Reconciling China’s Regional Input-Output Tables

David Roland-Holst and Muzhe Yang
UC Berkeley

Lecture I
Presented to the
Development Research Centre
State Council of the PRC
Beijing, 6 June 2005

Lectures on Data and Model Development

1. Regional Data Reconciliation
2. Multi-regional Trade Flow Estimation
3. Integrated Micro-Macro Modeling

Objectives

- Implement an *efficient* econometric methods for reconciling provincial Input-output tables with national accounts.
- Establish coherent national standards for data harmonization

Motivation

- Provincial Input-output are available for China, but they exhibit a variety of consistency problems
  - Among the more serious of these is inconsistency with national-level tables, individually and collectively
- Consistent individual and aggregate tables are essential to implement detailed economic analysis within and across provinces and regions

Foundation – PRC Provincial IO Tables

- Already available
- Nationally comprehensive and consistent in terms of account definitions
- This work supports efforts already under way at the provincial and national (NBS) level, and also builds on existing DRC capacity for SAM and CGE research

Proposed Approach

- Using Bayesian econometric techniques to incorporate prior information when updating and reconciling economic accounts
- We show how to estimate a consistent provincial table with additional prior information at the national level.
- The estimation begins with a consistent national table that is assumed (for convenience only) to be known with certainty.

Overview of the Estimation Problem

The set-up of this matrix balancing problem follows Golan, Judge and Miller (1996). We focus on balancing schemes for provincial table.

The proposed approach is an extension of that usually applied to a national table.

Estimation Strategy

Consider one province, $g \in \{1, 2, \cdots, G\}$, a K -sector economy, represented by an input-output table, IO$^{(g)}$, where each entry indicates a payment by a column account to a row account:

$$IO^{(g)} = \begin{bmatrix} T^{(g)} & \mathbf{f}^{(g)} \\ \mathbf{v}^{(g)\prime} & 0 \end{bmatrix}_{(K+1)\times(K+1)}$$

where $T^{(g)}$ is a $K \times K$ matrix of intermediate sales, $\mathbf{f}^{(g)}$ is a K -vector of final demands, and $\mathbf{v}^{(g)}$ is a K -vector of sectoral value added. The table IO $^{(g)}$ is therefore a $(K+1)\times(K+1)$ matrix, where corresponding column and row sums are equal.

Estimation 2

Assume:
(1) Intermediate demands are determined by a $K \times K$ fixed coefficient matrix $A^{(g)}$;
(2) A K-vector, $\mathbf{x}^{(g)}$, represents sectoral sales to both intermediate and final demanders.
Then, we have the following standard Leontief input-output model:

$$A^{(g)}\mathbf{x}^{(g)} + \mathbf{f}^{(g)} = \mathbf{x}^{(g)}$$

Define $\mathbf{y}^{(g)} \equiv \mathbf{x}^{(g)} - \mathbf{f}^{(g)}$, as the sectoral sales to intermediate demanders. This transaction has double meanings: the column vector of $\mathbf{y}^{(g)}$ represents sectoral intermediate expenditures, while the row vector of $\mathbf{y}^{(g)}$ represents sectoral intermediate receipts.

Estimation 3

Now we transform the matrix balancing problem into the econometric problem of identifying the $a_{ij}^{(g)}$ elements of the $A^{(g)}$ matrix, based on the available economic information contained in the row and column sums IO table. This strategy takes the form

$$\mathbf{y}^{(g)} = A^{(g)}\mathbf{x}^{(g)}$$

$$\underset{K\times 1}{\mathbf{y}^{(g)}} = \sum_{j=1}^{K} \underset{K\times 1}{A_j^{(g)}} \underset{1\times 1}{x_j^{(g)}} \qquad \left(j = 1,\cdots,K\right)$$

$$\Rightarrow y_i^{(g)} = \sum_{j=1}^{K} a_{ij}^{(g)} x_j^{(g)} \quad \left(i, j = 1,\cdots,K\right)$$

$$\because T_{ij}^{(g)} = a_{ij}^{(g)} x_j^{(g)}$$

$$\Rightarrow \sum_{j=1}^{K} T_{ij}^{(g)} = y_i^{(g)} = \sum_{j=1}^{K} T_{ji}^{(g)} \quad \left(i, j = 1,\cdots,K\right)$$

Identification Strategy

To proceed, we transform the national table in precisely the same way [omit the (g) superscript in the last three slides].

Now we use the entropy principle to recover A and $A^{(g)}$ from the top down, under the row-column linear restrictions and the micro-macro consistency requirement.

Balancing Scheme for the National Table

Consider the standard formulation $\mathbf{y} = \mathbf{Ax}$, where $\mathbf{y}$ and $\mathbf{x}$ are K-dimentional vectors of known data and A is an unknown $K \times K$ matrix that must satisfy the following three conditions:

(1) **Consistency**:

$$\sum_{i=1}^{K} a_{ij} = 1 \quad \left(j = 1, \cdots, K \right)$$

(2) **Adding up**:

$$\sum_{j=1}^{K} a_{ij} x_j = y_i \quad \left(i = 1, \cdots, K \right)$$

(3) **Non-negativity**:

$$a_{ij} \geq 0 \quad \left(i, j = 1, \cdots, K \right)$$

Maximum Entropy Principle

Given the three conditions, the problem of identifying the a_{ij} elements of the A matrix is formulated as:

$$\max_{a_{ij}>0} -\sum_{i=1}^{K}\sum_{j=1}^{K} a_{ij} \ln a_{ij}$$

subject to:

$$\sum_{i=1}^{K} a_{ij} = 1 \quad \left(j = 1, \cdots, K \right)$$

$$\sum_{j=1}^{K} a_{ij} x_j = y_i \quad \left(i = 1, \cdots, K \right)$$

The solution to this problem is denoted as $\overset{\frown}{a}_{ij}^{\mathrm{ME}}$.

Balancing Scheme for Provincial Tables

Consider the previous formulation for province $g \in \{1, 2, \cdots, G\}$, i.e.

$$\mathbf{y}^{(g)} = A^{(g)}\mathbf{x}^{(g)}$$

where $\mathbf{y}^{(g)}$ and $\mathbf{x}^{(g)}$ are K -dimensional vectors of known data and $A^{(g)}$ is an unknown $K \times K$ matrix that must satisfy:

(1) Consistency:

$$\sum_{i=1}^{K} a_{ij}^{(g)} = 1 \quad (j = 1, \cdots, K)$$

(2) Adding up:

$$\sum_{j=1}^{K} a_{ij}^{(g)} x_j^{(g)} = y_i^{(g)} \quad (i = 1, \cdots, K)$$

(3) Non-negativity:

$$a_{ij}^{(g)} \geq 0 \quad (i, j = 1, \cdots, K)$$

Specification of Prior Information

The national level estimates provide information that may be used in recovering estimates of provincial SAMs. This information can be stated as a series of prior restrictions on estimating the new provincial IO. We give six examples:

(1) Links between national and provincial accounts:

$$\sum_{g=1}^{G} x_j^{(g)} = x_j \quad (j = 1, \cdots, K)$$

$$\sum_{g=1}^{G} y_i^{(g)} = y_i \quad (i = 1, \cdots, K)$$

(2) Properties of national level estimates $\overset{\frown}{a}_{ij}^{\mathrm{ME}}$:

$$\sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}} x_j = y_i$$

$$\overset{\text{(1 above)}}{\Leftrightarrow} \sum_{g=1}^{G} \sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}} x_j^{(g)} = \sum_{g=1}^{G} y_i^{(g)}$$

Priors 3-5

(3) Maximum entropy estimates for each provincial IO: $\overset{\frown}{a}_{ij}^{\mathrm{ME}(g)}$

(4) Provincial adding-up restrictions:

$$\sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}(g)} x_j^{(g)} = y_i^{(g)}$$

$$\sum_{g=1}^{G} \sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}(g)} x_j^{(g)} = \sum_{g=1}^{G} y_i^{(g)}$$

$$\overset{\text{(2 above)}}{\Leftrightarrow} \sum_{g=1}^{G} \sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}(g)} x_j^{(g)} = \sum_{g=1}^{G} \sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}} x_j^{(g)}$$

(5) If the set of provincial IO tables is not complete, we can assume the following:

$$\sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}(g)} x_j^{(g)} = \sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}} x_j^{(g)}$$

Prior 6

(6) Assume that the activity accounts represent homogeneous technologies in each province, with the technological coefficients denoted as $\bar{a}_{ij}$. Therefore, we have:

$$\sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}} x_j^{(g)} = \sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}(g)} x_j^{(g)} = \sum_{j=1}^{K_A} \bar{a}_{ij} x_j^{(g)} + \sum_{j=K_A+1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}(g)} x_j^{(g)}$$

$$\Rightarrow \sum_{j=1}^{K_A} \bar{a}_{ij} x_j^{(g)} = \sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}} x_j^{(g)} - \sum_{j=K_A+1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}(g)} x_j^{(g)}$$

$$\Rightarrow \bar{\bar{a}}_{i\cdot} = \frac{\sum_{j=1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}} x_j^{(g)} - \sum_{j=K_A+1}^{K} \overset{\frown}{a}_{ij}^{\mathrm{ME}(g)} x_j^{(g)}}{\sum_{j=1}^{K_A} x_j^{(g)}}$$

Other Prior Information

In addition to the examples given here, any specific prior information about the accounts or underlying technical relationships. These include:

1. Cell inequality or boundary constraints (><0, etc.)
2. Institutional budget constraints.
3. Fixed values or variance constraints.

Summation

With these conditions in mind, we can express prior information for estimating each provincial table as follows:

$$a_{ij}^{0(g)} = \begin{cases} \bar{\bar{a}}_{i\cdot} & \left(j = 1, \cdots, K_A\right) \\ \frown{a}_{ij}^{\mathrm{ME}(g)} & \left(j = K_A, \cdots, K\right) \end{cases} \left(i = 1, \cdots, K\right)$$

where $\frown{a}_{ij}^{\mathrm{ME}(g)}$ comes from the first-round maximum entropy estimation for each provincial IO table.

Cross Entropy Principle

Finally, the problem of identifying the $a_{ij}^{(g)}$ elements of provincial the $A^{(g)}$ matrix is formulated as:

$$\min_{a_{ij}^{(g)}>0} \sum_{i=1}^{K}\sum_{j=1}^{K} a_{ij}^{(g)} \ln a_{ij}^{(g)} - \sum_{i=1}^{K}\sum_{j=1}^{K} a_{ij}^{(g)} \ln a_{ij}^{0(g)}$$

subject to:

$$\sum_{i=1}^{K} a_{ij}^{(g)} = 1 \quad \left(j = 1, \cdots, K \right)$$

$$\sum_{j=1}^{K} a_{ij}^{(g)} x_j^{(g)} = y_i^{(g)} \quad \left(i = 1, \cdots, K \right)$$

the solution to which is denoted by $\overset{\frown}{a}_{ij}^{\mathrm{CE}(g)}$.

How to Proceed

I recommend four steps:

1. Encode the estimation methods (STATA, MATLAB)
2. Select a sample province to establish estimation standards
3. Apply to all provincial tables
4. Iterate with NBS, provincial sources

References

1. Golan, A. (2003), “Image Reconstruction of Incomplete and Ill-Posed Data: An Information -Theoretic Approach,” Working Paper, Department of Economics, American University, January.

2. Golan, A., G. Judge, and D. Miller (1996), Maximum Entropy Econometrics: Robust Estimation with Limited Data (John Wiley & Sons, New York).