$

The assumption $\lim_{n\to+\infty} m_n^{1+\delta} n^{-\delta/2} \to 0$ does not seem too technical. Indeed, a permutation π_n of $\{1,\ldots,n\}$ decomposes into cycles and a classical result in combinatorics lets us expect $1+\frac{1}{2}+\cdots+\frac{1}{n}$ cycles per permutation, and this harmonic series is approximately $\log(n)$.

Proof. To prove that $\widehat{\tau}^D_{j,n}$ is unbiased, we use the same arguments as the ones used to prove that $\widehat{\tau}^S_{j,n}$ defined by (28) in Theorem 9 is unbiased, additionally noting that $\pi_n(i) \neq i$ for all $i = 1,\ldots,n$. To prove the central limit theorem, we first decompose, for each n, the permutation π_n in cycles $C_{1,n},\ldots,C_{m_n,n}$. Let us arbitrarily fix the first element in each cycle. We then form p_n blocks, with $p_n = \max_{1\leq k\leq m_n} \ell_{k,n}$ and $\ell_{k,n}$ the length of cycle $C_{k,n}$. For each n, we re-order the X^is so that the first $b_{1,n} = m_n$ re-ordered variables are the first element of each cycle, the next $b_{2,n}$ re-ordered variables are the second element of each cycle with length at least two and so on until the $n - \sum_{v=1}^{p_n-1} b_{v,n}$ last $b_{p_n,n}$ re-ordered variables which are the last element in each cycle of length p_n. Here, $1 \leq b_{p_n,n} \leq \ldots \leq b_{1,n} = m_n$. For each n, we denote the re-ordered sequence of

X^is by $X^{i,n}$, $1 \leq i \leq n$, $n \geq 1$. We then define $S_{v,n} = \sum_{i=k_{v-1,n}+1}^{k_{v,n}} \tilde{V}_j^{i,n}$, with $k_0 = 0$, $k_{v,n} = \sum_{w=1}^{v} b_{w,n}$ and $\tilde{V}_j^{i,n}$ defined as $V_j^i - \tau_j/\sqrt{n}$ but with the $X^{i,n}$s in place of the X^is. More precisely, we have to use the trick in the proof of Theorem 9 by replacing first the last variable X^i in cycle $C_{1,n}$ by X^{n+1}, ... , the last variable X^i in cycle $C_{m_n,n}$ by X^{n+m_n}. As $\lim_{n\to+\infty} m_n/\sqrt{n} = 0$, we prove that the remaining term decreases fast enough not to perturb the result of the central limit theorem (see the proof of Theorem 9 for more details). Now, as $\lim_{n\to+\infty} m_n/\sqrt{n} = 0$, then $p_n \geq n/m_n \geq \sqrt{n}/m_n \to +\infty$. For each $n \geq 1$, the sequence $(S_{v,n})_{1\leq v\leq p_n}$ is a sequence of 1-dependent variables. Then applying [42, Theorem 2.1], we get the central limit theorem, as soon as there exists $\delta > 0$ such that $\mathbb{E}[|V_j^1|^{2+\delta}] < +\infty$ for some $\delta > 0$. Indeed, as variables inside each block are centered, independent and identically distributed,

$$\sum_{v=1}^{p_n} \mathbb{E}\left[S_{v,n}^2\right] = \sum_{v=1}^{p_n} \frac{b_{v,n}}{n} \mathbb{V}\left[V_j^1\right] = \mathbb{V}\left[V_j^1\right] < +\infty.$$

Then, Assumption (2.1) in [42, Theorem 2.1] is true due to the stationarity of $V_j^i V_j^{i+1}$. Indeed, due to 1-dependence,

$$\begin{aligned}
\mathbb{V}\left[\sum_{v=1}^{p_n} S_{v,n}\right] &= \sum_{v=1}^{p_n} \mathbb{V}\left[S_{v,n}\right] + 2 \sum_{1\leq v<w\leq p_n} \operatorname{cov}\left(S_{v,n}, S_{w,n}\right) = \sum_{v=1}^{p_n} b_{v,n} \frac{\mathbb{V}[V_j^1]}{n} \\
&\quad + 2 \sum_{v=1}^{p_n-1} \operatorname{cov}\left(S_{v,n}, S_{v+1,n}\right) = \frac{\mathbb{V}[V_j^1]}{n} \sum_{v=1}^{p_n} b_{v,n} + \frac{2}{n} \sum_{v=1}^{p_n-1} b_{v+1,n} \operatorname{cov}\left(V_j^1, V_j^2\right) \\
&= \mathbb{V}[V_j^1] + 2\frac{n-m_n}{n} \operatorname{cov}\left(V_j^1, V_j^2\right) \xrightarrow[n\to+\infty]{} \sigma_j^2 < +\infty.
\end{aligned}$$

Let us now prove that Assumption (2.2) of [42, Theorem 2.1] holds. For any $\varepsilon > 0$ we

have:

$$
\begin{aligned}
\sum_{v=1}^{p_n} \mathbb{E}\left[S_{v,n}^2 \mathbb{I}_{|S_{v,n}|>\varepsilon}\right] &\leq \sum_{v=1}^{p_n} \mathbb{E}\left[|S_{v,n}|^{2+\delta}\right]^{\frac{2}{2+\delta}} \left(\mathrm{pr}\left(|S_{v,n}|>\varepsilon\right)\right)^{\frac{\delta}{2+\delta}} \text{ (using Hölder Inequality)} \\
&\leq \varepsilon^{-\delta} \sum_{v=1}^{p_n} \mathbb{E}\left[|S_{v,n}|^{2+\delta}\right] \text{ (using Markov Inequality)} \\
&\leq \varepsilon^{-\delta} \sum_{v=1}^{p_n} \mathbb{E}\left[\Big(\sum_{i=k_{v-1,n}+1}^{k_{v,n}} n^{-1/2}|V_j^i-\tau_j|\Big)^{2+\delta}\right] \\
&= \varepsilon^{-\delta} \sum_{v=1}^{p_n} \frac{b_{v,n}^{2+\delta}}{n^{1+\frac{\delta}{2}}} \mathbb{E}\left[\Big(\sum_{i=k_{v-1,n}+1}^{k_{v,n}} b_{v,n}^{-1}|V_j^i-\tau_j|\Big)^{2+\delta}\right] \\
&\leq \varepsilon^{-\delta} \sum_{v=1}^{p_n} \frac{b_{v,n}^{2+\delta}}{n^{1+\frac{\delta}{2}}} \mathbb{E}\left[\sum_{i=k_{v-1,n}+1}^{k_{v,n}} b_{v,n}^{-1}|V_j^i-\tau_j|^{2+\delta}\right] \text{ (using Jensen Inequality)} \\
&\leq \frac{1}{\varepsilon^{\delta}} \frac{1}{n^{1+\frac{\delta}{2}}} \sum_{v=1}^{p_n} b_{v,n}^{1+\delta} \sum_{i=k_{v-1,n}+1}^{k_{v,n}} \mathbb{E}\left[|V_j^i-\tau_j|^{2+\delta}\right] \\
&\leq \frac{1}{\varepsilon^{\delta}} \frac{1}{n^{1+\frac{\delta}{2}}} m_n^{1+\delta} \sum_{v=1}^{p_n} b_{v,n}\, \mathbb{E}\left[|V_j^1-\tau_j|^{2+\delta}\right] \text{ (by stationarity of } \left(|V_j^i-\tau_j|\right)_{i\geq 1}) \\
&\leq \frac{1}{\varepsilon^{\delta}} \frac{m_n^{1+\delta}}{n^{\frac{\delta}{2}}} \mathbb{E}\left[|V_j^1-\tau_j|^{2+\delta}\right] \xrightarrow[n\to+\infty]{} 0
\end{aligned}
$$

as $\mathbb{E}\left[|V_j^1|^{2+\delta}\right] < +\infty$ and $\lim_{n\to+\infty} \frac{m_n^{1+\delta}}{\sqrt{n}^{\delta}} \to 0$. This concludes the proof of Theorem 11. □

Remark 12. It is possible to compute confidence intervals by block-bootstrapping. It is important to implement bootstrapping with blocks, in order to preserve the 1-dependence structure.

5 Nearest-Neighbour Estimation

Thus far, we have introduced and discussed designs for obtaining total effects that generalize Jansen's intuition. They are listed in rows 2 to 5 of Table 1. The costs of these estimators are an upper bound: Whenever the density quotient vanishes, the model output does not contribute to the estimation, so there is no need to evaluate the model at coordinates which have a density quotient of zero and we save computational time. However, these designs require a simulator in the loop as the X' points need to be reevaluated through the model. Also, they depend on the number of features (d) and thus can be exposed to the curse of dimensionality. However, for the same n a design that does not depend on d is nominally the computationally cheapest. We study a nearest-neighbour approach that achieves such a $d-$independent cost.

The construction is as follows. Consider two observations x and x', and consider the recombined point $(x'_j : x_{-j})$. The first step is the evaluation of the density quotient $\iota_j(x', x)$ at this point. If the density quotient is non-negligible, then select the point which is closest to $(x'_j : x_{-j})$ with respect to some predefined Euclidean metric from the available data. More precisely, the point is the solution of

$$g(x'_j : x_{-j}) \approx g(x_k), \quad k^* = \underset{k \in [n]}{\operatorname{argmin}} \left\{ \left\| x_{k,j} - x'_j \right\|_2^2 + \| x_{k,-j} - x_{i,-j} \|_2^2 \right\}. \tag{32}$$

Then, use the corresponding value of Y as a proxy for $g(x'_j : x_{-j})$ in (21). Otherwise, if the quotient is small, we set the contribution of $\iota_j(x', x) \times (g(x'_j : x_j) - g(x))^2$ equal to zero. This step can be implemented by setting a threshold on the value of the density quotient and considering as negligible all values of the density quotient below the threshold.
All the required information for applying this design is contained in a sample generated by a once-through pass of a Monte Carlo simulation.
The nearest-neighbour approach serves here as a metamodel, predicting model outputs for the mixed input sample. This is a different use of the nearest-neighbour intuition than in [5, 49], where nearest neighbours are used to select a conditional stratum (see [11, 12] for theoretical results). At this stage, also due the presence of the density quotient threshold, we do not furnish any theoretical results for our estimator based on (32). We will then evaluate this strategy based on empirical comparisons in a series of experiments in which we compare the performance of this design with the other estimators in Table 1.

6 The Link Between Total Indices and Breiman's Permutation Importance with Feature Constraints

There is a close link between MDA_j and τ_j visible in [26, Theorem 2] and discussed in great detail in [2]. In this section, we extend the relationship to the case in which inputs are constrained. Consider the problem of training an input-output mapping of the type $h(X, \theta)$, $h : \mathcal{X} \times \mathbb{R}^q \to \mathbb{R}$, where θ is a q-dimensional vector of auxiliary parameters. Let $\mathcal{L}(Y, g(X; \theta))$, $\mathcal{L} : \mathbb{R} \times \mathbb{R}$ a loss function. The training problem can be defined as finding

$$\theta^* = \operatorname{argmin}_\theta \mathbb{E}[\mathcal{L}(Y, g(X; \theta))]. \tag{33}$$

We let $\mathbb{E}[\mathcal{L}(Y, g(X; \theta^*))]$ denote the nominal (and minimal) expected loss function for the machine learning problem. Then, Breiman's importance of feature X_j is defined as

$$\mathrm{MDA}_j = \mathbb{E}[\mathcal{L}(Y, g(X'_j : X_{-j}; \theta^*))] - \mathbb{E}[\mathcal{L}(Y, g(X; \theta^*))], \tag{34}$$

where $\mathbb{E}[\mathcal{L}(Y, g(X'_j : X_{-j}; \theta^*))]$ is the expected loss that we incur if feature X_j is permuted. The intuition is that if the machine learning model relies heavily on X_j for its predictions, then the loss in predictive accuracy should be high and consequently the difference between the nominal loos and the loss after permutation should be significant. After feature permutation we expect a decrease in model prediction accuracy, so that

the expected loss after feature permutation is larger than the nominal expected loss, yielding $\text{MDA}_j \geq 0$.
Of relevance to us are Equations (3.1) and (6.2) of [15]. [15] formulate Breiman's importance in terms of model reliance, as a ratio between the expected loss after permutation over the expected loss before permutation. Using our notation, their definition of model reliance in their Equation (3.1) (p. 8) would read

$$MR_j = \frac{\iota_j(X', X)\,\mathbb{E}[\mathcal{L}(Y, g(X'_j : X_{-j}; \theta)]}{\mathbb{E}[\mathcal{L}(Y, g(X; \theta))]}. \tag{35}$$

Rewritten in MDA terms, (35) yields

$$MDA_j = \iota_j(X', X)\,\mathbb{E}[\mathcal{L}(Y, g(X'_j : X_{-j}; \theta)] - \mathbb{E}[\mathcal{L}(Y, g(X; \theta))]. \tag{36}$$

Proposition 13. *Consider MDA_j in (36). If $\mathcal{L}(Y, g(X))$ is a squared loss function and $g(X; \theta^*)$ is a perfect predictor, then $MDA_j = \tau_j$.*

Proof. Let us write the squared loss function as $\mathcal{L}(Y, g(X; \theta)) = (Y - g(X; \theta))^2$. Then, (36) becomes

$$\text{MDA}_j = \mathbb{E}[\iota_j(X', X)(Y - g(X'_j : X_{-j}; \theta))^2] - \mathbb{E}[(Y - g(X; \theta))^2]. \tag{37}$$

Using the assumption that $g(X; \theta)$ is a perfect predictor, we have $Y = g(X; \theta)$ for all values of X, so that $\mathbb{E}[(Y - g(X; \theta))^2] = \mathbb{E}[0] = 0$. Then, (37) becomes

$$\text{MDA}_j = \mathbb{E}[\iota_j(X', X)(g(X, \theta) - g(X'_j : X_{-j}; \theta))^2] = \tau_j.$$

Hence total effects and MDA with a weighted squared loss are the same. □

7 Analytical Total Effects for Linear Models with Gaussian Features

To provide analytical examples, we derive analytical expressions for the total effects to be used as benchmarks in numerical experiments. This discussion also gives us a formula for the density quotient in the case of Gaussian copulas
Let us consider a linear mapping between Y and X, $Y = \beta^0 + \beta^T X$, $X \in \mathbb{R}^d$, $\beta^0 \in \mathbb{R}$, $\beta \in \mathbb{R}^d$ and let the features be normally distributed with mean μ and variance-covariance matrix Σ. Under this assumption all conditional distributions are Gaussian and all conditional expectations are linear. The pair (X, Y) is then also Gaussian with augmented covariance matrix, $\Sigma' = \begin{pmatrix} \Sigma & \Sigma\beta \\ \beta^T\Sigma & \beta^T\Sigma\beta \end{pmatrix}$.
Let m be a positive integer. Let u, v, w be pairwise disjoint index sets from $[m]$. For any m-dimensional multivariate normal distribution $Z \sim \mathcal{N}(\mu, \Gamma)$, the conditional distribution of Z_{u+v} given $Z_w = z_w$ is

$$\mathcal{N}(\mu_{u+v} + \Gamma_{u+v,w}\Gamma_{w,w}^{-1}(z_w - \mu_w), \Gamma_{u+v,u+v} - \Gamma_{u+v,w}\Gamma_{w,w}^{-1}\Gamma_{w,u+v}). \tag{38}$$

The correlation matrix in (38) is given in form of a Schur complement. It is assumed that the submatrix selected by w is invertible.
For a multivariate Gaussian distribution, the statement $u \perp\!\!\!\perp v|w$ holds if and only if $\Gamma_{u,v} = \Gamma_{u,w}\Gamma_{w,w}^{-1}\Gamma_{w,v}$ holds, as in this case the correlation matrix in (38) is block-diagonal, i.e.

$$\begin{bmatrix}\Gamma_{u,u} & \Gamma_{u,v}\\ \Gamma_{v,u} & \Gamma_{v,v}\end{bmatrix} - \begin{bmatrix}\Gamma_{u,w}\\ \Gamma_{v,w}\end{bmatrix}\Gamma_{w,w}^{-1}\begin{bmatrix}\Gamma_{w,u} & \Gamma_{w,v}\end{bmatrix} = \begin{bmatrix}\Gamma_{u,u} - \Gamma_{u,w}\Gamma_{w,w}^{-1}\Gamma_{w,u} & 0\\ 0 & \Gamma_{v,v} - \Gamma_{v,w}\Gamma_{w,w}^{-1}\Gamma_{w,v}\end{bmatrix}.$$

Theorem 14. *Given $Y = \beta^0 + \beta^T X$, $X \sim \mathcal{N}(\mu, \Sigma)$, $X \in \mathbb{R}^d$, the unnormalized main and total effects are given by*

$$S_j = \beta^T\left(\frac{\Sigma_{[d],j}\Sigma_{j,[d]}}{\Sigma_{j,j}}\right)\beta = \beta^T\left(\frac{\Sigma_{[d],j}\Sigma_{[d],j}^T}{\Sigma_{j,j}}\right)\beta, \tag{39}$$

$$T_j = \beta_j^2\frac{\det(\Sigma)}{\det(\Sigma_{-j,-j})}. \tag{40}$$

The output variance is $\mathbb{V}[Y] = \beta^T\Sigma\beta$.

Alternative computations are offered in [38] for the case $d = 3$. These results can also be retrieved from the proof of [46, Theorem 4.1].

Proof. Main effect: We consider the Gaussian multivariate distribution of (X, Y). Setting $u = \{d+1\}$, $v = \emptyset$, $w = \{j\}$ in (38), then the variance of Y conditionally on X_j is

$$\Sigma_{Y,Y} - \Sigma_{Y,j}\Sigma_{j,j}^{-1}\Sigma_{j,Y} = \beta^T\Sigma\beta - \beta^T\left(\Sigma_{[d],j}\Sigma_{j,j}^{-1}\Sigma_{j,[d]}\right)\beta, \tag{41}$$

using here the special structure of the augmented matrix. This variance is constant (as it does not depend on the value of X_j), so that $\mathbb{E}[\mathbb{V}[Y|X_j]] = \mathbb{V}[Y|X_j]$. However, for main effects we are interested in the variance of the conditional expectation, so we have to subtract the value in (41) from the total variance which yields (39).
Total effect:
The Rosenblatt transform in the Gaussian case uses the Cholesky decomposition of the covariance matrix. The Cholesky matrix $C = \mathrm{chol}(\Sigma)$ is an upper triangular matrix such that $C^TC = \Sigma$. If $Z \sim \mathcal{N}(0, I)$ then $X = \mu + C^TZ \sim \mathcal{N}(\mu, \Sigma)$. One can define the Cholesky decomposition recursively,

$$\mathrm{chol}(\Sigma) = \begin{pmatrix}\Sigma_{1,1}^{1/2} & \Sigma_{1,1}^{-1/2}\cdot\Sigma_{1,-1}\\ 0 & \mathrm{chol}\left(\Sigma_{-1,-1} - \Sigma_{1,-1}^T\Sigma_{1,1}^{-1}\Sigma_{1,-1}\right)\end{pmatrix}$$

where the index -1 denotes all coordinates but the first (if Σ is a scalar then $\mathrm{chol}(\Sigma) = \Sigma^{1/2}$). By reordering the input factors, we may assume without loss of generality that $j = d$. Then define $Y = \beta^0 + \beta^T(\mu + C^TZ)$ and $Y' = \beta^0 + \beta^T(\mu + C^TZ')$ with $Z, Z' \sim \mathcal{N}(0, I)$, differing only in their last coordinate independent from each other. Then Y and Y' are identically (but not independently) distributed, and we know from

Section 2.2 that $\mathrm{cov}(Y, Y') = \frac{1}{2}\,\mathbb{E}[(Y - Y')^2] = \frac{1}{2}\beta_d^2 C_{d,d}^2\,\mathbb{E}[(Z_d - Z_d')^2]$. But Z_d and Z_d' are iid standard normal, i.e., $\frac{1}{2}\,\mathbb{E}[(Z_d - Z_d')^2] = 1$. We are left with the identification of the last diagonal entry of the Cholesky matrix. Because of its hierarchical triangular structure, it keeps subdeterminants intact, and $C_{d,d} = \sqrt{\frac{\det \Sigma}{\det \Sigma_{-d,-d}}}$. □

Theorem 14 provides analytical formulas for main and total effects for linear models with Gaussian features. When we want to estimate total effects under a linear model with Gaussian features with the reweighting approaches introduced above, we require the density quotient. For multivariate Gaussian input distributions, it reads

$$\iota_u(x, x) = \sqrt{\frac{\det(\Sigma_u)\det(\Sigma_{-u})}{\det(\Sigma)}}\exp\left(-\frac{1}{2}(x-\mu)^T(\Sigma^{-1} - (\Sigma_u^{-1} : \Sigma_{-u}^{-1}))(x-\mu)\right) \quad (42)$$

where $(\cdot : \cdot)$ is the out-of-order composition of block diagonal matrices. This readily generalizes to Gaussian copulas. In the bivariate case, setting $\mu = 0$ and $\Sigma = \left(\begin{smallmatrix} 1 & \varrho \\ \varrho & 1 \end{smallmatrix}\right)$ in (42) therefore yields (19).

8 Experiments

In this section we study the numerical implementation of the designs discussed in this work, using various examples which focus on different aspects. Subsection 8.1 exploits the analytical results obtained for Gaussian linear models to compare numerical estimates and analytical results for showing asymptotic consistency. Subsection 8.2 uses the Ishigami function to compare the reweighting estimators in greater detail. Subsection 8.3 considers the case in which inputs are constrained to lie in a circle. In this case, the winding stairs estimator cannot be used anymore and the section concerns the comparison of nearest-neighbours and double-loop approaches. Subsection 8.4 addresses the case in which inputs are constrained on a non-connected domain (two separate triangles). Subsection 8.5 reports results for the case in which inputs are constrained on a simplex, based on the case study of [20]. Subsection 8.6 reports results in which inputs are constrained on complex and non-connected domains which can be modeled as Sierpinski gaskets. Subsection 8.7 reports results for a realistic simulator. All experiments are run on personal computer with an Intel(R) i7-3770 CPU at 3.40GHz and 8GB RAM, using MatLab R2022a. We rely on the MatLab k-d-tree implementation for the nearest-neighbour search. In all the experiments we implemented the nearest-neighbour approach (see Section 5) with the low threshold equal to zero.
In these experiments we compare the estimators in Table 1.
As shown in the third column, the methods are associated with different computational costs. To make experiments comparable, we fix the overall budget of the experiment and then find back the block sample size n. To illustrate, in the case the budget is, say, $B = 10000$ model runs, for a $d = 3$ variable model we have sample block sizes respectively of $n = 2500$ for the generalized winding stairs and the shift(derange)-and-reweight designs, and $n = 58$ for the U-statistic design, while $B = n = 10000$ for the nearest-neighbour design.

Table 1: Computational costs for the estimation of all d total effects with constrained inputs, block sample size n.

Design	Reference	Cost
Generalized Winding Stairs	Equation (15)	$n(d+1)$
Mix-and-Reweight, U-statistics	Proposition 6	$n(1+d(n-1))$
Shift-and-reweight	Theorem 9	$n(d+1)$
Derange-and-reweight	Theorem 11	$n(d+1)$
Reweight with nearest neighbours	Equation (32)	n

8.1 Linear Model with Normal and Correlated Inputs

We consider the linear model $Y = g(X_1, X_2, X_3) = X_1 + X_2 + X_3$ with Gaussian inputs $X \sim \mathcal{N}(0, \Sigma)$ where the correlation matrix is $\Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & \varrho\sigma \\ 0 & \varrho\sigma & \sigma^2 \end{pmatrix}$. This example is also discussed in [35]. We also test the effect of using alternative sample generators, namely, crude Monte Carlo (MC) and randomized Quasi-Monte Carlo (RQMC). For the randomization we employ the MatLab Sobol' sequence scrambler, following [44, 40]. Applying Theorem 14, for this model we find the values of the total effects analytically. They can then be used to study convergence and uncertainty in the estimates. In our parametrization, we set $\sigma = 2$ and let ϱ vary between $[-1, 1]$.

For the numerical experiments, we hypothesize a computational budget restriction, with a fixed budget B. Fixed budgets are of interest for computationally expensive models, where one faces a trade-off between statistical accuracy and the needs to keep computational time under control. Given the designs in Table 1, once the total budget is fixed, one derives as previously explained the corresponding basic sample sizes n. We set $B = 1680$, which, with $d = 3$ input factors, leads to $n = 24$ for the U-statistics, $n = 420$ for the winding stairs and reweighting approaches, and $n = 1680$ for the nearest-neighbour approach.

Figure 1 shows the results for alternative values of ϱ (that is, we repeat the calculations with the fixed buget B for $\varrho \in \{-0.9, -0.75, -0.6, \ldots, 0.6, 0.75, 0.9\}$). Each graph reports the values of ϱ on the horizontal axis. On the vertical axis, the corresponding analytical values of τ_i $(i = 1, 2, \ldots, 3)$ are displayed using continuous lines, and the estimates by dashed lines. Confidence intervals are displayed as shaded areas.

For the U-statistics approach, we use the asymptotic normality result of Lemma 8 together with plug-in estimates for the estimator variance. The results for the winding stairs estimate follow from [21]. In the shift-and-reweight approach (panels C and D) and in the derange-and-reweight approach (panels G and H) we use the upper and lower 2.5% quantiles from a block-bootstrap. In our further tests, a normal approximation using the estimator variance from the asymptotic results of Theorem 9 or Theorem 11 offers comparable intervals.

Figure 1 shows that the point estimates from QMC designs (using scrambled sequences)

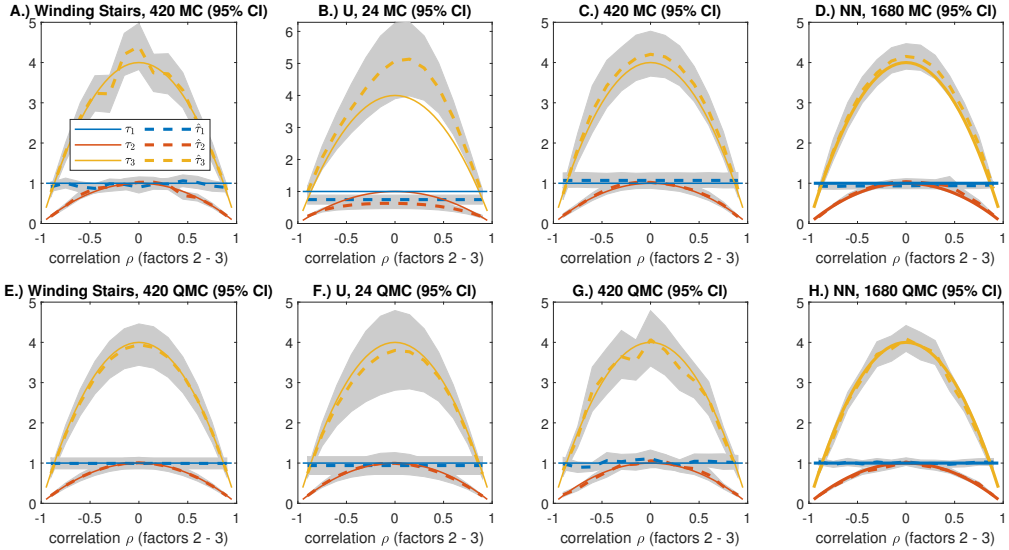

Figure 1: Total effects (unnormalized) for the linear model, including confidence bounds (gray areas): Winding stairs (panels A and E), U-statistics estimator (panels B and F), shift-and-reweight (panels C and D), derange-and-reweight (panels G and H). The analytical total effects are represented by thin lines.

generally perform better than those from a plain Monte-Carlo design, while the confidence bounds are comparable. Preliminary tests showed that the shift-and-reweight approach is not working well with QMC design. Instead, we replaced the shift with a permutation, and a derange-and-reweight approach has been used for panels G and H. Considering the mean squared error for different simulations one can conclude that for this example, a QMC design is an advantage for U-statistics and winding stairs methods, while the impact is not so clear on the derange-and-reweight methods (both with reevaluations and with nearest-neighbour approximations of the mixed sample).
In this example, investing the computational budget into one large sample and using a nearest-neighbour metamodeling approach seems to offer a good performance, only to be beaten by winding-stairs QMC design that, as already remarked, is not necessarily available for general dependence in form of constraints of the input features (these visual findings are corroborated by corresponding mean squared errors, see Table 2).

8.2 The Ishigami Function with Correlations under a Gaussian Copula

We further test the proposed designs on a well known test case, the Ishigami function with correlated inputs. This test case has been used in previous studies on total effects with dependent features in works such as [35]. We also use these experiments to continue our investigation about the effects of the random sample generator on the estimators. The input-output mapping is $Y = g(x_1, x_2, x_3) = \sin(x_1) + 7\sin^2(x_2) + 0.1x_3^4 \sin(x_1)$ with

Table 2: Averaged mean squared errors (with respect to ϱ) of the different methods and sampling designs for the Linear Model using 20 repetitions.

Factor	1	2	3
Winding Stairs MC	5.38e-03	2.52e-03	4.09e-02
Winding Stairs QMC	1.40e-04	7.61e-05	1.48e-03
U Statistics MC	6.19e-02	1.99e-02	5.91e-01
U Statistics QMC	1.43e-02	9.88e-03	9.14e-02
Derange&Reweight MC	1.04e-02	5.02e-03	5.73e-02
Derange&Reweight QMC	2.32e-03	2.17e-03	3.41e-02
NN Derange&Reweight MC	3.27e-03	1.28e-03	2.53e-02
NN Derange&Reweight QMC	7.47e-04	7.41e-04	1.11e-02

X_i uniformly distributed on $[-\pi, \pi]$. As in [35, Section 7.3], we introduce a statistical dependence between X_1 and X_3 by prescribing a pairwise Gaussian copula with a varying correlation coefficient.

For the winding-stairs approach we use the inverse Knothe–Rosenblatt transformation detailed in Example 5. For the reweight strategies, we implement the rank correlation using (17). We let the correlation coefficient $\rho(X_1, X_3)$ range from $\rho(X_1, X_3) = -0.9$ to $\rho(X_1, X_3) = 0.9$. While there is no closed form solution for the total indices at a generic value of $\rho(X_1, X_3)$, a reference value is available at $\rho(X_1, X_3) = 0$, for which the total indices are analytically known, with values $T_1 = 0.56$, $T_2 = 0.44$ and $T_3 = 0.24$, respectively.

We fix a budget of about $B = 8200$ simulations. This yields a basic sample size of $n = 2048$ for the generalized winding stairs and weight and derange approach, of $n = 53$ for the U-statistic and $n = 8192$ points for the nearest-neighbour estimator. We generate the sample first with crude MC and then with QMC.

Figure 2 shows the estimates of the normalized total effects T_1, T_2, T_3 as a function of the correlation between X_1 and X_3. The upper row displays estimates when crude MC generation is used, while the lower row displays estimates when QMC generation is used. When using crude MC, the winding-stairs, derange-and-reweight as well as the nearest neighbout designs perform comparably. At $\rho(X_1, X_3) = 0$, estimates for these designs are close to the analytical values.

When using QMC, the curves of the winding-stairs and U-statistic estimators exhibit increased regularity, while methods with a random derangement do not. This can be due to random derangements breaking the properties of low discrepancy sequences. Despite the greater regularity of the U-statistic estimates as a function of $\rho(X_1, X_3)$, its estimates are upward biased for T_1 and most notably for T_2, compared to the other approaches. A reason may be that at $n = 53$, the QMC Sobol' sequence does not populate a Latin Hypercube and the projections on the marginals are not uniform. To fill in a Latin Hypercube, we would need to increase the basic sample size to the next power of 2, $n = 64$. However, these additional eleven points in the sample block would propagate into a new budget of $B = 12160$, a nearly 50% increase in computational cost. Lastly,

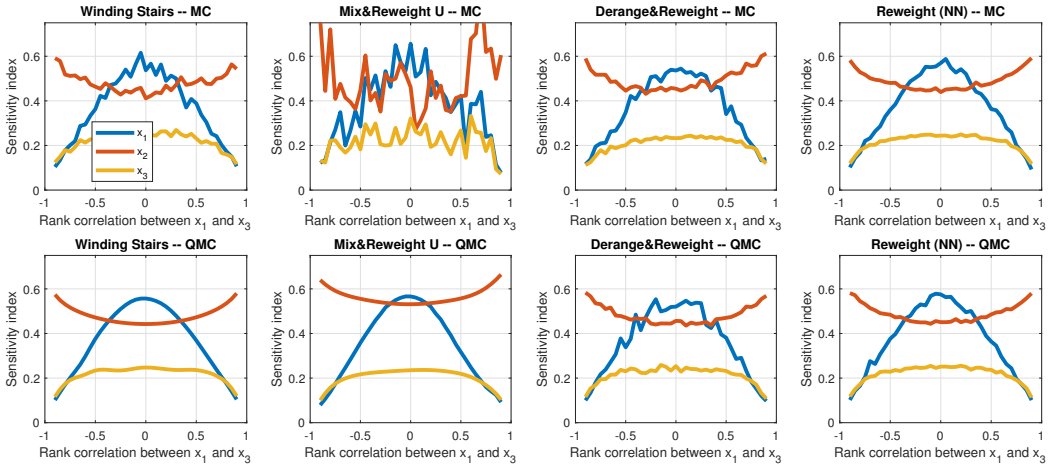

Figure 2: Ishigami function with rank correlation: normalized total (T_i) indices depending on the rank correlation $\rho^*(X_1, X_3)$. Total budget: 8200 model evaluations. Upper row: crude Monte Carlo sampling, lower row: Quasi Monte Carlo sampling.

the rightmost panels show that the nearest-neighbour estimator performs similarly with both sample generation methods.

8.3 Features Constrained on a Circle

While the previous examples featured dependence structures given by Gaussian copula correlation, this example is used to study a non-rectangular support. Let us consider the two-dimensional input-output mapping

$$Y = g(X_1, X_2) = (X_1 - 1) \cdot (X_2 - 1), \tag{43}$$

with X_1 and X_2 uniformly distributed within a circle of radius π centered at the origin. Figure 3 shows an input sample of size $n = 1024$ and the model output surface. As the output density can be computed analytically, calculation with a symbolic software (Mathcad Prime 8 in our case) yields unnormalized total effects $\tau_1 = \tau_2 = 6.527$. This geometry rules out the application of a design based on the Knothe-Rosenblatt transform for calculating total effects. In fact, we would need to find the quantile function corresponding to the marginal cdf of X_1,

$$x \mapsto \frac{2}{\pi R^2}\left(\frac{x}{2}\sqrt{\max\{0, R^2 - x^2\}} + \frac{R^2}{2}\arcsin\left(\min\left\{1, \max\left\{-1, \frac{x}{R}\right\}\right\}\right) + \frac{\pi R^2}{4}\right)$$

and plug this into the conditional quantile function of x_2, $(u, x) \mapsto 2(u - \frac{1}{2})\sqrt{R^2 - x^2}$. Even for this simple geometry, this seems to be a tantalizing task. In contrast, in case of a constrained domain C the joint density is $x \mapsto \frac{\mathbb{1}_C(x) f_X(x)}{\int_C f_X(x)dx}$ where we assume that

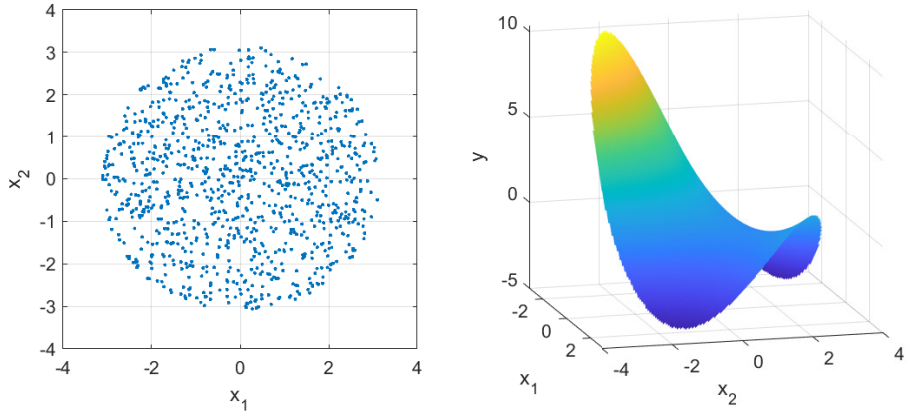

Figure 3: Uniform inputs constrained within a circle (left), model response (right).

the probability density without the constraint is f_X. The marginal distributions can be obtained by integrating over the joint distribution. The probability of $X \in C$ can be estimated from a rejection sample approach. In the case of an output constraint, D, in the context of target sensitivity, then the input constraint C becomes the inverse image of set D under g: $C = \{x : g(x) \in D\} = g^{-1}[D]$. In the example discussed here, C is a circle of radius R and the density quotient to be used in our proposed methods is

$$\iota_1(x_1, x_2) = \frac{\pi R^2}{4} \frac{\mathbb{1}\{x_1^2 + x_2^2 \leq R^2\}}{\sqrt{\max\{0, R^2 - x_1^2\}} \cdot \sqrt{\max\{0, R^2 - x_2^2\}}} = \iota_2(x_2, x_1). \tag{44}$$

Equation (44) is symmetric in its arguments, hence it can be used for computing the sensitivity of both input factors. We use the experiments to investigate the rate of convergence of the derange-and-reweight approach of Theorem 9 when using different sampling strategies: plain Monte Carlo, Quasi Monte Carlo and randomized Quasi Monte Carlo. The mean squared errors which are averaged over all factors follow a $O(n^{-1})$ convergence rate for all sampling strategies, as seen in Figure 4, where the diagonal of the dashed triangle evidences the n^{-1} convergence rate. The estimators therefore are of rate $O(n^{-1/2})$, as for standard Monte Carlo estimation. We observe a similar convergence rate across the alternative generators. A reason may be that the shifting strategy interferes with the regularity of the QMC structure, thus reducing the advantage of using a QMC generator in this context. Furthermore, a constraint may introduce discontinuities which are not compatible with functions of bounded variation in the sense of Hardy and Krause and in this case the Koksma-Hlawka Theorem does not provide an improved convergence rate.

8.4 Features Constrained on Two Disconnected Triangles

In this section, we consider experiments for features constrained on a disconnected domain. We hypothesize that the model is $g(x_1, x_2) = (x_1 - 1) \cdot (x_2 - 1)$ and the features

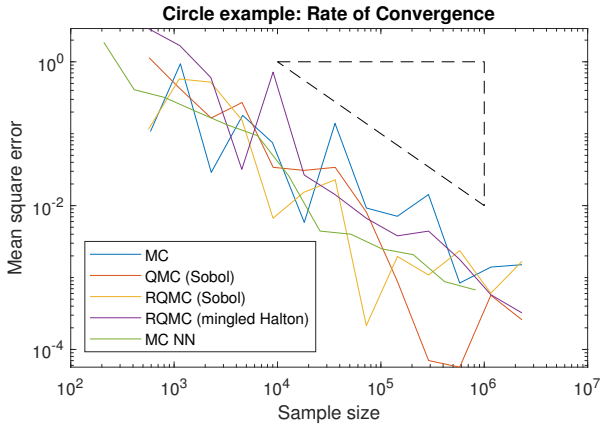

Figure 4: Mean squared errors for the unnormalized total effects over both factors with alternative sample generators, depending on the sample size (number of model evaluations) using basic sample sizes from 2^8 to 2^{20}.

lie in the 2-dimensional domain $\mathcal{X} = \{(x_1, x_2) \in [0,1]^2 : x_2 \leq \frac{1}{2} - x_1 \vee x_2 \leq 1 - \frac{1}{4}x_1\}$ (Figure 5). The joint input density, f_{12}, is 4 within the triangles and vanishes outside the triangles. With a symbolic calculator, one can compute analytical formulas for the marginal densities f_1 and f_2 by integrating f_{12}. In this case, also the sensitivity indices can be computed analytically. Theoretical computations yield (normalized) total effects of $T_1 = 0.037$ and $T_2 = 0.27$ (the total variance is 0.117).
We will now test these values against a numerical implementation. The sampling in this case is performed with rejection, starting from a uniform distribution on the square. The procedure generates dependence between X_1 and X_2. Figure 5 shows a sample of realizations consistent with this domain. We run experiments to test the proposed reweighting methods. Table 3 shows the results for the shift-and-reweight method of Theorem 9. Due to the use of rejection sampling and because mixture realisations which fall out the constraints are not evaluated, the sample sizes do not form a regular progression. The right part of the Table reports the results when processing the basic sample size generated before with nearest neighbours, i.e., the number of model evaluations is reduced to the size of the basic sample block. One observes only minor differences between both approaches. The estimator variances of Table 3 are computed from a plug-in estimator (see the discussion following Lemma 8).

8.5 Features Constrained on a Simplex

We implement a dependence between the last and the last-but-one input by restricting the input to satisfy a simplex condition $\{(x_1, \ldots, x_d) \in [0,1]^d : x_{d-1} \leq x_d)\}$, see Figure 6 for $d = 3$ and a regular set of points. The density quotient between all pairs of inputs is constantly one, except for the last two features where it is $w(x_{d-1}, x_d) = \frac{\mathbb{1}\{x_{d-1} \leq x_d\}}{2(1-x_{d-1})x_d}$. For this constraint, analytical values of the normalized total effects are available for a number of examples with different input dimensions, using symbolic calculations. For

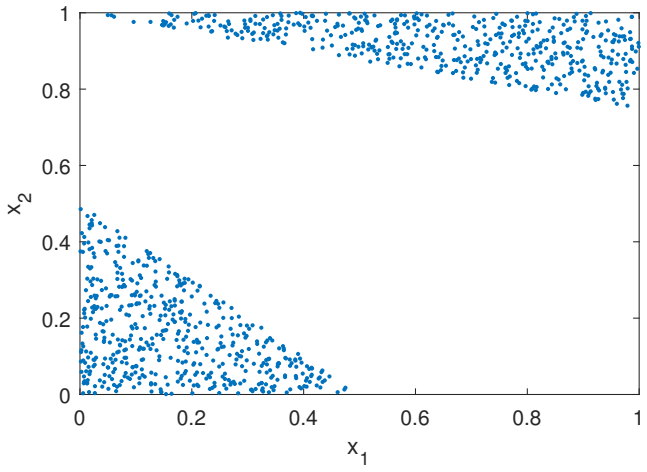

Figure 5: An input domain in form of two separate triangles.

Table 3: A product model on a support of two separate triangles, relative total effects, shift-and-reweight method of Theorem 9.

		Plain Monte Carlo					Nearest Neighbours			
Size	Runs	T_1	T_2	$\sigma(T_1)$	$\sigma(T_2)$	Runs	T_1	T_2	$\sigma(T_1)$	$\sigma(T_2)$
		0.037*	0.27**	$10^{-2}\cdot$			0.037*	0.27**	$10^{-2}\cdot$	
474	960	0.0328	0.2447	0.5575	3.7109	474	0.0330	0.2445	0.5545	3.7247
701	1445	0.0325	0.2401	0.5041	3.0781	701	0.0320	0.2396	0.5028	3.0652
990	2063	0.0376	0.2768	0.4063	2.4153	990	0.0375	0.2768	0.4060	2.4113
1969	3953	0.0471	0.2592	0.3825	1.7471	1969	0.0473	0.2595	0.3858	1.7512
3522	7138	0.0401	0.2475	0.2603	1.3221	3522	0.0400	0.2476	0.2580	1.3237

$d = 3$ and $g(x_1, x_2, x_3) = x_1x_2 - x_3$ one has analytical total effects given by $T = (0.125,\ 0.125,\ 0.375)$. For $d = 4$ and $g(x) = -x_1 + x_1x_2 - x_1x_2x_3 + x_1x_2x_3x_4$ analytical values are given by $T = (0.6300,\ 0.4861,\ 0.0064,\ 0.0064)$. A third variant uses the well-known g-function with $d = 4$ and parameter vector $a = (0, 1, 3, 6)$, again using the simplex constraint condition on the third and fourth argument. Here the analytical values are $T = (0.7659,\ 0.2357,\ 0.0522,\ 0.0172)$. These examples were introduced in [20] to test a new space-filling sampling strategy for estimating grouped Sobol' effects in the framework of constrained inputs (see [28]). We perform simulations using the shift-and-derange estimators of Theorem 9 with random shifts. The mean squared errors of all the examples (taking the average of 20 runs) are reported in Figure 7. Without preprocessing the QMC sample by permuting the order of realizations, the shift method fails. However, the use of (permuted) scrambled Sobol' sequences as input does not change the convergence rate which is consistent with the Monte Carlo approximation error. The nearest-neighbour approach presents even the best performance in the first two examples.

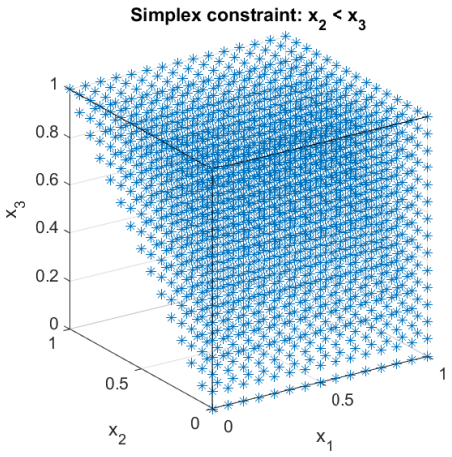

Figure 6: An input constraint in form of a simplex condition.

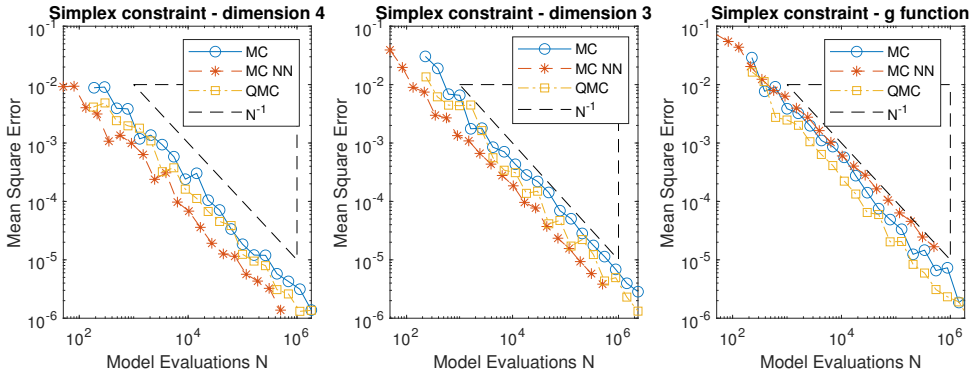

Figure 7: Convergence (MSE) of total effects for models with a simplex constraint in the inputs.

8.6 Features Constrained on Sierpinski Gaskets

In this section, we consider an experiment that starts with a Cartesian support and remove parts of the support, until we reach an almost fractal structure, an approximation of the Sierpinski gasket, that contains holes and is not star-shaped connected. Figure 8 demonstrates the regions where the two input features are constrained on.
For the model, we consider a two-dimensional input-output mapping of the form $Y = X_1 + \beta X_2$. In the experiments we consider alternative values for the model-parameter β, namely $\{-2, -1, 0, 1, 2\}$. Although this is a family of linear models, it can be used to demonstrate the subtleties which occur when working with dependent input data.
The first two domains are formed by cutting corners of square: a cut-one-corner-constraint consisting of all points on the $[0, 1]$ with exclusion of the pairs (x_1, x_2) such that $x_1 + x_2 > 3/2$, an exclude-two-corners constraint, where additionally points satisfying $x_1 + x_2 < 1/2$ are excluded. Then, we approximate a fractal structure, the Sierpinski gasket, by exclud-

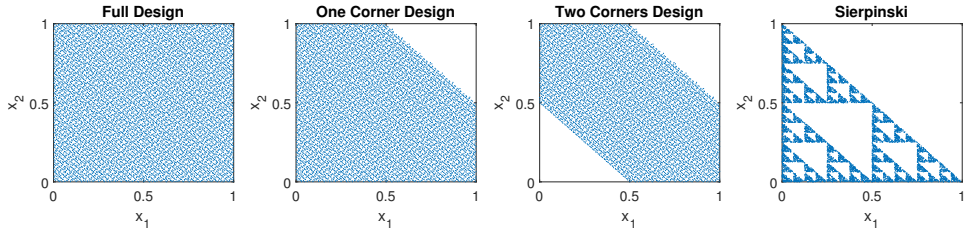

Figure 8: Feature constraints: left one-corner, center two-corner, right Sierpinski with depth $k = 5$.

ing all realizations (x_1, x_2) that satisfy the following three conditions for $k = 1, 2, \ldots, 5$:

- $\mathrm{mod}(2^{k-1}(x_1 + x_2), 2) > 1$,
- $\mathrm{mod}(2^{k-1}x_1, 2) \leq 1$, and
- $\mathrm{mod}(2^{k-1}x_2, 2) \leq 1$,

where $x \mapsto \mathrm{mod}(x, 2) = x - 2\left\lfloor \frac{x}{2} \right\rfloor$ denotes the rest after an integer division by 2.
To use the reweighting approach with these constraints, we formulate the density quotient as composed of the indicator function of the acceptance region, divided by the product of the marginal densities. The required normalizing constant is related to the acceptance rate. It is possible to derive the expression of the density quotient analytically. For the cut-one-corner constraint the joint density is

$$f_{12}^{OC}(x_1, x_2) = \frac{7}{8} \cdot \mathbb{1}\left\{x_1 + x_2 \leq \frac{3}{2}\right\} \tag{45}$$

and the marginal densities by

$$f_1^{OC}(x) = f_2^{OC}(x) = \frac{7}{8} \cdot \left(1 - \left(x - \frac{1}{2}\right)^+\right). \tag{46}$$

For the two-corner design we have a joint density give by

$$f_{12}^{TC}(x_1, x_2) = \frac{3}{4} \cdot \mathbb{1}\left\{\frac{1}{2} \leq x_1 + x_2 \leq \frac{3}{2}\right\} \tag{47}$$

and marginal densities

$$f_1^{TC}(x) = f_2^{TC}(x) = \frac{3}{4} \cdot \left(1 - \left|x - \frac{1}{2}\right|\right). \tag{48}$$

Here $x_1, x_2, x \in [0, 1]$. The rejection rate is present in both the marginal distributions and the joint one, so that the inverse of the rejection rate is a multiplicative constant in the density quotient. For the Sierpinski gasket, the joint distribution is the product of indicator functions of the complements of the Sierpinski sets listed above.

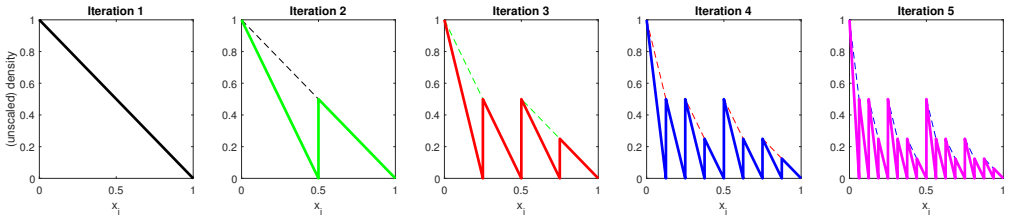

Figure 9: Sierpinski marginal density derivation: evolution of the marginal integral with rejections.

Table 4: Estimates of the total effects (in percent) for the linear model $Y = X_1 + \beta X_2$. (SR: shift-and-reweight, NN: nearest-neighbour method)

	Full Design				One Corner				Two Corners				Sierpinski			
	SR		NN		SR		NN		SR		NN		SR		NN	
β	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1	T_2
-2	20	80	20	80	15	63	16	63	11	45	11	44	11	42	11	43
-1	50	50	50	50	37	38	38	38	25	26	25	25	25	25	26	25
0	101	0	100	0	92	0	93	0	74	0	76	0	75	0	77	0
1	50	50	50	50	60	61	60	60	72	77	74	74	77	76	76	75
2	20	80	20	80	22	92	23	91	24	102	25	99	26	101	25	100

To proceed with numerical experiments, we use a rejection method. We generate data using crude Monte-Carlo sampling, and reject feature realizations that do not satisfy the constraints. For the cut-one-corner and cut-two-corners constraints, we start with a sample of size $n = 10240$ on the $[0, 1]^2$ full Cartesian domain. Following rejection of realizations falling outside the domain, for the cut-one-corner domain, we are left with $n = 8938$ realizations and for the cut-two-corners domain, we are left with $n = 7690$ realizations. These numbers are in line with the theoretical acceptance rates of $7/8$ and $3/4$, respectively.

For the Sierpinski gasket, we start with a base random sample size of $n_0 = 50000$. The first step in the rejection process retains half of the observations, while each further step retains $3/4$ of the observations. Hence after five iterations we reach a sampling size of $n \approx \frac{1}{2} \cdot \left(\frac{3}{4}\right)^4 n_0$. After rejection, we are left with $n = 7932$ observations. Figure 9 shows the empirical marginal densities of the features. Iteration $k = 1$ removes $1/2$, the further iterations $k > 1$ remove each $1/4$ of the mass. One needs to rescale these curves to obtain the marginal probability densities for both factors which are then piecewise linearly defined. The two-corner design and the Sierpinski gasket both satisfy $\mathrm{cov}(X_1, X_2) = -0.5$ and have roughly the same input variances. In a linear model, only the input variances and covariances enter into the computation of total effects, so that the estimates for the last two cases should be similar. Table 4 reports the results for the calculation of the total effects. The shift-and-reweight method of Theorem 9 was used. The nearest-neighbour approach shows only small differences in the results. Table 4 shows that for the full design $\sum T_i = 1$ as to be expected for a linear model, however, the inequality $1 \leq \sum T_i$ does not generally hold true under constraints. Note also the model behaviour

of for the case $\beta = 0$. The total effect of X_2 correctly asserts that in this case the model behaviour can be explained without recurring to information from X_2. However, the total effect of X_1 being less than one must then be due to input data dependence, and not by factor interaction in the model. The sum of totals is large if there are competing effects, i.e., when the negatively correlated features are fed into a simulation model with positive monotonicity ($\beta > 0$).

8.7 Application: the Flood Model of De Rocquigny (2006)

We study the flood model in [10], assuming the same dependence structure in the input features as in [7]. The model calculates the maximum annual overflow, given eight input features (Table 5). The correlation between the pair of features (1,2) is set

Table 5: Flood model feature list.

Feature	Symbol	Description	Unit	Distribution and Truncation
1	Q	Maximal annual flow rate	$\frac{m^3}{s}$	Gumbel(1013,558) on (500,3000)
2	K_s	Strickler coefficient	$-$	Normal(30,8) on (15,+∞)
3	Z_v	River downstream level	m	Triangular(49,50,51)
4	Z_m	River upstream level	m	Triangular(54,55,56)
5	H_d	Dyke height	m	Uniform(7,9)
6	C_b	Bank level	m	Triangular(55,55.5,56)
7	L	Length of river stretch	m	Triangular(4990,5000,5010)
8	B	River width	m	Triangular(295,300,305)

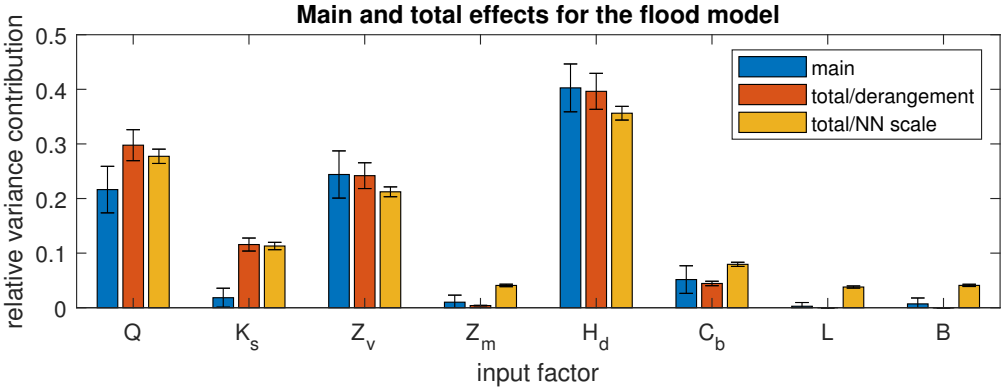

Figure 10: Sensitivity results for the flood model (error bars represent the 95% confidence band).

to 0.5, the correlation between (3,4) and (7,8) to 0.3 each, via Gaussian copula. The density quotients for each pair are given in (19). We calculate total indices with the derange-and-reweight approach of Theorem 9 and nearest neighbours, fixing a budget of $B = 9,000$ model evaluations. The basic sample size is then $n = 1000$ Monte Carlo

realisations. The correlation structure in the basic sample is implemented via the Iman-Conover method [27, 37]. Main effects are estimated from this basic sample using a discrete cosine transformation [48] with 8 harmonics. A jackknife is used to derive confidence bounds. For nearest neighbours, a Monte Carlo sample of size $n = 9,000$ (same budget) is used. Because of the different scales in the inputs, we standardize the input sample using the empirical standard deviations as scaling factors. Figure 10 displays the results in the form of a barplot, reporting the main and total effects for each feature, as well as the error bands on top of the bars. The error bands for the main effects and for the shift/derangement approach are calculated as 1.96 times the square root of the plug-in variance estimates, using a normal approximation for a 5% confidence bound. Regarding the feature ranking, Feature 6 (H_d) is identified as the most important, followed by Features 1 (Q), 3 (Z_v), 2 (K_s) and 7 (C_b); the remaining features play a minor role. This result is in accordance with the findings in [7]. The values of the main effects and total effects (estimated with the derange-and-reweight approach) are close. Because Feature 6 is probabilistically independent of the remaining features, this equality signals that H_d is not involved in relevant interactions. In contrast, features 1 and 2 are noticeably different under main and total effects, with total effects larger than their main effects. Also, if ranked according to main effects, Feature 3 would rank second most important, switching place with Feature 1. Overall, the derange-and-reweight approach that evaluates the model at the mixed realizations shows comparable results to the nearest-neighbour approach. However, the nearest-neighbour approach seems to exhibit a bias for the least important inputs. By the theoretical results of [12] for dimensions larger than four, the nearest-neighbour estimator bias dominates the estimator variance. One insight gained from this application is to use both a derangement and a nearest-neighbour approach, when possible, to confirm the estimates of both methods. This is because the former is less affected by bias as dimensionality increases.

9 Conclusion

Estimating total effects under feature dependence and constraints is a challenging task. We have first proposed a generalized winding stairs approach that relies on a Knothe-Rosenblatt transformation for the case of dependent features. However, this estimator becomes impractical for general dependence structures imposed by feature constraints. We then offered new approaches for total effects under feature constraints with a formal theoretical investigation of convergence properties. The intuition is based on pairing Jansen's estimator with a reweighting factor. We have first proposed a U-statistic estimator, for which a central limit theorem is immediately derived from classical results. The estimator, however, turns out to be computationally expensive. We have then studied two alternatives based on shifts and derangements that abate computational burden. We have derived corresponding central limit theorems. The proposal of an estimator based on the nearest-neighbour technique has been studied. The behavior of the estimators has been tested on numerous experiments with feature constraints of increas-

ing complexity. We have formally derived the link between total indices and Breiman's permutation feature importance measures under constraints.
From a machine learning perspective, our estimators prevent the flaw of permute-and-predict methods identified by [26], not only for dependent features, but also in the more difficult setting of non-Cartesian domains: the density quotient, in fact, places a zero weight on points that violate the constraint after the permutation.

References

[1] A. Badea and R. Bolado. Milestone M.2.1.D.4: Review of sensitivity analysis methods and experience. Technical report, PAMINA Project, Sixth Framework Programme, European Commission, 2008. http://www.ip-pamina.eu/downloads/pamina.m2.1.d.4.pdf.

[2] C. Bénard, S. D. Veiga, and E. Scornet. Mean decrease accuracy for random forests: inconsistency, and a practical solution via the Sobol-MDA. *Biometrika*, 2022. DOI:10.1093/biomet/asac017.

[3] A. Bose and S. Chatterjee. *U-Statistics, M_m-Estimators and Permutations.* Springer Verlag, Singapore, 2018.

[4] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[5] B. Broto, F. Bachoc, and M. Depecker. Variance reduction for estimation of Shapley effects and adaptation to unknown input distribution. *SIAM/ASA Journal on Uncertainty Quantification*, 8(2):693–716, 2020.

[6] K. Chan, A. Saltelli, and S. Tarantola. Winding stairs: A sampling tool to compute sensitivity indices. *Statistics and Computing*, 10(3):187–196, 2000.

[7] G. Chastaing, F. Gamboa, and C. Prieur. Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis. *Electronic Journal of Statistics*, 6:2420–2448, 2012.

[8] S. Chatterjee. A new coefficient of correlation. *Journal of the American Statistical Association*, 116(536):2009–2022, 2021.

[9] S. Da Veiga, F. Gamboa, B. Iooss, and C. Prieur. *Basics and Trends in Sensitivity Analysis: Theory and Practice in R.* SIAM, Philadelphia PA, 2021.

[10] E. de Rocquigny. La maîtrise des incertitues dans un contexte industriel. 1re partie: une approche méthodologique globale basée sur des exemples. *Journal de la Société Française de Statistique*, 147(3):33–71, 2006.

[11] L. Devroye, P. G. Ferrario, L. Györfi, and H. Walk. Strong universal consistent estimate of the minimum mean squared error. In *Empirical Inference*, pages 143–160. Springer Verlag, 2013.

[12] L. Devroye, L. Györfi, G. Lugosi, and H. Walk. A nearest neighbor estimate of the residual variance. *Electronic Journal of Statistics*, 12:1752–1778, 2018.

[13] P. H. Diananda. The central limit theorem for m-dependent variables asymptotically stationary to second order. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 50, pages 287–292. Cambridge University Press, 1954.

[14] B. Efron and C. Stein. The jackknife estimate of variance. *The Annals of Statistics*, 9(3):586–596, 1981.

[15] A. Fisher, C. Rudin, and F. Dominici. All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. *Journal of Machine Learning Research*, 20:1–81, 2019.

[16] M. Fréchet. Sur le coefficient, dit de corrélation et sur la corrélation en géneral. *Revue de l'Institut International de Statistique*, 1(4):16–23, 1934.

[17] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur. Statistical inference for Sobol pick-freeze Monte Carlo method. *Statistics*, 50(4):881–902, 2016.

[18] D. Gatelli, S. Kucherenko, M. Ratto, and S. Tarantola. Calculating first-order sensitivity measures: A benchmark of some recent methodologies. *Reliability Engineering&System Safety*, 94:1212–1219, 2009.

[19] R. Genuer, V. Michel, E. Eger, and B. Thirion. Random forests based feature selection for decoding fmri data. In *Proceedings Compstat*, volume 267, pages 1–8, 2010.

[20] L. Gilquin, C. Prieur, and E. Arnaud. Replication procedure for grouped sobol'indices estimation in dependent uncertainty spaces. *Information and Inference: A Journal of the IMA*, 4(4):354–379, 2015.

[21] T. Goda. A simple algorithm for global sensitivity analysis with Shapley effects. *Reliability Engineering&System Safety*, 213:107702, 2021.

[22] J. Hart and P. A. Gremaud. An approximation theoretic perspective of Sobol' indices with dependent variables. *Int. J. Uncertainty Quantification*, 8(6):483–493, 2018.

[23] R. Helmers. On the Edgeworth expansion and the bootstrap approximation for a Studentized U-statistic. *The Annals of Statistics*, 19:470–484, 1991.

[24] W. Hoeffding. A Class of Statistics with Asymptotically Normal Distribution. *The Annals of Mathematical Statistics*, 19(3):293 – 325, 1948.

[25] T. Homma and A. Saltelli. Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering&System Safety*, 52(1):1–17, 1996.

[26] G. Hooker, L. Mentch, and S. Zhou. Unrestricted permutation forces extrapolation: variable importance requires at least one more model, or there is no free variable importance. *Statistics and Computing*, 31(6):82:1–16, 2021.

[27] R. L. Iman and W. J. Conover. A distribution-free approach to inducing rank correlation among input variables. *Communications in Statistics - Simulation and Computation*, 11(3):311–334, 1982.

[28] J. Jacques, C. Lavergne, and N. Devictor. Sensitivity analysis in presence of model uncertainty and correlated inputs. *Reliability Engineering&System Safety*, 91(10–11):1126–1134, 2006.

[29] A. Janon, T. Klein, A. Lagnoux, M. Nodet, and C. Prieur. Asymptotic normality and efficiency of two Sobol index estimators. *ESAIM: Probability and Statistics*, 18:342–364, 2014.

[30] M. J. Jansen, W. A. Rossing, and R. A. Daamen. Monte carlo estimation of uncertainty contributions from several independent multivariate sources. *Predictability and nonlinear modelling in natural sciences and economics*, pages 334–343, 1994.

[31] M. J. W. Jansen. Analysis of variance designs for model output. *Computer Physics Communications*, 117(1–2):35–43, 1999.

[32] H. Joe. *Dependence Modeling with Copulas*. CRC Press, Boca Raton, 2014.

[33] H. Knothe. Contributions to the theory of convex bodies. *Michigan Math. J.*, 4(1):39–52, 1957.

[34] S. Kucherenko, O. V. Klymenko, and N. Shah. Sobol´ indices for problems defined in non-rectangular domains. *Reliability Engineering&System Safety*, 167:218–231, 2017.

[35] S. Kucherenko, S. Tarantola, and P. Annoni. Estimation of global sensitivity indices for models with dependent variables. *Computer Physics Communications*, 183(4):937–946, 2012.

[36] G. Li and H. Rabitz. Relationship between sensitivity indices defined by variance- and covariance-based methods. *Reliability Engineering&System Safety*, 167:136–157, 2017.

[37] G. Mainik. Risk aggregation with empirical margins: Latin hypercubes, empirical copulas, and convergence of sum distributions. *Journal of Multivariate Analysis*, 141:197–216, 2015.

[38] T. A. Mara and S. Tarantola. Variance-based sensitivity indices for models with dependent inputs. *Reliability Engineering&System Safety*, 107:115–121, 2012.

[39] T. A. Mara, S. Tarantola, and P. Annoni. Non-parametric methods for global sensitivity analysis of model output with dependent inputs. *Environmental Modelling&Software*, 72:173–183, 2015.

[40] J. Matoušek. On the L^2-discrepancy for anchored boxes. *Journal of Complexity*, 14(4):527–556, 1998.

[41] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu. Definitions, methods and applications in interpretabile machine learning. *Proceedings of the National Academy of Sciences*, 116(44):22071–22080, 2019.

[42] M. H. Neumann. A central limit theorem for triangular arrays of weakly dependent random variables, with applications in statistics. *ESAIM: Probability and Statistics*, 17:120–134, 2013.

[43] J. E. Oakley and A. O'Hagan. Probabilistic sensitivity analysis of complex models: A Bayesian approach. *Journal of the Royal Statistical Society, Series B*, 66(3):751–769, 2004.

[44] A. B. Owen. Scrambled net variance for integrals of smooth functions. *The Annals of Statistics*, 25(4):1541–1562, 1997.

[45] A. B. Owen and C. R. Hoyt. Efficient estimation of the anova mean dimension, with an application to neural net classification. *SIAM/ASA Journal on Uncertainty Quantification*, 9(2), 2021.

[46] A. B. Owen and C. Prieur. On Shapley value for measuring importance of dependent inputs. *SIAM/ASA Journal on Uncertainty Quantification*, 5(1):986–1002, 2017.

[47] K. Pearson. Notes on regression and inheritance in the case of two parents. *Proc. Royal Soc. London*, 58:240–242, 1895.

[48] E. Plischke. How to compute variance-based sensitivity indicators with your spreadsheet software. *Environmental Modelling&Software*, 35:188–191, 2012.

[49] E. Plischke, G. Rabitti, and E. Borgonovo. Has the spell been broken? Estimating global sensitivity measures via nearest neighbors. 2022. In Preparation.

[50] C. Prieur and S. Tarantola. Variance-based sensitivity analysis: Theory and estimation algorithms. In *Handbook of Uncertainty Quantification*, pages 1217–1239. Springer Verlag, Cham, 2017.

[51] H. Rabitz and Ö. F. Alış. General foundations of high-dimensional model representations. *J. Math. Chem.*, 25(2–3):197–233, 1999.

[52] M. Rosenblatt. Remarks on a multivariate transformation. *Ann. Math. Statist.*, 23(3):470–472, 1952.

[53] A. Saltelli, K. Chan, and E. M. Scott. *Sensitivity Analysis*. John Wiley&Sons, Chichester, 2000.

[54] A. Saltelli and S. Tarantola. On the relative importance of input factors in mathematical models: Safety assessment for nuclear waste disposal. *Journal of the American Statistical Association*, 97(459):702–709, 2002.

[55] I. M. Sobol'. Sensitivity estimates for nonlinear mathematical models. *Mathematical Modelling & Computational Experiments*, 1:407–414, 1993.

[56] I. M. Sobol', S. Tarantola, D. Gatelli, S. S. Kucherenko, and W. Mauntz. Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering&System Safety*, 92:957–960, 2007.

[57] D. M. Sparkman, J. E. Garza, H. R. Millwater, Jr., and B. P. Smarslok. Importance sampling-based post-processing method for global sensitivity analysis. In *18th AIAA Non-Deterministic Approaches Conference. 4-8 January 2016, San Diego, California, USA*. AIAA SciTech, 2016. Paper #AIAA 2016-1444.

[58] X. Sun, W. Zhong, and P. Ma. An asymptotic and empirical smoothing parameters selection method for smoothing spline anova models in large samples. *Biometrika*, 108(1):149–166, 2021.

[59] A. W. van der Vaart. *Asymptotic Statistics*, volume 3 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, 1998.