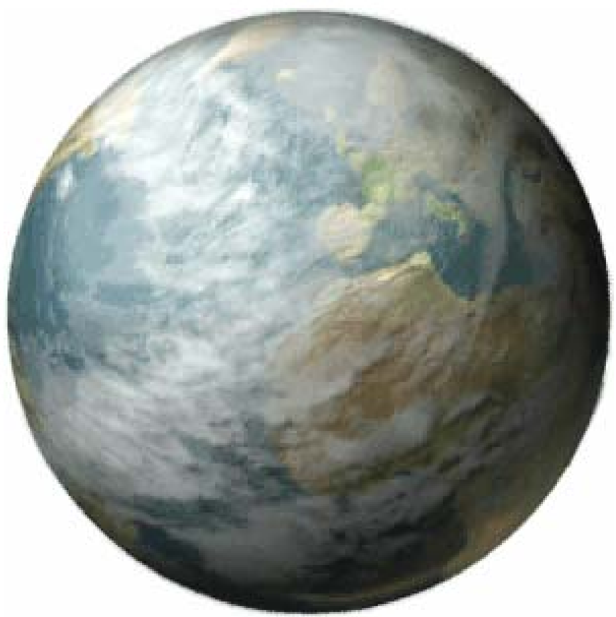
# Provincial Trade Flow Estimation for China



David Roland-Holst and Muzhe Yang

UC Berkeley

Lecture II

Presented to the

Development Research Centre

State Council of the PRC

Beijing, 6 June 2005

# Objectives

- Implement an *efficient* procedure for estimating a multi-regional trade flows across China
- Integrate this with a complete set of consistent provincial SAMs to form a national 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 integrated policy analysis
  - CGE
  - Economic integration studies
- Generate 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

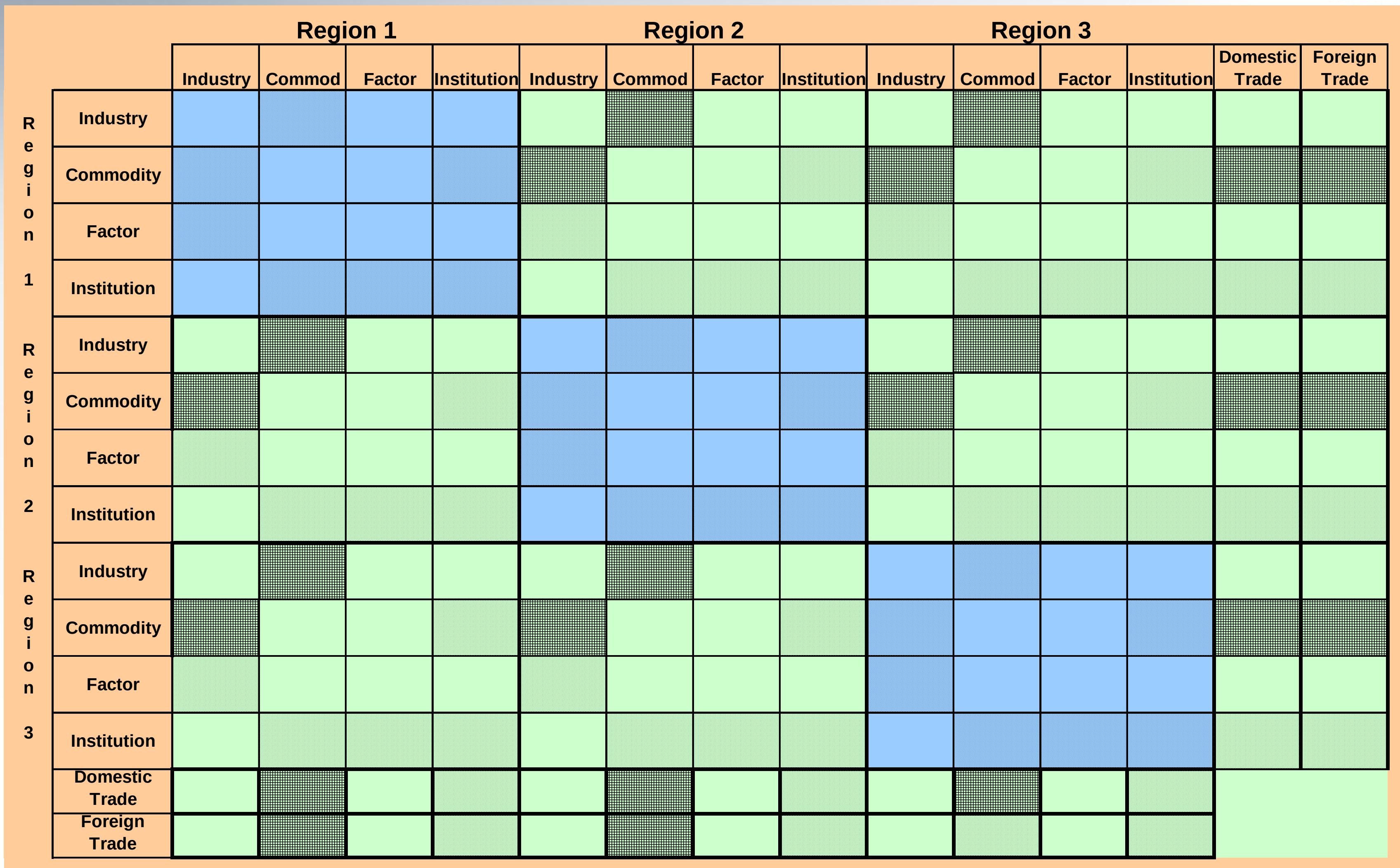
# Procedure

- 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 $n$ ;

$\alpha_m$ is the home regional effect, $\gamma_n$ is the foreign regional effect, and $\lambda_t$ 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;

$\varepsilon_{mnt}$ is a white noise disturbance term.

# Comments

- From an econometric point of view, the $\alpha$ , $\gamma$ and $\lambda$ 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 $\alpha$ , $\gamma$ , $\lambda$ , $\beta_1$ , $\beta_2$ , $\beta_3$ 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 $D_\alpha$, $D_\gamma$ and $D_\lambda$ are dummy variable matrices for $\alpha$, $\gamma$ and $\lambda$.

# Trade Flow Estimation 2

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

$$Y=\left(\mathbf{y}^{(1)\prime},\mathbf{y}^{(2)\prime}\cdots,\mathbf{y}^{(I)\prime}\right)'_{(N\times N\times T\times I)\times 1}$$

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

# Three Alternative Estimators

Modern econometrics has developed a large set of alternative estimation strategies, each performing differently under different data conditions.

For the present application, we recommend three alternative estimators, to be evaluated ex post in terms of statistical performance:

1. Ordinary Least Squares (OLS)– the most traditional approach
2. Seemingly Unrelated Regressors (SUR) – an generalization of OLS that imposes less prior assumptions on the data structure
3. Generalized method of moments (GMM)

# Ordinary Least Squares 1

OLS estimates for Y and X above can be obtained as follows:

Regressand: vector $\mathbf{y}^{(i)}$ consists of bilateral trade flows of commodity $i \in \{1, 2, \cdots I\}$ between region $m$ ( $m = 1, \cdots, N$ ) and region ( $n = 1, \cdots, i-1, i+1, \cdots, N+1$ ) sorted by $t$ ( $t = 1, \cdots, T$ ).

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

$$Y = \left( \mathbf{y}^{(1)'}, \mathbf{y}^{(2)'} \cdots, \mathbf{y}^{(I)'} \right)'_{(N \times N \times T \times I) \times 1} \qquad (\text{where } i = 1, \cdots, I)$$

# OLS 2

Regressors: matrix $X^{(i)}$ consists of dummy variables for home region $m$, foreign region $n$ and time $t$:

$$D_\alpha = \begin{cases} 1 & \text{for region } m \\ 0 & \text{else} \end{cases}$$

$$D_\gamma = \begin{cases} 1 & \text{for region } n \\ 0 & \text{else} \end{cases}$$

$$D_\lambda = \begin{cases} 1 & \text{for time } t \\ 0 & \text{else} \end{cases}$$

$$X^{(i)} = \left[ D_\alpha, D_\gamma, D_\lambda, Y_{mt}^{(i)}, Y_{nt}^{(i)}, d_{mn} \right]_{(N\times N\times T)\times 6}$$

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

# OLS 3

Disturbances are given by $$\varepsilon = \left(\varepsilon_1', \varepsilon_2', \cdots, \varepsilon_I'\right)'_{(N\times N\times T\times I)\times 1}$$

with
$$E\left(\varepsilon \mid X\right) = 0$$
$$E\left(\varepsilon\varepsilon' \mid X\right) = \sigma^2 I_{N\times N\times T\times I}$$

Now formulate the model as

$$Y = X\beta + \varepsilon$$

and estimate with

$$\widehat{\beta}_{OLS} = \left(X'X\right)^{-1}\left(X'Y\right)$$

# Seemingly Unrelated Regressions 1

In this case we use a technique called Feasible Generalized Least Squares (FGLS), with

Regressand: vector $\mathbf{y}^{(i)}$ consists of bilateral trade flows of commodity $i \in \{1, 2, \cdots I\}$ between region $m$ ($m = 1, \cdots, N$) and region ($n = 1, \cdots, i-1, i+1, \cdots, N+1$) sorted by $t$ ($t = 1, \cdots, T$).

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

$$Y = \left( \mathbf{y}^{(1)'}, \mathbf{y}^{(2)'} \cdots, \mathbf{y}^{(I)'} \right)'_{(N \times N \times T \times I) \times 1} \qquad \text{(where } i = 1, \cdots, I\text{)}$$

# SUR 2

Regressors: matrix $X^{(i)}$ consists of dummy variables for home region $m$, foreign region $n$ and time $t$:

$$D_\alpha = \begin{cases} 1 \text{ for region } m \\ 0 \quad \text{else} \end{cases}$$

$$D_\gamma = \begin{cases} 1 \text{ for region } n \\ 0 \quad \text{else} \end{cases}$$

$$D_\lambda = \begin{cases} 1 \text{ for time } t \\ 0 \quad \text{else} \end{cases}$$

$$X^{(i)} = \left[ D_\alpha, D_\gamma, D_\lambda, Y_{mt}^{(i)}, Y_{nt}^{(i)}, d_{mn} \right]_{(N \times N \times T) \times 6}$$

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

# SUR 3

Disturbances in this case are given by

$$E(\varepsilon \mid X) = 0$$

with

$$E(\varepsilon\varepsilon' \mid X) = \Omega_{(N\times N\times T\times I)\times(N\times N\times T\times I)}$$

$$= \Sigma_{(I\times I)} \otimes I_{(N\times N\times T)\times(N\times N\times T)}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1I} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{I1} & \sigma_{I2} & \cdots & \sigma_{II} \end{bmatrix}$$

# SUR 4

Now formulate the model as in OLS, i.e. $Y = X\beta + \varepsilon$

but estimate with FGLS as

$$\hat{\beta}_{FGLS} = \left(X'\hat{\Omega}^{-1}X\right)^{-1}\left(X'\hat{\Omega}^{-1}Y\right)$$

$$= \left[X'\left(\hat{\Sigma}^{-1} \otimes I\right)X\right]^{-1}\left[X'\left(\hat{\Sigma}^{-1} \otimes I\right)Y\right]$$

where $\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1I} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{I1} & \hat{\sigma}_{I2} & \cdots & \hat{\sigma}_{II} \end{bmatrix}$

The least squares residuals $\mathbf{e} = Y - X\hat{\beta}_{OLS}$ can be used to estimate consistently the elements of $\Sigma$ with

$$\hat{\sigma}_{ij} = \frac{\mathbf{e}_i'\mathbf{e}_j}{N \times N \times T} \quad \left(i, j = 1, \cdots, I\right)$$

# Generalized Method of Moments 1

Using information of aggregate provincial/regional trade flows by commodity, we can add additional moment restrictions:

$$\sum_{m=1}^{N} \mathbf{y}_{\cdot nt}^{(i)} = IM^{(i)} \quad (i = 1\cdots, I)$$

$$\begin{bmatrix} \sum_{m=1}^{N} \mathbf{y}_{\cdot nt}^{(1)} \\ \vdots \\ \sum_{m=1}^{N} \mathbf{y}_{\cdot nt}^{(I)} \end{bmatrix} = \begin{bmatrix} IM^{(1)} \\ \vdots \\ IM^{(I)} \end{bmatrix}$$

where *IM* denotes provincial/regional domestic import demand.

# GMM 2

Regressand: The vector $\mathbf{y}^{(i)}$ consists of bilateral trade flows of commodity $i \in \{1, 2, \cdots I\}$ between region $m = 1, \cdots, N$, and region $n = 1, \cdots, i-1, i+1, \cdots, N+1$, sorted by $t$ ( $t = 1, \cdots, T$ ).

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

$$Y = \left( \mathbf{y}^{(1)\prime}, \mathbf{y}^{(2)\prime} \cdots, \mathbf{y}^{(I)\prime} \right)'_{(N \times N \times T \times I) \times 1} \qquad \text{(where } i = 1, \cdots, I\text{)}$$

# GMM 3

Regressors: matrix $X^{(i)}$ consists of dummy variables for home region $m$, foreign region $n$ and time $t$:

$$D_\alpha = \begin{cases} 1 \text{ for region } m \\ 0 \quad \text{else} \end{cases}$$

$$D_\gamma = \begin{cases} 1 \text{ for region } n \\ 0 \quad \text{else} \end{cases}$$

$$D_\lambda = \begin{cases} 1 \text{ for time } t \\ 0 \quad \text{else} \end{cases}$$

$$X^{(i)} = \left[ D_\alpha, D_\gamma, D_\lambda, Y_{mt}^{(i)}, Y_{nt}^{(i)}, d_{mn} \right]_{(N \times N \times T) \times 6}$$

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

# GMM 4

Disturbances in this case are given by

$$E\left(\varepsilon \mid X\right) = 0$$

with $$E\left(\varepsilon\varepsilon' \mid X\right) = \Omega_{(N\times N\times T\times I)\times(N\times N\times T\times I)}$$

$$= \Sigma_{(I\times I)} \otimes I_{(N\times N\times T)\times(N\times N\times T)}$$

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1I} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2I} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{I1} & \sigma_{I2} & \cdots & \sigma_{II} \end{bmatrix}$$

# GMM 5

Again we formulate the model as in OLS, i.e. $Y = X\beta + \varepsilon$

but use the GMM estimator given by

$$\widehat{\beta}_{GMM} = \arg\min_{\widehat{\beta}} \left( \frac{\sum_{i=1}^{N\times N\times T\times I} \psi_{i\cdot}(X, Y \mid \widehat{\beta})}{N \times N \times T \times I} \right)' W \left( \frac{\sum_{i=1}^{N\times N\times T\times I} \psi_{i\cdot}(X, Y \mid \widehat{\beta})}{N \times N \times T \times I} \right)$$

# GMM 6

where

$$\underset{(K+I)\times 1}{\psi_{i\cdot}(X,Y\,|\,\widehat{\beta})} = \begin{bmatrix} \frac{1}{N\times N\times T\times I}\sum_{i=1}^{N\times N\times T\times I} x_{i1}\left(y_i - x_{i1}'\widehat{\beta}\right) \\ \vdots \\ \frac{1}{N\times N\times T\times I}\sum_{i=1}^{N\times N\times T\times I} x_{iK}\left(y_i - x_{i1}'\widehat{\beta}\right) \\ \frac{1}{N}\sum_{m=1}^{N}\mathbf{y}_{\cdot nt}^{(i)} - \frac{1}{N} IM^{(1)} \\ \vdots \\ \frac{1}{N}\sum_{m=1}^{N}\mathbf{y}_{\cdot nt}^{(I)} - \frac{1}{N} IM^{(I)} \end{bmatrix}$$

and

$$\widehat{W} = \widehat{\Delta}^{-1}$$

$$\widehat{\Delta} = \widehat{\mathrm{Var}}\left(\psi_{i\cdot}(X,Y\,|\,\widehat{\beta})\right)$$

$$= \frac{\sum_{i=1}^{N\times N\times T\times I}\left(\psi_{i\cdot}(X,Y\,|\,\widehat{\beta})\right)\cdot\left(\psi_{i\cdot}(X,Y\,|\,\widehat{\beta})\right)'}{N\times N\times T\times I}$$

# Estimator Selection

- After generating estimates by all three methods, we can use a variety of criteria to choose between them.
- In traditional econometric analysis, one would use the goodness of fit measure, adjusted $R^2$ as the selection criterion.
- For our primary objective is imputing missing bilateral trade flows, we would choose the estimator with the largest $\bar{R}^2$.

# References


- He, Jianwu, and Shantong Li (2004), “A Three-regional CGE Model for China,” presentation to the DRC, Beijing, November.
- He, Jianwu (2004), “A Social Accounting Matrix for China: 2000,” processed, DRC, Beijing, November.
- Mátyás, L. (1997), “Proper Econometric Specification of the Gravity Model,” *The World Economy* 20(3): 363--368.