# Bayesian Updating of Input-Output Tables

Oleg Lugovoy (EDF, RANE)
Andrey Polbin (RANE)
Vladimir Potashnikov (RANE)

18th Annual Conference on Global Economic Analysis
June 18, 2015
Melbourne

# Outline

- Motivation
- Conceptual framework
- Real data experiments: updating IRIOS tables
- Some conclusions

# Motivation

**Basic problem of updating IOT:**

- Find an unknown matrix *A* of IO coefficients with known sums of rows and columns and a known IO matrix for some previous year.
- Restrictions on the coefficients of matrix *A*:

$$Y = AX,$$

$$\sum_i a_{i,j} = \bar{a}_j, \quad a_{i,j} \geq 0$$

where *Y, X* are known vectors of intermediate demand and total output, $\bar{a}_j$ are known sums of columns

# Motivation

- Existing methods for updating IO tables **RAS** (Stone, 1961), **TRAS** (Gilchrist, St Louis, 1999), **GRAS** (Günlük-Şenesen, Bates, 1988), **KRAS** (Lenzen et al., 2009), **CEM** (Golan et al., 1994), **WLS** (Byron, 1978), **NLS** (Friedlander, 1961) among thers are concentrated on the getting one point estimate of the unknown table

# Motivation

- For example, the problem of Cross Entropy Minimization method:

$$\widehat{A} = \arg\min \left\{ \sum_{i,j} a_{i,j} \ln \frac{a_{i,j}}{a_{i,j}^0} \right\}$$

$$s.t.: Y = AX,$$

$$\sum_i a_{i,j} = \overline{a}_j, \quad a_{i,j} \geq 0$$

where $a_{i,j}^0$ IO coefficients for some previous year

- In this study we propose to follow Bayesian method for updating IO tables

# Motivation

- In this study we propose to follow Bayesian method for updating IO tables
- The method provide full density profile on estimated parameters with covariates
- It is useful for **sensitivity** analysis of results of applied general equilibrium models to **uncertainty** in IO tables

# Conceptual framework

$$p(\mathrm{z} \mid \Upsilon) = \frac{L(\Upsilon \mid \mathrm{z})\, p(\mathrm{z})}{\int L(\Upsilon \mid \mathrm{z})\, p(\mathrm{z}) dz} \propto L(\Upsilon \mid \mathrm{z})\, p(\mathrm{z})$$

$z$ - unknown parameters $\Upsilon$ - data

$p(\mathrm{z})$ - prior distribution of parameters

$L(\Upsilon \mid \mathrm{z})$ - likelihood function

$p(\mathrm{z} \mid \Upsilon)$ - posterior distribution (combination of prior information and data)

# Conceptual framework

**Candidates for prior distribution:**

- Truncated normal distributions for each IO coefficient
- Uniform distributions
- Beta distributions

# Conceptual framework

- We consider equality and inequality constraints of the system of restriction separately.
- Inequality constrains are simply introduced in prior distribution by assigning zero value of density in inadmissible domain.
- We specify likelihood function as an indicator function of that all linear constraints are satisfied

# Conceptual framework

- let us consider linear equality constraints and rewrite it in the following form:

$$\boxed{Bz = T}$$

where *B* is the known matrix, *T* is the known vector,

$z = vec(A)$ - vectorization of matrix *A*

- This system represents undetermined linear system of equations with the following solution:

$$\boxed{z = \tilde{z} + F^{(1)}\xi^{(1)}}$$

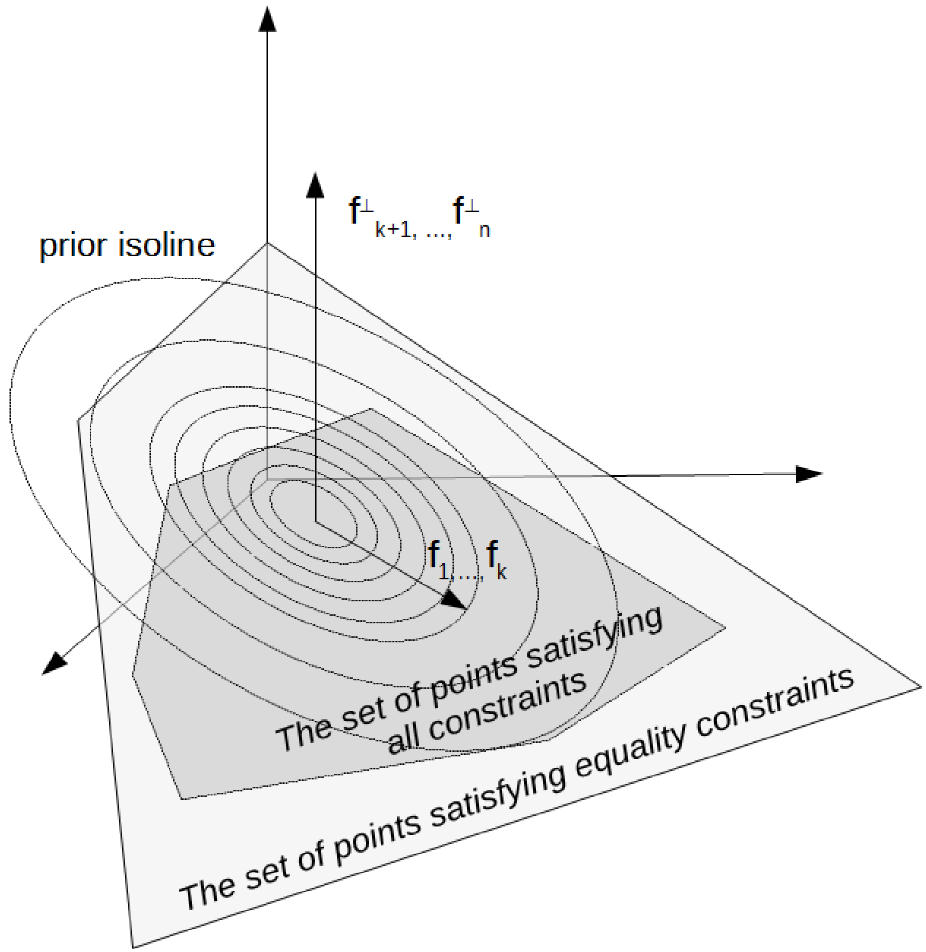
- This equation determine subspace in the original Euclidean space.

# Conceptual framework

To estimate posterior distribution we:

- change initial coordinate system and produce corresponding transformation of prior density function
- posterior distribution is simply a conditional distribution given the zero value of coordinates in orthogonal subspace
- perform the Metropolis sampling algorithm to generate sample from posterior distribution
- move to the original space

# Conceptual framework

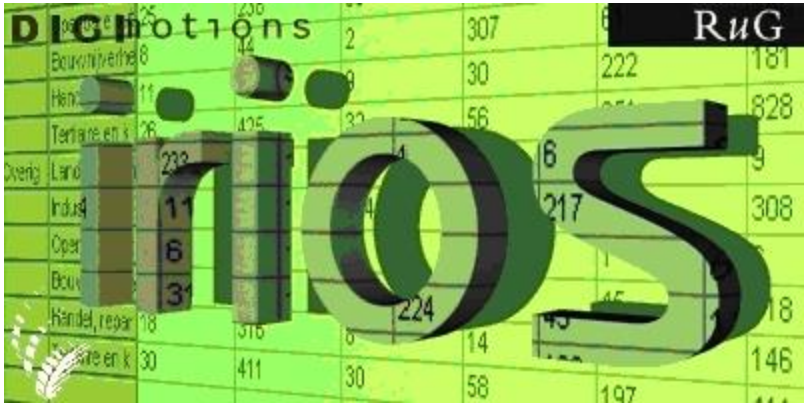


# Real data experiments

## Estimate IO table 1985 based on previous years tables

- 5 country
  - Belgium, Denmark, France, Italy, Netherland
- 25 sectors

| | |
|---|---|
| 1Agri | Agriculture, forestry and fishery products |
| 2Ener | Fuel and Power products |
| 3Meta | Ferrous and non-ferrous ores and metals |
| 4Mine | Non-metallic mineral products |
| 5Chem | Chemical products |
| 6MetP | Metal products except machinery and transport equipment |
| 7AIMa | Agricultural and industrial machinery |
| 8ODMa | Office and data processing machines |
| 9ElGo | Electrical goods |
| 10TrEq | Transport equipment |
| 11Food | Food, beverages, tobacco |
| 12Text | Textiles and clothing, leather, footwear |
| 13Pape | Paper and printing products |
| 14Rubb | Rubber and plastic products |
| 15OMan | Other manufacturing products |
| 16Buil | Building and construction |
| 17ReTr | Recovery, repair services, wholesale and retail trade |
| 18Lodg | Lodging and catering services |
| 19InTr | Inland transport services |
| 20MATr | Maritime and air transport services |
| 21Auxi | Auxiliary transport services |
| 22Comm | Communication services |
| 23Cred | Credit and Insurance |
| 24OMSe | Other market services |
| 25PSer | Non-market services |



# Real data experiments

## Measures of closeness of estimate to true matrix

| | |
|---|---|
| $RMSE = \sqrt{\frac{1}{n^2}\sum_{ij}\left(x_{ij}^{true} - x_{ij}\right)^2}$ | Root mean square error |
| $MAE = \frac{1}{n^2}\sum_{ij}\left\|x_{ij}^{true} - x_{ij}\right\|$ | Mean absolute error |
| $MAPE = \frac{1}{n^2}\sum_{ij}\frac{\left\|x_{ij}^{true} - x_{ij}\right\|}{\left\|x_{ij}^{true}\right\|} \cdot 100$ | Mean absolute percentage error |
| $WAPE = \sum_{ij}\left(\frac{\left\|x_{ij}^{true}\right\|}{\sum_{kl} x_{kl}^{true}}\right) \cdot \frac{\left\|x_{ij} - x_{ij}^{true}\right\|}{\left\|x_{ij}^{true}\right\|} \cdot 100$ | Weighted absolute percentage error (Minguez et al., 2009) which weights each percentage deviation of $x_{ij}$ from $x_{ij}^{true}$ by the relative size of the corresponding true element in the overall sum of the actual elements. |
| $SWAD = \frac{\sum_{ij}\left\|x_{ij}^{true}\right\| \cdot \left\|x_{ij} - x_{ij}^{true}\right\|}{\sum_{kl}\left(x_{ij}^{true}\right)^2}$ | Standardized weighted absolute difference (Lahr, 2001b) which is effectively similar to WAPE with the difference being that absolute deviations are weighted by the size of the true transactions. |
| $PSI = \frac{1}{\sum_{kl} x_{kl}^{true}} \cdot \sum_{ij}\left[\left\|x_{ij}^{true}\right\| \cdot \left\|\ln\left(\frac{x_{ij}^{true}}{s_{ij}}\right)\right\| + \left\|x_{ij}\right\| \cdot \left\|\ln\left(\frac{x_{ij}}{s_{ij}}\right)\right\|\right],$ $where\ s_{ij} = \frac{\left\|x_{ij}\right\| + \left\|x_{ij}^{true}\right\|}{2}$ | Knudsen and Fotheringham (1986) concluded that PSI is one of the most useful goodness-of-fit measures for comparative purposes because this indicator shows a linear relation between its value and the level of error. |
| $RSQ = cor\left(x_{ij}^{true}, x_{ij}\right)^2$ | (or coefficient of determination) – the square of the correlation coefficient between the elements of the actual and predicted matrices of $x_{ij}$ true and $x_{ij}^{true}$, respectively, when at least one of them is different from zero. |
| $AED = \sum_{ij}\left\|x_{ij}^{true} \cdot \log x_{ij}^{true} - x_{ij} \cdot \log x_{ij}\right\|$ | Average entropy distance |

# Real data experiments

| Country | Value | Differences between method and Bayes | | | | |
|---|---|---|---|---|---|---|
| Belgium | Bayes | Cross-Entropy | Least square | Normalized least square | Weighted least square | RAS |
| RMSE | 0.005 | 2.29E-04 | 2.51E-03 | 3.34E-04 | 4.83E-03 | -7.09E-04 |
| MAE | 0.002 | -2.15E-04 | 1.09E-03 | -1.51E-04 | 2.06E-03 | -5.72E-04 |
| MAPE | 0.155 | -3.15E-02 | 4.66E+00 | -2.34E-02 | 2.01E+01 | -7.37E-02 |
| WAPE | 9.452 | -9.99E-01 | 5.03E+00 | -6.98E-01 | 9.53E+00 | -2.65E+00 |
| SWAD | 0.050 | 8.06E-04 | 3.41E-02 | 3.21E-03 | 4.09E-02 | 2.14E-04 |
| Psi | 0.093 | -8.65E-03 | 4.54E-02 | -5.81E-03 | 7.65E-02 | -2.50E-02 |
| RSQ | 0.993 | 1.14E-03 | 8.60E-03 | 1.43E-03 | 2.08E-02 | -1.51E-03 |
| AED | 2.638 | -3.74E-01 | 1.75E+00 | -3.04E-01 | 3.25E+00 | -1.01E+00 |
| Denmark | Bayes | Cross-Entropy | Least square | Normalized least square | Weighted least square | RAS |
| RMSE | 0.008 | 7.46E-04 | 2.80E-03 | 1.30E-03 | 4.64E-03 | 2.77E-03 |
| MAE | 0.004 | 2.61E-04 | 1.30E-03 | 5.29E-04 | 2.21E-03 | 1.09E-03 |
| MAPE | 0.562 | 1.15E-02 | 4.46E-01 | 2.49E-02 | 8.48E-01 | 4.01E-02 |
| WAPE | 16.119 | 1.16E+00 | 5.78E+00 | 2.36E+00 | 9.84E+00 | 4.85E+00 |
| SWAD | 0.072 | 9.71E-03 | 3.69E-02 | 1.69E-02 | 4.34E-02 | 3.30E-02 |
| Psi | 0.158 | 1.25E-02 | 5.29E-02 | 2.41E-02 | 7.82E-02 | 4.76E-02 |
| RSQ | 0.982 | 3.45E-03 | 1.53E-02 | 6.28E-03 | 2.95E-02 | 1.42E-02 |
| AED | 5.486 | 3.30E-01 | 1.72E+00 | 6.54E-01 | 2.64E+00 | 1.31E+00 |
| France | Bayes | Cross-Entropy | Least square | Normalized least square | Weighted least square | RAS |
| RMSE | 0.009 | 2.68E-03 | 3.97E-03 | 2.79E-03 | 5.77E-03 | 2.47E-03 |
| MAE | 0.004 | 7.69E-04 | 1.57E-03 | 8.48E-04 | 2.91E-03 | 6.06E-04 |
| MAPE | 0.722 | -1.20E-01 | 5.26E+00 | -1.51E-01 | 1.30E+01 | -3.05E-02 |
| WAPE | 21.414 | 3.69E+00 | 7.54E+00 | 4.07E+00 | 1.39E+01 | 2.91E+00 |
| SWAD | 0.104 | 2.92E-02 | 4.03E-02 | 3.35E-02 | 5.35E-02 | 1.81E-02 |
| Psi | 0.204 | 3.87E-02 | 6.03E-02 | 4.24E-02 | 9.93E-02 | 3.15E-02 |
| RSQ | 0.969 | 2.07E-02 | 3.34E-02 | 2.13E-02 | 5.01E-02 | 1.95E-02 |
| AED | 6.302 | 9.34E-01 | 1.68E+00 | 1.03E+00 | 2.92E+00 | 7.39E-01 |

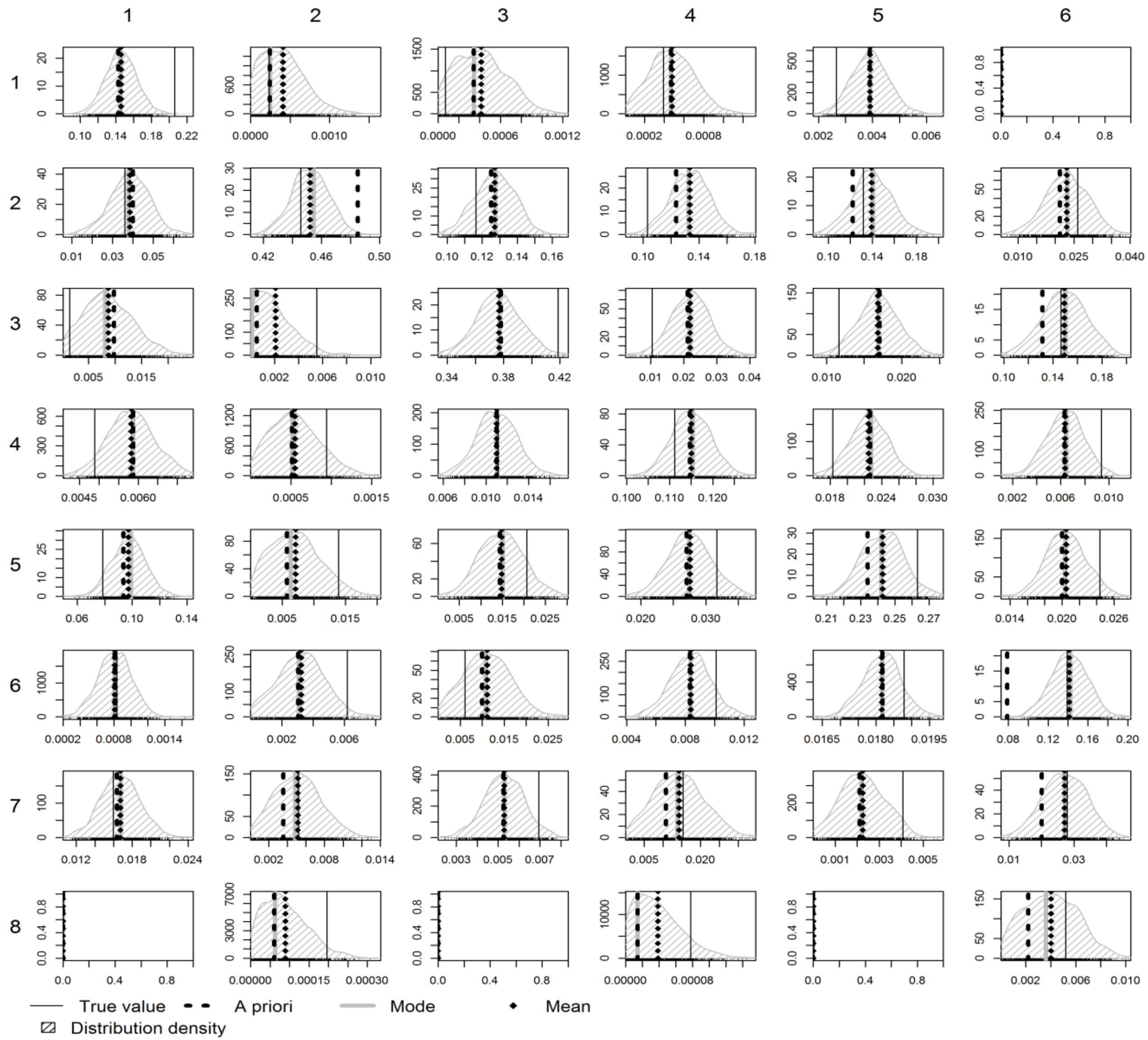
| Country | Value | Differences between method and Bayes | | | | |
|---|---|---|---|---|---|---|
| Italy | Bayes | Cross-Entropy | Least square | Normalized least square | Weighted least square | RAS |
| RMSE | 0.007 | 1.08E-03 | 2.09E-03 | 1.06E-03 | 3.75E-03 | 3.98E-04 |
| MAE | 0.003 | 4.74E-04 | 1.20E-03 | 4.71E-04 | 2.26E-03 | 2.67E-04 |
| MAPE | 0.749 | 4.28E-02 | 1.81E-01 | 3.94E-02 | 7.36E-01 | -1.01E-02 |
| WAPE | 15.447 | 2.15E+00 | 5.45E+00 | 2.13E+00 | 1.03E+01 | 1.21E+00 |
| SWAD | 0.063 | 2.01E-03 | 2.92E-02 | 1.50E-03 | 4.08E-02 | 5.87E-03 |
| Psi | 0.152 | 2.08E-02 | 4.38E-02 | 2.07E-02 | 7.25E-02 | 1.17E-02 |
| RSQ | 0.986 | 4.60E-03 | 9.20E-03 | 4.44E-03 | 1.80E-02 | 1.67E-03 |
| AED | 5.358 | 9.10E-01 | 1.42E+00 | 9.21E-01 | 2.42E+00 | 5.14E-01 |
| Netherland | Bayes | Cross-Entropy | Least square | Normalized least square | Weighted least square | RAS |
| RMSE | 0.021 | 4.21E-02 | 2.60E-03 | 5.22E-03 | 1.41E-02 | 3.76E-04 |
| MAE | 0.008 | 1.49E-02 | 1.67E-03 | 1.49E-03 | 3.13E-03 | 3.92E-04 |
| MAPE | 1.545 | -6.06E-01 | 1.56E+00 | 8.25E-02 | 3.30E+00 | -1.48E-01 |
| WAPE | 35.642 | 6.49E+01 | 7.30E+00 | 6.51E+00 | 1.36E+01 | 1.71E+00 |
| SWAD | 0.164 | 8.36E-01 | 7.31E-02 | 6.31E-02 | 9.08E-02 | 2.97E-02 |
| Psi | 0.323 | NA | 6.51E-02 | 5.90E-02 | 8.10E-02 | 1.77E-02 |
| RSQ | 0.876 | NA | 4.36E-02 | 6.75E-02 | 1.48E-01 | 1.05E-02 |
| AED | 10.632 | -1.06E+01 | 1.96E+00 | 1.54E+00 | 1.44E+00 | 4.01E-01 |

Bayes have better characteristics than compare

Bayes have worse characteristics than compare

# Real data experiments

## Posterior distribution for France



# Some conclusions

- Bayesian approach is a flexible and natural tool to incorporate uncertainties in data into estimation process
- The experimental estimates demonstrate a way of application of Bayesian inference for updating and estimating IOT
- The results of the estimates – multidimensional distribution of the estimated parameters might be used as an input information for sensitivity analysis on a stage of implementation of the analysis.

# Thank you for your attention!

olugovoy@gmail.com

apolbin@gmail.com

potashnikov.vu@gmail.com