Estimating China's Inter-provincial Commodity Trade Coefficients[*]

Zhang Yaxiong[**], Li Jifeng, Zhao Kun and Liu Yu
(Department of Economic Forecasting, State Information Center, P.R.China)

Abstract: This paper discussed the method and procedures to estimate China's inter-provincial commodity trade coefficients in the context of the inter-regional input-output framework. Based on the literature review and our previous studies, we propose a more detailed model, which has the uniform format with gravity model, entropy maximizing model etc. and develop a distance deterrence equation to effectively employ the reliable provincial input-output and detailed transportation data. Then we estimate 2002 China's inter-provincial commodity trade coefficients at 29-sector classification. Since the compilation technique of 2002 provincial input-output table has been broad improved and the available of transport data and Customs statistical data, they become the main primary data sources. Accordingly, we could give more reliability to the 2002 provincial input-output tables, from which, the inflow, outflow, import and export data are adjusted.

Key words: inter-regional input-output, inter-regional trade, provincial input-output table, gravity model, entropy model

---

[*] We thanks Meng Bo, Institute of Developing Economies, Japan, Okamoto Nobuhiro, Daito Bunka University, Japan, and Qi Shuchang, Liu Huiping, National Bureau of Statistics, China, for their valuable contributions in this project.

[**] zhangyx@mx.cei.gov.cn, +86-10-6855-7129, Department of Economic Forecasting, State Information Center, No.58 Sanlihe Road, Beijing P.R.China, 100045.

1. **Introduction**

Accompanied with the fast development of China's economy, regional economic relation and disparity study becomes one of the important issues, and thus many policy makers as well as researchers have paid attentions to deal with. In order to carry out the studies in this field, the estimation of inter-regional trade flow matrix by commodities is the essential fundamental work, which is also the important part to compile the multi-regional input-output table. But most of countries in the world, even developed countries, do not have enough statistical files to build such matrix, which is usually estimated by some non-survey methods, like gravity model, entropy model and others. In China, there are quite few studies to carry out such matrix compilation till now.

Since the wide applications of inter-regional input-output model, we has compiled 1997 China multi-regional input-output model (SIC, 2005). The inter-regional transaction matrix was estimated at 8 regions and 30 sectors level. Based on the literature review and our previous studies, this paper tries to estimate 2002 China's inter-provincial commodity trade coefficients at 30-province and 29-sector level. We propose a model, which is initialed from the uniform format of gravity model and entropy maximizing model, but with a more detailed formula description on distance deterrence based on the reliable 30 provincial input-output and detailed transportation data.

2. **Literature review**

2.1 The definition of the trade coefficient in the input-output context

Isard (1951) initially developed the inter-regional input-output model with detailed inter-regional trade coefficient matrix:

$$X_i^r = \sum_s \sum_i a_{ij}^{rs} X_j^s + \sum_s Y_i^{rs} \tag{1}$$

Where, X_i^r is the total output of industry i in region r, a_{ij}^{rs} is the technical coefficients of region s defined the intermediate demand of industry j in region s on commodity i from region r, Y_i^{rs} is the final demand of region s from industry i in region *r*.

In the reality, especially within an economy, the parameters a_{ij}^{rs} in this non-competitive inflow type inter-regional input-output model are difficult to compile because of the shortage of the primary data. Under this circumstance, Moses (1955) proposed a competitive inflow type model for inter-regional trade, which is so-called Chenery-Moses model to simplify this problem:

$$X_i^r = \sum_s \sum_i t_i^{rs} (a_{ij}^s X_j^s + Y_i^s) \tag{2}$$

Where t_i^{rs} is the inter-regional trade coefficient, which is ratio of the movement of commodity *i* in region *r* to region *s*. Comparing with a_{ij}^{rs}, the assumption is the same distribution ratio for each sector in region *s* of the commodity *i* from region *r*. In this paper, we estimate China's inter-provincial trade coefficients defined within this model.

2.2 Literature review of the trade coefficient estimation

Carey(1858) made the first contribution to estimate the human spatial interaction by applying the Newton's law of the gravitational force F_{ij} between two masses m_i and m_j, which possesses the distance of d_{ij}:

$$F_{ij} = r \frac{m_i m_j}{d_{ij}^2} \tag{3}$$

It means that the attraction force is driven by the concentration of the people size and the inverse of distance. Following Carey's pioneer work, since the late 1940s, the gravity model were introduced and promoted in terms of spatial interaction studies by Stewart(1948), Zipf(1949), Dodd(1950) and so on. Based on their contributions, Anderson(1955) proposed a classical form of unconstrained gravity model:

$$T_{rs} = kO_r^{\gamma} D_s^{\alpha} d_{rs}^{\beta} \tag{4}$$

Where T_{rs} is the trade flow from region *r* to *s* , *k* is a scaling constant needed for normalization, α , β and γ are predominated parameters. Assuming α and γ =1 ,

β =-2 ,equation (4) will be the same as equation (3). The unconstrained gravity model has an obvious weakness. Suppose O_r and D_s doubled and keeping d unchanged, the total amount of T_{rs} will increase 4 times, which is not consistent with the fact that it should be 2 times, which equals to the total of O_r and D_s (Batten and Boyce 1986). In order to overcome this problem, the double-constraint was added into this model:

$$\sum_s T_{rs} = O_r \tag{5}$$

$$\sum_r T_{rs} = D_s \tag{6}$$

Wilson (1967) merged these two restrictions into the equation (4) by introducing the balancing factors A_r and B_s, so the equation of the double-constraint gravity model is as follows：

$$T_{rs} = A_r B_s O_r D_s f(d_{rs}) \tag{7}$$

$$A_r = [\sum_s B_s D_s f(d_{rs})]^{-1} \tag{8}$$

$$B_s = [\sum_r A_r O_r f(d_{rs})]^{-1} \tag{9}$$

Gravity model has been applied in various studies, such as human behavior, transportation movement and inter-regional interaction. Leontief and Strout（1963）proposed following equation to estimate the inter-regional trade coefficients based on the idea of the gravity model：

$$T_i^{rs} = \frac{X_i^r D_i^s}{\sum_r X_i^r} Q_i^{rs} \tag{10}$$

The coefficient Q_i^{rs} is the distance decay parameter, which can be viewed as empirical constant and is negatively related to per-unit transportation cost. In this case, this equation can be considered as a special instance or simplified gravity model as it merged the distance variable into the pre-determined constant. In practice, how to estimate Q_i^{rs} becomes a difficult problem. In general, the transportation primary data are employed to estimate Q_i^{rs} as a proxy. Ihara (1979 and 1996) defined proportional distribution coefficient of inter-regional commodity flows, thus to calculate Q_i^{rs} for each commodity by using fright exchange between regions, which is so called origin and destination (OD) table.

It was thought that those models following the idea originally from the gravitational force is lack of the explanation capacity under the background of economic theory. Besides, the gravity model cannot reflect the individual's behavior, which can be explained by probabilistic theory. In this case, different from the gravity model, Wilson (1970) adopted another approach called entropy-maximizing model to estimate the inter-regional trade flows. The general form of the model is as follows:

$$\text{maximize} \quad E = -\sum_r \sum_s T_{rs} \ln T_{rs} \tag{11}$$

$$s.t. \quad \sum_s T_{rs} = O_r \tag{12}$$

$$\sum_r T_{rs} = D_s \tag{13}$$

$$\sum_r \sum_s T_{rs} c_{rs} = C \tag{14}$$

Where E is the entropy, c_{rs} is the cost to transport one unit of goods i from region r to region s, C is the total transportation cost of goods i , which represents the total budget constrain. Solving the programming (11)-(14), we can calculate the inter-regional trade transaction:

$$T_{rs} = A_r B_s O_r D_s \exp(-\eta c_{rs}) \tag{15}$$

$$A_r = [\sum_s B_s D_s \exp(-\eta c_{rs})]^{-1} \tag{16}$$

$$B_s = [\sum_r A_r O_r \exp(-\eta c_{rs})]^{-1} \tag{17}$$

Where A_r and B_s are also the balancing factors , which have the same format with equations (8) and (9), and η is the Lagrangian multipliers of the constrain equation (14).

It is clear that the entropy model (15) and gravity model (7) have the same function form, while the ideas on how to define the deterrence between origin and destination are different. The gravity model uses the distance to represent the deterrence, while the transportation cost is introduced into entropy maximizing model.

Actually, the similarity between gravity and entropy-maximizing model is not unique. The multinomial logit model from random utility maximization theory developed by McFadden (1974)，the potential model from deterministic profit maximization theory developed by Amano and Fujita (1970) are both similar to gravity model in terms of the formation of trade coefficient t_i^{rs} , as well as in the updated work done by Meng and Ando (2005), the model was developed based on the assumption of supplier profit maximizing. These works on one hand contribute the approaches estimating inter-regional trade flow or trade coefficient, on the other hand provide comprehensive and rational explanations for the application of gravity model, which is one of the most useful non-survey methods to estimate the inter-regional trade flow. Since the individual behavior information in the inter-regional trade is hard to attain, and the data of producer price of commodities produced in region *r* and the transportation cost data from *r* to *s* are also more difficult to get than the time-distances, the gravity model is relative feasible to use than other various models.

In terms of the gravity model, there has been some theoretical improvement. Alonso(1973, 1978) followed Wilson (1970) and expanded the double constrained gravity model with the so-called “systemic variable”, which make gravity model more flexible. Fotheringham(1984) proved that the “systemic variable” are inverse to the weighted distance between the origin and other destinations or between the destination and other origins, which gave the explanations on that important variables.

Comparing with the theoretical improvement, there are relative more application studies on the inter-regional trade flowing. Chisholm and O’Sullivan (1973), using UK 1962 and 1964 commodities flow data, covering 78 zones and 13 commodity groups, estimated a production-constrained gravity model. Black (1971, 1972) analyzed the properties and determinants of the distance exponent in the gravity model using the 1967 US commodities flow statistical data for 24 major shipper groups. Ashtakala and Murthy (1988), using data from a 1977 survey of commodity flows between communities within Alberta, estimated a production-constrained gravity model, where the independent variables are the production (origin) and consumption (destination) of a commodity, and the distance between a pair of communities. There are still many other works, like Smith(1975), Sasaki etc(1987), Okuda (2003), Okubo (2004), Celika and Guldmann(2007) and so on.

In China, there are few studies on estimation of trade flow matrix. Liu and Okamoto (2002) have tested the gravity model (Leontief and Strout model) and econometric model to estimate the trade flow matrix by 3 regions and 10 sectors for 1997. Zhang and Zhao (2005) estimated a more detailed inter-regional trade flow matrix by 8 regions and 30 sectors, and compiled the first China multiregional input-output table by survey and non-survey methods. Canning and Wang (2004) suggest using the statistical methods like minimizing the cross-entropy theory to optimize an initial inter-regional trade flow matrix in the framework of multiregional input-output table, which could introduce more effective information into the adjustment process to improve the matrix, and this suggestion had been adopted by Xu and Li (2008) in their recent work.

3. **The model**

Among each kind of model, the main difference is how to define the deterrence between origin and destination. We believe it is important to base on the availability of the primary data source. In this paper, the following model is proposed to estimate China's inter-provincial trade coefficient matrix:

$$T_i^{rs} = A_i^r B_i^s X_i^{ro} X_i^{os} f({}^k D_i^{rs}) \tag{18}$$

Where,

$$A_i^r = [\sum_s B_i^s X_i^{os} f({}^k D_i^{rs})]^{-1} \tag{19}$$

$$B_i^s = [\sum_r A_i^r X_i^{ro} f({}^k D_i^{rs})]^{-1} \tag{20}$$

Where X_i^{ro} and X_i^{os} are respectively the total outflow of commodity *i* from region *r* and the total inflow to region *s*. For applying the above model, the important issue is how to determine the function form of $f(d)$. Based on the existing literature reviews and data, the following distance deterrence equation can be considered:

$$f({}^{k}D_i^{rs}) = \sum_k {}^{k}M_i ({}^{k}D_i^{rs})^{-{}^{k}\alpha_i^{r}} \tag{21}$$

Where ${}^{k}M_i$ represents the share ratio of freight mode k in the transportation of commodity i from r to s. ${}^{k}D_i^{rs}$ is the time-distance from region r to s by commodity and transportation mode. ${}^{k}\alpha_i^{s}$ is the power parameter, reflecting the influence magnitude of time-distances for trade activities.

We propose the follow method to estimate the parameter ${}^{k}\alpha_i^{s}$.

The average transport distance of transportation mode k from region r to other regions for commodity i can be written as,

$$\frac{\sum_r {}^{k}C_i^{rs}\, {}^{k}d_i^{rs}}{\sum_r {}^{k}C_i^{rs}} = {}^{k}\bar{d}_i^{s} \quad \text{s.t.} \quad C_i^{rs} \geq 0 \tag{22}$$

Where ${}^{k}C_i^{rs}$ is the quantity of freight from region r to s for commodity i by transportation mode k, ${}^{k}d_i^{rs}$ is the corresponding distance variable, ${}^{k}C_i^{rs}\, {}^{k}d_i^{rs}$ shows the freight turnover for commodity i from region r to s by transportation mode k, ${}^{k}\bar{d}_i^{s}$ is the average transport distance of transportation mode k to region s from other regions for commodity i, which equals to the total turnover of freight divided by the total quantity of transportation to region s from other regions for commodity i by transportation mode k.

Then ${}^{k}C_i^{rs}$ was substituted by the following destination-constraint gravity model:

$${}^{k}C_i^{rs} = {}^{k}A_i^{s} \cdot {}^{k}C_i^{ro} \cdot {}^{k}C_i^{os} \cdot ({}^{k}d_i^{rs})^{-{}^{k}\alpha_i^{s}} \tag{23}$$

Where ${}^{k}C_i^{ro}$ is the total outflow of commodity i from region r by transportation mode k, ${}^{k}C_i^{os}$ is the total inflow of commodity i of region s through transportation mode k, we can have:

$$\frac{\sum_r {}^kC_i^{ro}({}^kd_i^{rs})^{1-{}^k\alpha_i^s}}{\sum_r {}^kC_i^{ro}({}^kd_i^{rs})^{-{}^k\alpha_i^s}} = {}^k\bar{d}_i^s \tag{24}$$

Since we know ${}^kC_i^{ro}$, then we can solve this equation to calculate ${}^k\alpha_i^s$.

4. **Estimating China's inter-provincial trade coefficients**

Comparing with other countries, the related primary data on the inter-provincial trade flow is much less in China. The basic reliable data are from 2 sources, the first one is the provincial input-output tables, and the second one is the OD tables of railway transportation by 15 and other kinds of commodities and the dispatch freight of road and waterways by 16 and other kinds of commodities. Each province except for Tibet compiled the provincial input-output table. Currently, the updated provincial input-output table is at year 2002.

4.1 The calculation of outflow and inflow vectors for each province

The 2002 provincial input-output tables supply reliable primary data. Although there have various differences between provincial level input-output table and national input-output table, we can consider the provincial input-output tables reflect the economic structure of each province. Particularly, provincial input-output tables provide basic information on outflow and inflow transaction by sector. It is a good starting point to calculate the outflow and inflow vectors for each province. There are 3 kinds of tables in term of the treatment of the outflow and inflow vectors：

(1) The four-column table. It means that those tables contain (domestic) outflow and inflow vectors and (international) export and import vectors by sector. The following 12 provinces process this kind of table: Beijing, Tianjin, Hebei, Liaoning, shanghai, Jiangsu, Zhejiang, Guangdong, Hainan, Anhui, Guangxi, Xinjiang.

Ideally, we can directly apply the domestic outflow and inflow vectors in the provincial input-output tables to estimate inter-provincial trade coefficients. However there has the discrepancy between Customs data and provincial input-output table. For example, in Hebei province, the value of export of coal mining sector is 50.98 billion yuan in Customs data, while the total value of export plus outflow of coal mining sector in the Hebei input-output table is only 32.19 billion yuan. In this case, we adopt the Customs export value in coal mining sector, and set zero for the domestic outflow.

(2) The two-column table. The table only has outflow and inflow vectors, where the outflow is the sum of the domestic outflow and export, the inflow is the sum of the domestic inflow and import. The following 13 provinces process this kind of table: Jilin, Fujian, Shanxi, Henan, Jiangxi, Hunan, Hubei, Inner-Mongolia, Shaanxi, Yunnan, Ningxia, Gansu, Qinghai.

In this case, we need to split these two-column tables to four-column tables. We set up a converter between China input-output sector classification (here 26 commodities sectors) and HS 10-digit code (7163 commodities in 2002); then we aggregate the provincial export/import by input-output sectors. It should be note that in order to compile the producer/purchaser price export/import vectors, we need to estimate the trade and transportation margins (TTM) and tariff to modify the vectors we have from previous step.

Because the Customs statistics do not have the service import and export, we estimate the import and export data in service sectors assuming that the services import and export are proportional to the total amount of import and export of commodities.

(3) The one-column table. The table only has net outflow vector, which is the difference between the outflow (domestic outflow + export) and inflow (domestic inflow + import). The following 5 provinces process this kind of table: Heilongjiang, Shandong, Sichuan, Chongqing, Guizhou.

The net outflow vector can provide us little information to split to four-column table, because of no total control of inflow and outflow, and even including some error items inside in some provincial tables. In this case, we give up the net outflow data in the provincial input-output tables and estimate the four vectors for these provinces directly. The import and export vectors are estimated from the Customs statistical data. Then we estimate the inflow and outflow by introducing the structural information in neighboring province. For example, we consider the ratio of inflow to total use in Heilongjiang is similar to the ratio in Liaoning province, and Shandong is similar to Hebei province.

4.2 The estimation of deterrence equation

According to the deterrence equation (21), there are three parameters should be defined firstly, which are the time-distance ${}^{k}d_{i}^{rs}$, share ratios of freight mode ${}^{k}M_{i}$, and power parameter ${}^{k}\alpha_{i}^{s}$.

4.2.1 The estimation of distance between each province

We calculate three kinds of transportation distances between each province, which are the railway, road and waterway distance. For the calculation of railway and road distance between each province, we simply use the capital as the proxy point to estimate the corresponding distance.

For the railway distance, we try to calculate "time distance" between capitals of the provinces by applying the passenger railway transport timetable. But for some provinces' capitals, there not exist the direct passenger train between them. In this case, we select the most efficient connections through national railway network and add reasonable transfer time.

For the road distance, we use the length of the expressway between each capital of the province to measure the distance between each province. But there not exist direct expressway between some of the provinces' capitals. In this case, similarly, we choose the most efficient connections through national expressway network.

For the waterway distance, we select 17 provinces, because only those provinces possess important ports with large shipping fright volume. The 17 provinces are Heilongjiang, Liaoning, Tianjin, Hebei, Shandong, Shanghai, Jiangsu, Zhejiang, Fujian, Guangdong, Hainan, Anhui, Jiangxi, Hubei, Hunan, Guangxi and Chongqing.

The biggest port in each province is selected. Then we use the distance between each pair of these ports to represent the waterway distance between each province.

4.2.2 The estimation of share of freight mode ${}^{k}M_{i}$

There are five transportation modes for commodity, which include railway, road, waterway, air and pipeline. Since the ratios of freight by air and pipeline are rarely small, we only calculate the share of railway, road and waterway transportation. For the big difference between classification of transport commodities and input-output sector, we can not estimate the share of freight mode by commodities. In this paper, we just calculate the average share of freight mode ${}^{k}M$, which are 13% for railway, 76% for road transport, and 11% for water transport.

4.2.3 The estimation of power parameters ${}^{k}\alpha_{i}^{s}$

According to the present data, it is difficult to estimate power parameters by different mode, for each province at commodity level ${}^{k}\alpha_{i}^{s}$. For the data limitation, we just calculate the power parameter by using railway OD table, and the average railway transport distance. We find the values for power parameters of different commodities among provinces are about 1.9, which is close to the value 2.0 estimated by previous studies.

5. **Closing remarks**

This paper discussed the method and procedures to estimate China's inter-provincial commodity trade coefficients in the context of the inter-regional input-output framework. Since the compilation technique of 2002 provincial input-output tables has been broad improved, we could give more reliability to the provincial input-output tables and Customs statistical data. In this case, we reviewed the popular non-survey approaches, such as the gravity model, entropy model and other related models. Based on these theoretical models, we propose a more detailed model using the above data effectively to estimate the inter-provincial commodity trade coefficients. Particularly, we developed a distance deterrence equation based on the initial data and found a way to get the solution by using the detailed transportation primary data, which is another reliable data source.

References:

A.G.Wilson. 1967. A statistical theory of spatial distribution models. *Transportation Research*, 1: 253-269.

A.G.Wilson. 1970. *Entropy in Urban and Regional modeling*. London: Pion.

Anderson, T.R., Intermetropolitan migration: a comparison of the hypotheses of Zip and Stouffer, *American Sociological Review*, 20, pp.285–291, 1955.

Ashtakala B, Murthy ASN. 1988. Optimized gravity models for commodity transportation. *Journal of Transportation Engineering* (ASCE),114:393–408.

Batten, D.F. and Boyce, D.E. 1986. Spatial interaction, transportation, and interregional commodities flow models. The Chapter 9 in *Handbook of regional and urban economics* Vol.1, edited by Peter Nijkamp. North-Holland, 1986.

Black WR. 1971. The utility of the gravity model and estimates of its parameters in commodity flow studies. *Proceedings of the Association of American Geographers*, 3:28–32.

Black WR. 1972. Interregional commodity flows: some experiments with the gravity model. *Journal of Regional Science*, 12:107–118.

Chisholm M, O'Sullivan P. 1973. *Freight flows and spatial aspects of the British economy*. New York and London: Cambridge University Press.

Dogg, S.C. 1950. The interactance hypothesis: a gravity model fitting physical masses and human groups. *American Sociological Review*, 15:245-256.

E.C.Young. 1924. *The movement of farm population*, Bulletin 426, Ithaca, Cornell Agricultural Experiment Station.

Harris, C.D. 1954.The market as a factor in the localization of industry in the United States, *Annals of the Association of American Geographers*, 44: 315–348.

H. Murat Celika, Jean-Michel Guldmann. 2007. Spatial interaction modeling of interregional commodity flows. *Socio-Economic Planning Sciences* 41:147–162.

Isard, W. 1956.*Location and Space Economy*, Gambridge, The M.I.T. Press.

Leontief, W. and A. Strout. 1963. Multi-regional input-output analysis, in T. Barna, (ed.). *Structural Interdependence and Economic Development*, St. Martin's Press.

Liu Qiang, Okamoto Nobuhiro.2002. Establishing Inter-regional Input-Output Model of China and Its problem. *Statistical Research*. No. 131:58-64. (in Chinese)

McFadden, D. 1974.Conditional logit analysis of qualitative choice behavior, *Frontiers in Econometrics*, Academic Press: 105–142.

Meng, B. and A. Ando. 2005. An economic derivation of trade coefficients under the framework of multi-regional IO analysis, *IDE Discussion Paper*, 29.

Okuda, T. 2003. An updating method of inter-regional input-output tables utilising entropy

optimization, *17th Applied Regional Science Conference*(ARSC), in Japan.(in Japanese)
Patrick Canning, Zhi Wang. 2004. A flexible modeling framework to estimate interregional trade patterns and input-output accounts[R]. *World Bank Policy Research Working Paper 3359*.
Sasaki, K., M. Shinmei, and S. Kunihisa. 1987. Multi-regional model with endogenous price system for evaluating road construction projects, *Environment and Planning* A, 19: 1093–1114.
Smith, T.E. 1975. A choice theory of spatial interaction, *Regional Science and Urban Economics*, 5, pp.137–176.
State Information Center. 2005. *Multi-regional Input-Output Model for China*. Social Sciences Academic Press (China)
Sterwart, J.Q.1948. Demographic gravitation: evidence and applications. *Sociometry*, 11:31-58.
Toshihiro Okubo.2004.The border effect in the Japanese market:A Gravity Model analysis. *Journal of Japanese and International Economies* 18: 1–11.
Xu Zhaoyuan, Li Shantong. 2008. Estimating the Inter-Regional Trade Based on 2002 IO Tables. Report on *China-Holland Input-Output Workshop.Beijing*. 2008.10
Zipf, G.K. 1947. The hypothesis of the Minimum Equation as a unifying social principle: with attempted synthesis, *American Sociological Review*, 12: 646–647.
Zhang Yaxiong, Zhao Kun.2005. *Interregional Input-Output Analysis*. Social Sciences Academic Press: 94-114. (in Chinese)