A Bayesian Approach to Conflicting Input-Output Data

Rodrigues, João F. D.*

January 12, 2012

Abstract

In this paper we apply a Bayesian approach to the reconciliation of conflicting data in Input-Output (IO) tables. In a Bayesian context IO transactions are treated as nonnegative random variables of truncated Gaussian distribution with known best guess and uncertainty. From the Maximum Entropy Principle we derive an analytical expression that obtains a consistent set of posteriors from a set of conflicting priors. We report a numerical approximation of the general solution and compare this Bayesian algorithm to conventional techniques (least squares and biproportional update methods) using an empirical example.

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 with 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 the data is inconsistent (i.e., row and column sums do not add up) and the information quality of the data is different (e.g., a row or column sum is known for

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

the target year while a matrix element is known from another year). Many methods exist for the IO estimation problem, in which numerical constraints are used to balance data of lower information quality, such as the biproportional or least-squares families (Mesnard, 2004). Within this subject, there has been recent attention devoted to the problem of reconciling numerical constraints that are themselves mutually inconsistent (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 present a method to 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 applying 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 consists in moving from a set of mutually inconsistent priors to a set of mutually consistent posteriors, where the IO transactions are connected to one another through topological constraints (such as row and column sums).

We present the general solution, which does not allow for an explicit formula, and a numerical algorithm, which takes the form of an iterative weighted least square method. We make use of invariance considerations (Jaynes, 1973) to derive the sequence in which the MEP algorithm is applied. The Bayesian method thus derived allows for an information hierarchy, in which the number of IO entries which can be adjusted is progressively increased, until a solution is found.

We present a numerical example using the symmetric IO tables of Portugal. We use this example to compare the Bayesian method with the recently proposed KRAS (Lenzen et al., 2009) and SWLS (Rampa, 2008) methods. We also use this example to examine the behaviour of the MEP solution and the validity of the numerical algorithm.

The structure of this paper is as follows. In Section 2 we review current methods for IO estimation and present the background Bayesian theory. In Sections 3 to 5 we derive the Bayesian theory of IO uncertainty. In Section 6 we report the numerical algorithms. In Section 7 we present a real-world example and in Section 8 we draw conclusions.

2 Literature review

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 that RAS is the solution of a maximum entropy (MEP) problem, the minimization of relative entropy (Kullback and Leibler, 1951). In this context 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 these constraints' initial uncertainty and its current inconsistency. One characteristic of this method, which derives 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 have been proposed (Lieu et al., 1987; Lieu and Hicks, 1994).

However, 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, which 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. That is, a set of default assumptions should be available to deal with incomplete information.

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.

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 entropy of a discrete probability distribution $\{p_i\}_{i=1}^N$ is $-\sum_{i=1}^N p_i \ln p_i$ and in our case the available information are the j-th moments of the distribution $\sum_{i=1}^N i^j p_i = M_j$. If a prior probability distribution $\{\pi_i\}_{i=1}^N$ is available, then the posterior probability distribution is obtained by minimizing relative entropy $\sum_{i=1}^N p_i \ln(p_i/\pi_i)$, 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). (Notice that entropy is maximized while relative entropy is minimized, because the former is defined with a minus sign and the latter not).

Entropy maximization is familiar in IO analysis. However, what is 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, which can be naturally adjusted (if we so wish).

All input data to the estimation problem consist in the properties (expectation and standard-deviation) 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 that long thin needles are dropped randomly over a small circle. What is the probability that a chord (i.e., the line segment defined by a needle touching the circle in two points) will have a given length? This question poses a paradox because there are different ways of choosing a chord at random, leading to different distributions. Jaynes (1973) solved this paradox by noting that 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 Maximum entropy priors

If an IO quantity is known with some degree of uncertainty, then its true (unknown) value can take different realizations, which are described by a probability distribution. In this Section our goal is to determine the properties of the probability distribution describing such an IO quantity.

According to the Bayesian paradigm (Jaynes, 2003), the probability distribution of an unknown quantity is obtained by using all available information and no other.

All available information, in this context, means both numerical and logical information. The numerical information we possess usually takes the form of a best guess and some estimated degree of uncertainty (related, for example, to the sample size of a survey). The logical information is related to the physical properties of the object considered. In this case, an IO quantity represents an economic transaction which is a nonnegative real number.

In an IO table, transactions sometimes appear as negative quantities (e.g., services provided by margins in supply tables) but these quantities can be simply reallocated as positive values in another region of the table. Balancing items, such as a change in stocks or net taxes, on the other hand, can indeed take both positive and negative values. We address this situation again in Section 6.

In a Bayesian context, using no other information besides the one that is available means the application of the Maximum Entropy Principle (MEP). That is, we search for the least informative (or maximally entropic) distribution that is consistent with the available information.

We follow the example of Weise and Woger (1992) and interpret the positive real-valued *best guess*, μ, and *uncertainty*, σ, of the source data as the expected value, $\mathrm{E}(\theta) = \mu$, and standard deviation, $\mathrm{Var}(\theta) = \sigma^2$, of an yet unspecified random variable Θ, with probability $\pi(q)$, which represents an IO quantity that takes values in the range $[0, q_{\max}]$. We used the conventional notation $\mathrm{E}(f(\theta)) = \int_0^{q_{\max}} dq \pi(q) p(q)$ and $\mathrm{Var}(\theta) = \mathrm{E}(\theta^2) - \mathrm{E}(\theta)^2$. Throughout this and the following Section we shall use q to represent an event or realization of a random variable.

An important assumption we make is that the possibility of a negative transaction is zero because it is economically meaningless, but the possibility of a very large transaction is not zero, although it may be very small. That is, although transactions must take a finite value, the maximum possible value may be much larger than the best guess, where $\mu \ll q_{\max}$.

The Maximum Entropy Principle (Jaynes, 1983), states that a *posterior* distribution is obtained by minimizing the entropy of the posterior relative to the *prior* distribution subject to the known constraints (recall that entropy maximization implies relative entropy minimization). By the end of this Section our goal is to obtain a prior distribution $\pi(q)$. However, at this stage we treat $\pi(q)$ as a posterior, considering a more "fundamental" prior $\psi(q)$. Notice that the distinction between a prior and a posterior is positional. A posterior is obtained by combining a prior and some other information. Under this light the same distribution can be both a prior and posterior, depending on the context.

Under the above conditions the Hamiltonian or objective function is:

$$H = \int_0^{q_{\max}} dq\pi(q) \ln\left(\frac{\pi(q)}{\psi(q)}\right) + \lambda_0(\mathrm{E}(1) - 1) + \\ \lambda_1(\mathrm{E}(\Theta) - \mu) + \lambda_2\left(\mathrm{E}(\Theta^2) - \mathrm{E}(\Theta)^2 - \sigma^2\right). \tag{3.1}$$

The first term in the right hand side of Eq. 3.1 is the differential entropy of the unknown distribution. The remaining terms in the right hand side of Eq. 3.1 are the set of known constraints: the *zeroth 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.

At this stage assume that the prior distribution $\psi(q)$ is uniformly distributed in the range $[0, q_{\max}]$. At the end of the Section we review this assumption. Differentiation of Eq. 3.1 with respect to $\pi(q)$ leads to:

$$0 = -\ln\left(\frac{\pi(q)}{\psi(q)}\right) - 1 + \lambda_0 + \lambda_1 q + \lambda_2\left(q^2 - 2\mu q\right). \tag{3.2}$$

Since Eq. 3.1 defines a concave function, differentiation yields a unique maximum. Now let us consider three cases. First, if we only know the zero-th order constraint, $\lambda_1 = \lambda_2 = 0$, Eq. 3.2 leads to a uniform distribution, $p(q) = 1/q_{\max}$. 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.

Second, if we also know the first order constraint, only $\lambda_2 = 0$, and Eq. 3.2 leads to a truncated exponential distribution, $p(q) = (\lambda e^{-\lambda q})/(1 - e^{-\lambda q_{\max}})$. The parameter λ is determined by the best guess μ.

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

$$\pi(q) = \frac{1}{Z_0}\frac{1}{\sqrt{2\pi\tilde{\sigma}^2}} \exp\left(-\frac{(q - \tilde{\mu})^2}{2\tilde{\sigma}^2}\right), \tag{3.3}$$

with the substitution $2\lambda_2 = 1/\tilde{\sigma}^2$ and $\lambda_1 - 2\mu\lambda_2 = -\tilde{\mu}/\tilde{\sigma}^2$, where Z_0 is a normalization constant. Note that since this distribution is truncated, the Gaussian parameters $\tilde{\mu}$ and $\tilde{\sigma}^2$ are *not* the *observable* expectation and variance of the distribution, μ and σ^2.

The forms of the zeroth, 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 $\mu = q_{\max}/2$ has uniform solution, and is therefore equivalent to the zeroth order solution. In essence, knowing an expectation that lies exactly in the middle of the range without knowing the variance is equivalent to not knowing that expectation.

At this point it is convenient to remember that the observable best guess is much lower than the maximum economically possible value, $\mu \ll q_{\max}$, and so the first order solution is well approximated by an exponential without truncation, $q_{\max} \simeq \infty$ and $\pi(q) = \exp(-q/\mu)/\mu$. An important property of the nontruncated exponential distribution is that the expected value and the standard deviation are identical, $\mu = \sigma$. So, if we only know the best guess of an IO quantity, but we do not know its uncertainty, we are in the same condition of knowing that its uncertainty is exactly identical to the best guess. Therefore, there is an upper bound of one for the relative uncertainty, ν, defined as $\nu = \sigma/\mu$, such that $0 \leq \nu \leq 1$.

We expect a smooth transition from the second to the first order solution, just as we found a smooth transition from the first order to the zeroth order solution. Unfortunately, there is no closed form analytical expression to connect the observables, μ and σ, and the Gaussian parameters, $\tilde{\mu}$ and $\tilde{\sigma}$ in the truncated Gaussian distribution (Tallis, 1961). However, we can perform numerical simulations and observe that such a smooth transition exists. Making use of the assumption that $\mu \ll q_{\max}$ and its implication that $q_{\max} \simeq \infty$, we can study the Gaussian distribution truncated on the left side at 0, and nontruncated on the right side. If relative uncertainty is small, $\nu < 0.3$, the truncated and the nontruncated Gaussian distributions are indistinguishable. As the relative uncertainty increases, the peak of the distribution slides to the left, until after $\nu \simeq 0.75$ the distribution becomes monotonically decreasing. And in the limit of $\nu > 0.98$ the truncated Gaussian becomes indistinguishable from the exponential distribution.

The limit behaviour when relative uncertainty approaches unity can be deduced analytically. We observed that in this case $\tilde{\mu} \to -\infty$ and $\tilde{\sigma} \to \infty$. We now perform the expansion of Eq. 3.3 under these conditions:

$$\pi(q) = C_1 \exp\left(-\frac{(q-\tilde{\mu})^2}{2\tilde{\sigma}^2}\right) = C_1 \exp\left(-\frac{q^2}{2\tilde{\sigma}^2} + \frac{2q\tilde{\mu}}{2\tilde{\sigma}^2} - \frac{\tilde{\mu}^2}{2\tilde{\sigma}^2}\right) \simeq$$
$$\simeq C_1 \exp\left(0 + \frac{2q\tilde{\mu}}{2\tilde{\sigma}^2} - C_2\right) = C_3 \exp\left(-\frac{|\tilde{\mu}|}{\tilde{\sigma}^2} q\right),$$

where the C's are appropriately chosen constants. That is, as expected the tail of a truncated Gaussian distribution tends to the exponential distribution and we have found an explicit expression that links the Gaussian parameters to observables $|\tilde{\mu}|/\tilde{\sigma}^2 = 1/\mu = 1/\sigma$.

In this Section we have observed that starting from a uniform prior and introducing information on the zeroth, first and second moments we obtained, respectively, a uniform, an exponential and a truncated Gaussian distribution. After that we observed that there is a smooth transition between these distributions. If the second moment is known, the shape of the distribution can be approximated by a nontruncated Gaussian, in the limit of low relative uncertainty, or by an exponential, in the limit of high relative uncertainty. Furthermore, relative uncertainty itself is bounded by zero and one. These are the properties of the priors use in the data reconciliation problem of the next Section.

At this point, the interested reader can repeat the derivation of Eq. 3.1 with the prior $\psi(q)$ having either exponential or truncated Gaussian form. In either case, if the first and second moment are known, the posterior is also a truncated Gaussian. That is, the transformation from prior to posterior implies either an increase in or the maintenance of the level of information, in the sense that a truncated Gaussian is more informative than an exponential that in turn is more informative that a uniform distribution.

This observation is important because in the data reconciliation problem to be dealt with in the following Section we expect that all best guess priors are available and at least some best guess uncertainties. Under these conditions we know a priori that all posteriors will have a truncated Gaussian distribution, even if some of them fall on the exponential limit.

4 Maximum entropy posteriors

In this Section we want to calculate an analytical expression that links a set of conflicting priors and a corresponding set of balanced posteriors.

The properties of priors were determined in Section 3, that is, they are positively valued continuous random variables with MEP distributions with known best guess and uncertainty. We now consider multiple random variables so it is necessary to consider covariances. For the purpose of this Section we assumed that the covariance of each pair of priors is known. In Section 6 we discuss covariances again.

The transactions are connected to one another and to numerical constraints through topological constraints, i.e., rules that indicate how transactions are linked to one another. The simplest example of a topological constraint is a row sum of an IO table. In this case the numerical value of the sum is the numerical constraint and the topological constraint is the rule specifying which transactions are summed.

The set of balanced posteriors is obtained using the MEP, as in the previous Section, but we now consider that both the prior and the posterior are multivariate instead of univariate random variables. The posterior configuration is obtained by minimizing entropy relative to the prior configuration, subject to the constraint that both first and second moments must be balanced via the topological constraints.

We consider that the *prior transactions* are the components of a n_T-dimensional truncated multivariate normally distributed random variable, $\boldsymbol{\theta}$ with probability density $\boldsymbol{\pi}$, best guess vector $\boldsymbol{\mu}$ and 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 a n_T-dimensional multivariate truncated 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 $s_{jk} = s_{kj}$. Whenever one of the previous symbols is represented with a tilde, $\tilde{}$, it means it is not an observable quantity but the corresponding Gaussian parameter.

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

$$\mathbf{0} = \mathbf{G}\mathbf{t} + \mathbf{k}, \tag{4.1}$$

where $\mathbf{t}$ (the vector of posteriors) and $\mathbf{k}$ (the vector of numerical constraints) have 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. Vectors are in column format by default and $\mathbf{0}$ is a vector of zeros.

We consider that every topological constraint (i.e., a row of $\mathbf{G}$) connects at least one disaggregate transaction (an entry with a positive sign) and at least an aggregate transaction (an entry with a negative sign) or a numerical constraint. This is a logical requirement because a topological constraint is a link between two quantities. If a topological constraint has only one nonzero entry, then that transaction must be set to zero and removed from the reconciliation problem.

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 4 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 variance of the numerical constraints, respectively. In Section 5 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. If the random variables thus defined are identical, their first and second moments must also be identical.

The constraint on best guesses is:

$$\mathbf{0} = \mathbf{G}\mathbf{m} + \bar{\mathbf{m}}. \tag{4.2}$$

If diag denotes the main diagonal of a matrix and $'$ denotes transpose, the constraint on uncertainties is:

$$\mathbf{0} = \operatorname{diag}\left(\mathbf{G}\mathbf{S}\mathbf{G}'\right) + \bar{\mathbf{s}}^{\mathbf{2}}. \tag{4.3}$$

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

$$\begin{aligned} H = & -\int_\Omega dq\, p(\mathbf{q}) \ln\left(\frac{p(\mathbf{q})}{\pi(\mathbf{q})}\right) + \lambda\left(\int_\Omega dq\, p(\mathbf{q}) - 1\right) + \\ & + \sum_{i=1}^{n_K} \alpha_i \left(\sum_{j=1}^{n_T} G_{ij} \int_\Omega dq\, p(\mathbf{q}) q_j + \bar{m}_i\right) + \\ & + \sum_{i=1}^{n_K} \beta_i^* \left(\sum_{j=1}^{n_T} G_{ij} \int_\Omega dq\, p(\mathbf{q}) (q_j - m_j)^2 + \right. \\ & \left. + 2 \sum_{j=1}^{n_T} \sum_{k=1}^{j-1} G_{ij} G_{ik} \int_\Omega dq\, p(\mathbf{q}) (q_j - m_j)(q_k - m_k) + \bar{s}_i^2 \right). \end{aligned} \tag{4.4}$$

In Eq. 4.4 the expression $\int_\Omega dq$ is a shorthand for the product $\prod_{j=1}^{n_T} \int_0^\infty dq_j$. Each q_j is the realization of the random variables t_j and θ_j. The first term in Eq. 4.4 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, defined as $m_j = \int_\Omega dq\, q_j p(\mathbf{q})$. The λ, α's and β^*'s are, respectively, the Lagrange multipliers of the normalization, best guess and uncertainty constraints. Derivation of Eq. 4.4 with respect to $p(\mathbf{q})$, yields:

$$0 = -\left(\ln p(\mathbf{q}) + 1\right)\frac{1}{\ln \pi(\mathbf{q})} + \lambda + \sum_{j=1}^{n_T}\left(\sum_{i=1}^{n_K} G_{ij}\alpha_i\right) q_j +$$
$$+ \sum_{j=1}^{n_T}\left(\left(\sum_{i=1}^{n_K} G_{ij}\tilde{\beta}_i\right)\left(q_j^2 - 2q_j m_j\right)\right) +$$
$$+ \sum_{j=1}^{n_T}\sum_{k=1}^{j-1}\left(2\left(\sum_{i=1}^{n_K} G_{ij}G_{ik}\tilde{\beta}_i\right)\left(q_j q_k - q_j m_k - q_k m_j\right)\right) + C.$$

The C's in the previous and subsequent expressions denote different appropriately chosen constants. The previous expression can be rewritten in the form:

$$p(\mathbf{q}) = \pi(\mathbf{q}) C \exp\left(\sum_{j=1}^{n_T}\left(\sum_{i=1}^{n_K} G_{ij}\tilde{\beta}_i\right) q_j^2 + \sum_{j=1}^{n_T}\sum_{k=1}^{j-1} 2\left(\sum_{i=1}^{n_K} G_{ij}G_{ik}\tilde{\beta}_i\right) q_j q_k + \right.$$
$$\left. + \sum_{j=1}^{n_T}\left(\sum_{i=1}^{n_K} G_{ij}\alpha_i - 2\sum_{k=1}^{n_T} m_k\left(\sum_{i=1}^{n_K} G_{ij}G_{ik}\tilde{\beta}_i\right)\right) q_j\right).$$

Notice that the exponent in the previous expression is a polynomial whose coefficients are linear combinations of Lagrange multipliers. 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:

$$p(\mathbf{q}) = C \exp\left(-\frac{1}{2}(\tilde{\mathbf{q}} - \tilde{\mathbf{m}})'\tilde{\mathbf{S}}^{-1}(\tilde{\mathbf{q}} - \tilde{\mathbf{m}})\right). \tag{4.5}$$

The exponent of the prior and posterior probability densities can be expanded in a polynomial form. In particular, Eq. 4.5 becomes:

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

and the polynomial expansion of the prior distribution displays a similar pattern. In the previous expression $\tilde{s}^{-1}_{jk}$ 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. 4.4. We therefore obtain:

$$\tilde{\mathbf{S}}^{\mathbf{-1}} = \tilde{\mathbf{\Sigma}}^{\mathbf{-1}} + (\mathbf{G})'\hat{\boldsymbol{\beta}}(\mathbf{G}); \tag{4.6}$$

$$\tilde{\mathbf{S}}^{-1}\tilde{\mathbf{m}} = \tilde{\mathbf{\Sigma}}^{-1}\tilde{\boldsymbol{\mu}} + \mathbf{G}'\boldsymbol{\alpha} + \left(\tilde{\mathbf{S}}^{\mathbf{-1}} - \tilde{\mathbf{\Sigma}}^{\mathbf{-1}}\right)\mathbf{m}, \tag{4.7}$$

where we have made the substitution $\beta_i = -2\beta_i^*$ and $\hat{}$ denotes a diagonal matrix. Equations 4.6-4.7 and Eqs. 4.2-4.3 define the solution of the maximum entropy problem. Note however that Eqs. 4.6-4.7 contain Gaussian parameters (denoted with $\tilde{}$) and an observable on the left hand side of 4.7 while Eqs. 4.2-4.3 contain only observables.

As desired, we have obtained analytical expressions that define the configuration of mutually consistent posteriors that is obtained by adjusting a configuration of mutually conflicting priors so that all relevant topological constraints are satisfied.

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 occurs despite 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, which can be stretched in any direction. 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, $r_{\min} < r_{jk} < r_{\max}$. In the limit case in which both transactions have unitary observable relative uncertainty, if $u_j = u_k = 1$, 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 transactions.

5 Information hierarchy

In this Section we depart from the line of inquiry developed in Sections 3 and 4 to clarify the nature of numerical constraints, introduced in the previous Section.

In principle, all IO data is subject to empirical error and should be subject to adjustment, if it conflicts with other data. However, indepedently of the uncertainty assigned to a data point, it is reasonable to consider that source data has multiple vintages of information quality. Consider for example that we construct a multi-regional table using both survey data from national statistical offices and secondary data obtained by a non-survey method (Oosterhaven et al., 2008).

Irrespective of the uncertainty reported in the priors, we consider that the information quality of the survey data is better than that of the non-survey data. In this case, it is reasonable to impose that the survey data be adjusted only if by adjusting the non-survey it was not possible to find a consistent table.

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, irrespectively 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 4 is a tool to operationalize this principle: data of higher information quality is held fixed while we try to reconcile data of lower information quality. If there is no solution because the numerical constraints are inconsistent, we relax the following level of information quality.

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 is a hierarchy of H levels of information quality, and the data are indexed by increasing level of information quality. That is, all points in the range $(n_{L-1} + 1, n_L)$ have information quality of level L, where $n_0 = 0$ and $n_H = n_T$.

We look for a consistent solution of information quality L, by holding fixed all data points $i > n_L$, and the best guess and variance of the numerical constraints, $\bar{\mu}$ and $\bar{\sigma}^2$, are:

$$\bar{\mu}_i = \sum_{j=n_L+1}^{n_T} G_{ij}\mu_j; \tag{5.1}$$

$$\bar{\sigma}_i^2 = \sum_{j=n_L+1}^{n_T} G_{ij}\sigma_j^2 + 2\sum_{j=n_L+1}^{n_T}\sum_{k=1}^{j} G_{ij}G_{ik}\sigma_j\sigma_k\rho_{jk} + \tag{5.2}$$
$$+ 2\sum_{j=n_L+1}^{n_T}\sum_{k=1}^{n_L} G_{ij}G_{ik}\sigma_j\sigma_k\rho_{jk}.$$

Notice that not only covariance σ_{jk}, where $j, k > n_L$ is held fixed, but covariance σ_{jk}, where $j > n_L$ and $k \leq n_L$ is also held fixed. Since only the first n_L transactions are being adjusted it is necessary to truncate the dimension of all relevant vectors and matrices from n_T to n_L.

Below the highest information level (H) there may be no solution, due to higher order inconsistencies. That is there may be no posterior configuration for which all best guess and covariance topological constraints are satisfied. In this case it is necessary to remove the inconsistencies, and one way to achieve this goal is to perform a LU factorization to the aggregation matrix $\mathbf{G}$ (Golub and Van Loan, 1996). For the sake of clarity let:

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

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 $n_R < n_K$ (n_K is the number of topological constraints), then the first n_R entries along the main diagonal are nonzero and its last $n_K - n_R$ rows are zero. If we introduce the LU factorization in Eq. 4.1:

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

Permutation and triangular matrices are special matrices that are easy to invert. Now let $\mathbf{L}_*^{-1}$ be the last $n_K - n_R$ rows of $\mathbf{L^{-1}}$. The system is consistent at information level L if, at that level,

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

where ϵ is the cutoff value (typically the lowest nonzero source data point). If the system is inconsistent, it is necessary to ignore the last $n_K - n_R$ topological constraints in order to obtain a consistent solution for the current information level. 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:

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

It is now possible to determine the best guess and uncertainty of the posterior distribution for the current information level. Since this procedure involves losing some topological information, it is convenient to permute the original data so that the most informative topological constraints are kept. In the absence of additional information, this can be guaranteed if they are ordered by decreasing best guess magnitude.

6 Numerical approximation

In this Section we derive a numerical approximation of the general solution reported in Section 4 in two steps. First, we obtain a generalized least square solution by making assumptions about the relative uncertainty of the priors. Second, we obtain a weighted least square solution by making assumptions about the topology of typical IO data.

There is no analytical explicit solution to the maximum entropy problem (Eqs. 4.2-4.3 and Eqs. 4.6-4.7). The difficulties lie in the absence of an analytical conversion from the multivariate truncated Gaussian 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. 4.7. However, it is possible to obtain a simple numerical approximation for the best guess posteriors.

Using the results of Section 3, if all data points have a small relative uncertainty ($u < 0.3$), the truncated Gaussian parameters are observable best guesses and uncertainties. Under these conditions, Eq. 4.7 simplifies to:

$$\mathbf{m} = \boldsymbol{\mu} + \boldsymbol{\Sigma}\mathbf{G}'\boldsymbol{\alpha}. \tag{6.1}$$

Combining Eq. 6.1 and Eq. 4.2 we determine the best guess Lagrange multipliers as the solution of:

$$\left(\mathbf{G}\boldsymbol{\Sigma}\mathbf{G}'\right)\boldsymbol{\alpha} = -\left(\mathbf{G}\boldsymbol{\mu} + \bar{\mathbf{m}}\right). \tag{6.2}$$

Equations 6.1-6.2 represent a generalized least square (LS) solution, which is rigorous when relative uncertainties are moderately small. We do not expect all data points to fulfill these conditions, but we believe that most of them will. We therefore consider that this approximation can be used in any real-world IO application.

At this point it is convenient to expresss the covariances as a product of uncertainties and correlations. That is, $\sigma_{jk} = \rho_{jk}\sigma_j\sigma_k$, where ρ_{jk} is the prior *correlation* between j and k, $\rho_{jj} = 1$ and $\rho_{jk} = \rho_{kj}$. The prior correlation matrix is $\mathbf{P}$, such that $\boldsymbol{\Sigma} = \hat{\boldsymbol{\sigma}}\mathbf{P}\hat{\boldsymbol{\sigma}}$, where $\hat{}$ denotes diagonal matrix. Likewise, r_{jk} and $\mathbf{R}$ are, respectively, a posterior correlation and the posterior correlation smatrix.

If all prior uncertainties and correlations are known, Eqs. 6.1-6.2 define the solution, keeping in mind that it is only an approximation when uncertainties are high. Although we can make an educated guess of what the prior uncertainties are, it is highly unlikely that we possess information on prior correlations.

We do not wish to discuss correlations here because the matter is non-trivial and we discuss it at length in a forthcoming paper. Here we shall only consider the effect of considering two extreme cases, zero and unitary correlations, and we argue that, in the absence of further information, correlations should be assumed to have maximal value, i.e., to be close to one.

If all correlations are zero, then Eqs. 6.1-6.2 define a weighted least square solution, in which the weights are covariances. If all correlations are one, then Eqs. 6.1-6.2 define a generalized least square solution, which is computationally much more complex. However, if we take into account the typical topology of IO tables, a substantial simplification can be obtained.

A typical IO table contains many sectors and therefore most constraints (row or column sums) aggregate many transactions. However, each transaction is only affected by few constraints (typically only two, the row and column sum). Under these conditions (many transactions per constraint, few constraints per transaction and correlations different from zero) we can make simplifications.

Consider a dense IO matrix, such that every entry (ij) is affected by the row and column constraints. The expansion of $\mathbf{G}'\boldsymbol{\alpha}$ becomes a vector where each entry is the sum of two Lagrange multipliers, $\alpha_i^R + \alpha_j^C$, corresponding to the constraints of the i-th row and j-th column. For simplicity we shall use (ij) to denote a single transaction. The expansion of an entry of Eq. 6.1 becomes:

$$m_{ij} = \mu_{ij} + \sigma_{ij}\left(\alpha_i^R\left(\sigma_{ij} + \sum_{k\neq j}\rho_{(ij,ik)}\sigma_{ik}\right) + \alpha_j^C\left(\sigma_{ij} + \sum_{k\neq i}\rho_{(ij,kj)}\sigma_{kj}\right)\right.$$
$$\left. + \sum_{k\neq j}\alpha_k^R\rho_{(ij,ik)}\sigma_{ik} + \sum_{k\neq i}\alpha_k^C\rho_{(ij,kj)}\sigma_{kj}\right).$$

If all correlations are unitary we find that:

$$m_{ij} = \mu_{ij} + \sigma_{ij}\left(\alpha_i^R\sum_k\sigma_{ik} + \alpha_j^C\sum_k\sigma_{kj} + \sum_{k\neq j}\sigma_{ik}\alpha_k^R + \sum_{k\neq i}\sigma_{kj}\alpha_k^C\right).$$

If we make the substitutions $\alpha_i^{R*} = \alpha_i^R\sum_k\sigma_{ik}$ and $\alpha_j^{C*} = \alpha_j^C\sum_k\sigma_{kj}$ the previous expression becomes:

$$m_{ij} = \mu_{ij} + \sigma_{ij}\left(\alpha_i^{R*} + \alpha_j^{C*} + \sum_{k\neq j}\alpha_k^{R*}\frac{\sigma_{ik}}{\sum_l\sigma_{il}} + \sum_{k\neq i}\alpha_k^{C*}\frac{\sigma_{kj}}{\sum_l\sigma_{lj}}\right).$$

If there are many transactions per constraint, it is reasonable to consider that $\sigma_{ik} \ll \sum_l\sigma_{il}$ and that $\sigma_{kj} \ll \sum_l\sigma_{lj}$. Introducing these considerations in the previous expression we find that:

$$m_{ij} = \mu_{ij} + \sigma_{ij}\left(\alpha_i^{R*} + \alpha_i^{C*}\right).$$

In the above example we considered a particular (but typical) setting (dense matrix and only row and column constraints), but the result obtained holds in the general conditions considered (many transactions per constraint, few constraints per transaction, and most correlations close to unity). Generalizing the previous expression to matrix format we find the numerical solution of the best guess posteriors:

$$\mathbf{m} = \boldsymbol{\mu} + \hat{\boldsymbol{\sigma}}\mathbf{G}'\boldsymbol{\alpha}, \tag{6.3}$$

and:

$$\left(\mathbf{G}\hat{\boldsymbol{\sigma}}\mathbf{G}'\right)\boldsymbol{\alpha} = -\left(\mathbf{G}\boldsymbol{\mu} + \bar{\mathbf{m}}\right). \tag{6.4}$$

The solution is a weighted least square method in which the weights are prior uncertainties.

Care must be taken to ensure that the solution is meaningful, which means it cannot change sign. We suggest to constrain the adjustment $\boldsymbol{\alpha}$, so that $|m_j - \mu_j| < \mu_j$ for all entries, iterating until a consistent solution is obtained. That is, we make the minimal requirement that that the relative displacement from prior to posterior must be smaller than 100% at every iteration, not allowing an entry to change sign or to double in magnitude. Of course, the reader can implement more stringent requirements, (for example to impose relative displacement to be smaller that 10% or 1%) but we do not expect this to alter results significantly. Each intermediate posterior uncertainty must also be adjusted, so that the solution remains meaningful (i.e., relative uncertainties remain between zero and one). The simplest option is to impose that relative uncertainty does not change, $s_j = m_j \sigma_j / \mu_j$.

At this point we must address the problem of IO entries that are not economic transactions but balancing items. Such terms are described by a non-truncated Gaussian distribution, which means that relative uncertainty has no upper bound and that the quantities can change sign. The simplest way to introduce balancing items in the framework described above is to separate the balancing item into an input and an output component, each of which is positive.

For example, to distinguish taxes from subsidies, the former may be described as an outflow of currency from a company, and the latter as an inflow. Consider for example that taxes exceed subsidies. The relative uncertainty assigned to the taxes flow is either the relative uncertainty (if provided by the source data with value smaller than one) or is unitary otherwise. The best guess of the subsidies flow should be a nonzero residual value, e.g., a few orders of magnitude smaller than the smallest best guess prior, with unitary relative uncertainty. Following this approach it is possible for the balancing item to change sign, if consistency so requires. Of course, the requirement that the relative displacement should be smaller than 100%, $|m_j - \mu_j| < \mu_j$, is not required for these quantities. In this case what is necessary is to perform the necessary adjustment, for example transferring the negative taxes to the subsidies entry.

In Section 7 we shall compare the Bayesian approximation with two related methods. Rampa (2008) suggested a subjective weighted least square (SWLS) method, in which the solution is:

$$\mathbf{m} = \boldsymbol{\mu} + \hat{\mathbf{a}}\hat{\boldsymbol{\mu}}\mathbf{G}'\boldsymbol{\alpha};$$
$$(\mathbf{G}\hat{\mathbf{a}}\hat{\boldsymbol{\mu}}\mathbf{G}')\boldsymbol{\alpha} = -\left(\mathbf{G}\boldsymbol{\mu} + \bar{\mathbf{m}}\right).$$

In this method, $\mathbf{a}$ is a vector of inverse reliability indexes, such that if the practicioner considers entry 1 more reliable than entry 2, then $a_1 < a_2$. The choice of indexes is completely arbitrary. This method should be applied in a single step and therefore best guesses can become negative (or be set to zero). No information hierarchy is considered, although a similar result can be achieved by an appropriate choice of reliability indexes, forcing higher quality data to adjust very little compared to lower quality data.

We shall also consider the recently proposed KRAS method (Lenzen et al., 2009), which incorporates the uncertainty of numerical constraints (i.e., IO row and column sums). With minor adjustments, Equations 13 and 23 of that article become:

$$m_j = \mu_j \prod_{i=1}^{n_K} \alpha_i^{G_{ij}};$$
$$\bar{m}_i = \bar{m}_i - a\epsilon_i\sigma_i,$$

where α_i is a biproportional adjustment parameter and a should be chosen by the user. That is, using this method the RAS technique is applied to interior points of an IO table and, alternately, the row and column sums are linearly adjusted in proportion to uncertainties and the error of the corresponding topological constraint.

According to Rampa (2008) a second order Taylor expansion of the RAS method yields a weighted least square method in which the weights are proportional to the best guess. Therefore, this method allows for the incorporation of the uncertainty of numerical constraints, but assumes that all interior points have the same relative uncertainty, if viewed from our Bayesian perspective.

This method also allows for a sort of information hierarchy, but different from the Bayesian one. In KRAS, either interior points or numerical constraints are adjusted, so this is a two-tier alternate hierarchy. In our Bayesian method, the number of levels in the hierarchy is arbitrary, and whenever a higher quality level data point is adjusted all points of lower quality data are also adjusted.

In the typical biproportional problem, in which row and column sums are of higher quality than interior points, we expect all these methods to deliver similar results. Typical IO data spans several orders of magnitude so the scaling of absolute uncertainty to best guess is flat, $\sigma_j \simeq \nu_j\mu_j$, where all data in the same information level has the same (or very similar) relative uncertainty ν_j. Thus, if the the reliability indices of of the SWLS method are bound between zero and one, we are in the conditions of the MEP solution, but all data is adjusted in a single step. If all data in the worst information level has a similar relative uncertainty, then the first step in the application of the MEP method is actually independent of the relative uncertainty, since the solution of Eqs. 6.3-6.4 is not affected if the diagonal matrix $\hat{\boldsymbol{\sigma}}$ is multiplied by a scalar, and so $\hat{\boldsymbol{\sigma}}$ can be replaced by $\hat{\boldsymbol{\mu}}$ in Eqs. 6.3-6.4. Therefore, the KRAS method (in which interior points are adjusted in proportion to best guess and not uncertainty) yields the same solution as the MEP method.

Thus, the Bayesian method combines features of both SWLS and KRAS, but within a more coherent framework. It explicitly considers the uncertainty of all data, it provides clear bounds for the relative uncertainty of all data, and it allows the usage of an arbitrary information hierarchy. In the conventional biproportional problem the results of the three methods are similar, but the MEP method can be applied in a wider range of circumstances, such as considering multiple relative uncertainties in the same information level, having more than two information levels or arbitrary topological constraints.

7 Case-study

In this Section we consider a simple case-study to illustrate the behaviour of the maximum entropy estimation method. We try to reconcile an inconsistent table with two information levels. After describing the data we compare different estimation methods and study the properties of the Bayesian solution.

We use the 59-sector national symmetric IO tables in current prices for Portugal for the years 1995, 1999 and 2005, available from EUROSTAT (http://epp.eurostat.ec.europa.eu/portal/page/portal/esa95_supply_use_input_tables/data/workbooks). We consider three scenarios. In scenario 1 we use the original 1995 table, in scenarios 2 and 3 total input/output, factor payments and final demand are taken from the year 1995 while the inter-industry transactions have the production structure of the years 1999 and 2005 respectively. That is, the inter-industry tables in scenarios 2 and 3 were obtained, respectively, as $\mathbf{Z}_{99}\hat{\mathbf{x}}_{99}^{-1}\hat{\mathbf{x}}_{95}$ and $\mathbf{Z}_{05}\hat{\mathbf{x}}_{05}^{-1}\hat{\mathbf{x}}_{95}$, where $\mathbf{Z}$ is the inter-industry matrix, $\mathbf{x}$ is the vector of total output and the subscript denotes the year. The global inconsistency in the three scenarios is 0.165, 2014.4 and 3784.2 $\times 10^6$ Euro, respectively and the ratio of the inconsistency to total output is 0.00005%, 2.5% and 5%. Thus, Scenarios 1-3 pose a problem of increasing inconsistency. In all scenarios we consider that the points of the inter-industry matrix are of information level 1 (low information quality) and the remaining points (final demand, primary inputs and total output) are of information level 2 (high information quality), where a higher information level means that the data is more trusted and is only adjusted if lower information data is inconsistent.

The total of non-empty IO entries in the three scenarios are, respectively, 2559, 2623 and 3109 (these numbers differ because different cut-off values have been used). The total number of topological constraints is 180, accounting for all row and column sums and the identity between total input and total output. There are two constraints per transaction in the inter-industry matrix or total input/output, and one constraint per transaction in final demand or primary input transactions. Of the topological constraints corresponding to row or column sums, 90% aggregate more 20 transactions and 72% aggregate more than 40 transactions, so we are under the conditions of the application of the Bayesian, described in Section 6.

According to Lenzen (2001) and Lenzen et al. (2010), the relative uncertainty of empirical IO data often decreases with the magnitude of the best guess in a power-law fashion. Nhambiu (2004) reports that the relative uncertainty of the Portuguese 1995 IO table varies from 27% for the smaller entries to 13% for the larger entries. Since the magnitude of best guesses ranges from 0.001 to 15000 $\times 10^6$ Euro, the best fit of a power-law function to the relation between the prior relative uncertainty, ν, and the prior best guess, μ, yields:

$$\nu = 0.2\mu^{-0.045}.$$

This expression is valid for all data of information level 2. In scenarios 2 and 3, the entries of the inter-industry matrix must be assigned higher uncertainty than in scenario 1, since they have been obtained by a non-survey method. We set their uncertainty as the average between the survey data uncertainty and the maximum admissible uncertainty. Therefore, we consider that the uncertainty of the inter-industry matrices in scenarios 2 and 3 (data of information level 1) is:

$$\nu = 0.5\left(1 + 0.2\mu^{-0.045}\right).$$

It must be checked that $0 < \nu < 1$ for all priors. Later on we make a sensitivity analysis to the choice of relative uncertaity of data of information level 1.

Besides the Bayesian method, which we shall refer to as MEP, we shall consider the SWLS and KRAS methods, also described in Section 6. In the SWLS method we consider that the data of information level 2 has an inverse reliability index of 1, and that the data of information level 1 has an inverse reliabililty index of 10000. We observed empirically that with these index values it was possible to simulate the information hierarchy.

In the KRAS method we set parameter $a = 1$ and applied the RAS procedure to the data of information level 1, with the data of information level 2 aggregated as numerical constraints. When no further improvement could be obtained, data of information level 2 was disaggregated and data information level 1 was aggregated. For each topological constraint the disaggregated data (the conflicting numerical constraints in the original problem) were now adjusted as:

$$m_j = \mu_j - G_{ij} a \epsilon_i \frac{\sigma_j}{\sum_k |G_{ik}| \sigma_k},$$
$$\epsilon_i = \sum_j G_{ij} \mu_j + \bar{m}_i.$$

That is, each numerical constraint is linearly adjusted in proportion to the error of the topological constraint, ϵ_i, and to its standard-deviation, σ_j, but this weight is now normalized, so that if $a = 1$ the error of the topological constraint is eliminated.

We considered that a consistent solution had been found when the average mean quadratic error, ϵ, was lower than one Euro:

$$\epsilon = \frac{1}{n_K} \sqrt{\sum_i \epsilon_i^2},$$

where ϵ_i is the error of constraint i and n_K is the total number of constraints. The initial best guess average mean quadratic error in the three scenarios was, respectively, 866 EUR, 11.7×10^6 EUR and 28.5×10^6 EUR. Notice that ϵ is defined per constraint, so that it is possible compare the results for systems with different numbers of constraints.

In the following analysis we define the distance between two solutions as:

$$\delta = \frac{1}{n_T} \sqrt{\sum_j \delta_j^2},$$

where δ_j is the distance between two individual data points and n_T is the number of data points. Notice that δ is defined per transaction, so that it is possible compare the results for systems with different numbers of transactions.

All programming was made in Octave, with linear algebra performed in sparse format. The calculations were performed in a personal computer with a dual core CPU at 2.6 GHz and 4 GB of RAM.

All the methods considered are constructed to deal with potentially inconsistent constraints and they all did in fact produced balanced tables. In the MEP and KRAS methods the solution is always meaningful (i.e., positive best guess and positive and less than unitary relative uncertainty), while in the SWLS method such is not guaranteed. In fact, in scenario 3 the SWLS method generated 52 negative entries, while in all other scenarios and for all other methods no entry changed sign.

In general the computation time increases with the amount of initial inconsistency (i.e, from scenario 1 to 3), and it differs substantially between methods. SWLS is the fastest method (0.13 to 0.20 seconds), since it makes use of highly optimized routines to solve linear systems and does not require any adjustment to data. The computation time of the MEP method is one order of magnitude slower ($1.3 - 1.6s$), and most of this time is spent in the aggregation or disaggregation of data required by the information hierarchy. The computation time of the KRAS method is still an order of magnitude slower ($17.8 - 45.9s$), due to its iterative nature. The RAS routine invoves nested FOR statements which are time consuming in a high-level language, but would not be so in a low-level one.

We calculated the distances, respectively, to the target configuration (the prior of scenario 1) and to the source configuration (the prior in the corresponding scenario), and found that they exhibited significant differences between scenarios but not between methods. In scenario 1 the distances to source and target are in the range of 47 to 55 EUR. The distances to target are 260.7 to 261.9 $\times 10^3$ EUR in scenario 2 and 630.5 to 658.3 $\times 10^3$ EUR in scenario 3. The distances to source are 393.1 to 399.1 $\times 10^3$ EUR in scenario 2 and 767.4 to 837.6 $\times 10^3$ EUR in scenario 3. For the sake of comparison the initial distance between scenarios 2 and 1 and scenarios 3 and 1 is, respectively, 497.04×10^3 and 1086.20×10^3 EUR. Therefore, in both scenarios 2 and 3 the posterior is roughly half as close to the target as the prior was, and the distance from prior to posterior is roughly 80% of distance from prior to target, for all methods.

The displacement (from prior to posterior) for an individual entry was strongly correlated with the magnitude of the prior. We calculated log-log linear regressions for all methods and scenarios, $|m_j - \mu_j| = a\mu_j^b$, and found that the determination coefficients were high for all methods and scenarios, in the case of low quality data, with $R^2 > 0.6$ in scenario 1 and $R^2 > 0.8$ in scenarios 2 and 3. In the case of high quality data, the displacement was overall small and the correlations were somewhat lower, dropping to $R^2 = 0.4$

for KRAS in scenario 3. This means that in most cases, more than half of the adjustment from prior to posterior can be explained by the magnitude of the prior best guess.

For low quality data the slopes are in the ranges $0.80 < b < 0.85$, $0.95 < b < 0.97$ and $0.93 < b < 0.94$, respectively for the three scenarios. The corresponding values for high quality data are $0.85 < b < 0.90$, $0.82 < b < 0.91$ and $0.72 < b < 0.85$. That is, for low quality data all methods yield very similar results. For high quality data there is a higher variation in results, but we should not forget that overall high quality data was very little adjusted, the same as low quality data in scenario 1 – in all these cases the scalar coefficient a is very small, $a < 10^{-5}$, while for low quality data in scenario 2 we observe that $0.11 < a < 0.12$ and in scenario 3 that $0.28 < a < 0.30$.

To check the effect of the relative uncertainty of low quality data in the results we considered several cases, in which the median of this data is shifted from 20% (identical to high quality data) to 100% (the worst case scenario) and the slope is shifted accordingly. We found that the relative difference in results, for low quality data in scenarios 2 and 3 was less than 2% while for low quality data in scenario 1 it was 60% and for high quality data it was close to 12%. The values for low quality data in scenario 1 and high quality data must be regarded with caution since the overall adjustment in all these cases was very small. When the relative adjustment was meaningful (low quality data in scenarios 2 and 3) the difference between results was minimal, as we expected from our discussion in Section 6: if the relative uncertainty is identical for all data in the first information level to be adjusted, the results are independent of that value.

8 Conclusions

In this paper we presented a Bayesian estimation method for Input-Output (IO) Analysis, which can reconcile possibly conflicting data of arbitrary form, taking into account the uncertainty of the source data.

In a first part of the paper (Sections 3-4) we derived the Bayesian properties of IO quantities and an analytical expression of the maximum entropy consistent posterior solution, given an inconsistent prior configuration. IO quantities are strictly positive quantities, of which only best guess and uncertainty may be known. In this circumstance, application of the maximum entropy principle (MEP) shows that the underlying distribution of an IO quantity is a truncated Gaussian, whose relative uncertainty is bounded by zero and one.

We allow each data point to have a given level of information quality and impose that lower quality inconsistencies should not affect higher quality data, if the latter is fully consistent. As discussed in Section 5 the MEP method should respect this information hierarchy, and therefore the method must be applied recursively, so that at each step the higher quality data is held fixed and lower quality data is adjusted. The information quality level currently adjusted is relaxed until a consistent solution is obtained.

In Section 6 we found that the MEP analytical solution has a simple and elegant form in the limit of low relative uncertainties given by Eqs. 6.1-6.2. This solution is a generalized least square problem that can be applied if all prior information is available, including correlations. The latter are unlikely to be known so we derive a simple approximation, Eqs. 6.3-6.4 which does not involve correlations and it is valid in typical IO situations (correlations different from zero, a high number of transactions per constraint and a low number of constraints per transaction).

In Section 7 we considered a typical biproportional problem with mildly conflicting row and column sums. We observed that, in this particular context and as we expected from theoretical considerations, the MEP and currently existing methods yielded very similar results. The MEP approximation combines and generalizes features of the recently proposed SWLS and KRAS methods. The MEP method allows for the specification of the uncertainty of each data point within defined bounds, the consideration of a multi-tiered information hierarchy and of arbitrary topological constraints.

The MEP method provides a consistent solution that takes into account all available information and whose displacement from the available data is minimally informative. However, there is no guarantee that the solution will be close to an unknown target solution. If the prior is very different from the target, it is likely that the same will happen to the posterior. A good data reconciliation method is no substitute for the gathering of accurate source data.

Acknowledgments

This work was partially financed by Fundação para a Ciência e Tecnologia through the PETE project (grant PTDC/AMB/64762/2006). This work has benefitted from comments by Tiago Domingos, Gonçalo Marques and others, including two anonymous referees. Any errors it may contain are the sole responsibility of the author.

References

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.

E. T. Jaynes. *Probability Theory: The Logic of Science*. Cambridge University Press, Cambridge, UK, 2003.

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 inputâĂŞoutput 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. Second Edition.

J. O. P. Nhambiu. *AvaliaÃğÃčo do ciclo de vida de produtos com recurso a quadros econÃşmicos de entrada-saÃηda: aplicaÃğÃčo a Portugal.* Instituto Superior TÃľcnico, Universidade TÃľcnica de Lisboa, Lisboa, Portugal, 2004.

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*, 27:379âĂ§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.

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.