Bayesian input-output table update using a benchmark LASSO prior

Mike G. Tsionas

December 16, 2019

**Abstract**

We propose updating a multiplier matrix subject to final demand and total output constraints, where the prior multiplier matrix is weighted against a LASSO prior. We update elements of the Leontief inverse, from which we can derive posterior densities of the entries in input-output tables. As the parameter estimates required by far exceed the available observations, many zero entries deliver a sparse tabulation. We address that problem with a new statistical model wherein we adopt a LASSO prior. We develop novel numerical techniques and perform a detailed Monte Carlo study to examine the performance of the new approach under different configurations of the input-output table. The new techniques are applied to a 196 ×196 U.S. input-output table for 2012.

**Key Words**: input-output tables; Bayesian inference; Markov Chain Monte Carlo; LASSO priors.

**JEL Codes**: C11, C13, C30, D57, E17.

**Acknowledgments**: The author is indebted to the Editor Michael Lahr, three anonymous reviewers and Gerald Steele for providing useful comments on earlier versions. The usual disclaimer applies.

## 1. Introduction

The problem that concerns us has been stated clearly by Golan, Judge, and Robinson (1994, p.541):

> Commonly, one starts with complete data for a particular period and seeks to estimate the matrix of flows for a later period, based only on row and column sum information. In general, the problem is to recover, from the incomplete data, a new matrix that satisfies a number of linear restrictions

A well-known means of solving this problem is RAS (Bacharach, 1970); a prominent alternative is the entropy method (Golan, Judge, and Robinson, 1994). RAS is the most well-known and widely used method if we include its variants, such as generalized RAS algorithm (Günlük-Senesen and Bates, 1988; Junius and Oosterhaven, 2003; Lahr and De Mesnard, 2004). The basic idea of RAS and its extensions is minimizing the dissimilarity between the existing table and the target table. RAS is based on the assumption that we know the row and column sums of the matrix to be estimated. Some techniques also employ a range of partial information (Lenzen et al., 2006, 2009, 2013; Wood, 2011; Tukker et al., 2013). Of the RAS alternatives, perhaps that by Lenzen et al. (2006, 2007, 2009) is the most general; called '*Konfliktfreies* RAS' (KRAS), it balances and reconciles input-output (IO) tables under conditions of conflicting external information and inconsistent constraints. Another prominent technique is using least squares to minimize the relative difference between a prior and the target table (for a description and variants see Miller and Blair, 2009, Chapter 7; Robinson et al., 2001; and Canning and Wang, 2005).

For econometric procedures in IO analysis, see, for example, Gerking (1976) and Kockläuner (1989). See Temursho (2017), for a more detailed discussion of this topic, where, for example, the probability density of the Leontief inverse is spelled out for any given probability density of the direct input coefficients matrix as per Fox and Quirk (1985Kop Jansen and ten Raa (1990) characterize the problem in an axiomatic fashion. which may be of considerable empirical importance. Rueda-Cantuche and Amores (2010) also use least squares in the context of IO analysis; their number of observations is greater than that of estimated parameters—the usual case in least-squares regressions. In ten Raa and Rueda-Cantuche (2007), a "large $n$, small $p$" paradigm is at work as well: Specifically, there are more activities than commodities and output multipliers can be estimated as regression coefficients. Counter to the usual

econometric approach, in this paper the number of elements in the input-output table is greater than the number of observations, the well-known "small $n$, large $p$" paradigm (where $n$ denotes the number of observations and $p$ is the number of parameters). This disables the use of ordinary least-squares techniques. For other studies using the econometric approach, see Lynch (1986), ten Raa and Steel (1994) and the references therein. Least squares and instrumental variables estimation also have been discussed by Gerking (1976).

So the present problem is different and much more difficult than anything that can be handled by RAS or regular econometric techniques. Moreover, we want to estimate/update the IO table elements, not multipliers. In this vein, related studies include those by Rodrigues, Amores and Paulo (2019), Torres-González and (2019), and Valderas-Jaramillo et al. (2019) among others

In this paper, we propose a novel Bayesian approach to estimate an IO table given a prior table exists, along with a LASSO ("least absolute shrinkage and selection operator") prior. A LASSO prior is used since sparsity is typically a characteristic feature of IO tables. The use of a Bayesian analysis is fairly novel in IO analysis. Rodrigues (2014) has proposed also a Bayesian approach, but it focuses on entropy-based techniques. The advantage of the LASSO relative to, for example, a normal or $L_2$ prior, is that when a coefficient is zero, then it is estimated exactly as zero, not as a "small" number. Given the abundance of zero elements in typical IO matrices, this is a significant advantage of the LASSO prior. More precisely, another alternative is to use ridge regression, which is equivalent with an $L_2$-norm prior on the regression coefficients. The problem is that $L_2$-norm never enforces zeros when appropriate, whereas the LASSO is based on the $L_1$-norm, which does enforce such constraints (Tibshirani, 1996; Figueiredo, 2003; Bae and Mallick, 2004; Yuan and Lin, 2005; Park and Casella, 2008). One may think that as a result, some nonzero elements of updated matrix are transformed to zero, indicating that the corresponding variables are not contributing to the model; so, using this approach may lead to the misrepresentation of the interindustrial flow structure. It is rare that elements of small IO tables are zero, whereas in much larger tables this is quite often the case. As we use a prior table to revise the IO matrix, the $L_1$-approach is better suited when prior tables contains zeros, whereas the $L_2$-approach should prove better when such zeros are absent. It is also

possible to use both approaches and compare them formally in the light of the data using the concepts of marginal likelihood and Bayes factors.

As I mentioned earlier, the use of least squares and other econometric techniques in the context of IO analysis is well-established. In fact, in this paper I show that the problem of updating IO matrices reduces to a least-squares problem in which the prior on the coefficient vector has to account for (1) the abundance of zeros in the original IO table and (2) the fact that, in this problem, the number of observations is much less the number of observations. For example, in the case of a $10 \times 10$ table, the number of unknown elements (or "$\beta$" in the least-squares problem) is 100. If final demand is available for 10 years and the assumption of approximate constancy of IO coefficients can be made, then we the sample has a size of 10. Clearly, estimating 100 parameters from 10 observations is an ill-posed problem. I will use "ill-posed" in this sense throughout the paper. I should mention, however, that incorporating prior information in least-squares problems is possible through a "mixed estimation" approach (Theil and Goldberger, 1961). But some elements of the prior must be selected by the analyst; This can be avoided, for the most part, by using formal Bayesian analysis.

There are special Bayesian techniques that deal with ill-posed problems of this sort. A prominent one is the LASSO prior, which can deliver at most 10 nonzero elements out of the 100 that we seek. This is not an outcome that most analysts would consider to be reasonable. The purposes of this paper are as follows. First, it delivers posterior inferences for a new IO table (or Leontief inverse). Second, in doing so, I propose a way to craft the prior using information from the LASSO, a normal prior as well as a benchmark IO table (or Leontief inverse). The prior is crafted so that with probability $\varpi$ we choose the benchmark, and with probability $1 - \varpi$ we choose the LASSO prior. Naturally, $\varpi$ is unknown and has to be determined from the data.

Another novelty of the paper is that I introduce a prior IO table (or Leontief inverse) corresponding to a benchmark year. The table to be estimated and the prior table need not be the same or similar vintage. If they are, however, this clearly increases the precision of the estimated table. If not, the user can control the effect of the prior table through a precision parameter that shows how close the tables are. This parameter can be estimated from the data given some other underlying prior information about its prior mean and prior standard deviation. To see how powerful Bayesian analysis is, we

focus on Theil and Goldberger's (1961) approach. We have a linear regression model given by

$$y = X\beta + u,\ u \sim N(\mathbf{0}, \sigma^2 I_n),$$

where $X$ is the $n \times k$ matrix of regressors and $y$ is the $n \times 1$ vector of observations on the dependent variable. A frequentist believes that vector $\beta$ has mean $b$ and covariance $\omega^2 I_k$ so that $\beta \sim N(b,\ \omega^2 I_k)$, where $b$ and $\omega$ are known. The frequentist does not use Bayesian analysis but rather by a "leap of faith" perhaps he can be convinced that $b$ is part of the data as it represents his "prior belief". So, we can write the model as follows:

$$y = X\beta + u,$$
$$b = I_k\beta + v,$$

where $v \sim \mathcal{N}(\mathbf{0},\ \omega^2 I_k)$. If we use the generalized least-squares estimator,[1] we end up with the following estimator, after a bit of algebra:

$$\hat{\beta} = (\omega^2 X'X + \sigma^2 I_k)^{-1}(\omega^2 X'y + \sigma^2 b) = (X'X + \frac{\sigma^2}{\omega^2} I_k)^{-1}(X'y + \frac{\sigma^2}{\omega^2} b).$$

Clearly, as $\omega \to \infty$, $\hat{\beta}$ converges to the least-estimator $\tilde{\beta} = (X'X)^{-1}X'y$. If $\omega \to 0$ then we obtain the prior mean, $b$. The important lesson here is that if $\sigma^2$ can be estimated (which is, of course, possible) and the frequentist can provide $\lambda = \frac{\sigma^2}{\omega^2}$ (which is the inverse of a noise-to-signal ratio) then, it does not matter if $X'X$ is perfectly collinear or whether $n > k$ or $n < k$! So, the reader may wonder why we do not make use of this simple device ("mixed estimation") to address the problems of this prior. First of all, we do not want to assume outright a value for $\lambda$ (which when $b = \mathbf{0}$ corresponds to ridge regression!). Second, there is no reason to assume a normal prior for $\beta$. Third, one can combine a normal prior and a Laplace prior as the latter is more robust. Fourth, the importance of data versus priors and normal versus Laplace priors, cannot be set outright by the user but a more-subtle procedure should be used to determine such key parameters of the problem from the data themselves.

Given the ill-posed nature of the problem of updating IO tables, this paper contributes to the literature by showing how Bayesian analysis can update IO tables,

[1] Writing $b = I_k\beta + v$ instead of $\beta \sim N(b, \omega^2 I_k)$, which is equivalent to $\beta = I_k b + v$ is, of course, unacceptable from the point of view of both frequentist and Bayesian purists. Taking this "leap of faith" can, actually, be traced back to Fisher (1939) who, in a simpler context, called this produce by the name of "fiducial inference". Nonetheless, $\hat{\beta}$ is exactly equal to the posterior mean through formal Bayesian analysis! As a matter of fact, in the author's view, this is a good way for students or novices to be introduced into the "apocrypha" of Bayesian analysis.

given a prior benchmark input-output table. This techniques performs well in extensive Monte Carlo experiments and an actual application to the $196 \times 196$ table for the U.S. in 2012. A novel Markov Chain Monte Carlo (MCMC) technique is applied to implement Bayesian inference. We examine their performance through a detailed Monte Carlo study involving large IO tables. The new technique performs well even when sparsity is not prevalent.

## 2. Bayesian updating of input-output coefficients

### *2.1. Model*

Consider the basic IO model:

$$y_i = \sum_{j=1}^{n} z_{ij} + f_i, i = 1, \dots, n, \tag{1}$$

where $y_i$ is production of sector $i \in \{1, \dots, n\}$, $z_{ij}$ represents intersectoral flows flows from sector $i$ to sector $j$, and $f_i$ is final demand. Suppose

$$z_{ij} = \alpha_{ij} y_j, \tag{2}$$

where $\alpha_{ij}$ are input-output coefficients. From (9) and (2) we have:

$$y_i = \sum_{j=1}^{n} \alpha_{ij}\, y \quad + f_i, i = 1, \dots, n. \tag{3}$$

In vector notation:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{f}, \tag{4}$$

where $\mathbf{f} = [f_1, \dots, f_n]'$ and, therefore:

$$\mathbf{y} = (\mathbf{I} - \mathbf{A})^{-\mathbf{1}}\mathbf{f}. \tag{5}$$

Define

$$\mathbf{B} = (\mathbf{I} - \mathbf{A})^{-\mathbf{1}}, \tag{6}$$

the Leontief matrix, so that

$$\mathbf{y} = \mathbf{B}\mathbf{f}. \tag{7}$$

Notationally, I use $\mathbf{B}$ instead of the more familiar $\mathbf{L}$ to represent the Leontief inverse since parameters of $\mathbf{A}$ will be denoted by $\alpha_{ij}$ and elements of $\mathbf{B}$ by $\beta_{ij}$, which is the standard in statistical literature. One advantage of the Bayesian approach

is that given the posterior distribution of entries in $\mathbf{A}$ we can derive easily the posterior distribution of entries in $\mathbf{B}$ and vice versa. Let $\mathbf{B} = [\beta_{ij}, i, j = 1, \ldots, n]$.

Typically, a benchmark IO table is known and the problem is to update it (Bacharach, 1970; Golan, Judge and Robinson, 1994). In the approach presented here, the system in (7) is rewritten as follows:

$$\begin{gathered} y_1 = \beta_{11} f_1 + \beta_{12} f_2 + \cdots + \beta_{1n} f_n, \\ y_2 = \beta_{21} f_1 + \beta_{22} f_2 + \cdots + \beta_{2n} f_n, \\ (\ldots) \\ y_n = \beta_{n1} f_1 + \beta_{n2} f_2 + \cdots + \beta_{nn} f_n, \end{gathered} \tag{8}$$

where $\beta_{ij}$s are elements of the Leontief inverse, $\mathbf{B} = (\mathbf{I} - \mathbf{A})^{-1}$ as in (5) and (6). Define $\mathbf{y} = [y_1, \ldots, y_n]'$.

Therefore, we can write (8) as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \tag{9}$$

where[2]

$$\begin{aligned} \boldsymbol{\beta} &= (\beta_{11}, \beta_{12}, \ldots, \beta_{1n}, \ldots, \beta_{n1}, \beta_{n2}, \ldots, \beta_{nn})' = \text{vec}(\mathbf{B}) \\ &= \text{vec}[(\mathbf{I}_n - \mathbf{A})^{-1}], \end{aligned} \tag{10}$$

is the $n^2 \times 1$ vector of stacked rows of the Leontief matrix,

$$\mathbf{X} = \begin{bmatrix} \mathbf{f}' & \mathbf{0}' & \ldots & \mathbf{0}' \\ \mathbf{0}' & \mathbf{f}' & \ldots & \mathbf{0}' \\ & \ldots & \ldots & \\ \mathbf{0}' & \mathbf{0}' & \ldots & \mathbf{f}' \end{bmatrix} = \mathbf{I}_n \otimes \mathbf{f}', \tag{11}$$

and $u$ represents a statistical error term. In (9) we have a standard linear model where *the number of coefficients $n^2$ exceeds the number of observations $n$ for a given time period*. In the statistical literature, the situation is known as "small $n$, large $p$", where, now, $n$ stands for the number of observations and $p$ denotes the number of unknown parameters. In this instance, it is clear that application of standard least-squares or other econometric techniques is not possible. A multi-year formulation of the problem is possible and details are provided in Appendix A in the Supplementary file.

Given an existing $n \times n$ input-output matrix $\mathbf{A}_o$ that results in a Leontief matrix whose rows are stacked in the $n^2 \times 1$ vector $\alpha_o$, we expect that $\alpha$ and $\alpha_o$ are not very dissimilar. Suppose also that elements of $u$ are normally distributed with zero mean and variance $\sigma^2$. Let

[2]For any matrix $\mathbf{M}$, the vec operator stacks rows of $\mathbf{M}$ to a vector..

$$\boldsymbol{\alpha} = \text{vec}(\mathbf{A}), \tag{12}$$

so that $\boldsymbol{\alpha}$ is $n^2 \times 1$. Our prior information is about $\boldsymbol{\alpha}$, not $\boldsymbol{\beta}$.

Before proceeding, I should note that the social accounting matrix (SAM) approach is a special case if we set $\mathbf{f} = \mathbf{0}$, a zero vector, and impose the additional constraints: $\sum_{j=1}^{n} a_{ij} = 1, \forall i = 1, \ldots, n$ (Golan, Judge and Robinson, 1994). In this case, we may or may not have prior information about $\alpha$, depending on whether a benchmark matrix is available. In fact, if there is prior information in the form of additional constraints, this is quite easy to impose using Bayesian analysis.

*2.2. Priors in input-output models*

As mentioned before, the prior information is mostly about $\boldsymbol{\alpha}$, see (12), not $\boldsymbol{\beta}$, viz. the elements of the Leontief matrix. Specifically, for the elements of $\mathbf{A}$, we must have:

$$0 \leq a_{ij} \leq 1 \; \forall i, j = 1, \ldots, n. \tag{13}$$

In SAMs, in addition, we must have (Golan, Judge, and Robinson, 1994):

$$\sum_{j=1}^{n} a_{ij} = 1 \; \forall i = 1, \ldots, n. \tag{14}$$

Any prior for $\boldsymbol{\alpha}$ must be subject to these restrictions which establish nonlinear relationships between $\beta$ and the elements of $\mathbf{A}$, since $\mathbf{B} = (\mathbf{I}_n - \mathbf{A})^{-1}$, see Temursho (2017) and Kop Jansen and ten Raa (1990). In a Bayesian framework, the issue has an easy solution taking full account of measurement errors.

Before proceeding, it is important to describe briefly the LASSO procedure (Tibshirani, 1996; Park and Casella, 2008). Suppose a linear model of the form $\mathbf{y} = \mathbf{X}\boldsymbol{\theta} + \mathbf{e}$ where $\mathbf{X}$ is $n \times p$, $\boldsymbol{\theta}$ is $p \times 1$, $\boldsymbol{e}$ is an $n \times 1$ error term, and $\mathbf{y}$ denotes the $n \times 1$ vector of observations on the dependent variable. We use bold symbols for the dependent variable and the matrix of regressors, as these may not correspond to (9). A major reason is that elements of $\boldsymbol{\beta}$ in (9) depend on $\boldsymbol{\alpha}$ and sparsity information relates to $\boldsymbol{\alpha}$. Here, $\boldsymbol{\theta}$, however, is a usual regression parameter.

We believe that certain or most elements of $\boldsymbol{\theta}$ are zero as many predictors in matrix $\mathbf{X}$ may be irrelevant. Clearly, a penalization term must be used in least-squares estimation. The problem solved by LASSO is: $\min_{\boldsymbol{\theta}}: (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) + \lambda \sum_{j=1}^{p} |\theta_j|$ where $\lambda \geq 0$ is the LASSO penalization parameter. Larger values of $\lambda$ place more

emphasis on the penalization term and imply that more elements of the regression coefficient zero are likely to be zero. Notice that a quadratic penalty of the form: $\min_{\boldsymbol{\theta}}: (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\theta}) + \lambda \sum_{j=1}^{p} \theta_j^2$ would not be consistent with placing (unknown) zero restrictions on $\boldsymbol{\theta}$ contrary to the $L_1$ penalty term used in LASSO. The reason is that the LASSO problem is equivalent to: $\min_{\boldsymbol{\theta}}: (\mathbf{y}-\mathbf{X}\boldsymbol{\theta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\theta}), \text{s.t.} \sum_{j=1}^{p} |\theta_j| \leq u$ for a certain value of $\tau$ which determines the amount of "regularization". Ridge regression is based on minimizing the same objective with a different constraint: $\sum_{j=1}^{k} \theta_j^2 \leq \tau'$ for a certain value of $\tau'$, viz. ridge regression uses an $L_2$ penalty or regularization term. Unlike $L_2$, $L_1$ −penalization results in certain zeros in the optimal solution. In revising an IO table, it is clear that many elements will actually be zero. Therefore, the LASSO prior seems appropriate in this setting.

The LASSO prior (Tibshirani, 1996; Figueiredo, 2003; Bae and Mallick, 2004; Yuan and Lin, 2005; Park and Casella, 2008) leaves unanswered the question of how to use a benchmark $\alpha_o$ resulting from an existing $\mathbf{A}_o$. LASSO stands for "least absolute shrinkage and selection operator". It is true that we could have used ridge regression as well in our context. To anticipate some of our results in Monte Carlo simulations, as expected, ridge regression results are worse compared to the benchmark – LASSO approach, as ridge regression corresponds to $L_2$-regularization. Therefore, if there are many zero entries in an IO table, these cannot be estimated as exactly zero and, as a result, the mean squared error of ridge regression is inflated, particularly when the dimensionality of the IO matrix is large and the degree of sparsity (number of zero entries) is moderate to large. These results are available in Appendix C of the Supplementary file. Moreover, in the statistical literature it has been shown that the LASSO has an "oracle property" in the sense that, asymptotically, model selection / variable selection work as if we knew the true model / true variables in the model. In the context of IO analysis, this means that this approach does not lead to the misrepresentation of the interindustry flow structure. This result is of considerable interest as it implies that the true interindustry flow structure can be recovered due to the "oracle property" of the LASSO (Zou, 2012).

To illustrate the situation, suppose $n = 2$ so that, in the context of (4) or (8), we have:

$$y_1 = \beta_1 f_1 + \beta_2 f_2 + u_1,$$
$$y_2 = \beta_3 f_1 + \beta_4 f_2 + u_2,$$

as in (8) with the error terms in (9). The parameters $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_4]'$ are elements of the Leontief inverse. Clearly, we have two observations but four parameters. Therefore, we must use prior information as the least squares problem:

$$\min_{\beta} \ (y_1 - \beta_1 f_1 - \beta_2 f_2)^2 + (y_2 - \beta_3 f_1 - \beta_4 f_2)^2,$$

does not have a solution. A normal prior $\boldsymbol{\beta} \sim N_4(\boldsymbol{\beta}_o, \mathbf{V})$ with mean $\boldsymbol{\beta}_o = [\beta_{o,j}, j = 1, \ldots, 4]'$ and covariance matrix $\mathbf{V} = \lambda_1^{-1}\mathbf{I}_4$ can certainly be used. In this case, the corresponding problem would be:

$$\min_{\beta} \ (y_1 - \beta_1 f_1 - \beta_2 f_2)^2 + (y_2 - \beta_3 f_1 - \beta_4 f_2)^2 + \lambda_1 \sum_{j=1}^{4} \left(\beta_j - \beta_{o,j}\right)^2.$$

In this problem, we are trying to make fit as best as possible and keep the $\beta$s as close to $\beta_o$ as possible with a weight which is given by prior precision (viz. the inverse of the variance, $\lambda_1$).

If we use a LASSO prior the corresponding problem would be:

$$\min_{\beta} \ (y_1 - \beta_1 f_1 - \beta_2 f_2)^2 + (y_2 - \beta_3 f_1 - \beta_4 f_2)^2 + \lambda_2 \sum_{j=1}^{4} |\beta_j|.$$

In this problem, we are trying to make fit as best as possible but impose the notion of sparsity in the sense that we expect "many" elements of $\boldsymbol{\beta}$ to be zero. In the context of IO tables, we know that sparsity usually prevails particularly when $n$ increases. A modification of the LASSO formulation would be:

$$\min_{\beta} \ (y_1 - \beta_1 f_1 - \beta_2 f_2)^2 + (y_2 - \beta_3 f_1 - \beta_4 f_2)^2 + \lambda_2 \sum_{j=1}^{4} |\beta_j - \beta_{o,j}|.$$

In this problem we are trying to make fit as best as possible and impose the notion that certain elements of β are *exactly* equal to the benchmark $\boldsymbol{\beta}_o$. Clearly, at most two elements of $\boldsymbol{\beta}$ would be updated in this instance, as $n = 2$ and $p = 4$. This approach does not seem reasonable particularly when the benchmark refers to the distant past, and the sparsity prior seems more reasonable.

Our strategy is to modify the prior of $\boldsymbol{\alpha}$ as follows:

With probability $\varpi$, $\alpha$ follows a multivariate normal distribution:

$$\alpha \sim \mathcal{N}_{n^2}(\alpha_o, \eta \tilde{A}_o),$$

and with probability $1-\varpi$, $\alpha$ follows the LASSO prior:

$$p(\alpha|\eta) = \prod_{j=1}^{n^2} \frac{\lambda}{2\eta} e^{-\lambda|\alpha_j|}. \qquad (15)$$

where $\eta > 0$ is a smoothing parameter, $\lambda$ is a parameter that controls sparsity in the LASSO prior. We call this the "benchmark-LASSO prior" since it combines a benchmark matrix $\mathbf{A}_o$ with the LASSO. Here, $\widetilde{\mathbf{A}}_o = [\tilde{a}_{ij}^o]$, where

$$a_{ij} = \tilde{a}_{ij}^o \text{ if } \tilde{a}_{ij}^o > 0, \text{and } 1 \text{ otherwise}, i, j = 1, \ldots, \text{n}.$$

The introduction of matrix $\widetilde{\mathbf{A}}_o$ is necessary as (i) we want to make the covariance matrix in the first line of (15) dependent on elements of $\mathbf{A}_o$ to control for size differences and scaling issues, instead of being $\eta I_{n^2}$, and (ii) for zero entries in $\mathbf{A}_o$ we do not wish to have zero elements in $\mathbf{A}$ as well but, instead, allow for prior variation. Unfortunately, it does not seem a good idea to use different $\eta$s and / or $\lambda$s for each entry in $\boldsymbol{\alpha}$ as the proliferation of parameters would compromise what we are trying to do via both prior components in (15).

A natural question is whether we use (15) instead of simply using $\boldsymbol{\alpha} \sim \mathcal{N}_{n^2}(\alpha_o, \eta I_{n^2})$. This is certainly possible but we want the "best of two options": The first option is a LASSO prior and the second option is a standard normal prior as in the first line of (15). More specifically, in the so-called "elastic net" approach, the objective is:

$$\min_{\boldsymbol{\theta}} \ (\mathbf{y} - \mathbf{X}\boldsymbol{\theta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\theta}) + \lambda_1 \sum_{j=1}^{p} \theta_j^2 + \lambda_2 \sum_{j=1}^{p} |\theta_j|,$$

for certain weights $\lambda_1$ and $\lambda_2$. Therefore, if we set $\varpi = 1$ we use a normal prior and if $\varpi = 0$ we use the LASSO prior. Intermediate values of $\varpi$ allow for a model which is in the "elastic network" spirit. Additionally, in the "large $p$, small $n$" paradigm, the LASSO delivers at most $n$ non-zero entries. In general, the quadratic penalty (normal prior) delivers non-zero entries when, in fact, such entries are precisely zero. Therefore, (15) corrects the drawbacks of both a normal prior and the LASSO. This correction is made with an eye towards an update of IO tables as precisely as possibly, realizing that deviations between actual and estimated entries is due to both (i) ignoring the difference between non-zero entries that should, in fact, be zero, when using a normal prior, as well as (ii) having at most $n$ non-zero entries when, in fact, many more could have

been non-zero, as in LASSO. To the best of our knowledge, there is no easy or automatic procedure to determine the weights $\lambda_1$ and $\lambda_2$. Using (15) we have such a procedure, which seems to be novel in the literature.

Notice that this prior is *not* conjugate as in typical LASSO applications. The probability $\varpi$ is assumed unknown, and follows a *beta* distribution:

$$\varpi \sim Be(P, Q). \tag{16}$$

A relatively "uninformative" prior is when $P = Q = 0.01$. As the smoothing parameter, $\eta$, is unknown, we assume a *gamma* prior:

$$\eta \sim Ga(r_\eta, d_\eta). \tag{17}$$

We set $r_\eta = d_\eta = 0.1$ so that the mean is 1 and the prior standard deviation is 3.16, which is fairly diffuse or "uninformative". Therefore, we have the model in (9) which we rewrite as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}_{\boldsymbol{\alpha}} + u, u \sim \mathcal{N}_n(0, \sigma^2 I_n), \tag{18}$$

where

$$\boldsymbol{\beta}_{\boldsymbol{\alpha}} = \text{vec}(\mathbf{I}_n - \mathbf{A})^{-1}, \boldsymbol{\alpha} = \text{vec}(\mathbf{A}), \tag{19}$$

and the prior on elements of $\boldsymbol{\alpha}$ is described in (15) and (16)-(17). This notation makes it clear that $\boldsymbol{\beta}$ depends on $\boldsymbol{\alpha}$ so, although we update the Leontief inverse, in fact, we can obtain an update for input-output coefficients in α. Although the point may seem trivial, it is important as the posterior density of entries in $\boldsymbol{\beta}$ can be derived easily based on the posterior density of entries in $\boldsymbol{\alpha}$.

Combining the likelihood of (18) and the prior we have the following posterior:

$$\begin{aligned} p(\beta_\alpha, \eta, \varpi, \sigma | y, X) \\ \propto \sigma^{-(n+1)} \exp\{ -\frac{1}{2\sigma^2}(y - X\beta_\alpha)'(y \\ - X\beta_\alpha)\} \cdot p(\beta_\alpha | \eta, \varpi) \cdot p(\eta) \cdot p(\varpi), \end{aligned} \tag{20}$$

where $p(\beta_\alpha | \eta, \varpi, \sigma)$ is the implied prior of $\boldsymbol{\beta}$ given the benchmark LASSO prior on $\boldsymbol{\alpha}$, conditional on the parameters $\eta$, $\sigma$ and $\varpi$, and $p(\eta)$, $p(\varpi)$ denote the prior densities of $\eta$ and $\varpi$ from (17) and (16) respectively. The prior of $\sigma$ is $p(\sigma) \propto \sigma^{-1}$, the standard "reference prior" for regression problems. The prior of $\boldsymbol{\beta}_{\boldsymbol{\alpha}}$ is defined over the set of restrictions in (13) and (14).

To set up a MCMC scheme, we first notice that it is easy to obtain draws from the posterior conditional distribution of $\sigma$ :

$$\frac{(y - X\beta_\alpha)'(y - X\beta_\alpha)}{\sigma^2}|y, X, \beta_\alpha, \eta, \varpi \sim \chi^2(n^2). \quad (21)$$

Moreover, we can obtain draws from the posterior conditional distributions of $\eta, \varpi$ following univariate inversion of the respective cumulative density functions (cdfs).[3] We obtain draws for $\boldsymbol{\beta}_\alpha$ using an efficient Girolami and Calderhead (2011) Langevin diffusion MCMC scheme using first- and second-order derivative information from the log posterior. For details about the Girolami-Calderhead algorithm the interested reader is referred to Technical Appendix B.

An alternative would have been the Metropolis-Hastings algorithm to obtain a long sample $\left\{\beta_\alpha^{(s)}, s = 1, \ldots, S\right\}$ which converges to the distribution whose unnormalized density is $p(\beta_\alpha|y, X)$. More details are provided in Appendix B of the Supplementary file. But the main problem is that with a large number of parameters we have not been able to maintain an acceptance rate close to 20-30% and, in fact, the rate has been zero or one. Another variation of the Metropolis-Hastings algorithm is to draw each parameter in $\beta_\alpha$ individually. Unfortunately, maintaining this rate is impossible for a large number of parameters, and it is, again, zero or unity. Moreover, the autocorrelation in the Metropolis-Hastings draws is substantial which means that it is, nearly, impossible to explore the posterior in finite time with certain numerical accuracy. The restrictions in (13) and (14) are enforced via rejection sampling. A Monte Carlo experiment to examine the behavior of new techniques, is presented in Technical Appendix C in the Supplementary file.

## 3. Empirical application

We use data from the Bureau of Labor Statistics[4] and specifically the $196 \times 196$ (Real, Use) Table for 2012. We use the 1993 version as a benchmark. In Figure 1 and panel (a) reported are actual and estimated ordered mod eigenvalues of **A**. In panel (b) we show actual versus estimated elements $\alpha_{ij}$ along with the 45-degree line (dotted). In panel (c) reported is the histogram of the percentage approximation error of $\alpha_{ij}$.

[3] For each parameter, say $\varpi$ we construct a grid consisting of 100 values on a support, which is adapted every 500 MCMC iterations during the burn-in phase. The cdf is computed using density values and then normalizing. In turn, we invert the cdf to obtain a random draw from the respective posterior conditional distribution. This is also known as "griddy Gibbs sampler".

[4] Obtained from http://www.bls.gov/emp/ep_data_input_output_matrix.htm.

**Figure 1: Posterior statistics for the 2012 US table (using 2010 as benchmark)**

**[Picture: *"Figure_1.EPS"* , Please insert here]**

**[Picture: *"Figure_2.EPS"* , Please insert here]**

**[Picture: *"Figure_3.EPS"* , Please insert here]**

We perform MCMC using as (alternative) benchmarks the tables for 1993, 2000, 2005 and 2010. The marginal posterior densities of $\varpi$ are reported in Figure 2. Evidently, tables that are closer to 2012 receive greater weight. However, the LASSO prior is still receiving weight due to the ill-posed nature of the problem and the significance of the LASSO in dealing with it. For example, using the 1993 table as benchmark, produces a posterior mean of $\varpi$ close to 0.35 with a minor mode close to 0.72. Using the 2000 table as benchmark, the posterior mean is 0.32 and ranges from 0.2 to 0.55. The table for 2005 has a posterior mean of $\varpi$ close to 0.5 and ranges from 0.35 to 0.7. Using 2012 as benchmark, $\varpi$ ranges from slightly over 0.8 to unity with a posterior mean close to 0.90. Therefore, the benchmark tables for 1993, 2000 and 2005 produce, on the average, values of $\varpi$ less than about 0.5, implying that the LASSO receives considerable weight. On the contrary, using the 2010 table as benchmark, places a weight (posterior probability) near 0.90 for the normal prior and nearly 0.10 for the LASSO. As the 2010 table should be close to the table for 2012, this is a reasonable outcome. However, it is quite interesting that LASSO still has posterior probability between 0 and 0.2 (viz. $1-\varpi$) which implies that zero elements in the 2010 table need to be replicated in the 2012 table. This is not possible using the normal prior as the update would be close to zero but not exactly zero.

**Figure 2. Marginal posteriors of $\varpi$ for different benchmark priors, $A_o$**

**[Picture: *"Figure_4.EPS"* , Please insert here]**

The important questions that we need now to address are the following:

- How well does the method perform?
- How well does it compare to other methods and priors?
- What happens when the benchmark table is more distant? How wide are the posterior intervals of large and small elements?

The first question is answered in Appendix C of the Supplementary Information file in which I report details of Monte Carlo experiments. To answer the second and third questions in a concise way, and in the context of the empirical application, I present

Figure 3, kernel densities of percentage approximation errors of RAS, LASSO ($\varpi = 0$), normal prior ($\varpi = 1$), and a simplified prior where we set $\varpi = 0.5$. Panel (a) is based on the 2010 table as benchmark and panel (b) on the 1993 table. This suggests that benchmarking and the use of a LASSO prior can produce accurate estimates of tables delivering smaller errors compared to what is known for methods such as RAS and other priors. As a matter of fact, the only density which is centered close to zero is the one corresponding to the prior in (15). Although using 1993 as the benchmark table (panel (b) of Figure 3) has much larger errors (ranging from -0.1 to 0.1% relative to -0.04% to 0.04% when the benchmark is the table for 2010) it still is the best performing method relative to RAS and the other priors.

**Figure 3: Densities of approximation errors of different techniques / priors**

**[Picture: *"Figure_6.EPS"* , Please insert here]**

**[Picture: *"Figure_7.EPS"* , Please insert here]**

**Concluding remarks**

In this paper, I propose new techniques in connection with updating input-output (IO) tables and social accounting matrices (SAMs). I use a stochastic representation of the IO model along with LASSO priors to derive posterior means of the updated IO tables and associated matrices. I present traditional measures for matrix comparison (like SRMSE, $\Psi$ and $\varphi$ statistics) are used and their posterior distributions. I use MCMC methods for the computations. The new methods appear to perform well in a Monte Carlo study in which sparsity of IO matrices is controlled. An empirical application to U.S. illustrates the new techniques. I apply data from the U.S. Bureau of Labor Statistics, specifically the $196 \times 196$ (Real, Use) Table for 2012 and use as alternative benchmarks and priors tables for 1993, 2000, 2005 and 2010. As Bayesian models organized around MCMC and a benchmark LASSO prior perform very well; the methods could thus well be of considerable practical use in empirical IO studies.

One could also apply this approach to estimate output multipliers or emission multipliers in the vein of Kop Jansen and ten Raa (1990), Rueda-Cantuche and ten Raa (2009), and Rueda-Cantuche and Amores (2010). In this context, least-squares techniques can be used as well since the multipliers can be represented as a parameter vector in a regression equation where the number of observations exceeds the number of parameters. Admittedly least-squares techniques provide unbiased estimators, but

they may still be of considerable interest to adopt Bayesian techniques as developed in the present paper due to the "oracle" properties of the LASSO but also because the degrees of freedom do not always yield the precision required to estimate such multipliers. Thus, Bayesian techniques based on MCMC and the LASSO may well provide much better finite-sample performance.

**References**


Bacharach, M. (1970) *Biproportional Matrices and Input-Output Change*. Cambridge University Press. Cambridge, U.K.

Bohlin, L., and L. M. Widell. (2006) Estimation of Commodity-by-Commodity Input-Output Matrices. *Economic Systems Research* 18, 205-215.

Brown, P. J., T. Fearn, and M. Vannucci. (1999) The Choice of Variables in Multivariate Regression: A Bayesian Non-Conjugate Decision Theory Approach. *Biometrika*, 86, 635–648.

Canning, P. and Z. Wang. (2005) A Flexible Mathematical Programming Model to Estimate Interregional Input–Output Accounts. *Journal of Regional Science*, 45, pp. 539–563.

Fisher, R. A. (1939) A Note on Fiducial Inference. *Annals of Mathematical Statistics*, 10, 383-388.

Fox, G. and J. Quirk. (1985) Uncertainty and Input-Output Analysis. *Economic Research Series, no. 23*, Jet Propulsion Laboratory, Pasadena, CA, October.

Gerking, S. D. (1976) Input-Output as a Simple Econometric Model. *Review of Economics and Statistics*, 58, 274-282.

Geweke, J. (1992) Evaluating the Accuracy of Sampling-based Approaches to the Calculation of Posterior Moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (eds.), *Bayesian Statistics 4*. New York: Oxford University Press, pp. 169-193.

Girolami, M. and B. Calderhead. (2011) Riemann Manifold Langevin and Hamiltonian Monte Carlo Methods. *Journal of the Royal Statistical Society B*, 73, 123-214.

Golan, A., G. Judge, and S. Robinson. (1994) Recovering information from incomplete or Partial Multisectoral Economic Data. *Review of Economics and Statistic*s. 76, 541-549.

Günlük-Senesen, G. and J. M. Bates. (1988) Some Experiments with Methods of Adjusting Unbalanced Data Matrices. *Journal of the Royal Statistical Society, Series A*, 151, 473–490.

Ichimura, S. and H. Wang, eds. (2003) *Interregional Input-Output Analysis of the Chinese Economy*. Singapore: World Scientific Publishing Co.

Junius, T. and J. Oosterhaven. (2003) The Solution of Updating or Regionalizing a Matrix with Both Positive and Negative Entries. *Economic Systems Research*, 15, 87–96.

Knudsen, D. C. and A. S. Fotherington. (1986) Matrix Comparison, Goodness-of-Fit, and Spatial Interaction Modeling. *International Regional Science Review*, 10, 127-147.

Kockläuner, G. (1989) Econometric Problems in Cross–sectional Stochastic Input-Output Analysis, *Economic Systems Research*, 1, 311–316.

Kop Jansen, P. and T. ten Raa. (1990) The Choice of Model in the Construction of Input-Output Coefficients Matrices. *International Economic Review*, 31, 213–227.

Lahr, M. L. and L. De Mesnard. (2004) Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis, *Economic Systems Research*, 16, 115–134.

Lenzen, M., B. Gallego, and R. Wood. (2006) A Flexible Approach to Matrix Balancing under Partial Information. *Journal of Applied Input-Output Analysis*, 11, 1–24.

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

Lenzen, M., D. Moran, K. Kanemoto, and A. Geschke. (2013) Building Eora: A Global Multiregion Input-Output Database at High Country and Sector Resolution. *Economic Systems Research*, 25, 20–49.

Lynch, R. G. (1986) An Assessment of the RAS Method for Updating Input-Output Tables. In I. Sohn (ed.), *Readings in Input-Output Analysis*. New York: Oxford University Press, pp. 271–284.

Miller, R. E. and P. D. Blair. (2009) *Input-Output Analysis: Foundations and Extensions*, 2nd ed. New York: Cambridge University Press.

Park, T. and G. Casella. (2008) The Bayesian Lasso. *Journal of the American Statistical Association*, 103(482), 681-686.

Robinson, S., A. Cattanbo, and M. El-Said. (2000) Updating and Estimating a Social Accounting Matrix using Cross-entropy Methods. *Economic Systems Research*, 13, 47–67.

Rodrigues, J. F. D. (2014) A Bayesian Approach to the Balancing of Statistical Economic Data. *Entropy*, 16(3), 1243–1271.

Rodrigues, J. F. D., A. F. Amores, and R. Paulo. (2019) Bayesian Selection of Technology Assumptions for the Transformation from Supply-Use to Input-Output Tables. *Economic Systems Research*, 31, 551-573.

Rueda-Cantuche, J. M. and A. F. Amores. (2010) Consistent and Unbiased Carbon Dioxide Emission Multipliers: Performance of Danish Emission Reductions via External Trade. *Ecological Economics*, 69, 988-998.

Temursho, U. (2017) Uncertainty Treatment in Input-Output Analysis, Chapter 12 in T. ten Raa (ed.) *Handbook of Input-Output Analysis.* Cheltenham, U.K.: Edgar Elgar, pp. 407–463.

ten Raa, T. and J. M. Rueda-Cantuche. (2007) Stochastic Analysis of Input-Output Multipliers on the Basis of Use and Make Matrices, *Review of Income and Wealth*, 53, 1–17.

ten Raa, T., and M. F. J. Steel. (1994), Revised Stochastic Analysis of an Input-Output Model, *Regional Science and Urban Economics*, 24, 361–371.

Theil, H., and A. S. Goldberger. (1961) On Pure and Mixed Statistical Estimation in Economics. *International Economic Review*, 2, 65–78.

Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society B*, 58, 267–288.

Tierney, L. (1994) Markov Chains for Exploring Posterior Distributions. *Annals of Statistics*, 22, 1701–1728.

Torres-González, L. D. and J. Yang. (2019) The Persistent Statistical Structure of the U.S. Input-Output Coefficient Matrices: 1963–2007. *Economic Systems Research*, 31, 481-504.

Tukker, A., A. de Koning, R. Wood, T. Hawkins, S. Lutter, J. Acosta, and J. Kuenen. (2013) EXIOPOL–Development and Illustrative Analyses of a Detailed Global MR EE SUT/IOT. *Economic Systems Research*, 25, 50–70.

West, M. (2003) *Bayesian Factor Regression Models in the "Large* p, *Small* n*" Paradigm*. Technical Report, Duke University.

Zou, H. (2012) The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476), 1418–1429.

Zou, H., H. Hastie, and R. Tibshirani. (2006) Sparse Principal Component Analysis. *Journal of Computational and Graphical Statistics*, 15, 265-286.

Wood, R. (2011) Construction, Stability and Predictability of an Input-Output Time-series for Australia. *Economic Systems Research*, 23, 175–211.

Valderas-Jaramillo, J. M. J. M. Rueda-Cantuche, E. Olmedo, and J. Beutel. (2019) Projecting Supply and Use Tables: New Variants and Fair Comparisons. *Economic Systems Research*, 31, 423-444.

Verdinelli, I. and L. Wasserman. (1995) Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association*, 90(430), 614-618.