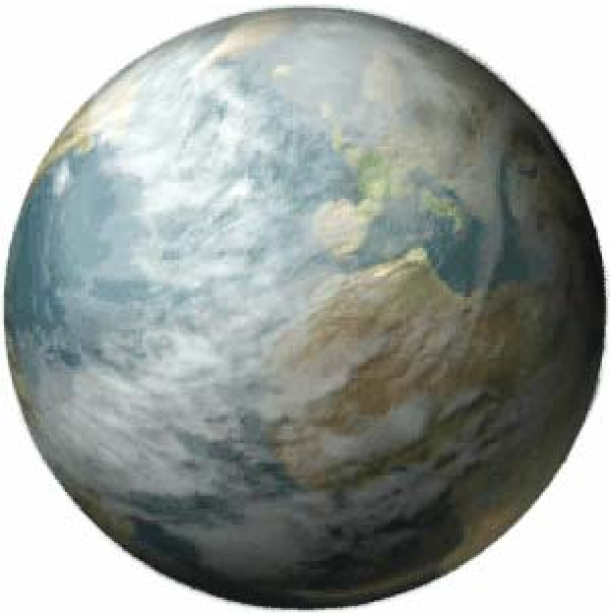
Data Development for Regional Policy Analysis



David Roland-Holst
UC Berkeley

ASEM/DRC Workshop on
Capacity for Regional Research on Poverty and Inequality in China

Monday-Tuesday, March 27-28, 2006

Contents

1. Introduction and Motivation

- Detailed and rigorous accounting practices always have been at the foundation of sound and sustainable economic policy.
- A consistent set of real data on the economy is likewise a prerequisite to serious empirical work with economic simulation model.
- For this reason, a complete general equilibrium modeling facility stands on two legs: a consistent economywide database and modeling methodology.

Multi-Sectoral Development Analysis

- Macro policy is important, but so are economic structure and interactions.
- Indeed, linkages and indirect effects are often more important than the direct targets of policy.
- To improve visibility for policy makers and make appropriate recommendations, we need to understand these interactions.

1. What is a SAM?

- An economy-wide accounting device to capture detailed interdependencies between institutions and sectors/regions. An extension of input-output analysis.
- A SAM is a form of double entry book keeping that itemizes detailed income and expenditure linkages across the economy.
- It is a closed form accounting system, reflecting the general equilibrium structure of the underlying economic relationships.

What is needed?

To successfully develop a detailed, consistent, and up-to-date SAM, four ingredients are needed:

1. Official commitment
2. Component data resources
3. Methodology
4. Expertise and, where this is lacking, talent
5. Computer hardware and software

Fortunately, we are in a strong position in all these areas.

SAM Concepts

- A SAM is a square matrix that builds on the input-output table - but it goes further.
- A SAM considers not only production linkages, but tracks income-expenditure feedbacks (institutions are introduced).
- Each transactor (such as factors of production, households, enterprises, the government and the ROW) has a row (income sources) and a column (expenditures) – double entry national income accounting.
- A SAM is consistent data system that provides a snapshot of the economy – note that the SAM reconciles data from different sources.
- Detail is on the the biggest virtues of the SAM approach, but we actually build SAMs from the top down.

SAMs from a Macroeconomic Perspective

A macroeconomic SAM is also an extension of basic national income identities:

1. $Y + M = C + G + I + E$ (GNP)
2. $C + T + Sh = Y$ (Income)
3. $G + Sg = T$ (Govt. Budget)
4. $I = Sh + Sg + Sf$ (Savings-Investment)
5. $E + Sf = M$ (Trade Balance)

Schematic Macroeconomic SAM

			Expenditures			
Receipts	1	2	3	4	5	Total
1. Suppliers	-	C	G	I	E	**Demand**
2. Households	Y	-	-	-	-	**Income**
3. Government	-	T	-	-	-	**Receipts**
4. Capital Acct.	-	S_h	S_g	-	S_f	Savings
5. Rest of World	M	-	-	-	-	Imports
Total	Supply	Expenditure	Expenditure	Investment	ROW	

Disaggregation

Detail is interesting for research, but essential for policy for two reasons.

1. Economic policy may be made from the top down, but the political consequences of economic activity are ultimately felt from the bottom up.
2. In today’s complex market economies, policy makers relying on intuition and rules-of-thumb alone are unlikely to achieve anything approaching optimality.

For this reason, it is essential to improve understanding of incidence effects that arise from complex linkages in the economic structure. CGE models, supported by detailed data, can elucidate these linkages and improve visibility for policy

Receipts	Expenditures								
	1. Activities (124)	*2. Commodities (124)*	*3. Factors (13)*	*4. Private Households (5)*	*5. Enterprises (3)*	*6. Recurrent State (1)*	*7. Investment Savings (1)*	*8. Rest of World (94+1)*	*9. Total*
1. Activities (124)		Marketed Production							Total Sales
2. Commodities (124)	Intermediate Consumption			Private Consumption		State Consumption	Investment	Exports	Total Commodity Demand
3. Factors (13)	Value Added								Value Added
4. Private Households (5)			Wages, Salaries and Other Benefits		Distributed Profits and Social Security	Social Security and Other Current Transfers to Households		Net Foreign Transfers to Households	Private Household Income
5. Enterprises (3)			Gross Profits					Net Foreign Transfers to Enterprises	Enterprise Income
6. Recurrent State (1)	Indirect Taxes	Consumption Taxes plus Import Tariffs	Factor Taxes	Income Taxes	Enterprise Income Taxes			Net Foreign Transfers to State	State Revenue
7. Investment Savings (1)				Household Savings	Retained Earnings & Enterprise Savings	State Savings		Net Capital Inflows (=Foreign Savings)	Total Savings
8. Rest of World (94+1)		Imports							Imports
9. Total	Total Payments	Total Commodity Supply	Total Factor Payments	Allocation of Private Household Income	Total Enterprise Expenditure	Allocation of State Revenue	Total Investment	Total Foreign Exchange	

2. Developing Regional SAM Accounts

Three core components of a regional SAM database:

1. National SAM
2. Individual regional/provincial SAMs
3. Inter-regional Flow Data
    1. Trade flows
    2. Private and public distribution margins

Regional/Provincial SAMs

- These are very similar to national SAMs, but may pose special data challenges
- IO tables may be less reliable/detailed
- NIPA accounts are rarely complete at the regional level
- Capital and transfer accounts are likely to be incomplete (financial flows, remittances)

Inter-regional Flow Data

- Very few countries have reliable regional trade data
- This may be imputed from data on administrative taxes, transport, or other proxies
- The results are usually balanced against aggregate control totals, and very approximate

Development Strategy I

Database development should proceed in four steps:

1. An up-to-date national SAM
2. Individual regional/provincial SAMs, including a Residual Economy SAM to account for omitted regions
3. National aggregation balancing
4. Trade flow imputation

Development Strategy II

This approach would support two tiers of model implementation:

1. Individual regional/provincial models.
2. A multi-region national model.

Both types of model will be useful for different kinds of policy research. Generally, both types 2 will be implemented at the ministerial level, while only type 1 will be implemented at the regional level.

Direct SAM Analytical Methods

- In addition to its role as a static database for national accounting and CGE model calibration, the SAM can be used for direct estimation with a variety of multiplier methods.
- We describe one example here.

Regional Multiplier Decomposition

- While trade flow data are revealing, they only capture direct bilateral effects.
- In the real economy, a myriad of interactions delineate the path from initial expenditure to ultimate incomes.
- This is particularly the case with trade in an era of globalization, where international supply chains are ever more elaborate and indirect linkages can represent the majority of value creation.
- To assess these effects empirically, we use the international SAM for multiplier analysis.

Social Accounting Matrix

Consider an example of three regions, each represented by a social accounting matrix of the form

$$T_k = \begin{bmatrix} T_{kk} & F_k \\ V_k & X_k \end{bmatrix}$$

where the component matrices denote commodity flows (T), final demand (FD), value added (VA), and other domestic accounts (X).

Multilateral Social Accounting Matrix

Consider SAMs for three regions, compiled into a multi-regional transactions table

T_{11}	$\underline{T_{12}}$	$\underline{T_{13}}$	F_1
$\underline{T_{21}}$	T_{22}	$\underline{T_{23}}$	F_2
$\underline{T_{31}}$	$\underline{T_{32}}$	T_{33}	F_3
V_1	V_2	V_3	X

where the off-diagonal T matrices (underlined) are bilateral trade flows.

Block Decomposition

To elucidate **multi-lateral** regional trade linkages, we carry out the following block multiplier decomposition:

T_{11}	T_{12}	T_{13}	F_1
T_{21}	T_{22}	T_{23}	F_2
T_{31}	T_{32}	T_{33}	F_3
V_1	V_2	V_3	X

$$M = M_3 M_2 M_1$$

Block Decomposition (cont.)

$$M_1 = \begin{bmatrix} (I-A_{11})^{-1} & 0 & 0 \\ 0 & (I-A_{22})^{-1} & 0 \\ 0 & 0 & (I-A_{33})^{-1} \end{bmatrix}$$

$$M_2 = \begin{bmatrix} I & (I-A_{11})^{-1}A_{12} & (I-A_{11})^{-1}A_{13} \\ (I-A_{22})^{-1}A_{21} & I & (I-A_{22})^{-1}A_{32} \\ (I-A_{33})^{-1}A_{31} & (I-A_{33})^{-1}A_{32} & I \end{bmatrix}$$

$$M_3 = \begin{bmatrix} I-D_{12}D_{21}-D_{13}D_{31} & D_{21}D_{12} & D_{31}D_{13} \\ D_{12}D_{21} & I-D_{21}D_{12}-D_{23}D_{32} & D_{23}D_{32} \\ D_{13}D_{31} & D_{23}D_{32} & I-D_{31}D_{13}-D_{23}D_{32} \end{bmatrix}$$

Linkages

Intra-region

Inter-region (bilateral)

Equilibrium Indirect

Note: $D_{ij} = (I-A_{ii})^{-1}A_{ij}$

3. Reconciling China’s Regional Input-Output Tables

Motivation

- Provincial Input-output data 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

Objectives

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

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.

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 an 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:

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

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.

4. Multi-regional Trade Flows

- The availability of global trade flow data has dramatically advanced trade policy analysis
- Here we propose an *efficient* procedure for estimating a multi-regional trade flows across China
- Integrating this with a complete set of consistent provincial SAMs would create an integrated Multi-regional Social Accounting Matrix (MrSAM)

Motivation

- Single-region IO tables are already accessible, but neither mutually consistent not integrable
- MrSAM is of interest for its own sake, but can also support more coherent economywide policy analysis
  - CGE
  - Economic integration studies
- We propose creation of a prototype data set as a template for more standardized regional data reporting and management

Foundation – PRC Provincial IO Tables

- Already available
- Nationally comprehensive and consistent in terms of account definitions
- Builds on DRC capacity for SAM and CGE research at the national level

Consistency Issues

- Provincial trade statistics are maintained independently
- Domestic imports and exports are not consistently distributed across other sub-national regions
- There is very little accounting of margins arising from distribution costs and administrative measures

Proposed Approach

- Uses a new gravity specification to estimate bilateral trade econometrically
- Integrates the steps necessary to
  – Generate the interregional trade flow portions of the China MrSAM, while
  – insuring the consistency of the province accounts, regional aggregations, and the national system as a whole

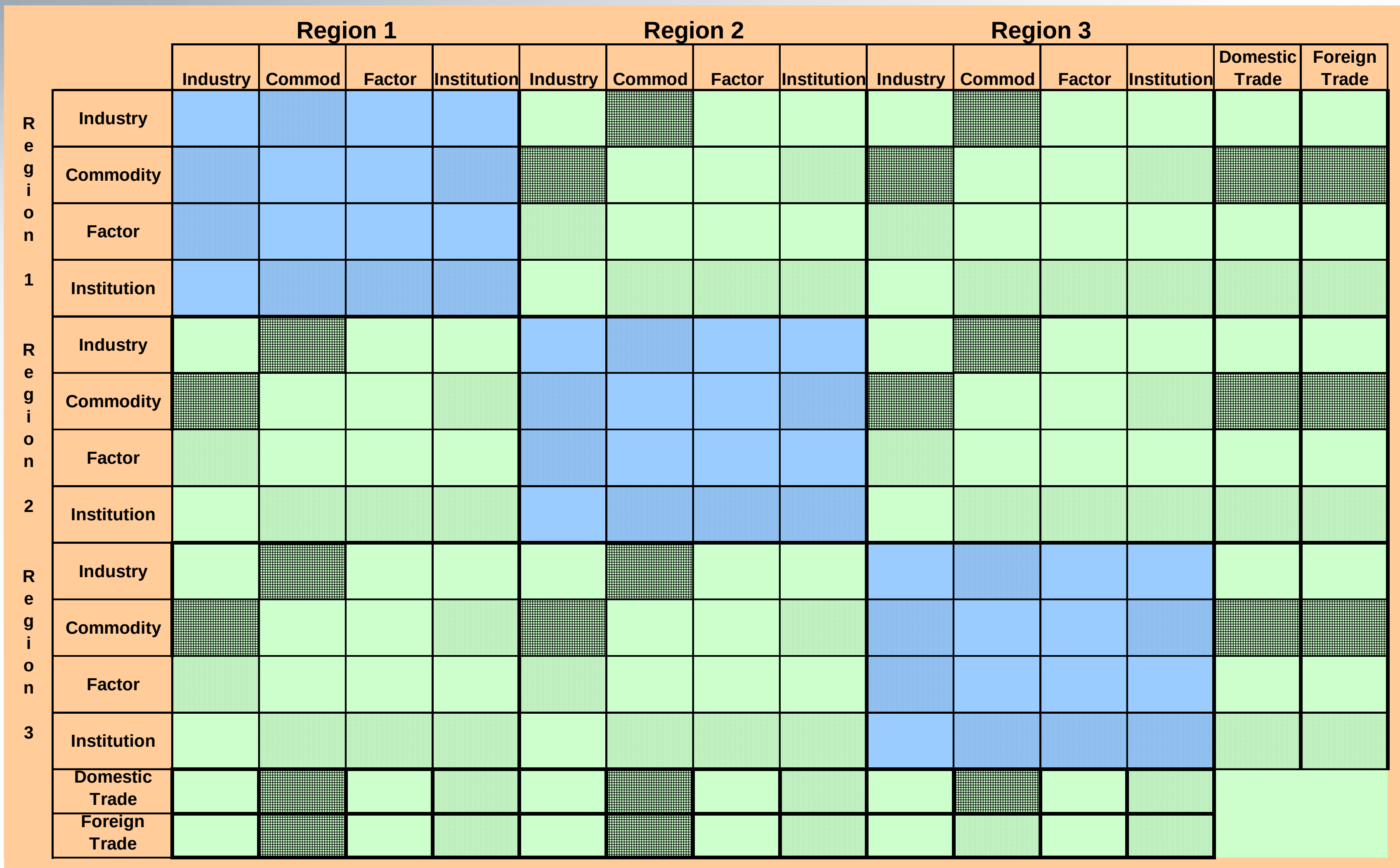
Procedure

- Definitional Framework
  - Define the provinces
  - Define sectoral classifications and detail
- Generate single-region and national tables
- Estimate interregional trade distributions by commodity

Overview of the Estimation Problem

- Extending prior DRC work (He and Li: 2004) we propose a new gravity model specification of bilateral trade.
- We then propose three alternative estimators.
- Each of these can be implemented with standard statistical software, and the most attractive estimates used for multi-regional analysis

Schematic Trade Matrix



Estimation Technique

- The gravity type model has been commonly used in estimating trade flows in international economics.
- We apply this approach to modeling and predicting regional trade flows with a variation of an international strategy proposed by Mátyás (1997).

Generic Gravity Model

- Consider the following specification

$$\ln y_{mnt}^{(i)} = \alpha_m + \gamma_n + \lambda_t + \beta_1 \ln Y_{mt}^{(i)} + \beta_2 \ln Y_{nt}^{(i)} + \beta_3 d_{mn} + \varepsilon_{mnt}$$

where:

$y_{mnt}^{(i)}$ is the volume of commodity i 's trade (exports) from region m to region n at time t ;

$Y_{mt}^{(i)}$ is the GDP for commodity i in region m at time t , and the same for $Y_{nt}^{(i)}$ for region n ;

d_{mn} is the distance between the regions m and n ;

α_m is the home regional effect, γ_n is the foreign regional effect, and λ_t is the time effect;

$m = 1, \cdots, N$, $n = 1, \cdots, i-1, i+1, \cdots, N+1$, where the $N+1$ -th element is the rest of the world, $t = 1, \cdots, T$;

$i = 1, \cdots, I$, the number of tradable goods;

ε_{mnt} is a white noise disturbance term.

Comments

- From an econometric point of view, the α , γ and λ specific effects can be treated as either random effects or fixed effects. In this analysis, we assume those specific effects associated with regions are time-invariant, and adopt the fixed effects approach.
- Also note that our main goal is prediction, so the parameter estimates for α , γ , λ , β_1 , β_2 , β_3 only bear the meaning of best linear predictor, not estimates for latent structural parameters.
- In addition, we could also add other terms to the right hand side, such as $\ln \text{POP}_{mt}$, and $\ln \text{POP}_{mt}$, the population for region m and region n at time t respectively.

Estimating Bilateral Trade Flows

Consider commodity $i \in \{1, 2, \cdots, I\}$, the explained variable, $\mathbf{y}^{(i)}$, in the model (1-1) is an $N \times N \times T$ -vector of observations, arranged in the form:

$$\mathbf{y}^{(i)} = \left(y_{121}^{(i)}, \cdots, y_{12T}^{(i)}, y_{131}^{(i)}, \cdots, y_{13T}^{(i)}, \cdots, y_{N11}^{(i)}, \cdots, y_{N1T}^{(i)}, \cdots, y_{N(N+1)1}^{(i)}, \cdots, y_{N(N+1)T}^{(i)} \right)'$$

The explanatory variables are arranged accordingly:

$$X^{(i)} = \left[D_\alpha, D_\gamma, D_\lambda, Y_{mt}^{(i)}, Y_{nt}^{(i)}, d_{mn} \right]$$

where D_α, D_γ and D_λ are dummy variable matrices for α, γ and λ.

Trade Flow Estimation 2

Then we stack these I ($i = 1, \cdots, I$) vectors to construct an I -good trade-flow (demand) system:

$$Y = \left(\mathbf{y}^{(1)\prime}, \mathbf{y}^{(2)\prime} \cdots, \mathbf{y}^{(I)\prime} \right)^{\prime}_{(N \times N \times T \times I) \times 1}$$

$$X = \begin{bmatrix} X^{(1)} & 0 & 0 & 0 \\ 0 & X^{(2)} & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & X^{(I)} \end{bmatrix}_{(N \times N \times T \times I) \times (6 \times I)}$$

Conclusions

- SAMs are critically important descriptive tools and resources for more advanced, evidence based policy analysis
- While they must be macroeconomically consistent, their biggest virtue is detail.
  - In most cases, indirect effects of economic policy outweigh direct ones, but these are often difficult to ascertain without deeper insight into linkages.
  - Data development for SAMs should be correspondingly ambitious.
- Overall goal: Improve ex ante visibility for policy makers about the detailed incidence of economic decisions and external events.

Discussion