SPIQ-FS Dataset

Version 0

Joint sampling of marginal damage costs

Steven Lord

Environmental Change Institute, University of Oxford

Citation: Lord, S. (2022) Joint sampling of marginal damage costs. Documentation of the SPIQ-FS Dataset Version 0. Environmental Change Institute, University of Oxford.

Summary

The SPIQ-FS calculates marginal damage costs in US$2020 PPP for GHG emissions, nitrogen emissions to air and surface waters, land-use changes, blue water withdrawal, number of undernourished and headcount of extreme poverty occurring in the baseline year of 2020 [1-5]. Samples and a parametric fit of the marginal damage costs are provided as output from SPIQ-FS.

This document examines three sets of joint sampling of marginal damage costs from the SPIQ-FS cost model. The three sets are used in SPIQ-FS studies to examine whether economic risk of aggregated damages increases due to interaction in the impact pathways of the quantities associated to the marginal damage costs, and what are possible ranges for the increase in economic risk.

Contents

Joint sampling of marginal damage costs

The SPIQ-FS model calculates marginal damage costs in US$2020 PPP for GHG emissions, nitrogen emissions to air and surface waters, land-use changes, blue water withdrawal, number of undernourished and headcount of extreme poverty occurring in the baseline year of 2020 [1-5]. Each of the marginal damage costs are generated with uncertainty based on the modelled relationships and data in the models. SPIQ-FS models output samples of the marginal damage costs, and a log-normal fit to the samples based on the samples and based on arguments using algebra of the random variables involved in the cost model, see Annex A.

SPIQ-FS datasets of marginal damages costs are customised for different studies, but the generic SPIQ-FS dataset has 14 marginal damage cost items for 158 countries, or 2212 marginal damage costs in total (Table 1).

In most economic studies using SPIQ-FS we want to calculate total damages. We exemplify the process of obtaining uncertainty in total costs in SPIQ-FS. For concreteness we assume we are using the generic SPIQ-FS dataset. In this case, from the output of the individual SPIQ-FS cost models, we start with 2212 random variables given by the marginal damage costs.

Uncertainty in quantities and total costs

Uncertainty in total costs is a compound random variable between an uncertain costing function and uncertainty in quantities. The marginal damage costs are the partial derivatives of the costing function. A first order approximation to total costs as a compound random variable is the dot product distribution of uncertain marginal costs and uncertain quantities being costed. Uncertainty in quantities and marginal costs are separated in this case. The tail toward higher total costs is more informed as the dot product of two multi-variate random variables.

Most macro-economic models do not produce stochastic output, in which case the vector of quantity changes determined by the model is treated as a certain multivariate random vector. Uncertainty in the marginal damage costs translates into uncertainty in total costs by taking the dot product of the vector of random variables of marginal damage costs with the deterministic vector of quantity changes.

Dependency in marginal costs

Historical data and interactions of the impact pathways of the NH3 and NOx emissions in the nitrogen cost modelling create empirical correlations in the marginal costs between the two gases in a single country and across countries. Loss of crop value across countries from water scarcity in the blue water cost model also shows correlations across countries due to correlations in variation in FAO producer price indices. The existence of other interaction between the impact pathways of the marginal cost items, such as climate change created by GHG emissions inducing heat stress and higher levels of water vapour, which can makes air pollution more damaging to human health, creates the possibility that the 2212 marginal damage items across impact quantities and across countries, treated as random variables, may have correlations due to interaction or common dependencies on natural, human, produced or financial capital.

Net damages in a SPIQ-FS study using the SPIQ-FS generic dataset would be estimated by a dot product of the 2212 marginal damage random variables and the corresponding 2212 quantity changes. The shape of the distribution of the dot product, particularly the shape and percentiles of the tails of the dot product random variable, can be sensitive to dependency in the vectors of random variables in the product.

Analysis of economic risk from food system impacts will depend on the joint structure of the 2212 marginal damage costs treated as marginals of a 2212-variate random variable. Ignoring the dependency and treating the 2212 marginal damages costs as 2212 independent univariate random variables may underestimate risk. Overestimating the dependency may overestimate risk.

Shape of the marginals

SPIQ-FS version 0 outputs two parameters for log-normally distributed random variable for each marginal damage cost. Later versions of SPIQ-FS will include parameters for a log-normal mixture.

One reason discussed in Annex A to use log-normally distributed marginal costs is algebra – impact pathways involve a sequence of conditional events from emissions, resource use, etc., to environmental, human, and social consequences, which leads to economic effects. Conditional sequences of effects, where each effect in the sequence has uncertainty, leads to a product of random variables. Products of random variables are approximated by log-normal distributions, and the longer the sequence the better the approximation, up to caveats on dependency of the uncertainty in each step of the sequence. Most of the marginal costs in Annex A are log-normally distributed according to maximum likelihood criteria, illustrating empirically the influence of the way emissions, resource use, and consumption leads to economic consequences through a sequence of effects.

Many features of the conditional sequence are the same for each emission – the same chemistry, the same dose-dependent responses, etc. for each emission. Some of the features of the sequence are global, such as the fundamental chemistry, other are more local, such as concentrations of reactants or receivers. Global characteristics are modelled in the cost models for each cost category using inter-country correlations based on historical data. Variance in location of emission, receivers, etc. is partly captured by fitting a distribution for each country. It is assumed for the sake of calculation the random variables for each unit of quantity emitted in a country are identical to capture common features, that is identically distributed and perfectly correlated.

Central limit arguments and dependency in unit quantity changes

Some environmental damage cost references argue that dependency can be ignored. They treat the marginal cost for each unit of quantity as identically distributed and independent[1]. The 2018 CE Delft environmental prices EU-28 handbook briefly addresses uncertainty in Annex C.2 [6]. A centrality argument was made for total damages in the CE Delft handbook. That is, if the environmental cost per unit of pollution is incurred frequently and the variation in the valuation uncertainty was independently allocated to the occurrences, e.g. 10000 independent unit emissions in total of 10000 units of emission, then the central limit theorem applies and the total costs to society are approximated well by the mean environmental price for that pollutant times the total quantity of emission.

The centrality argument is unlikely to hold for each unit emission given the common mechanisms in the impact pathway. For damage costs of GHG, as an example, it is unclear why emissions from production in The Netherlands in 2020 get reabsorbed into the global economy in a future world where a tonne of carbon dioxide had 2020 USD PPP 20/tCO2 impact, and emission from production in Germany in 2020 gets reabsorbed into the global economy in a future world where a tonne of carbon dioxide had 2020 USD PPP 200/tCO2 impact. The emissions from The Netherlands and Germany in the same year contribute to the same causal mechanism in a globally common

[1] Units do not need to be independent for a central limit argument to hold, only that the correlation is sufficiently weak.

atmosphere for the same future world. Their impact is drawn from the same lottery not different lotteries for the social cost of carbon.

More sophisticated risk modelling could make assign weaker correlations to unit emissions. SPIQ-FS rejects the premise that central limit arguments will remove variance from total damages and errs on the side of greater dependence by assuming unit quantity changes are identically distributed and perfectly correlated. Perfect correlation also simplifies the modelling of uncertainty to 2212 random variables since each unit quantity change in an impact quantity and in a country is treated as identical. Treating each unit in $n_{i,j}$ units of quantity change, where i indexes the 14 impact quantities and j indexes 158 countries, as a random variable requires an ensemble of hundreds of thousands, or millions, of random variables.

Joint sampling problem in SPIQ-FS

Obtaining estimates of total costs in SPIQ-FS that acknowledges dependency in marginal costs requires

1) A representation of the expected dependency in the marginal costs as a joint 2212-variate random variable with 2212 marginals corresponding to the 2212 marginal costs[2]
2) A joint sample of the 2212-variate random variable reflecting that dependency

We describe possible methods to generate a joint sample of a multi-variate random variable with thousands of (log-normal) marginals, given the information provided by the SPIQ-FS model and the considerations of interaction between the impact quantities provided in the documentation in Annex A. We use the generic SPIQ-FS dataset for concrete numbers on the description of the process.

Quantities in the generic SPIQ-FS dataset

The individual cost models in SPIQ-FS produce joint samples within the cost categories, see Table 1 column 1 for cost categories. Joint samples are produced for the 3 GHG social costs as marginals by the GHG cost model. Joint samples are produced across the 4 Nitrogen marginal damage costs as marginals by the Nitrogen costing model. The joint samples from the GHG cost model and the joint samples from Nitrogen cost model are not jointly joint however and are not sampled considering interactions between N and GHG impact pathways.

Table 1: The SPIQ-FS version 0 generic dataset has marginal damage costs of 14 impact quantities in 6 cost categories. The 14 marginal damage costs are calculated for each of the 158 countries in the dataset.

Cost Category	Unit	No.	Marginal Cost of
GHG Emissions	Metric ton	1	CH4
GHG Emissions	Metric ton	2	N2O
GHG Emissions	Metric ton	3	CO2
Water use	Cubic metre	4	Blue water withdrawal
Land use	Ha	5	Forest Habitat Loss
Land use	Ha	6	Forest Habitat Return
Land use	Ha	7	Other Habitat Loss
Land use	Ha	8	Other Habitat Return

[2] The univariate distribution obtained by integrating over all but one of the random variables in a multivariate distribution is called a marginal of the multivariate distribution. The multi-variate distribution is the joint distribution of the marginals. If the multi-variate distribution involves *n* joint random variables, then there are *n* marginals. The term marginal of a multi-variate distribution is distinct from the term marginal applied to a damage cost. The *m*-th marginal damage cost with respect to a vector of *m* quantities is the *m*-th partial derivative of the damage function on a domain in *m* dimensional Euclidean space.

Nitrogen Emissions	Kg N-weight	9	NH3 emission to air
Nitrogen Emissions	Kg N-weight	10	NOx emission to air
Nitrogen Emissions	Kg N-weight	11	NO3- leached to groundwater
Nitrogen Emissions	Kg N-weight	12	Nr run-off to surface water
Undernourishment	Person	13	Undernourishment
Poverty	Person	14	Average income shortfall from $1.90 a day (2011 PPP) (ppl)

Therefore, the collection of 2212 sets of samples together are not univariate, they can be considered as 6 joint distributions of order (3x158,1x158,4x158,4x158,1x158,1x158). We seek to specify the dependency structure across cost categories to turn the 6 multivariate random variables of order (3x158,1x158,4x158,4x158,1x158,1x158) into a full multivariate random variable of order 2212.

Joint sampling approaches

Pearson's correlation coefficient is used to specify cross cost category dependence. It has a simple interpretation as explanation of variation and requires only pairwise cross category observations. For this reason, it is useful for a low number of literature sources and observations of explanation of variance from linear regression often made in the scientific literature.

Many joint observations allow examination of a scatterplot and fitting of a copula directly. Joint observations across all the cost categories are rare in data and literature, however.

Given the information from SPIQ-FS of 2212 random variables that are jointly sampled with cost categories, one joint sampling approach is to generate a 2212 x 2212 correlation matrix, where blocks within the correlation matrix for cost categories are already approximated by the correlation matrix of the joint samples within cost categories. This provides a block diagonal matrix of order (3x158,1x158,4x158,4x158,1x158,1x158) for the GHG, Water, Land-Use, Nitrogen, Undernourishment and Poverty costs category block respectively. Cross cost category dependence is encoded in the choice of off-diagonal blocks.

The off-diagonal blocks are assessed in Annex A documentation and assigned an expert assessment of strength of correlation as summarised by the block correlation matrix in Table 4. With the diagonal and off-diagonal blocks of the 2212 x 2122 correlation matrix specified, one could attempt to determine the 2212-multivariate joint distribution that has the specified correlation matrix and the 2212 individual damage cost distributions as marginals.

A second approach is to form a joint sample with correlation matrix that approximates the 2212 x 2212 correlation matrix with the given marginals by reordering blocks of joint samples. Re-ordering blocks of samples in the same cost category retains both the marginals and the block diagonal correlations, since the sort order of the existing samples within categories is not changed.

We review methods available for both the first and second approach. In practice both are computationally intensive and both approaches resort to approximations. For log-normally distributed marginals we ended up using a block correlation sort method described below.

Joint distribution of arbitrary marginals for a given correlation matrix

Given marginals and a correlation matrix, there are several approaches to construct a joint distribution of marginal costs.

Specifying marginals and a correlation structure sets up a minimal information problem. With the information that is known, an entropy functional can be places on the set of all joint distributions with the specified marginals and correlation matrix. A unique minimum of the entropy function on such a set would be the natural choice of joint distribution to use. Maximal entropy joint distributions given only marginals and correlation matrices are not computational efficient to solve for a 2212-multivariate random variable however, though approximations exist [9]. This approach was not used due to the computation resources required and the resources required to efficiently code approximation algorithms.

An order-statistics method from [7] is based on fitting univariate marginals and a given correlation matrix. The sampling method is efficient, in that solving inverse problems for conditional distributions is not required. The joint distribution is a special case of a Bernstein copula, and it minimises the distance as measured by chi-squared divergence between joint distributions with the given correlation matrix and the joint distribution assuming all random univariates are independent. The sampling method is, therefore, a conservative estimate for error (and hence risk) due to joint correlation. Though the method is efficient in that it involves reordering, to solve in practice optimisation is required on the space of orderings of the samples. As the dimensions increase, the space of orderings of samples blows up. It becomes a more difficult problem to then effectively choice minimising directions in the high dimensional space, and to avoid local minima. This approach was not used due to the computation resources required and the resources required to efficiently code approximation algorithms. This method was used as the second default method for the block sort order method, where block sort ordering reduces the dimension of the sort-ordering space.

An alternative, and similar, sampling method could be derived from [8] where the assumption on the cross sectional structure (the conditionals) is simplified. Linear conditionals make sampling efficient.

Note that, under the conservative sampling method from [7] where distance from a joint distribution with independent marginals is minimised, zero correlations are realised as independence. The block correlation method also uses this assumption. Approximating zero-correlations as independence in marginal costs results in an underestimate of variance in total costs. Zero cross-correlations would not have the same result as independence: the marginal costs are not being summed, the marginal costs are weighted by quantities and then summed into a total (dot product).

Joint distribution of log-normal marginals for a given correlation matrix

The parametric output of SPIQ-FS assumes that marginal costs are well approximated by log-normal distributions. This can be exploited by joint sampling methods based on a Gaussian copula, or a multivariate normal distribution. Given log-normal marginals and a correlation matrix, the covariance matrix of the assumed joint distribution can be transformed exactly under the log-normal function. The problem is reduced to log-transformed marginal random variables with the log-transformed covariance matrix. Log-transformed log-normal marginals are normal marginals, and the covariance matrix can be reduced by the standard deviations of the normal marginals into a correlation matrix.

Hence the problem is reduced to finding the joint distribution with given normal marginals and correlation matrix. The ability to reduce the joint sampling problem to normal marginals is another rationale to approximate marginal costs by log-normal distributed random variables. Some of the methods in the previous section can now be increased for their computational efficiency due to the symmetry of the normal distribution, and the fast algorithms for inverse to the cumulative distribution.

The most direct method for solving the joint sampling problem given normal marginals is to test the log-transform covariance matrix. If it positive semi-definite, then the maximal entropy solution is the multi-variate normal distribution with the given positive semi-definite covariance matrix. The multivariate normal distribution is very efficient to sample, and the samples can be exponentiated to solve the original joint sampling problem.

Unfortunately, high dimensional covariance matrices with mixed off-diagonal block terms are seldom positive-definite. Approximations to a positive definite matrix involve a difficult trade-off since they involve adjustments to the standard deviation of the marginals. The adjustment might no be large in log-transformed space, but the uncertainty can be doubled or greater when the adjustment is transformed into the standard deviation of log-normal marginals. This was the case when working with correlation matrices in studies involving the SPIQ-FS generic dataset, and the approach of approximation if the covariance matrix was not positive semi-definite was rejected.

Even with the additional computation efficiency of working with normal marginals, the general approaches mentioned in the last section were not used for a 2212-multivariate random variable due to the computational resources required and the resources required to efficiently code approximation algorithms.

The method chosen for joint sampling in SPIQ-Fs calculations reduced the dimension of the joint sampling required by exploited the fact that samples from SPIQ-FS were already sort-order as joint samples within cost categories.

Block correlation sort ordering

Block correlation sort-ordering exploits the assumptions of SPIQ-FS output that samples from SPIQ-FS were already sort-ordered as joint samples within cost categories, and that marginals were log-normally distributed. Let A_0 denote the given 2212 x 2122 correlation matrix with the diagonal and off-diagonal blocks specified. Assume SPIQ-FS output provides 1000 samples of 2212 marginal cost items that are jointly sampled in cost categories. Let S denote the 2212x1000 matrix of samples. Let $S_{(o)}$ denote the 2212x1000 matrix such that $S_{(o)}(i,j) = k$ if $S(i,j)$ is the k-th smallest sample in the i-th row.

The method involves:

1) Take 1000 2212-univariate log-normal samples using the 2212 mu and sigma parameters for each marginal cost item. This provides a 2212x1000 matrix C_U with rows each cost item and column the sample number.
2) Sort ordering the univariate log-normal samples in cost categories in the same sort-order as the SPIQ-FS samples to obtain a 2212x1000 matrix C_J. Recall the cost items are categories into the blocks (3x158,1x158,4x158,4x158,1x158,1x158). So, the samples in the first 3x158 rows of the 2212x1000 are rearranged into the sort order of the SPIQ-FS output samples for GHG, the rows from 3x158+1 to 3x158 +1x158 are sort ordered to match the SPIQ-outputs samples for blue water, etc. That is, if $S_{(o)}(i,j) = k$ then $C_J(i,j)$ is the k-th smallest sample in the i-th row of C_U.

Re-ordering does not change the marginals, so each row in C_J is still a sample of the marginal distributions. At this stage the 2212x1000 matrix C_J is a joint sample within cost categories – it solves the joint sampling problem for the given marginals and a correlation matrix that has block diagonals from Table 3 and zero off diagonals. The next steps approximate joint sampling methods for non-zero off diagonals. Let b_c, $c = 1, \dots, 6$, denote the partition of {1,..,2212} into cost categories (the blocks).

3) Each block is now assigned a sort order. The first 3x158 rows correspond to samples for the cost category of GHG. The columns in the 474x1000 matrix are reduced to a 1x1000 vector which indicates the ascending order of the columns for lowest to highest. The easiest way to assign a partial order to each column is to sum the values in the column. Doing this for each cost category results in a 6x1000 matrix containing the sort order of each column in the cost-category blocks. That is, define the matrix $X_{(o)}(c,j) = k$ if $\sum_{i \in b_c} S_{(o)}(i,j)$ is the k-smallest element as j ranges from 1 to 1000. The value
4) Reordering the elements in the rows of the 6x1000 matrix $X_{(o)}$ to a new matrix X represents changing the sort order of a column in each block. That is, if $X(c,s) = k$, then column *j* in the matrix $C_J(b_c,:)$ is moved to be column s in a new matrix $C_X(b_c,:)$. Forming the matrix C_X from vertically stacking the blocks $C_X(b_c,:), c = 1 \dots 6$, produces a 2212x1000 where each of the blocks is sort-ordered to match the matrix X.
5) The joint sample structure in each block is unchanged, but the block sort-order now introduces correlations. Let $A_X = corr(C_X)$ be the 2212 x 2212 correlation matrix of the 1000 joint samples of 2212 random variables given by C_X.
6) Find X such that A_X approximates A_0.

A sort-ordering method always preserves the marginals, and in this case always preserves the dependency structure within cost-categories as determined by a SPIQ-FS cost model. The matrix X found in step 6 provides a joint sample C_X of the marginals, with exact joint sampling within cost categories, and an approximation of the desired correlations across cost categories.

Step 6 is the computational challenge. How to reorder the elements in the 6x1000 matrix to minimise the distance between A_X and A_0. A reorder of a 6x1000 matrix is relatively fast to perform in a scientific programming language like R or Python that has logical indexing and linear algebra packages, as is the reorder of the blocks in the 2212 x 1000 matrix to form C_X. Determining the correlation matrix of C_X is also relatively fast. Conceptually, optimisation is possible due to the fast calculation of the objective function. The challenge is that the space of reorders of a 6x1000 matrix is still large (~1.37×10^{15}), and the choice of a distance metric on matrices by which to minimise the distance between A_X and A_0. Future versions of the SPIQ-FS calculation code could exploit global optimisation methods in permutation problems, but this optimisation method was not used because of the resources required to implement it.

A less accurate but efficient approximation method involves assigning samples from each column of the cost category block in step 3 to represent the cost category itself as a random variable. Let U be a 6x1000 matrix such that $U(c,:)$ is a 1x1000 uniform sample of integers in b_c. Let D be the 6x1000 matrix $D(c,j) = C_J(U(c,j),j)$. Each row in D represents a mixture of the random variables in block b_c. By construction the 6x6 block correlation matrix B in (Table 4) is the desired correlation matrix of the mixtures. By construction the rows of D are independently sampled and $corr(D) \sim 0$. In this approximation method we replace 6) by taking a joint sample D_J such that $corr(D_J) \sim B$. Then set $X(c,s) = k$, if $D(c,s)$ is the k-smallest element as s ranges from 1 to 1000. The 2212x1000 matrix C_X of the original marginals formed from X by step 5) is taken as a joint sample of the marginal cost distributions.

By weakening the correlation constraint to the mixtures of each block, the order-statistics method from [7] can be used on the 6 mixture marginals with correlation matrix B. This is the default method used in SPIQ-FS.

The fact that the marginals are log-normally can be exploited further. Joint sampling in SPIQ-FS is implemented to explore risk, and so is most concerned with the effect of correlations on the tail of the total cost distribution. The tail of a mixture of log-normal distributions is approximated reasonably well by a log-normal distribution. A log-normal fit to parameters of the mixture distribution of each cost category associates the samples in D to 6 log-normally distributed marginals that can be jointly sampled to match the correlation matrix B. When log-transformed to normal marginals and a covariance matrix B', SPIQ-FS tests whether the covariance matrix B' is positive semi-definite. If so, the matrix X is obtained as the sort order of the joint samples from the multivariate normal distribution corresponding to the normal marginals and covariance matrix B'. This method is computationally fast when applicable and has proven a feasible method in SPIQ-FS studies with 4000+ marginal cost items. The joint samples it produces are a reasonable approximation of off-diagonal correlation terms for SPIQ-FS builds and allows the exploration of additional risk from cross-category dependency in the marginal damage cost items.

Block correlation matrix

The follow block correlation matrices are derived from information in Annex A.

Scientific literature was reviewed in Annex A, and through a consideration of dependency in the impact pathways for a unit change in quantities associated to food system impacts a correlation coefficient was assigned pairwise using an expert-based assignment of weak, moderate, or strong dependency. A numerical value for the correlation coefficient was assigned to the weak, moderate and strong qualitative assessments (Table 2).

Table 2: assignment of quantitative values to assessment of the strength of dependency

Dependency	**Pearson correlation**
Strong negative	-0.8
Moderate negative	-0.5
Weak negative	-0.2
None	0
Weak positive	0.2
Moderate positive	0.5
Strong positive	0.8

It was argued above that Pearson correlation coefficients are easier for experts to assess for multiple random variables and based on explanation of variance in literature. For example, interaction is often stated in explanation of variance terms such as soil erosion increases nitrogen and phosphorous run-off from fertiliser application by 36%, [10]. Or that DALYs from chronic and hidden hunger under climate change are estimate to increase by 10%, [11].

Diagonal blocks

Marginal costs in the SPIQ-FS in the generic dataset break up into (3x158,1x158,4x158,4x158,1x158,1x158) items for the GHG, Water, Land-Use, Nitrogen, Undernourishment and Poverty costs categories respectively.

Without any dependency across cost categories, the 2212 x 2212 correlation matrix of a joint sample can be obtained as a block diagonal matrix. The blocks are a 474x474, 158x158, 624x624, 624x624, 158x158, 158x158 in size, respectively. The blocks are generated directly from the correlation matrices of SPIQ-FS cost models for each cost category.

Table 3 summarises the dependence within cost categories from the modelling described in Annex A.

Table 3: dependence within cost categories

Category	**Dependence**
GHG emissions	Perfectly correlated samples across country within each gas. No correlation across gases.
Blue water use	Correlation between countries arises from historical variation in food producer prices
Land use	No correlation between countries. No correlation between biomes.
Nitrogen emissions	Correlation between NH3 and NOx from pathway interactions. Correlation between countries based on EASIUR model of US counties. No correlation with run-off of leaching, no correlation across countries.
Undernutrition	No correlation across countries.
Poverty	No correlation across countries.

Off diagonal blocks

Marginal costs in the SPIQ-FS in the generic dataset break up into (3x158,1x158,4x158,4x158,1x158,1x158) items for the GHG, Water, Land-Use, Nitrogen, Undernourishment and Poverty costs categories respectively.

To capture dependency across cost categories, the off-diagonal blocks in the 2212 x 2212 correlation can be specified. There are 6 cost categories and 15 off-diagonal blocks in the upper triangular corner of the correlation matrix. The blocks are 474x158, 474x624, 474x624, 474x158, 474x158, 158x624, 158x624, 158x158, 158x158, 624x624, 624x158, 624x158, 624x 158, 624 x 158, 158 x 158 in size, respectively, and are specified by the matrices of those sizes with each entry given by the coefficient in Table 4.

The correlations were assessed according to the dependency scale in Table 2.

Table 4: the 6x6 block correlation matrix B in colour used in SPIQ-FS studies with joint sampling of the SPIQ-FS generic dataset version 0. Greyed correlation terms in an 8x8 matrix are presented for the version 1 dataset.

		Environmental changes				Socio-economic Changes	Health changes		
		GHG emission	Water use/ depletion	Land use	Nr emission	Poverty	Under-nourish-ment	NCDs	Obesity
Environ-mental changes	GHG emission		+0.5	+0.5	0	*+0.5*	*+0.2*	+0.2	+0.2
	Water use/ depletion	+0.5		+0.2	+0.2	+0.2	+0.2	0	0

	Land use	+0.5	+0.2		+0.5	-0.2	-0.2	0	0
	Nr emission	0	+0.2	+0.5		0	-0.2	0	0
Socio-economic Changes	Poverty	*+0.5*	+0.2	-0.2	0		*+0.5*	+0.5	+0.5
Health changes	Under-nourish-ment	*+0.2*	+0.2	-0.2	-0.2	*+0.5*		+0.2	0
	NCDs dietary intake	+0.2	0	0	0	+0.5	+0.2		+0.2
	Obesity	+0.2	0	0	0	+0.5	0	+0.2	

Autocorrelation over time

The SPIQ-FS generic dataset version 0 estimates marginal damages costs for a change of quantity in 2020. Marginal damages for quantity changes in 2030, and 2040, may be correlated over time. Later studies using SPIQ-FS will factor autocorrelation into the joint sampling to assess economic risk.

Sensitivity analysis

Sensitivity analysis: 1) no correlations, 2) use the estimate matrix below, 3) sign of correlation (-1 or +1).

The sources for cross category dependency listed in Annex A are incomplete. They are indicative of second order relationships that ideally would be modelled directly, for example coupling between the carbon and nitrogen cycles. They are derived from expert assessment and applied indiscriminately across countries. The purpose of joint sampling based on cross category dependency is to risk in total economic damages resulting from changes in environmental pollutants and resources, dietary intake, and poverty. Three sets of correlations (Table 5) are used to explore the joint nature of environmental, health and social conditions on total economic costs: no correlation, an expert assessed set of correlations (Table 4), and perfect correlation.

These three representations in risk from joint effects can be used to contrast ignoring joint effects of environment and health with the case where higher than expected environmental damage costs will always coincide with higher than expected health damage costs. The middle, expert-derived, set of correlations in Table 4 represents a best estimate of the additional economic risk from joint effects.

Table 5: sensitivity tests for cross category dependency

Category	Dependence
No correlation	All block diagonal terms in the block correlation matrix B in Table 4 are set to 0
Expert assessed	Matrix B in Table 4 is used as the block correlation matrix

Full correlation	All block diagonal terms in the block correlation matrix B in Table 4 are set to 1.

Another sensitivity analysis specifically for risk, not conducted, would examine higher cross-correlation at high values. Unless the sampling procedure is known, joint sampling in software packages can underestimate joint tail risk [12]. Joint sampling for higher tail risk would replace the Bernstein copula with a copula with higher correlation in tails of the univariate marginals. If evidence from scatterplots, or biophysical arguments, shows a stronger correlation at higher values, the so-called heavy-right tailed copula would be more representative of higher risk [13].

References

1. Lord, S., *Estimations of marginal social costs for GHG emissions*, in *SPIQ-FS Dataset Version 0*. 2021, University of Oxford.
2. Lord, S., *Estimation of marginal damage costs from reactive nitrogen emissions to air, surface waters and groundwater*, in *SPIQ-FS Dataset Version 0*. 2021, University of Oxford.
3. Lord, S., *Estimation of marginal damage costs from water scarcity due to blue water withdrawal*, in *SPIQ-FS Dataset Version 0*. 2021, University of Oxford.
4. Lord, S., *Estimation of marginal damage costs for loss of ecosystem services from land-use change or ecosystem degradation*, in *SPIQ-FS Dataset Version 0*. 2021, University of Oxford.
5. Paulus, E. and S. Lord, *Estimation of marginal damage costs from consumption related health risks*, in *SPIQ-FS Dataset Version 0*. 2022, University of Oxford.
6. de Bruyn, S., et al., *Environmental Prices Handbook EU28 Version*. 2018, CE Delft: Delft, The Netherlands.
7. Baker, R., *An order-statistics-based method for constructing multivariate distributions with fixed marginals.* Journal of Multivariate Analysis, 2008. **99**(10): p. 2312-2327.
8. Cuadras, C.M., *Probability distributions with given multivariate marginals and given dependence structure.* Journal of Multivariate Analysis, 1992. **42**(1): p. 51-66.
9. MacKenzie, G.R., *Approximately maximum-entropy multivariate distributions with specified marginals and pairwise correlations*. 1994, ProQuest Dissertations Publishing.
10. Cordell, D., J.-O. Drangert, and S. White, *The story of phosphorus: Global food security and food for thought.* Global Environmental Change, 2009. **19**(2): p. 292-305.
11. Sulser, T.B., et al., *Disability-adjusted life years due to chronic and hidden hunger under food system evolution with climate change and adaptation to 2050.* The American Journal of Clinical Nutrition, 2021.
12. Haas, C.N., *On Modeling Correlated Random Variables in Risk Assessment.* Risk Analysis, 1999. **19**(6): p. 1205-1214.
13. Ibragimov, R. and A. Prokhorov, *Heavy Tails and Copulas: Topics in Dependence Modelling in Economics and Finance*. 2015, Singapore: World Scientific. 304.