A Bayesian Method to Reconcile Conflicting Input-Output Data DRAFT VERSION

Rodrigues, João F. D.*

April 15, 2011

Abstract

At the heart of a CGE model lies an input-output (IO) model, i.e., the set of transactions taking place in the economy in the base-line year. For example, the GTAP 7 database consists, to some extent, to a large mult-regional IO model for the year 2004. The compilation of this large data set has to deal with the problem of reconciling potentially inconsistent data. Currently this procedure is performed using simple simple methods (RAS or least square minimization) and a good deal of expert knowledge. In this paper we derive a Bayesian method to obtain a consistent and fully determined set of posteriors when the priors are inconsistent and there is a set of known topological constraints of arbitrary structure. The expression relating priors (inconsistent data) and posteriors (reconciled data) is obtained by the application of the maximum entropy principle and its properties are discussed. In general, the posteriors form a vector of multivariate truncated Gaussian random variables. This distribution approaches the Gaussian distribution in the limit of low uncertainty and the exponential distribution in the limit of high uncertainty. Using invariance considerations we derive a numerical iterative implementation that takes the form of a generalized least squares method. We present a real application (60x60 symmetric IO table) to illustrate the behavior of the method and to compare its performance against conventional methods. The numerical implementation is slightly more complex than conventional methods. However, the adjustment from prior to posterior is more strongly correlated to

*Instituto Superior Técnico, IN+, DEM, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal. E-mail: joao.rodrigues@ist.utl.pt.

the initial uncertainty of the data points in the Bayesian method. If the data set is large, it is necessary to store data and perform algebra in sparse format.

KEYWORDS: Input-Output (IO) Analysis; Bayesian approach; maximum entropy principle (MEP); conflicting data; uncertainty; truncated Gaussian distribution.

1 Introduction

Input-Output (IO) Analysis is the field that deals with the compilation of macro-economic transaction data in IO tables, and the use of those tables to compute indirect effects such changes in employment or carbon emissions embodied in final consumption (Miller and Blair, 2009).

In the compilation of an IO table it is often the case that, for a particular element, there is only partial information in the form of row and column sums and the value of that element in a previous year. Many methods exist for the IO estimation problem, such as the biproportional or least-squares families (Mesnard, 2004). Within this subject, there has been recent attention devoted to the problem of reconciling conflicting constraints (Lenzen et al., 2009; Rampa, 2008). These problems are usually addressed by taking into account only the best (available) guess of a given value, and not its uncertainty. The uncertainty of source data is sometimes but not often reported in IO analysis (Lenzen, 2001; Lenzen et al., 2010; Oosterhaven et al., 2008).

The goal of the present paper is to contribute to this literature by presenting a method that can reconcile inconsistent entries in an IO table, taking into account conflicting information of arbitrary form and the uncertainty of the source data. We achieve this goal by taking a Bayesian approach (Jaynes, 1983) to uncertainty in IO analysis (Weise and Woger, 1992).

The elements of an IO table are aggregate economic transactions, non-negative quantities of which a best guess and an associated uncertainty are known. Under these conditions, the Maximum Entropy Principle (MEP) of Jaynes (1957) imposes that IO transactions are random variables of truncated Gaussian distribution. In a Bayesian context, the problem of reconciling conflicting constraints is addressed by applying the MEP to all IO transactions, which are connected by topological constraints (row and column sums). We discuss the connection between the Bayesian approach and the conventional biproportional (RAS) and least square (LS) methods, which can also be derived using the MEP. The reconciliation algorithm we propose turns out to combine features of the recently proposed KRAS (Lenzen et al., 2009) and SWLS (Rampa, 2008) methods. We make use of invariance considerations

(Jaynes, 1973) to derive the sequence in which the MEP algorithm is applied. After deriving the Bayesian theory of the reconciliation of conflicting data we present a simplified computational algorithm, and illustrate its application to a national 60x60 symmetric IO table.

The structure of the paper is as follows. In Section 2 we review current methods for IO estimation and present the background Bayesian theory. In Section 3 we derive the Bayesian theory of IO uncertainty. In Section 4 we report the numerical algorithms to determine the posteriors (reconciliation of conflicting constraints). In Section 5 we present examples, to illustrate the proposed methods and to compare them with conventional methods. Section 6 concludes. An Appendix reports useful numerical approximations for the truncated Gaussian distribution.

2 Review

2.1 Estimation in IO Analysis

Estimation occurs in IO Analysis under different circumstances, of which the most thoroughly explored, is the case of known numerical constraints (row and column sums for the current year) and an initial guess (from a previous year) for the economic transaction.

The most popular strategy to address this problem is the use of biproportional methods in which the original matrix is iteratively multiplied by a left and a right perturbation diagonal matrices, until the row and column sums are satisfied. The first such technique to be used in IO analysis was the RAS method (Stone et al., 1942), which has been extended in many ways, as reviewed in Lahr and Mesnard (2004). An important step was taken by Bacharach (1970), which noted hat RAS is the solution of a maximum entropy (MEP) problem, the minimization of relative entropy (Kullback and Leibler, 1951), in which a transaction is viewed as a probability, and thus the IO table as a whole is viewed as a probability distribution.

A recent development of a transaction-as-probability method is Lenzen et al. (2009), whose purpose is to solve conflicting constraints, and which works by first running an RAS-like method adjusting only transactions, and then, when no further improvement can be performed, by adjusting the constraints. This adjustment is additive and proportional to the product of that constraints' initial uncertainty and its current inconsistency. One characteristic of this method, which follows from the transaction-as-probability approach, is that there is no way to use information on the relative uncertainty of the transactions in the adjustment process (since it makes no

sense to talk about the uncertainty of a probability). In fact, there is no theoretical sound technique to reconcile constraints in such a case, although a combination of entropy maximization for the unknowns and least square (LS) minimization for the constraints has been proposed (Lieu et al., 1987; Lieu and Hicks, 1994).

But there are alternative formulations to entropy maximization (Jackson and Murray, 2004) and one such popular approach is least square (LS) minimization. Rampa (2008) presents a subjective weighted LS method, in which the uncertainty of each constraint is used as a weight, and the practitioner should specify subjectively the uncertainty of constraints for which no baseline information is available. This paper introduces an important concept into the problem of IO estimation: the idea of a topological constraint, that links the numerical constraint and the aggregated transactions, in such a way that both can be adjusted simultaneously. The topological constraints are rows of an aggregation matrix, which can have an arbitrary shape - as opposed to strict row and column sums or more complex intermediary cases (Gilchrist and Louis, 2004).

The choice of the weights and uncertainties in Rampa (2008) is arbitrary. We consider that there should be some scope for the practitioner to use his knowledge about the quality of the data, but also that his discretion should be bounded by plausibility, and a set of standard assumptions should be provided, for the use of the less skilled IO practitioner.

As Rampa (2008) shows, LS minimization is a second order Taylor approximation to the maximum entropy maximization and, therefore, its results should not be very different from RAS. However, in LS the objective function is symmetric around the initial guess, and thus there is no guarantee of sign preservation, an issue that is addressed by Junius and Oosterhaven (2003) in the context of biproportional methods. Another difference is that LS is direct while RAS is an iterative method.

In this paper we shall contribute to this literature by providing a method to compile an IO table that can take into account inconsistent priors, aggregations of arbitrary shape and that uses the uncertainty of the source data to reconcile conflicting information.

2.2 Approach

According to the Bayesian paradigm, a probability is a degree of belief about the likelihood of an event, and should reflect all relevant available information about that event (Lee, 1989). Therefore, an unknown probability distribution should be assumed to have the minimum information (or maximum entropy) that is consistent with the available information (Jaynes, 1983). The en-

tropy of a discrete probability distribution $\{p_i\}_{i=1}^N$ is $H = -\sum_{i=1}^N p_i \ln p_i$, and the available information in our case takes the form of the j-th moment of the distribution $\sum_{i=1}^N i^j p_i = M_j$. If a prior probability distribution $\{\tilde{p}_i\}_{i=1}^N$ is available, then the posterior probability distribution is obtained by minimizing relative entropy $H = \sum_{i=1}^N p_i \ln p_i / \ln \tilde{p}_i$, subject to the available information in the form j-th moments, $\sum_{i=1}^N i^j p_i = M_j$, through the method of Lagrange multipliers (Shannon, 1948).

Entropy maximization is familiar in IO analysis. However, what is perhaps not so familiar is the context in which we shall apply maximum entropy. Following Weise and Woger (1992) we shall treat every entry in an IO table, which represents an economic transaction, as a non-negative random variable whose expectation is the best guess and whose standard-deviation is the uncertainty. This is different from the standard approach in which a transaction is a probability and the whole IO table is a probability distribution.

Transactions are connected to one another through topological constraints, or equations that state how transactions sum up. In a Bayesian context, the numerical constraints of biproportional methods (row and column sums) are just like other transactions, that can be naturally adjusted (if we so wish).

All input data to the estimation problem consist in the properties (expectation and standard-variation) of the prior distribution of the transactions (if they are known) and the aggregation rules of the topological constraints. The prior distribution of the set of transactions is obtained using additional considerations. Recall that according to the Bayesian paradigm all relevant information should be used, and the structure of the system under consideration may also be relevant information. An invariance consideration is a method to make use of information that does not conform to the MEP paradigm.

We briefly look at Jaynes' solution to the Bertrand paradox to show how invariance considerations work. Consider an equilateral triangle inscribed in a circle. Suppose that a chord of the circle is chosen at random. What is the probability that the chord is longer than a side of the triangle? This question poses a paradox because there are different ways of chosing a chord at random, leading to different distributions. Jaynes (1973) solved this paradox by noting that the in the original problem there is no reference to the position or size of the circle, and thus the resulting distribution should be invariant to the rescaling or displacement of the circle. Imposing invariance solves the paradox, leading to a unique solution.

In the context of IO analysis, geometric transformations do not make sense, since we are not dealing with spatial objects. However, it makes sense to talk about the information quality of the data. In the table update

problem, for example, the initial guess from the previous year is of lower quality than the row and column sums, which are known for the current year. The table update itself is a transformation of the data, in which the topological transactions incorporate information from the initial best guesses. In order to determine the missing priors we consider that data of higher information quality should remain unaffected if combined with data of lower information quality in the topological constraints.

3 The maximum entropy principle

3.1 The prior distribution

Assume that all the information we possess about a nonnegative IO quantity T (an aggregate economic transaction), is that it takes values in the range $[0, t_{\max}]$, and we may know or not its *best guess*, μ, and its *uncertainty*, σ. In a Bayesian framework (Weise and Woger, 1992), T is a random variable, with probability density $p(t) = \Pr(T = t)$, whose best guess is the expectation, $m = \mathrm{E}(T)$, and whose uncertainty is the standard deviation, such that $s^2 = \mathrm{Var}(T) = \mathrm{E}(T^2) - \mathrm{E}(T)^2$. We use the notation $\mathrm{E}(f(T)) = \int_0^\infty f(t)p(t)dt$.

The Maximum Entropy Principle (Jaynes, 1983), states that the *posterior* distribution of T, $p(t)$, is obtained by maximizing the differential entropy of the distribution subject to the *prior* distribution, $\tilde{p}(t)$, and to the known constraints:

$$H = -\int_0^{t_{\max}} p(t)\frac{\ln p(t)}{\ln \tilde{p}(t)}dt + \lambda_0(\mathrm{E}(1) - 1) + \lambda_1(\mathrm{E}(T) - m) + \lambda_2\left(\mathrm{E}(T^2) - \mathrm{E}(T)^2 - s^2\right). \tag{3.1}$$

The first term in the right hand side of Eq. 3.1 is the differential entropy of the unknown distribution. We assume the prior distribution $\tilde{p}(t)$ is uniformly distributed in the range $[0, t_{\max}]$. The remaining terms in the right hand side of Eq. 3.1 are the set of known constraints: the *zero-th order* constraint is the normalization, the *first order* constraint is the expected value and the *second order* constraint is the variance. The λ's are the respective Lagrange multipliers. Differentiation of Eq. 3.1 with respect to $p(t)$ leads to:

$$0 = -\frac{\ln p(t)}{\ln \tilde{p}(t)} - 1 + \lambda_0 + \lambda_1 t + \lambda_2\left(t^2 - 2mt\right). \tag{3.2}$$

Now let us consider three cases. If we only know the zero-th order constraint, $\lambda_1 = \lambda_2 = 0$, and Eq. 3.2 leads to a uniform distribution,

$p(t) = 1/t_{\max}$. That is, the zero-th order maxent posterior is identical to the prior: we have introduced no information and, as expected, no further information was gained.

If we also know the first order constraint, only $\lambda_2 = 0$, and Eq. 3.2 leads to a truncated exponential distribution, $p(t) = (\lambda e^{-\lambda t})/(1 - e^{-\lambda t_{\max}})$. The parameter λ is determined by m.

Finally, if we also know the second order constraint, we need to solve the full Eq. 3.2 and obtain a truncated Gaussian distribution:

$$p(t) = \frac{1}{Z_0} \frac{1}{\sqrt{2\pi\tilde{s}^2}} \exp\left(-\frac{(t-\tilde{m})^2}{2\tilde{s}^2}\right), \tag{3.3}$$

with the substitution $2\lambda_2 = 1/\tilde{s}^2$ and $\lambda_1 - 2m\lambda_2 = -\tilde{m}/\tilde{s}^2$, where Z_0 is an integration constant. Note that since this distribution is truncated, the Gaussian *parameters* $\tilde{m}$ and $\tilde{s}^2$ are *not* the *observable* expectation and variance of the distribution, m and s^2.

The forms of the zero-th, first and second order maxent solutions are well known (Cover and Thomas, 1991) but the following observations are not. First, although the solutions of different orders are qualitatively different, there is a smooth transition between them. A first order solution for which $m = t_{\max}/2$ has uniform solution, and therefore is equivalent to the zero-th order solution. In essence, having an expectation that lies exactly in the middle of the range is equivalent to not knowing that expectation.

Now consider the following expansion of Eq. 3.3, when $\tilde{m} \ll 0$ and $\tilde{s} \ll 0$:

$$\begin{aligned} p(t) &= C_1 \exp\left(-\frac{(t-\tilde{m})^2}{2\tilde{s}^2}\right) = C_1 \exp\left(-\frac{t}{2\tilde{s}^2}(2|\tilde{m}| + t) + C_2\right) \simeq \\ &\simeq C_3 \exp\left(-\frac{|\tilde{m}|}{\tilde{s}^2} t\right), \end{aligned}$$

where the C's are arbitrary constants. That is, the tail of a truncated Gaussian tends to the exponential distribution, where $|\tilde{m}|/\tilde{s}^2 = 1/\lambda$. We shall see in a moment that $\tilde{m}$ and $\tilde{s}^2$ tend to $-\infty$ when the *observable relative uncertainty*, $u = s/m$, is large. Therefore, if we know the variance but the relative uncertainty is large, we are in a similar position to that of not knowing the variance.

As a second observation, note that if the prior is from a lower order, the posterior has the order corresponding to the order of the constraints. For example, if the prior is first order and we know the second order constraint, then the posterior is also second order. When we deal with the full estimation problem, at least some of the priors involved will be of second order, and thus will all posteriors.

In the remainder of the paper, we assume that $t_{\max} \gg m+s$, and therefore the first order solution is well approximated by an exponential distribution and the second order by a Gaussian distribution truncated at 0. If $t_{\max} > m + 24s$ this approximation is accurate. We also saw above that in the limit of a "high" relative uncertainty, the bounded Gaussian collapses to an exponential. The exponential distribution with parameter λ has expected value $1/\lambda$ and variance $1/\lambda^2$, and therefore $u = 1$, which is the upper bound of the relative uncertainty of the truncated Gaussian. The lower bound is $u = 0$, which occurs when there is no uncertainty.

We now describe how the Gaussian parameters $\tilde{m}$ and $\tilde{s}$ are connected to the observables m and s. There is no closed form analytical expression to connect these quantities (Tallis, 1961) but in the Appendix we present numerical expressions that convert observables to parameters and parameters to observables with an error of less than 0.5%, in the whole range.

Assume that T and X are random variables whose probability distribution is given by Eq. 3.3, and that T is truncated at 0 while X is not. Assume that m is fixed, and let the value of u, the relative uncertainty, vary. If $u < 0.3$, then almost all the probability mass of X stays in the positive range, and the observables are similar to the parameters, $\tilde{m} \sim m$ and $\tilde{s} \sim s$. That is, if u is low, both distributions are similar. If $u > 0.3$, then the parameter mean becomes increasingly smaller than the observable, $\tilde{m} < m$, and if $u \gtrsim 0.75$ the parameter mean becomes negative, $\tilde{m} < 0$, implying that the probability density $p(t)$ is no longer peaked and becomes monotonically decreasing. If $u \sim 1$, then $\tilde{m} \ll 0$ and, as expected, $\tilde{m}/\tilde{s}^2 \sim -m$.

The previous analysis, in terms of observable relative uncertainty u can be inverted, and we can see how observables depend on parameters, as a function of *parameter relative uncertainty* $\tilde{u} = \tilde{m}/\tilde{s}$ (note that the numerator and denominator are reversed, if compared to observable relative uncertainty). If $\tilde{u} > 3$ both distributions are similar. If $\tilde{u} < 0$ the truncated Gaussian becomes single peaked. If $\tilde{u} < -7$, the truncated Gaussian becomes approximately exponential, at an observable relative uncertainty of $u \sim 0.98$.

If the case of the Gaussian distribution without truncation, the median is m and the range $(m - s, m + s)$ contains 67% of probability mass. In the case of the exponential distribution, the median is $\ln(2)m$ and the range $(m - s, m + s)$ contains 86% of probability mass. The truncated Gaussian distribution takes values between these extremes, depending on u. Keeping these values in mind, we shall refer to m and s as best guess and uncertainty of the unknown distribution T.

The relationship between parameters and observables can be examined visually. In Figure 1 we look at the shape of the probability density $p(t)$, as a function of u. In Figures 2-3 we can see, respectively, the shape of the

functions of the observables as a function of the parameters and vice-versa.

3.2 The posterior distribution

In Section 3.1 we have found the properties of a positively valued continuous maxent distribution with, at best, known expected value and standard-deviation. We shall assume that the priors of the IO estimation process satisfy these conditions, and that the covariance of each pair of priors is also known, and shall use relevant partial information to obtain the posterior distributions.

We consider that the *prior transactions* are the components of an n_T-dimensional truncated multivariate normally distributed random variable, $\tilde{\mathbf{t}}$ with probability density $\tilde{\mathbf{p}}$, observable mean vector $\boldsymbol{\mu}$ and observable covariance matrix $\boldsymbol{\Sigma}$, where $\sigma_{jj} = \sigma_j^2$ is the variance and $\sigma_{jk} = \sigma_{kj}$. Likewise, the *posterior transactions* are the components of an n_T-dimensional truncated multivariate normally distributed random variable, $\mathbf{t}$ with probability density $\mathbf{p}$, observable mean vector $\mathbf{m}$ and observable covariance matrix $\mathbf{S}$, where $s_{jj} = s_j^2$ is the variance and $s_{jk} = s_{kj}$. We define $\tilde{\boldsymbol{\mu}}$, $\tilde{\boldsymbol{\Sigma}}$, $\tilde{\mathbf{m}}$ and $\tilde{\mathbf{S}}$ as the parameters of the corresponding normal distributions without truncation.

Furthermore, we consider that there is a total of n_K *topological constraints*, summarized in an *aggregation matrix* $\mathbf{G}$ that satisfies:

$$\mathbf{G}\mathbf{t} + \mathbf{k} = 0, \tag{3.4}$$

where $\mathbf{t}$ (the vector of posteriors) and $\mathbf{k}$ (the vector of numerical constraints) are column vectors of length n_T and n_K and every entry G_{ij} is either 1 or -1 if the constraint i aggregates transaction j or 0 otherwise.

Let $(\mathbf{G}_+) = \max\{0, \mathbf{G}\}$ and $(\mathbf{G}_-) = \max\{0, -\mathbf{G}\}$, such that $\mathbf{G} = (\mathbf{G}_+) - (\mathbf{G}_-)$. and let $\mathbf{g}_i$, $\mathbf{g}_i^+$ and $\mathbf{g}_i^-$ be, respectively, the i-th row of $\mathbf{G}$, $(\mathbf{G}_+)$ and $(\mathbf{G}_-)$. We consider that, as inputs to the problem, every topological $\mathbf{g}_i^+$ has at least two nonzero entries and that $\mathbf{g}_i^+$ has at most one nonzero entry. That is, each topological constraint connects multiple disaggregate transactions and an aggregate transaction and/or a numerical constraint.

The numerical constraints are random variables with known best guess and uncertainty that are *not* allowed to be adjusted by the maximum entropy method. Unless stated otherwise, in the remainder of Section 3.2 any expression with subscript i is valid in the range $i = 1, \ldots, n_K$ and every expression with subscript j is valid in the range $j = 1, \ldots, n_T$. All the partial information to be used in the estimation method is summarized in $\mathbf{G}$, $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$ and $\bar{\mathbf{m}}$ and $\bar{\mathbf{s}}^{\mathbf{2}}$, where the latter two are the vectors of the best guess and vari-

ance of the numerical constraints. In Section 3.3 we introduce the concept of information hierarchy and clarify the role of the numerical constraints.

A topological transaction i states that a partial sum of the components of the jointly distributed posterior subtracted to another such partial sum must be identical to the numerical constraint: $(\mathbf{g}_i^+)'\mathbf{t} - (\mathbf{g}_i^-)'\mathbf{t} + k_i = 0$, where $'$ denotes transpose and unless stated otherwise vectors are in column format. If the three random variables thus defined (numerical constraint and both positive and negative aggregations) are identical, their first and second moments must also be identical.

The expectation of a sum of random variables is identical to the sum of the expectations, $\mathrm{E}((\mathbf{g}_i^+)'\mathbf{t} - (\mathbf{g}_i^-)'\mathbf{t} + k_i) = \mathrm{E}((\mathbf{g}_i^+)'\mathbf{t}) - \mathrm{E}((\mathbf{g}_i^-)'\mathbf{t}) + \mathrm{E}(k_i)$. Therefore, $(\mathbf{g}_i^+)'\mathbf{m} - (\mathbf{g}_i^-)'\mathbf{m} + \bar{m}_i = 0$ or, more concisely:

$$0 = \mathbf{G}\mathbf{m} + \bar{\mathbf{m}}. \tag{3.5}$$

The variance of a sum of correlated random variables implies a sum over all terms of the covariance matrix, $\mathrm{Var}((\mathbf{g}_i^+)'\mathbf{t} - (\mathbf{g}_i^-)'\mathbf{t} + k_i) = \mathrm{Var}((\mathbf{g}_i^+)'\mathbf{t}) + \mathrm{Var}((\mathbf{g}_i^-)'\mathbf{t}) + \mathrm{Var}(k_i)$. Therefore, $(\mathbf{g}_i^+)'\mathbf{S}\mathbf{g}_i^+ - (\mathbf{g}_i^-)'\mathbf{S}\mathbf{g}_i^- + \bar{s}_i^2 = 0$. Notice that the numerical constraint must be independent of the unknown transactions. If diag denotes the main diagonal of a matrix, the second order constraint is:

$$0 = \mathrm{diag}\left(\mathbf{G}\mathbf{S}|\mathbf{G}|'\right) + \bar{\mathbf{s}}^{\mathbf{2}}. \tag{3.6}$$

Notice that the covariance matrix is multiplied on one side by the aggregation matrix and on the other side by the (transpose of) the module of the aggregation matrix. This is necessary because each topological constraint represents the difference of two sums, the positive and the negative aggregations. It is also worth noticing that there are only covariances between two positive or two negatively aggregated transactions. That is, $s_{jk} \neq 0$ only if transactions j and k are aggregated by at least one transaction i. It is convenient to define $G^*_{ijk} = 0$ if $G_{ij} > 0$ and $G_{ik} > 0$ or $G_{ij} < 0$ and $G_{ik} < 0$, or $G^*_{ijk} = 0$ otherwise.

We introduce the information about the first two moments into the Hamiltonian of the system in scalar form as:

$$H = -\int_0^\infty d\mathbf{t} p(\mathbf{t}) \frac{\ln p(\mathbf{t})}{\ln \tilde{p}(\mathbf{t})} + \lambda \left(\int_0^\infty d\mathbf{t} p(\mathbf{t}) - 1 \right) + \tag{3.7}$$
$$+ \sum_{i=1}^{n_K} \alpha_i \left(\sum_{j=1}^{n_T} G_{ij} \int_0^\infty d\mathbf{t} p(\mathbf{t}) t_j + \bar{m}_i \right) + \sum_{i=1}^{n_K} \tilde{\beta}_i \left(\sum_{j=1}^{n_T} G_{ij} \int_0^\infty d\mathbf{t} p(\mathbf{t}) (t_j - m_j)^2 + \right.$$
$$\left. + 2 \sum_{j=1}^{n_T} \sum_{k=1}^{j-1} G^*_{ijk} \int_0^\infty d\mathbf{t} p(\mathbf{t}) (t_j - m_j)(t_k - m_k) + \bar{s}_i^2 \right).$$

In Eq. 3.7 the expression $\int_0^\infty d\mathbf{t}$ is a shorthand for the product $\prod_{j=1}^{n_T} \int_0^\infty dt_j$. The first term in Eq. 3.7 contains the entropy of all unknown distributions, the second term contains the normalization constraint, the third term contains the best guess constraints, and the fourth term the uncertainty constraints. Note that m_j is the marginal expectation of T_j, itself defined as $m_j = \int_0^\infty t_j p(\mathbf{t}) d\mathbf{t}$. The λ, α's and $\tilde{\beta}$'s are, respectively, the Lagrange multipliers of the normalization, best guess and uncertainty constraints. Derivation of Eq. 3.7 with respect to $p(\mathbf{t})$, yields:

$$0 = -\left(\ln p(\mathbf{t}) + 1\right) \frac{1}{\ln \tilde{p}(\mathbf{t})} + \lambda + \sum_{j=1}^{n_T} \left(\sum_{i=1}^{n_K} G_{ij} \alpha_i \right) t_j +$$
$$+ \sum_{j=1}^{n_T} \left(\left(\sum_{i=1}^{n_K} G_{ij} \tilde{\beta}_i \right) \left(t_j^2 - 2 t_j m_j \right) \right) +$$
$$+ \sum_{j=1}^{n_T} \sum_{k=1}^{j-1} \left(2 \left(\sum_{i=1}^{n_K} G^*_{ijk} \tilde{\beta}_i \right) \left(t_j t_k - t_j m_k - t_k m_j \right) \right) + C.$$

All C's in the previous or subsequent expressions denote constants. This expression can be rewritten in the form:

$$p(\mathbf{t}) = \tilde{p}(\mathbf{t}) C \exp \left(\sum_{j=1}^{n_T} \left(\sum_{i=1}^{n_K} G_{ij} \tilde{\beta}_i \right) t_j^2 + \sum_{j=1}^{n_T} \sum_{k=1}^{j-1} 2 \left(\sum_{i=1}^{n_K} G^*_{ijk} \tilde{\beta}_i \right) t_j t_k + \right.$$
$$\left. + \sum_{j=1}^{n_T} \left(\sum_{i=1}^{n_K} G_{ij} \alpha_i - 2 \sum_{k=1}^{n_T} m_k \left(\sum_{i=1}^{n_K} G^*_{ijk} \tilde{\beta}_i \right) \right) t_j \right).$$

Notice that the exponent in the previous expression is a polynomial whose coefficients are linear combinations of Lagrange parameters. If the prior is a

multivariate truncated Gaussian and the constraints are of second order, the posterior is also a truncated multivariate Gaussian whose probability density is:

$$p(\mathbf{t}) = C \exp\left(-\frac{1}{2}(\tilde{\mathbf{t}} - \tilde{\mathbf{m}})' \tilde{\mathbf{S}}^{-1} (\tilde{\mathbf{t}} - \tilde{\mathbf{m}})\right). \tag{3.8}$$

The exponent of the prior and posterior probability densities can be expanded in polynomial form. In particular, Eq. 3.8 becomes:

$$p(\mathbf{t}) = C_1 \exp\left(-\sum_{j=1}^{n_T} \frac{\tilde{s}_{jj}^{-1}}{2} t_j^2 - 2\sum_{j=1}^{n_T}\sum_{k=1}^{j-1} \frac{\tilde{s}_{jk}^{-1}}{2} t_j t_k \right.$$
$$\left. +2\sum_{j=1}^{n_T}\left(\sum_{k=1}^{n_T} \frac{\tilde{s}_{jk}^{-1}}{2} \tilde{m}_k\right) t_j + C_2\right),$$

and the polynomial expansion of the prior distribution displays a similar pattern. In the previous expression $\tilde{s}_{jk}^{-1}$ is the (i,j) entry of matrix $\tilde{\mathbf{S}}^{-1}$. An explicit expression for the parameters of the posterior can be obtained by solving expressions of the form $C_{\text{post}} = C_{\text{prior}} + C_{\text{constraint}}$, where each constant is the coefficient of the corresponding polynomial expansion for the posterior and prior distributions and the expressions containing the Lagrange multipliers that result from differentiating the Hamiltonian, Eq. 3.7. We therefore obtain:

$$\tilde{\mathbf{S}}^{-\mathbf{1}} = \tilde{\mathbf{\Sigma}}^{-\mathbf{1}} + (\mathbf{G}_+)' \hat{\boldsymbol{\beta}} (\mathbf{G}_+) + (\mathbf{G}_-)' \hat{\boldsymbol{\beta}} (\mathbf{G}_-); \tag{3.9}$$

$$\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{m}} = \tilde{\mathbf{\Sigma}}^{-\mathbf{1}} \tilde{\boldsymbol{\mu}} + \mathbf{G}'\boldsymbol{\alpha} + \left(\tilde{\mathbf{S}}^{-\mathbf{1}} - \tilde{\mathbf{\Sigma}}^{-\mathbf{1}}\right)\mathbf{m}, \tag{3.10}$$

where we have made the substitution $\beta_i = -2\tilde{\beta}_i$ and $\hat{}$ denotes a diagonal matrix. Equations 3.9-3.10 and Eqs. 3.5-3.6 define the solution of the maximum entropy problem. Note however that Eqs. 3.9-3.10 contain Gaussian parameters (denoted with $\tilde{}$) and an observable on the left hand side of 3.10 while Eqs. 3.5-3.6 contain only observables. Since we are dealing with positively distributed random variables, these quantities may differ.

The properties of the truncated multivariate Gaussian distribution are not arbitrary. As in the univariate case, the observable relative uncertainty, $u_j = s_j/m_j$, is bounded by unity, $0 \leq u_j \leq 1$, and the observable best guess is strictly positive, $m_j > 0$. This is in spite of the fact that the mean and

variance of the non-truncated distribution can take any value. When the relative uncertainty is high, the mean of the non-truncated distribution lies deep in the negative range, $\tilde{m}_j \simeq -\infty$.

If the relative uncertainty of the pair of transactions (j,k) is small, then the probability isoquants in the positive (j,k)-hyperquadrant are ellipses, that can be stretched in any direction, and therefore the correlation, $r_{jk} = s_{jk}/s_j s_k$, can take any value in the range $-1 < r_{jk} < 1$. However, if the relative uncertainty of either of the transactions is high, then the isoquant is an ellipse seen from a long distance, i.e., a straight line. This means that the correlation is itself bounded, $r_{\min} < r_{jk} < r_{\max}$, and in the limit case in which one of the transactions has unitary observable relative uncertainty, $u_j = 1$ or $u_k = 1$, then the transactions must be uncorrelated, $r_{jk} = 0$. Therefore, if only first order information is known about a particular transaction (its best guess), then the prior of that transaction must be uncorrelated with all other transaction.

3.3 Information hierarchy

Consider that we know all priors and that no topological constraint has an associated numerical constraint. That is, $\mathbf{k} = \mathbf{0}$. Consider also that there is a *hierarchy of information quality*, such that among the n_T transactions there are L levels of information quality, and the data are indexed by increasing information quality. That is, all points in the range $(n_{h-1} + 1, n_h)$ have information quality of level h, where $n_0 = 0$ and $n_L = n_T$.

A hierarchy of information quality arises naturally in the compilation of IO tables. In the conventional table update problem the row and column sums have higher quality than the previous year estimate. Data collected from a national statistical office is likely to have higher quality than data processed by an international organization. And so forth.

The information hierarchy is distinct from the uncertainty level and more fundamental. We believe that in the presence of two priors of different information quality, the one of highest quality must be considered, irrespective of the uncertainty values of either one. That is to say, in the presence of higher quality information, the lower quality one is irrelevant.

We can formulate the general principle that *the estimation method should be invariant to the incorporation of irrelevant information.* The vector of numerical constraints introduced in Section 3.2 is a tool to operationalize this principle: higher information data is held fixed while we try to reconcile lower information data. If there is no solution because the numerical constraints are inconsistent, we relax the following information level.

We look for a consistent solution of information quality h, by holding fixed all data points $i > n_h$, and therefore:

$$\bar{\mu}_i = \sum_{j=n_h+1}^{n_T} G_{ij}\mu_j; \tag{3.11}$$

$$\bar{\sigma}_i^2 = \sum_{j=n_h+1}^{n_T} G_{ij}\sigma_j^2 + 2\sum_{j=n_h+1}^{n_T}\sum_{k=1}^{j} G^*_{ijk}\sigma_j\sigma_k\rho_{jk} + \tag{3.12}$$
$$+ 2\sum_{j=n_h+1}^{n_T}\sum_{k=1}^{n_h} G^*_{ijk}\sigma_j\sigma_k\rho_{jk}.$$

Notice that not only covariance σ_{jk} where $(j,k) > n_h$ is held fixed, but covariance σ_{jk} where $j > n_h$ and $k \leq n_h$ is also held fixed. Since only the first n_h transactions are being adjusted and the dimension of all relevant vectors and matrices must be truncated from n_T to n_h. Below the highest information level (L) there may be no solution, due to higher order inconsistencies. The level of information quality must be increased until a solution is found.

Moving along the information hierarchy means disaggregating the system, and thus potentially altering its topological structure. In Section 3.2 we considered that an entry of matrix $\mathbf{G}$ can be 0, 1 or -1. If aggregation is possible, as is now the case, then an entry can take any positive or negative integer value. This does not pose any problem, the practitioner only needs to ensure that at every information level all topological constraints aggregate two or more positive transactions and one negative transaction and/or numerical constraint.

To determine whether a solution of a given information level exists, it is necessary to find the topological constraints imposed on the numerical constraints and to check if they are satisfied. One way to address this problem is to perform an LU factorization to the aggregation matrix $\mathbf{G}$ (Golub and Van Loan, 1996):

$$\mathbf{G} = \mathbf{PLUQ},$$

where $\mathbf{P}$ and $\mathbf{Q}$ are (row and column) permutation matrices, $\mathbf{L}$ is a lower triangular matrix and $\mathbf{U}$ is an upper trapezoidal matrix. That is, matrix $\mathbf{U}$ is triangular, and if its rank is n_R, with $n_R < n_K$, then the first n_R entries along the main diagonal are full and its last $n_K - n_R$ rows are empty. If we introduce the LU factorization in Eq. 3.4:

$$\mathbf{UQt} = -\mathbf{L^{-1}P^{-1}k}.$$

The permutation matrix and the lower triangular matrix are easy to invert since they are both sparse. Let $\mathbf{L_*^{-1}}$ be the last $n_K - n_R$ rows of $\mathbf{L^{-1}}$. The system is consistent at information h if, at that level,

$$\mathbf{L_*^{-1}P^{-1}k} < |\epsilon|,$$

where ϵ is the cutoff value (typically the lowest nonzero source data point). If the system is inconsistent it is necessary to move one information level up. If it is consistent, now let $\mathbf{L^{*-1}}$ and $\mathbf{U^*}$ be the first n_R rows of $\mathbf{L^{-1}}$ and $\mathbf{U}$, and apply the following substitutions:

$$\begin{aligned}\mathbf{G} &:= \mathbf{U^*Q};\\ \mathbf{k} &:= \mathbf{L^{*-1}P^{-1}k}.\end{aligned}$$

Once the lowest consistent information level has been determined it is possible to determine the best guess and uncertainty of the posterior distribution.

4 Numerical considerations

4.1 Posterior best guess

There is no analytical explicit solution to the maximum entropy problem (Eqs. 3.5-3.6 and Eqs 3.9-3.10). The difficulties lie in the absence of an analytical conversion from the multivariate truncated parameters to observables (Horrace, 2005; Sharples and Pezzey, 2004), the need to invert matrices (Raveh, 1985) and the presence of the posterior best guess vector in the right hand side of Eq. 3.10. Fortunately, for practical purposes these difficulties can be circumvented. We now present a simple numerical approximations for the best guess posteriors.

Let us examine the limit cases of the general solution of the posterior best guess. Using the results of Section 3.1, if all data points have a small relative uncertainty ($u < 0.3$), we can treat all parameters as observables. Under these conditions, Eq. 3.10 simplifies to:

$$\mathbf{m} = \boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{G}'\boldsymbol{\alpha}. \tag{4.1}$$

Combining Eq. 4.1 and Eq. 3.5 we determine the best guess Lagrange parameters as the solution of:

$$(\mathbf{G}\boldsymbol{\Sigma}\mathbf{G}')\,\boldsymbol{\alpha} = -\,(\mathbf{G}\boldsymbol{\mu} + \bar{\mathbf{m}})\,. \tag{4.2}$$

Equations 4.1-4.2 represent a generalized least square (LS) solution, which is rigorous when relative uncertainties are moderately small.

Now consider the opposite limit case, in which only first order information is available, that is, all prior best guesses are known but no uncertainty is available. In this case, the relative uncertainty is unitary and the posteriors are independent and exponentially distributed. The expression for best guess posteriors is:

$$\mathbf{m^{-1}} = \boldsymbol{\mu}^{\mathbf{-1}} - \mathbf{G}'\boldsymbol{\alpha}.$$

where $\mathbf{m^{-1}}$ and $\boldsymbol{\mu}^{\mathbf{-1}}$ are vectors whose entry j is, respectively, $1/m_j$ and $1/\mu_j$. We assume that the displacement $dm_j = m_j - \mu_j$ is small, $|dm_j| \ll |\mu_j|$ and apply a first order Taylor expansion to the inverse of the best guess, $1/(\mu_j + dm_j) \simeq 1/\mu_j - dm_j/\mu_j^2$. Introducing this substitution in the previous expression yields:

$$\mathbf{m} = \boldsymbol{\mu} + \hat{\boldsymbol{\mu}}^{\mathbf{2}}\mathbf{G}'\boldsymbol{\alpha},$$

where $\hat{}$ denotes a diagonal matrix. This last expression is identical to Eq. 4.1 but under the simplified circumstances of the exponential limit, namely that $\sigma_j = \mu_j$ and $\sigma_{jk} = 0$ if $j \neq k$. It is interesting that in both limit cases a similar and simple solution was obtained. In this limit case, however, the solution is only an approximation.

A more general situation in which Eqs. 4.1-4.2 are valid is when the parameter covariances do not change, $\tilde{\mathbf{S}} = \tilde{\boldsymbol{\Sigma}}$, because in that case then the last term in the right hand side of Eq. 3.10 disappears. If the magnitude of all best guess Lagrange parameters, α, is small, the generalized LS solution applies.

We consider that, in general, second order information is of lower quality than first order information and we therefore assume that when the parameter best guess is adjusted by Eq. 3.10, $\tilde{\mu}_j \neq \tilde{m}_j$, the parameter uncertainty is already adjusted by Eq. 3.9, $\tilde{\sigma}_j = \tilde{s}_j$.

It is worth noticing that if the observable best guess changes, $\mu_j \neq m_j$, even if the parameter uncertainty is fixed, $\tilde{\sigma}_j = \tilde{s}_j$, then the observable

uncertainty also changes, $\sigma_j \neq s_j$. Let $u_j = s_j/m_j$ and $\nu_j = \sigma_j/\mu_j$ be the posterior and prior observable relative uncerainties. We found that in the absence of correlations the following expression is a good approximation to the posterior observable absolute uncertainty:

$$s_j = \sigma_j \left(\left(1 - \nu_j^2\right) + \frac{m_j}{\mu_j} \left(\nu_j^2\right) \right). \tag{4.3}$$

When the prior relative uncertainty is low, then the prior and posterior *absolute* uncertainty are close. When the prior relative uncertainty is high, then the prior and posterior *relative* uncertainty are close. Eq. 4.3, like Eq. 4.1, is only valid in a small range of α.

We studied numerically the behaviour of Eq. 3.10 in the absence of correlations and when the parameter uncertainties are fixed, which in scalar form is:

$$\tilde{m}_j = \tilde{\mu}_j + \tilde{\sigma}_j^2 \alpha_j,$$

where $\alpha_j = \sum_{i=1}^{n_K} G_{ij}\alpha_i$. We determined the range of α_j in which Eq. 4.1 and Eq. 4.3 would be accurate within an error margin of 1%. That is, we determined $|\alpha_j|$ for which $(m_j - m_j^*)/\mu_j < 0.01$ and $(u_j - u_j^*)/\nu_j < 0.01$, where m_j and u_j are determined using Eq. 4.1 and Eq. 4.3 and m_j^* and u_j^* are determined using Eq. 3.10. Notice that we want to take observable arguments and receive observable solutions, and that the analytical solution involves making conversions to and from parameters.

We found that the upper bound of 1% error is minorated by the empirical function $|\alpha_j| < \alpha^*/\mu_j$, where:

$$\alpha^* = \begin{cases} 0.63763(\nu_j)^{-2.0900} & \text{if } 0.01 \leq \nu_j < 0.05 \\ 0.033131(\nu_j)^{-3.0758} & \text{if } 0.05 \leq \nu_j < 0.7 \\ 0.1 & 0.07 \leq \nu_j < 0.99 \end{cases} \tag{4.4}$$

The displacement bound is narrow when the relative uncertainty is high. Notice also that Eqs. 4.4-4.3 are only accurate if correlations are zero. Due to the lack of a better alternative we suggest the use of the previous expressions even in the presence of correlations. However, we have seen at the end of Section 3.2 that not all potential correlation values are admissible. If the uncertainties are adjusted and the correlations are not, it may happen than a correlation moves outside the valid range. In Section Section 4.3 we address this problem.

4.2 Posterior uncertainty

We now obtain a numerical approximation for the posterior covariances. It is convenient to expresss the latter in terms of uncertainties and correlations. That is, $s_{jk} = r_{jk}s_j s_k$, where r_{jk} is the correlation between j and k. $r_{jj} = 1$ and $r_{jk} = r_{kj}$, so the posterior correlations can be arranged in the symmetrical matrix $\mathbf{R}$, whose main diagonal contains only 1's and whose (j,k) entry is r_{jk}, such that $-1 \leq r_{jk} \leq 1$. Prior correlations are ρ_{jk} and the prior correlation matrix is $\mathbf{P}$. For the moment consider that parameter and observable values are identical (the limit of low relative uncertainties). Substituting correlations in Eq. 3.9 leads to:

$$\hat{\mathbf{s}}^{-1}\mathbf{R}^{-1}\hat{\mathbf{s}}^{-1} = \hat{\boldsymbol{\sigma}}^{-1}\mathbf{P}^{-1}\hat{\boldsymbol{\sigma}}^{-1} + (\mathbf{G}_+)'\hat{\boldsymbol{\beta}}(\mathbf{G}_+) + (\mathbf{G}_-)'\hat{\boldsymbol{\beta}}(\mathbf{G}_-).$$

The explicit calculation of inverse matrices is highly undesirable. Let $\hat{\mathbf{s}}\mathbf{R}\hat{\mathbf{s}} = \hat{\boldsymbol{\sigma}}\mathbf{R}^*\hat{\boldsymbol{\sigma}}$. If the adjustment $\mathbf{dR} = \mathbf{R}^* - \mathbf{P}$ is small, $||\mathbf{dR}|| \ll ||\mathbf{P}||$ and a first order Taylor expansion is acceptable. Let $\mathbf{P} = \mathbf{I} - \mathbf{A}$ and recall that the expansion of the inverse of the correlation matrix is $\mathbf{P}^{-1} = \mathbf{I} + \mathbf{A} + \mathbf{A}^2 + \ldots$ Under these circumstances, the approximation is:

$$(\mathbf{P} + \mathbf{dR})^{-1} = \mathbf{P}^{-1} - \mathbf{P}^{-1}\mathbf{dR}\mathbf{P}^{-1}.$$

Substituting this Expression in Eq. 3.9 yields:

$$\begin{aligned}\mathbf{R}^* = \mathbf{P} &- \mathbf{F}\#\left(\mathbf{P}\hat{\boldsymbol{\sigma}}^*(\mathbf{G}_+)'\hat{\boldsymbol{\beta}}(\mathbf{G}_+)\hat{\boldsymbol{\sigma}}^*\mathbf{P}\right) \\ &- \mathbf{F}\#\left(\mathbf{P}\hat{\boldsymbol{\sigma}}^*(\mathbf{G}_-)'\hat{\boldsymbol{\beta}}(\mathbf{G}_-)\hat{\boldsymbol{\sigma}}^*\mathbf{P}\right),\end{aligned} \tag{4.5}$$

where (for the time being) $\sigma_j^* = \sigma_j$, $\mathbf{F}$ is the *filter matrix* and $\#$ is the Hadamard product. The filter matrix was introduced to account for the sensitivity of different entries of the covariance matrix to the Lagrange multipliers. The variances can be freely adjusted and so $F_{jj} = 1$. Correlation (j,k) can only be adjusted if there is some topological constraint i that aggregates transactions j and k. Therefore, if $G_{ijk}^* = 0$ for all i then $F_{jk} = 0$. If $G_{ijk}^* \neq 0$ for some i and $j \neq k$, then $F_{jk} = 1$.

Another limit situation that is easy to explore is when there are no correlations. In this case, Eq. 3.9 becomes, in scalar form:

$$\frac{1}{\tilde{s}_j^2} = \frac{1}{\tilde{\sigma}_j^2} + \beta_j,$$

where $\beta_j = \sum_{i=1}^{n_K} G_{ij}\beta_i$. We discovered that the following function is a good approximation of the previous expression, linking the posterior observable uncertainty directly to the Lagrange parameter:

$$\frac{1}{s_j^2} = \frac{1}{\sigma_j^2} + f(\nu_j)\beta_j,$$

which can in turn be linearized, leading to:

$$s_j = \sigma_j\sqrt{1 - \sigma_j^2 f(\nu_j)\beta_j}, \tag{4.6}$$

where:

$$f(\nu_j) = \begin{cases} 1 & \text{if } \nu_j < 1/\sqrt{2} \\ \sqrt{2}\nu_j & \text{if } \nu_j \geq 1\sqrt{2} \end{cases} \tag{4.7}$$

Notice that the above expressions are only a function of observables, and not of Gaussian parameters. When relative uncertainty is low, Eq. 4.6 becomes $s_j = \sigma_j\sqrt{1 - \sigma_j^2\beta_j}$, which is the scalar form of Eq. 4.5 in the absence of correlations. When relative uncertainty is high, the displacement induced by the Lagrange parameter is higher. We suggest to combine the numerical solutions when correlations are small, Eq. 4.6, and when relative uncertainties are low, Eq. 4.5, by setting the term σ_j^* in the latter expression as:

$$\sigma_j^* = \sigma_j f(\nu_j). \tag{4.8}$$

Combining Eq. 4.5 and Eq. 3.6 we the determine the uncertainty Lagrange parameters as the solution of:

$$\mathbf{B}\boldsymbol{\beta} = -\left(\hat{\mathbf{s}}^{\mathbf{2}} + \mathrm{diag}\left(\mathbf{G}\boldsymbol{\Sigma}|\mathbf{G}|'\right)\right), \tag{4.9}$$

where:

$$\begin{aligned}\mathbf{B} = &\left(\mathbf{G}(\mathbf{F}\#\boldsymbol{\Sigma}^*)(\mathbf{G}_+)'\right)\#\left(|\mathbf{G}|(\mathbf{F}\#\boldsymbol{\Sigma}^*)(\mathbf{G}_+)'\right) \\ &+ \left(\mathbf{G}(\mathbf{F}\#\boldsymbol{\Sigma}^*)(\mathbf{G}_-)'\right)\#\left(|\mathbf{G}|(\mathbf{F}\#\boldsymbol{\Sigma}^*)(\mathbf{G}_-)'\right).\end{aligned}$$

In the previous expression $\sigma_{jk}^* = \sigma_j\rho_{jk}\sigma_k^*$. Once Eqs. 4.5-4.9 have been solved, the posterior uncertainties and correlations are obtained as as $s_i =$

$\sigma_i\sqrt{r^*_{ii}}$, and $r_{ij} = r^*_{ij}\sigma_i\sigma_j/(s_i s_j)$. Notice that this new correlation may be outside the valid range and need to be adjusted.

In particular, $|r^*_{ii} - r_{ii}| < 1$ because otherwise the variance of i could become negative. Since Eq. 4.6 is only a local approximation, the vector of Lagrange parameters should be shortened until the displacement is small, $|r^*_{ii} - r_{ii}| < \delta$, and the least square minimization is valid at every step.

4.3 Correlation bounds

We now address the properties of the correlations. We studied the numerical behaviour of the bivariate truncated Gaussian distribution of transactions j and k. We discovered that the upper and lower bounds for the correlation, $r_{\max}$ and $r_{\min}$, are not affected by the ratio of the best guesses, m_j/m_k, and that they are symmetric functions of the relative uncertainties $(r_{\max}, r_{\max}) = f(u_j, u_k) = f(u_k, u_j)$. The limit cases are $f(u_j, 0) = (1, 1)$ and $f(u_j, 1) = (0, 0)$, that is, if a relative uncertainty is zero, then the bounds are unitary, and if a relative uncertainty is unitary, then the bounds are zero.

The isoquants of the bounds are of the form $u_j^n + u_k^n = 1$, where n is a function of $r_{\max}$ or $r_{\min}$. If we move along the identity line in the plane of relative uncertainties, $u_j = u_k$, then $r_{\min} \simeq -1$ if $u_j < 0.6$, and then converges smoothly to $r_{\min} = 0$ if $u_j = 1$ passing through $r_{\min} = -0.5$ if $u_j = 0.7$, $r_{\min} = -0.3$ if $u_j = 0.8$ and $r_{\min} = -0.15$, $u_j = 0.9$. The upper bound converges more abruptly, dropping from $r_{\max} \simeq 1$ if $u_j < 0.95$ to $r_{\max} = 0$ if $u_j = 1$. Note that this is the behaviour for the bivariate. In the general multivariate case, the situation is more complex. The presence of a third transaction of non-truncated Gaussian distribution does not affect the possible range of correlations between the first two. It is well known that a marginal multivariate Gaussian distribution is also Gaussian. If this third transaction is exponentially distributed, however, the valid correlation range is reduced because a higher proportion of probability mass is concentrated close to the origin.

We suggest the rule $r_{\max} = 1$ and:

$$r_{\min} = \sqrt{1 - \nu_j^2}\sqrt{1 - \nu_k^2}. \tag{4.10}$$

The previous expression has the qualitative behaviour required, although the precise shape is not accurate. We sugges that, if at some point $r_{jk} > r_{\max}$ then $r_{jk} := r_{\max}$ and $r_{jk} < r_{\min}$ then $r_{jk} := r_{\min}$.

The suggested rule Eq. 4.10 correctly addresses this limit cases of low and unitary uncertainty. In the intermediate region of high but less than unitary relative uncertainty, the rule is a qualitative approximation.

4.4 Numerical implementation

In this Section we have presented numerical approximations for the general solution of the posterior best guess, the posterior covariances, and correlation bounds. In the general solution, the best guess and Lagrange parameters, $(\boldsymbol{\alpha}, \boldsymbol{\beta})$, should be determined simultaneously. In our numerical approximation they are determined sequentially, in alternate steps. We consider that second order information is of lower quality than first order information, and that higher order inconsistencies should not affect the results if lower order information is consistent. Therefore, we consider that $\boldsymbol{\beta}$ should be determined, such that the second order information is well adjusted, while the observable best guesses are held fixed, $\mathbf{m} = \boldsymbol{\mu}$, before each step in $\boldsymbol{\alpha}$ is taken, in which case the best guesses are adjusted. At every step it must be checked if the correlations are within the reasonable bounds.

We consider the implementation of the more general and accurate method, MEP1, we shall also consider two simplifications, MEP2 and MEP3.

To summarize, in MEP1 we sequentially alternate two processes. Uncertainties and correlations are adjusted iteratively using Eqs. 4.5 and 4.9, where Eq. 4.8 and Eq. 4.7 determine σ^*, and correlations are adjusted by Eq. 4.10 if necessary, such that at every step $|r^*_{ii} - r_{ii}| < \delta$. Once the second order data is consistent, the best guesses are adjusted using Eqs. 4.1 and 4.2, taking a single step within the range Eq. 4.4 and the uncertainty is updated using 4.3 and correlations adjusted if necessary, using Eq. 4.10.

In MEP2 the update of uncertainties is performed in a single step, in between the iterative update of best guesses.

In MEP3 we avoid iterations altogether and proceed in three steps: update $\boldsymbol{\beta}$ in a single step, update $\boldsymbol{\alpha}$ in a single step, and finally readjust $\boldsymbol{\beta}$.

4.5 Comparison with other methods

At this stage it is interesting to compare the numerical Bayesian method described above with similar expressions found in the literature.

Rampa (2008) suggested the subjective weighted least square (SWLS) method, valid when there are priors for both the best guess and uncertainty:

$$\mathbf{m} = \boldsymbol{\mu} - \hat{\boldsymbol{\sigma}}^2 \mathbf{G}' \left(\mathbf{G} \hat{\boldsymbol{\sigma}}^2 \mathbf{G}' \right)^{-1} (\mathbf{G}\boldsymbol{\mu}) \,.$$

This is the solution of Eqs. 4.1-4.2 when the covariance matrix $\boldsymbol{\Sigma}$ is replaced by the diagonal matrix of variances, $\hat{\boldsymbol{\sigma}}^{\mathbf{2}}$. That is, using this method variances are considered but correlations are not. For the sake of compari-

son, the ordinary least square (OLS) method corresponds to the situation in which all weights are identical and $\mathbf{\Sigma}$ is replaced by $\mathbf{I}$.

The SWLS differs from the MEP solution in two different aspects. First, it does not consider an information hierarchy, but all transactions are adjusted simultaneously. In our Bayesian approach, a higher quality data point is adjusted only if it is not possible to reconcile lower quality data. Second, the weights are subjective and can be freely chosen by the practitioner. In our Bayesian approach, the relative uncertainty is bound by $0 < u < 1$ and can be obtained from empirical patterns, which may be helpful for the less skilled practitioner.

On the RAS side, although a large number of formulations is available (Stromman, 2009), to our knowledge there is none that incorporates the uncertainty of transactions. The recently proposed method KRAS (Lenzen et al., 2009) incorporates the uncertainty of numerical constraints (i.e., IO row and column sums). With minor adjustments, Equations 13 and 23 of that article become:

$$m_j = \mu_j \prod_{i=1}^{n_K} r_i^{G_{ij}};$$
$$\bar{m}_i^{k+1} = \bar{m}_i^k + a\epsilon_i^k \sigma_i.$$

The first expression is the conventional RAS algorithm, which is multiplicative in nature, while the LS methods are additive. If the displacement is small there is no difference between the two approaches but if the displacement is large, in the LS methods the cost function is linearly symmetric around the initial guess while in the RAS method it is log symmetric. The scaling parameters r_i's are determined iteratively, by minimizing the error at iteration k, $\boldsymbol{\epsilon}_k = \mathbf{G}\mathbf{m}_k + \bar{\mathbf{m}}_k$. The initial numerical constraints are $\bar{m}_i^0 = \bar{\mu}_i$ and they are adjusted if the errors do not converge to zero. Parameter a should be chosen by the user.

This method introduces a two tier information hierarchy, because the numerical constraints are adjusted less frequently than transactions. The transactions are all multiplicatively adjusted with the same weight, and the numerical constraints are adjusted additively, with the uncertainty as weight, and not the variance as is the case in LS methods.

We shall compare these four methods (OLS, SWLS, RAS and KRAS) against the MEP approximations in Section **??**.

5 Conclusions

In this paper we presented a Bayesian estimation method for IO Analysis, that can reconcile possibly conflicting data of arbitrary form, taking into account the uncertainty and the information hierarchy of the source data. We studied several theoretical and numerical properties of the application of the maximum entropy principle in this context, whose solution is a multivariate truncated Gaussian distribution.

The numerical implementation takes the form of a generalized least squares (LS) method, but if some data points have high relative uncertainty the method must be applied iteratively, such that displacements are small. When compared to conventional methods (either biproportional or LS), the proposed Bayesian method offers some additional complexity. However, the posteriors obtained by the Bayesian method have the property that data of higher uncertainty is proportionately more adjusted than in conventional methods.

The work reported in this paper is self-sufficient if we are interested in obtaining a detailed IO table for a reference year, and priors are available for all entries. This happens, to some extent, in the compilation of the GTAP database. However, if priors are missing, if we wish to determine the uncertainty of multipliers or to interpolate time-series data, other considerations are required.

A Appendix

A.1 Conversion of truncated Gaussian parameters

Let T be a truncated Gaussian distribution, $T > 0$, with Gaussian parameters $\tilde{m}$ and $\tilde{s}$ and observables m and s (best guess and uncertainty).

Let $X \sim N(\tilde{m}, \tilde{s})$ be a non-truncated normally distributed random variable with the same parameters as T, and let $Y \sim N(0, 1)$ be the standard normal distribution, whose probability density function is:

$$\phi(y) = \Pr(Y = y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}. \tag{A.1}$$

The probability of the non-truncated and standard normal distributions are linked as $\Pr(X = x) = \phi\left((x - \tilde{m})/\tilde{s}\right)/\tilde{s}$. Let $Z_n = \int_{y=-\tilde{u}}^{\infty} y^n \phi(y) dy$, where $\tilde{u} = \tilde{m}/\tilde{s}$. Using these expressions, the first three moments of T are as follows.

$$E(T^0) = \frac{1}{Z_0}\int_{x=0}^{\infty} \Pr(X = x)dx = \frac{1}{Z_0}\int_{y=-\tilde{u}}^{\infty} \phi(y)dy \qquad (A.2)$$
$$= \frac{Z_0}{Z_0} = 1.$$

From the normalization condition A.2 we find that indeed Z_0 is the correct normalization constant of 3.3.

$$m = E(T^1) = \frac{1}{Z_0}\int_{x=0}^{\infty} x \Pr(X = x)dx \qquad (A.3)$$
$$= \frac{1}{Z_0}\int_{y=-\tilde{u}}^{\infty} (\tilde{s}y + \tilde{m})\phi(y)dy = \tilde{s}\frac{Z_1}{Z_0} + \tilde{m}\frac{Z_0}{Z_0}$$

We see that the expected value of T is the expected value of the non-truncated distribution plus an additional term.

$$s^2 = E(T^2) - E(T^1)^2 = \frac{1}{Z_0}\int_{x=0}^{\infty} x^2 \Pr(X = x)dx - E(T^1)^2 \qquad (A.4)$$
$$= \frac{1}{Z_0}\int_{y=-\tilde{u}}^{\infty} (\tilde{s}^2y^2 - \tilde{s}\tilde{m}y + \tilde{m}^2)\phi(y)dy - E(T^1)^2$$
$$= \tilde{s}^2\frac{Z_2}{Z_0} + 2\tilde{s}\tilde{m}\frac{Z_1}{Z_0} + \tilde{m}^2\frac{Z_0}{Z_0} - (\tilde{s}^2\frac{Z_1^2}{Z_0^2} + 2\tilde{m}\tilde{s}\frac{Z_1}{Z_0} + \tilde{m}^2)$$
$$= \tilde{s}^2\left(\frac{Z_2}{Z_0} - \frac{Z_1^2}{Z_0^2}\right).$$

The variance of T is the product of the variance of X and an additional term. We have found by inspection that $Z_2 = Z_0 - \tilde{u}Z_1$, that $Z_1 = \phi(\tilde{u})$, where $\phi(z)$ is given by Eq. A.1, and that $Z_0 = \left(1 + \mathrm{erf}(\tilde{u})/\sqrt{2}\right)/2$, where $\mathrm{erf}(\tilde{u})$ is the error function. There is no analytical solution to Z_0 but we found the following approximation sufficient (Abramowitz and Stegun, 1964):

$$Z_0(\tilde{u}) = \begin{cases} 1 - \phi(|\tilde{u}|)\left(\sum_{i=1}^{5} \frac{c_i^z}{\left(1+c_0^z|\tilde{u}|\right)^i}\right) & \text{if} \quad \tilde{u} > 0 \\ \phi(|\tilde{u}|)\left(\sum_{i=1}^{5} \frac{c_i^z}{\left(1+c_0^z|\tilde{u}|\right)^i}\right) & \text{if} \quad \tilde{u} \leq 0 \end{cases} \qquad (A.5)$$

where the constants are:

i	c_i^z
0	0.2316419
1	0.319381530
2	- 0.356563782
3	1.781477937
4	- 1.821255978
5	1.330274429

Let us define $\eta(\tilde{z}) = Z_1/Z_0$. This function can be computed numerically usign Eqs. A.1 and A.5, since $Z_1 = \phi(\tilde{u})$. We now use Eqs. A.2-A.4 to write expressions that give us the observables (mean and variance) of T as functions of the parameters:

$$m = \tilde{m} + \tilde{s}\eta(\tilde{u}); \tag{A.6}$$

$$s = \tilde{s}\sqrt{1 - \eta(\tilde{u})\left(\eta(\tilde{u}) + \tilde{u}\right)}. \tag{A.7}$$

Note that $\tilde{s}$ is always nonnegative while $\tilde{m}$ can take negative values. The error function, and therefore η, are numerically inaccurate if $\tilde{u} << 0$. Thus, we use Eqs. A.6-A.7 only if $\tilde{u} > -2.95$, and otherwise use the approximations:

$$m = \frac{\tilde{s}^2}{\tilde{m}}\left(-1 + 1.97777\left(1 - u\right) + 6.8\left(1 - u\right)^2\right); \tag{A.8}$$

$$s = mu; \tag{A.9}$$

$$u = \left(1 - 0.014172\left(\tilde{u} + \sqrt{16.8 + \tilde{u}^2}\right)^2\right). \tag{A.10}$$

Let us define $u = s/m$. We are now interested in expressing the parameters as a functions of the observables. Since we do not have an analytical expression of the inverse of $\eta(\tilde{u})$, we had to search for numerical approximations and found that the following expressions have an absolute error of less than 0.5% in the range $0 < u < 1$. If $u \leq 0.3$, the conversion functions are defined as:

$$\tilde{m} = m;$$

$$\tilde{s} = s.$$

If $0.3 < u \leq 0.8$, the conversion functions are:

$$\tilde{m} = m g_m^{tp}(u);$$
$$\tilde{s} = \frac{s}{\sqrt{g_s^{tp}(u)}}.$$

Finally, if $u > 0.8$ we have:

$$\tilde{m} = m g_m^{tp}(u);$$
$$\tilde{s} = -s \frac{g_m^{tp}(u)}{\sqrt{g_m^{tp}(u) g_r^{tp}(u)}}.$$

The conversion functions are:

$$g_m^{tp}(u) = 1 - \frac{1}{1-u} e^{-(1-u)\left(c_4^m + c_3^m |u - c_1^m|^{c_2^m}\right)} + \frac{c_3^g}{c_2^g} \phi\left(\frac{u - c_1^g}{c_2^g}\right); \tag{A.11}$$

$$g_s^{tp}(u) = (1-u) \frac{c_3^s}{c_2^s} \phi\left(\frac{u - c_1^s}{c_2^s}\right); \tag{A.12}$$

$$g_r^{tp}(u) = (1-u) \frac{c_3^r}{c_2^r} \phi\left(\frac{u - c_1^r}{c_2^r}\right) - 1. \tag{A.13}$$

In the previous expressions $\phi(z)$ is given by Eq. A.1 and the parameters c_i^* are reported in the following table:

i	c_i^m	c_i^g	c_i^s	c_i^r
1	0.937492	0.468838	0.546626	0.83179
2	1.78863	0.118555	0.439319	0.617251
3	7.13173	0.00235939	1.83447	6.3836
4	5.42261			

The set of Eqs. A.6-A.9 and A.11-A.13 allow the conversion from observables to parameters and vice-versa, keeping the error below 0.5%.

References

M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions*. National Bureau of Standards, Washignton DC, USA, 1964.

M. Bacharach. *Biproportional Matrices and Input-Output Change*. Cambridge University Press, Cambridge, UK, 1970.

T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley and Sons, Inc., New York, 1991.

D. Gilchrist and L. St. Louis. An algorithm for the consistent inclusion of partial information in the revision of input-output tables. *Economic Systems Research*, 16 (2):149–156, 2004.

G. H. Golub and C. F. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, USA, 1996. 3rd edition.

W. C. Horrace. On ranking and selection from independent truncated normal distributions. *Journal of Econometrics*, 126:335–354, 2005.

R. W. Jackson and A. T. Murray. Alternative input-output matrix updating formulations. *Economic Systems Research*, 16 (2):135–148, 2004.

E. T. Jaynes. Information theory and statistical mechanics i. *Physical Review*, 106:620–630, 1957.

E. T. Jaynes. The well-posed problem. *foundations of Physics*, 3:477–493, 1973.

E. T. Jaynes. *Papers on Probability, Statistics and Statistical Physics*. Kluwer Academic Publishers, Dordrecht, Netherlands, 1983.

T. Junius and J. Oosterhaven. The solution of updating or regionalizing a matrix with both positive and negative entries. *Economic Systems Research*, 15 (1):87–96, 2003.

S. Kullback and R. A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22 (1):79–86, 1951.

M. L. Lahr and L. de Mesnard. Biproportional techniques in input–output analysis: Table updating and structural analysis. *Economic Systems Research*, 16 (2):115–134, 2004.

P. M. Lee. *Bayesian statistics: an introduction*. Oxford University Press, New York, NY, 1989.

M. Lenzen. Errors in conventional and input-output-based life-cycle inventories. *Journal of Industrial Ecology*, 4 (4):127–148, 2001.

M. Lenzen, B. Gallego, and R. Wood. Matrix balancing under conflicting information. *Economic Systems Research*, 21 (1):23–44, 2009.

M. Lenzen, R. Wood, and T. Wiedmann. Uncertainty analysis for multi-region input-output models - a case study of the uk's carbon footprint. *Economic Systems Research*, 22 (1):43–63, 2010.

R. Lieu and R. B. Hicks. A maximum entropy method when prior information consists of inexact constraints. *Astrophysical Journal*, 422 (2):845–849, 1994.

R. Lieu, R. B. Hicks, and C. J. Bland. Maximum-entropy in data-analysis with error-carrying constraints. *Journal of Physics A*, 20 (9):2379–2388, 1987.

L. de Mesnard. Biproportional methods of structural change analysis: a typological survey. *Economic Systems Research*, 16 (2):205–230, 2004.

R. E. Miller and P. D. Blair. *Input-Output Analysis: Foundations and Extensions*. Cambridge University Press, Cambridge, UK, 2009.

J. Oosterhaven, D. Stelder, and S. Inomata. Estimating international interindustry linkages: Non-survey simulations of the asian-pacific economy. *Economic Systems Research*, 20 (4):395–414, 2008.

G. Rampa. Using weighted least squares to deflate input-output tables. *Economic Systems Research*, 20 (3):1469–5758, 2008.

A. Raveh. On the use of the inverse of the correlation matrix in multivariate data analysis. *The American Statistician*, 39 (1):39–42, 1985.

C. E. Shannon. A mathematical theory of communication. *Bell System Technical Journal*, 27:379–423, 1948.

J. J. Sharples and J. C. V. Pezzey. Expectations of linear functions with respect to truncated multinormal distributions - with applications to uncertainty analysis in environmental modelling. *Economic Systems Research*, 16 (2):149–156, 2004.

R. Stone, D. G. Champernowne, and J. E. Meade. The precision of national income estimates. *Review of Economic Studies*, 9:111–125, 1942.

A. H. Stromman. A multi-objective assessment of input-output updating methods. *Economic Systems Research*, 21 (1):81–89, 2009.

G. M. Tallis. The moment generating function of the truncated multi-normal distribution. *Journal of the Royal Statistical Society, Series B*, 23:223–229, 1961.

K. Weise and W. Woger. A bayesian theory of measurement uncertainty. *Measurement Science and Technology*, 4 (1):1–11, 1992.