*Paola Antonello**

An Application of EM Algorithm to Analyse and Forecast Long-Run I-O Coefficient Changes

Paper to be presented at the
17th International Input-Output Conference
São Paulo, Brazil, July, 13-17, 2009
(Provisional, not to be quoted without the author's permission)

ABSTRACT

In this paper, the analogy between the set of demand coefficients of a closed Leontief model and the transitions probabilities matrix of a Markov chain is used to analyse the long-run determinants of I-O coefficient changes. It is assumed that the dynamics of the demand coefficients can be represented by a multinomial logistic function, parametrized in terms of time and of the gross output of the input sector. It is also assumed that gross outputs by sector are generated by AR(p) processes. Under these assumptions, by applying the EM algorithm, it is possible to produce Bayesian estimates of long-run input and demand coefficients, as well as estimates of the corresponding long-run inter-sectoral flows and of their variance-covariance matrix. This is also possible when the time series of the available I-O tables is not complete, i.e. when some observations are missing. As an example, this methodology is applied to the time series of 42 yearly, constant price I-O tables for the Italian economy, estimated by G. Rampa (Economic Systems Research, 20,3, 259-276), aggregated at a four sector level. The estimated long-run values of I-O coefficients and the corresponding information matrix are then used to produce GLS estimates of quarterly, four-sector I-O tables as in Antonello, P. (1990) (Economics Systems Research, 2, 2, 157-171).

JEL Classification: C46, C67

*Corresponding Author's Address:
Prof. Paola Antonello
Dept. of Economic Systems and Institutions
University of L'Aquila
Piazza del Santuario, 19
67040 ROIO POGGIO (AQ)
Italy
Tel. +390862434852
E-mail: paola.antonello@cc.univaq.it

2. Definitions and basic assumptions of the static model.

The symbols used to define the most relevant variables and constants of this model are listed in the following Table 1.

(Insert Table 1)

Set $t = \bar{t}$ and assume that *i* and Y_i are known constants. Hence, both the time variable and the sector index *i* can be provisionally ignored

The model's basic assumptions are:

Assumption 1:

$d_{i,l,j} = 0$, if the l^{th} unit of output i is not used by sector j

$d_{i,l,j} = 1$, if the l^{th} unit of output i is used by sector j

Consequently:

$$y_{i,j} = \sum_{l=1}^{Y_i} d_{i,l,j}$$

and

$$Y_i = \sum_{j=1}^{4} \sum_{l=1}^{Y_i} d_{i,l,j} = \sum_{j=1}^{4} y_{i,j}$$

.

Assumption 2:

Conditionally to the given level of gross output *Y*, the indicator functions $d_{l,j}$ are multinomial i.i.d., with distribution

$$\prod_{j=1}^{4} \pi_j^{d_{lj}}, \quad 0 \leq \pi \leq 1 , \qquad \sum_{j=1}^{4} \pi_j = 1$$

From assumption 2 we have that, conditionally to *Y* and π_j, each unit of i^{th} gross output is identically, independently distributed, with conditional density

$$f_l(x_l|\pi, Y) \sim_{iid} \prod_{j=1}^{4} \pi_j^{d_{l,j}} = \sum_{j=1}^{4} \pi_j^{d_{l,j}}$$

and that, conditionally to Y and π_j, the total consumption by sector j of i's gross product is a mixture of identically, independently distributed random variables:

The joint density function of x_l and Y is

$$f(x_l, Y|\pi, \mu_Y) = f_y(Y|\mu_Y) f_{x_l}(x_l|\pi, Y) = f_Y(Y|\mu_Y) \prod_{j=1}^{4} \pi_j^{d_{lj}}$$

and the likelihood function of μ_Y and π is

$$L(\mu_Y, \pi|Y, d) = f_Y(Y|\mu_Y) \prod_{l=1}^{Y} \prod_{j=1}^{4} \pi_j^{d_{lj}}$$

The corresponding log-likelihood is

$$\ln L(\mu_Y, \pi|Y, d) = \ln f_Y(Y|\mu_Y) \sum_{l=1}^{Y} \sum_{j=1}^{4} d_{i,j} \ln \pi_j$$

If π_j is given, the posterior probability that $d_{l,j} = 1$, i.e. that the x_l unit of output i is used by sector j is

$$z_{l,j} = \left[\pi_j f_{jY}(Y|\mu_Y)\right] \Big/ \left[f_Y(Y|\mu_Y) \sum_{j=1}^{4} \pi_j\right] = \pi_j = E(d_{l,j})$$

If an estimate $\widehat{z_{l,j}}$ of $E(d_{l,j})$ is available, then, conditional on this estimate, the log-likelihood function becomes

$$\ln L(\mu_Y, \pi|Y, \hat{z}) = \ln f_Y(Y|\mu_Y) + \sum_{l=1}^{Y} \sum\nolimits_{j=1}^{4} \widehat{z_{l,j}} \ln \pi_j$$

This corresponds to the E-step of the EM algorithm. The M-step maximizes the log-likelihood by solving the first order conditions

$$\frac{\delta \ln f_Y(Y|\mu_Y)}{\delta Y} = 0$$

and by setting

$$\widehat{\pi_J} = \frac{1}{Y}\sum_{l=1}^{Y}\widehat{z_{l.J}} = \frac{Y}{Y}\widehat{z_{l,J}} = \widehat{z_{l,J}}$$

since $E(d_{l,j})$ is the same for all l=1,…,Y.

3. Definitions and basic assumptions of the dynamic model.

Assumption 3.

The time series of the gross output of sector i, i = 1,.., 4, is generated by a first order autoregressive process, i.e.:

$$E\big(Y_i(t)\big) = \alpha_i Y_i(t-1), i = 1, \dots, 4$$

Define

$$\hat{Y}(t) = \big[\widehat{Y_\iota}(t)\big]$$

as a diagonal matrix that has $Y_i(t)$, i=1,…, 4, on its principal diagonal, and

$$\Pi(\mathrm{t}) = \big[\pi_{\mathrm{i,j}}(\mathrm{t})\big]$$

as a square matrix, the rows of which are the probabilities, defined in the previous section, that one unit of sector i output is sold to sector j at time t. Since the sum of all elements of each rows of $\Pi(t)$ is one, the following identity must be true:

$$\hat{Y}(t-1) = \Pi'(t)\hat{Y}(t-1)$$

and the expectation of $\hat{Y}(t)$ can be written as

$$E\big(\hat{Y}(t)\big) = \Pi'(t)\mathrm{A}\hat{Y}(t-1)$$

where A is a diagonal matrix that has α_i, i = 1,…,4, on its principal diagonal.

If we set

$$P(t) = \widehat{\Pi}'(t-1)\mathrm{A}$$

The data generating process of the four elements gross output vector is similar to a 1st order Markov process. However, *P(t)* can be identified with the **transition kernel** of a Markov chain only if

$$P'(t)e = e$$

i.e., if the row-sum of $A\Pi(t-1)$ equals *e*, a column vector of 1s. This implies that A is the identity matrix, i.e. that the time series of the sectoral gross outputs are generated by a non-stationary, integrated of order 1 process:

$$Y_i \sim I(1), \forall\, i$$

This condition, that must be satisfied if we want to consider the elements of *P(t)* as the probabilities of transition from the output sector *i* at time *(t-1)* to the input sector *j*, at time *t*, and that means that 1 is an eigenvalue of $\Pi(t)$ and that *e* is the right eigenvector associated to it, is tested, and the results are presented in paragraph 5.

4. The sectoral flows model

From section 2 we know that the joint density of $Y_i(t)$ and $y_{i,j}(t)$, is

$$f_y\left(y_{i,j}(t), Y_i(t) \middle| \pi, \mu_{Y_i}\right) = f_{Y_i}\left(Y_i(t) \middle| \mu_{Y_i}\right) \prod_{l=1}^{Y_i(t)} \pi_{i,j}^{d_{i,l,j}(t)}(t)$$

and that the conditional density of $y_{i,j}(t)$, given $Y_i(t)$ is

$$f_y\left(y_{i,j}(t) \middle| Y_i(t), \pi, \mu_{Y_i}\right) = \frac{f_y\left(y_{i,j}(t), Y_i(t) \middle| \pi, \mu_{Y_i}\right)}{f_{Y_i}\left(Y_i(t) \middle| \mu_{Y_i}\right)} = \prod_{l=1}^{Y_i(t)} \pi_{i,j}^{d_{i,l,j}(t)}$$

<u>Assumption 4</u>

Conditionally on $Y_i(t)$, the distribution of $y_{i,j}(t)$ is lognormal, and the expectation of $y_{i,j}(t)$ is a time varying function of the conditional expectation of the total output of the input sector *j*.

Hence

$$f_{\ln y}\big(\ln y_{i,j}(t)\,\big|\,\ln t, \ln Y_i(t), \pi, \mu_Y\big)$$

$$= \frac{1}{\sigma_{\varepsilon_{i,j}}\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_{\varepsilon_{i,j}}}\big(\ln y_{i,j}(t) - \beta_{i,j}^1 \ln t\right.$$

$$\left. - \beta_{i.j}^2 \ln E\big(Y_j(t)\big|Y_j(t-1)\big)\big)\right\}^2$$

Then, the log-likelihood of the parameter set β becomes:

$$\sum_{t=2}^{42}\sum_{i=1}^{4}\sum_{j=1}^{4}\left\{\ln y_{i,j}(t) - \beta_{i,j}^1 \ln t - \beta_{i,j}^2 \ln E\big(Y_i(t)\big|Y_j(t-1)\big)\right\}^2$$

Table 1

Symbols	Description
$t = 1,\ldots,42$	Time
$Y_i\ (t),\ i = 1,\ldots,4$	Gross output of sector i at t
$y_{i,j}(t),\ i,\ j = 1,\ldots,4$	Number of units of output i consumed as an input by industry j (j = 1, …, 3), or purchased by final users (j = 4)
$d_{i,l,j}(t)$	Indicator function
$x_{i,l}(t),\ l=1,\ldots,Y_i$	One unit of sector i gross output
$\pi_{i,j}(t)$, i=1,…,4	Probability that $d_{i,l,j}(t)$=1, i.e. conditional expectation of $d_{i,l,j}(t)$, given i and l.
$\mu_{Y_i}(t)$, $i = 1, \ldots, 4$	$E(Y_i(t)\|Y_i(t-1))$, i.e. conditional expectation of sector i gross output at t.