Estimating inter-regional trade flows in China: A sector-specific statistical model

LIU Weidong¹, LI Xin², LIU Hongguang³, TANG Zhipeng¹, GUAN Dabo^{2,4}

1. Institute of Geographic Sciences and Natural Resources Research, CAS, Beijing 100101, China;

2. Sustainability Research Institute, School of Earth and Environment, University of Leeds, LS2 9JT, UK;

3. College of Public Administration, Nanjing Agricultural University, Nanjing 210095, China;

4. ST Edmund's College, University of Cambridge, Cambridge, CB30BN, UK

Abstract: China has huge differences among its regions in terms of socio-economic development, industrial structure, natural resource endowments, and technological advancement. These differences have created complicated linkages between regions in China. In this study, building upon gravity model and location quotient techniques, we develop a sector-specific model to estimate inter-provincial trade flows, which is the base for making a multi-regional input-output table. In the model, we distinguish sectors with less intra-sector input from those with larger intra-sector input, and assume that the former sectors tend to compete among regions while the latter tend to cooperate among regions. Then we apply this new method of inter-regional trade estimation to three sectors: food and tobacco, metal smelting and processing, and electrical equipment. The results show that selection of bandwidth has a significant impact on the assessment of inter-regional trade. Trade flows are more scattered with the increase of bandwidths. As a result, bandwidth reflects the spatial concentration of geographical activities, which should be distinguishable for different industries. We conclude that the sector-specific spatial model can increase the credibility of estimates of inter-regional trade flows.

Keywords: multi-regional input-output analysis; trade flows; sector-specific statistical model; China

1 Introduction

Since the implementation of opening-up policy in the 1980s, China's regional development has followed two paths. The Eastern regions have developed rapidly due to preferential policies, which aim at improving economic efficiency and promoting economic growth in coastal regions at first and then move on to the development of interior regions (Fan, 1997). The uneven regional development policy has led to significant regional inequality in economic growth. For instance, in 2012, the wealthiest area, Tianjin, had a GDP per capita of 93,110 yuan (RMB), compared with 19,566 yuan in the poorest province, Guizhou (NBS,

Received: 2014-12-20 **Accepted:** 2015-01-20

Foundation: National Science Foundation for Distinguished Young Scholars of China, No.41125005

Author: Liu Weidong (1967–), PhD and Professor, specialized in economic geography and regional development.

E-mail: liuwd@igsrr.ac.cn

2012). The discrepancy between regions in China is also reflected in industrial structures and natural resource endowments. For example, the coastal regions have more advanced manufacturing industries, while the interior regions are endowed with coal, natural gas and metal ore, and are more engaged in primary industry and preliminary processing. As a result, spatial linkages are created in China between coastal-manufacturing regions and interior-resource-rich regions: natural resources from interior regions are delivered to the coastal regions and final products are transported from coastal to interior regions. Yet there are no statistical data on such linkages. Scholars studying these linkages have to rely on estimation of inter-regional trade by various methods.

This paper aims to develop a new method to estimate inter-regional trade flows by addressing the competitive and cooperative relations between industries as well as spatial dependence. It builds a sector-specific model by considering trade features of different sectors and to estimate China's inter-regional trade flows in food and tobacco, metal smelting and processing, and electrical equipment.

1.1 Regional input-output analysis

Input-output analysis has been used to evaluate the impacts of changes in final demand on the development of economic sectors at regional level. Miller and Blair(2009) pointed out that the significance of implementing a regional input-output study is determined by two fundamental features of a regional economy. First, the production structure of a specific region may not be identical to the national average. Second, contemporary economies are increasingly dependent on trade – both sales to and purchases from other economies.

Regional input-output models include single-regional models and multi-regional models. Single-regional input-output (SRIO) models deal with the impacts on economic sectors in a specific region (a country, a state or a city) of changes in final demand. SRIO models were initially used in economic accounting. Walter (1973), for example, used a regional input-output model to estimate the environmental discharge associated with American trade activity. In recent years, a number of studies have employed SRIO models to estimate environmental impacts at the regional level (Munksgaard *et al.*, 2000; Pan *et al.*, 2008). However, SRIO models assume that same production technology, energy structure and technology are used to produce goods and services in both import and export nations (Wiedmann, 2009; Lenzen *et al.*, 2004). As imported products come from a few countries, the disparities of production structure among countries may lead to disputed results if only domestic production structure is taken into consideration (Wiedmann, 2009). To cope with this problem, a number of studies have employed multi-regional models to estimate emissions or other environmental impacts embodied in trade between regions (Weber *et al.*, 2008; Shimoda *et al.*, 2008).

1.2 Multi-regional input-output models

Inter-regional input-output (IRIO) analysis, which was introduced by Walter Isard in the 1950s, has been applied to assess the interconnections between regions within a specific economy or among a few countries (Wiedmann, 2009). IRIO models attempt to capture the interconnections between all sectors in all regions. However, the large amount of data required in IRIO models make them difficult to implement (Miller and Blair, 2009). For ex-

ample, a two-region two-sector economy would require two trade flow matrices and eight inter-regional transaction elements; a three-region three-sector economy would need six trade flow matrices and fifty-four inter-regional transaction elements, and so on. Hence, a multi-regional input-output (MRIO) model, which is a simplified version of an inter-regional input-output model, is widely used in multi-regional IO studies. In this study, we focus on the multi-regional input-output model.

MRIO-models for particular countries were first introduced by Chenery (1953) and Moses (1955) in the 1950s, with the development of a two-region model for Italy and a nine-region model for the US, respectively. Polenske (1980) developed a MRIO framework, initiated at the Harvard Economic Research Project, with 51 regions and 79 economic sectors for the US. It is considered an early attempt at the large-scale implementation of a MRIO framework in the US. In recent years, Wiedmann *et al.* (2008) developed a global, uni-directional MRIO model for the UK. The model estimated the interactions between UK and three non-UK regions, i.e., OECD Europe, OECD non-Europe and non-OECD.

As concern with environmental issues has increased, MRIO models have been extended to assess greenhouse gas (GHG) emissions, ecological footprints and other environmental impacts in various locations (Norman *et al.*, 2007; Wiedmann *et al.*, 2007; Turner *et al.*, 2007). One of the most significant contributions of the MRIO approach is the estimation of environmental impacts embodied in trade. For instance, Minx *et al.* (2009) used the UK MRIO model to analyse changes in CO₂ emissions between 1992 and 2004. McGregor *et al.* (2008) employed a MRIO model to assess the CO₂ emissions embodied in trade between Scotland and the rest of the UK. Hayami and Nakamura (2007) used a MRIO model to estimate the greenhouse gas emissions embodied in trade activities between Japan and Canada. Furthermore, Lenzen *et al.* (2004) examined the CO₂ multipliers for multi-directional trade between Denmark, Germany, Norway, Sweden and rest of the world by using a MRIO model with feedback loop analysis. As argued by Wyckoff and Roop (1994), national CO₂ emission reduction policies may be insufficient if imported products account for a significant proportion of a country's consumption pattern. Furthermore, Peters and Hertwich (2006) pointed out that over 70% of exports from developing countries are used to satisfy the needs in developed countries. In other words, whether the producers (mainly developing countries) or the consumers (mainly developed countries) should be responsible for these emissions is questionable (Munksgaard and Pedersen, 2001). Shui and Harriss (2006) provided an empirical study in which they examine the CO₂ emission embodied in bilateral trade between China and the US using a MRIO model. This study involves the identification of CO₂ emissions avoided by importing Chinese goods in the US, the increase of CO₂ emission in China due to product export activity to the US and the contribution of CO₂ emissions embodied in US-China trade to global CO₂ emissions. The authors conclude that 7% to 14% of Chinese CO₂ emissions were attributed to product export to the US while the US reduced its CO₂ emissions by 3% to 6% as a result of the US-China trade. Consequently, the application of the MRIO framework to quantify the environmental impacts of trade is significant for planning national or global environmental policies.

A number of applications of MRIO modelling have focused specifically on China. For example, He and Duchin (2009) examined the impact of planned transportation infrastructure construction in western regions of China on changes in regional comparative advantage

and inter-regional trade. Guan and Hubacek (2007) developed an extended IO table by incorporating freshwater consumption and wastewater discharge into traditional IO table to evaluate the effects of the inter-regional trade structure on water consumption and pollution between north and south China. Liang *et al.* (2007) employed a MRIO model to assess the energy requirements and CO₂ emissions of eight economic regions in China. The regional classification and level of aggregation of economic sectors depend on the purpose of the study and the data in hand.

A few uncertainties have to be taken into consideration when implementing MRIO analysis. For example, sector aggregation may reduce the overall accuracy of the result (Lenzen *et al.*, 2004). Monetary exchange rates in different countries and omission of feedback effects in trade can also introduce errors into MRIO analysis (Weber, 2008).

1.3 Estimation of trade flows

Trade flows are fundamental to establishing the links among regions in MRIO analysis. In general, three approaches have been widely used in their estimation: survey-based methods; non-survey based methods; and hybrid methods. The survey-based approach provides more precise results. However, survey data is not always available, especially in the case of inter-regional trade flows within a country due to the significant workload and financial constraints (Idaha, 2005). Thus, simplified non-survey methods such as gravity models, entropy models, neural network models and behaviour-based models are used in the estimation of transactions between regions (Sargento *et al.*, 2012). The non-survey approach is more practical and makes it easier to operationalize the estimation of regional input-output data although its accuracy is always criticised. The hybrid method is popular as it combines the precision of survey-based approaches and the convenience of non-survey based approaches, benefitting from the advantages of these two approaches. The choice of estimation method is dependent on time, labour and funding availability. In this case, we adopt non-survey based approaches due to time and labour constraints.

One of the most widely used methods is the gravity model, which estimates migration, traffic, commodity shipping, and other flows (Bergstrand, 1985). Two main factors are considered in the gravity model, namely the size of the economy of two regions, of which flows are an increasing function, and the distance between two regions, of which flows are a decreasing function. A number of studies used traditional or augmented gravity models to estimate trade flows due to its simplicity and its good capacity to produce accurate results (Sargento, 2007). However, gravity models assume a universal relationship in trade flows among different industries, which is not always the case. For example, adjacent coal mining regions which both have a large amount of coal production are less likely to trade with one another. By contrast, two adjacent regions with large-scale telecommunication industries may tend to cooperate with each other due to supply chain trade. Thus, it is necessary to consider the characteristics of the relationships between different industries and whether these relations are competitive or cooperative. These considerations are not found in gravity models.

In addition, the estimation of the parameters in gravity models uses the ordinary least square (OLS) approach. It assumes that sample data do not have a relationship with the geographic location. Least-squares estimation therefore neglects spatial dependence. According

to Tobler's first law of geography, '*Everything is related to everything else but near things are more related than distant things*' (Tobler, 1970). Thus, both the variables and residuals¹ in the model might exhibit spatial dependence.

1.4 Industrial relations between regions

The Ricardian theory of comparative advantage argues that the foundations of international trade lie in discrepancies in production technologies, which in turn results in differences in cost, and that each country should concentrate on producing and exporting those products in which it has a comparative advantage, and should import products in which it has a comparative disadvantage. The Heckscher-Ohlin (H-O) model attributes comparative advantage to factor endowments and argues that countries that are relatively well-endowed with capital should export capital-intensive products and import labour-intensive products from countries relatively well-endowed with labour.

In the recent past, globalization has become one of the key drivers of increasing economic interconnection between regions. The comparative advantages of different regions are gradually evolving from inter-industry specialization into intra-industry specialization and further into intra-product specialization, promoting the interaction between regions within one industry. In addition, the development of industrial clusters as a product of globalization has increasingly become the main source of regional economic growth. The combination of industrial cluster formation and intra-industry specialization has resulted in increasing intra-industry trade flows. In other words, countries with similar factor endowments (either capital or labour) tend to import and export the same product at the same time. Leontief used the H-O theory to examine the import and export pattern of the United States, which was relatively endowed with capital. However, the results are opposite to the H-O theory, showing that the exported goods are more labour intensive than the imported goods. This result is known as the Leontief Paradox. The emergence of intra-industry trade has resulted in a growing discussion about the viability of the traditional trade theory. Thus, a number of studies have examined intra-industry trade in detail. For example, Krugman and Maurice (1997) incorporated transportation costs and multilevel demand parameters into a monopolistic competition model to establish the new international trade theory, explaining the phenomena of intra-industry trade.

Along with globalization and regional industrial specialization in regions, the same industry in different regions may not only compete but also cooperate with each other. However, different industries have distinctive competition-cooperation characteristics. For example, industries which have shorter supply chains may compete with other regions; examples might include agriculture, food manufacturing, textiles and clothing, etc. Other industries with longer supply chains, such as machinery and chemical products are more likely to cooperate. As a result, estimation of the interconnection of different regions needs to take into account the competition-cooperation characteristics of different industries. In order to reflect the competition-cooperation relations of industries between different regions, this research introduces the concept of the interconnection of the same industries in different regions and introduces a method for its estimation.

¹Residual is known as the difference between real value and estimated value of dependent variables.

1.5 Spatial dependence

The gravity model has been used widely in estimating inter-regional economic relations. As already mentioned, OLS estimates of the parameters of a gravity model rest on the assumption that the observations are not related to one another. However, in the spatial economy, observations are always related to geographical location. The closeness of geographical locations can result in spatial dependence among observations.

In order to incorporate the spatial character of observations into a regression model, Brunson *et al.* (1996) introduced the geographically weighted regression (GWR) model. GWR incorporates a geographically weighted matrix to represent the spatial character of each observation. In other words, in order to estimate the parameter at one specific location, the observations that are nearer to it are allocated a greater weight than the observations that are farther away. The basic principle of GWR is that the parameters can be estimated at any location if a dependent variable and one or more independent variables are known (Charlton and Fotheringham, 2009). The matrix of geographical weights is constructed using a spatial lag model or a spatial error model depending on whether spatial dependence is present in the variables in the model or the residual terms.

2 Methodology

2.1 Multi-regional input-output (MRIO) model

If we assume there are p regions and n sectors in a specific economy, the MRIO model is expressed mathematically as follows. The total output (T_i^s) of product i in region s is given by the equation

$$T_i^s = \sum_{j=1}^n a_{ij}^s x_j^s + f_i^s \quad (1)$$

where a_{ij}^s represents the requirement of input from sector i to produce the output in sector j in region s ; x_j^s is the total output of sector j ; f_i^s denotes final demand of sector i in region s . The total production of i in region r (x_i^r) is equivalent to the shipment of product i from region r to all the regions (x_i^{rs}) (including itself), and can be written as

$$x_i^r = \sum_{s=1}^p x_i^{rs} \quad (2)$$

The proportion of product i used in region s that comes from region r , c_i^{rs} is expressed as

$$c_i^{rs} = x_i^{rs} / T_i^s \quad (3)$$

By combining Equations (1)-(3), we can get

$$x_i^r = \sum_{s=1}^p c_i^{rs} \left(\sum_{j=1}^n a_{ij}^s x_j^s + f_i^s \right) \quad (4)$$

From which we can also further obtain

$$x^r = \sum_{s=1}^p \hat{c}^{rs} A^s x^s + \sum_{s=1}^p \hat{c}^{rs} f^s \quad (5)$$

Using matrix notation, let

$$x = \begin{bmatrix} x^1 \\ \vdots \\ x^s \\ \vdots \\ x^p \end{bmatrix}, \quad f = \begin{bmatrix} f^1 \\ \vdots \\ f^s \\ \vdots \\ f^p \end{bmatrix}, \quad A = \begin{bmatrix} A^1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A^p \end{bmatrix}, \quad C = \begin{bmatrix} \hat{c}^{11} & \cdots & \hat{c}^{1s} & \cdots & \hat{c}^{1p} \\ \vdots & & \vdots & & \vdots \\ \hat{c}^{r1} & \cdots & \hat{c}^{rs} & \cdots & \hat{c}^{rp} \\ \vdots & & \vdots & & \vdots \\ \hat{c}^{p1} & \cdots & \hat{c}^{ps} & \cdots & \hat{c}^{pp} \end{bmatrix},$$

the entire set of equations can be expressed as

$$x = CAx + Cf \quad (6)$$

2.2 A sector-specific spatial dependence model for estimating inter-regional trade flows

In constructing a multi-regional model within a country, a fundamental problem is to estimate inter-regional trade. In this paper, we build upon gravity model and location quotients techniques to develop a sector-specific spatial dependence model to estimate inter-provincial trade flows. The following subsection describes the development of a sector-specific spatial dependence model.

We start with the traditional gravity model, which is represented in Equation (7)

$$y_i^{gh} = e^{\beta_0} \frac{(x_i^{go})^{\beta_1} (x_i^{oh})^{\beta_2}}{(d^{gh})^{\beta_3}} \quad (7)$$

where y_i^{gh} represents trade flows of sector i from province g to province h ; e is a constant; x_i^{go} is total outflows of sector i from g ; x_i^{oh} is the total inflows of sector i to province h ; d^{gh} is the distance between province g to h (in this case, we use distance between the capital cities of two provinces g and h); β_0 , β_1 , β_2 and β_3 are parameters of the above variables, respectively.

If we express Equation (7) in logarithmic form, we obtain regression Equation (8)

$$\ln(y_i^{gh}) = \beta_0 + \beta_1 \ln(x_i^{go}) + \beta_2 \ln(x_i^{oh}) - \beta_3 \ln(d^{gh}) + \varepsilon \quad (8)$$

Assuming we have n regions, Equation (8) can be expressed in matrix form as Equation (9)

$$Y = \beta_0 L_n + \beta_1 X_1 + \beta_2 X_2 - \beta_3 X_3 + \varepsilon \quad (9)$$

Let $N = n \times n$ where Y is an $N \times 1$ matrix representing the trade flows between regions; L_n is an $N \times 1$ matrix with all elements equal to 1; X_1 and X_2 represent the total supply from supply regions and total demand from demand regions, respectively; X_3 is the distance between two regions. Given sufficient observations, Equation (9) can be solved by using multiple-regression analysis. However, it assumes that each observation is independent of the other observations. In reality, the contiguity of regions might have impacts on trade flows. If spatial dependence is present, the estimation of parameters may be biased, overestimating or underestimating the unknown true value. To take into account spatial dependence, we intro-

duce a spatial lag model.

The traditional regression model assumes that the relationship being modelled is the same for the entire data sample. Thus, the relationship fitted is an average level of the overall trends in the study area. This average may obscure many geographical, social and economic characteristics, even though it is known that there are often variations in relationships over space (LeSage, 1999). A number of approaches are used to estimate relationships that vary systematically over space such as Casetti's (1972) spatial expansion method which considers the parameters as a function of the locations of the observations, and Foster and Gorr's (1986) spatial adaptive filtering method.

GWR extends the basic regression model by allowing different relationships to exist at different locations rather than assuming the universal coefficient estimates. GWR has been applied in economic geography (Yang and Cai, 2010, Wang *et al.*, 2000). However, most of the studies use a single point or an area data source. The data used to estimate regional trade flows are different as they are flows between an origin and a destination. The spatial influencing factor of origin-destination flow data is relatively complicated. LeSage and Pace (2008) used GWR to estimate spatial OD flows, incorporating spatial lags for origins, destinations and flows. Spatial lags are similar to lag operators in time series analysis, which are used to produce a weighted average of the neighbouring observations (LeSage, 1999). LeSage and Pace (2008) argued that there are two main reasons to incorporate spatial lags. First, they permit consideration of spatial dependence as a long-run equilibrium of an underlying spatio-temporal process. Second, the omitted variables might be spatially correlated. Thus, we incorporate spatial lags in this study.

In principle, the spatial lag model incorporates the degree of spatial autocorrelation of the dependent variable (which is defined by ρ) to exhibit the impacts of location on regression parameters. It can be written as

$$Y = \rho WY + X\beta + \varepsilon \quad (10)$$

where Y represents the vector of dependent variables; W is an $n \times n$ non-negative spatial weight matrix; X is the matrix of independent variables; β represents the vector of parameters which shows the influence of independent variables on dependent variables; ε is a vector of disturbances.

However, this model focuses on each observation rather than the flows among observations. Trade flows from the origin to the destination might however impact on the trade flows from the neighbours of the origin to the neighbours of destination. LeSage and Pace (2008) estimated the origin-destination flows by taking into account this more complex set of spatial dependencies among observations, which was neglected in the traditional gravity model. Three spatial connectivity matrices were used to represent origin, destination, and origin-to-destination dependence, respectively.

In order to take into account spatial dependence, we therefore introduce spatial lags into Equation (9)

$$Y = \rho_0 W_0 Y + \rho_d W_d Y + \rho_w W_w Y + \beta_0 L_n + \beta_1 X_1 + \beta_2 X_2 - \beta_3 X_3 + \varepsilon \quad (11)$$

where W is an $n \times n$ diagonalized matrix representing spatial weights between provinces with zeroes on the principal diagonal; $W_0 = W \otimes I_n$, $\otimes$ is the Kronecker product, I_n is the identity matrix; $W_0 Y$ is an $N \times 1$ matrix which measures spatial dependence on the origin of the trade

flow; $W_d = I_n \otimes W$, $W_d Y$ is an $N \times 1$ matrix measuring spatial dependence on the destination of the trade flow; $W_w = W \otimes W$, $W_w Y$ is an $N \times 1$ matrix which measures the interdependence between the origin and the destination of trade flow. Spatial dependence on the origin can be considered as the exports from an origin to a destination, which might influence the exports from the neighbours of the origin to the neighbours of the destination. Spatial dependence on the destination refers to the imports to a destination from an origin which might influence the imports to the neighbours of the destination from the neighbours of the origin. Origin-destination dependence refers to the flows from the neighbours of the origins to the neighbours of the destinations.

The core of GWR is the choice of a spatial weighting function. A Gaussian function is the most frequently used method, which represents the relationship between distance and spatial weights by using a continuous and monotonically decreasing function as follows,

$$w_{ij} = \exp \left(- \left(\frac{1}{2} \right) \left(\frac{d_{ij}}{b} \right)^2 \right) \quad (12)$$

where d_{ij} is the distance between location i and location j , and b denotes the bandwidth which is a nonnegative monotonically decreasing function. For a given bandwidth, when $d_{ij}=0$, $w_{ij}=1$; as d_{ij} approaches infinity, w_{ij} approaches zero, which means that the observation has little impact on the estimated parameters. A number of approaches are used to select the bandwidth, such as the Akaike information criterion (AIC) and Bayesian information criterion (BIC). In this research, the selection of bandwidth is subject to the character of the industry. For example, for food manufacturing, trade usually happens with nearer regions, and so we select a smaller bandwidth; for metal smelting and electrical equipment, trade usually happens among regions that are further apart, and so we select a bigger bandwidth. After the selection of bandwidths, we estimate the spatial weights matrix. As a region does not have a distance to itself, $w_{ij}=0$. We unify each row of W to ensure that the sum of weights around one specific observation equals to 1.

In order to exhibit the competitive and cooperative relationships of different industries in different regions, we measure the degree of interaction of a particular production sector, as shown in Equation (13)

$$c_i^{gh} = \begin{cases} \frac{\mu_i^g + \mu_i^h}{|\mu_i^g - \mu_i^h| + \min_{r=1,2,\dots,n} \mu_i^r} & \text{if } g \neq h \\ 1 & \text{if } g = h \end{cases} \quad (13)$$

where c_i^{gh} represents the ‘degree of interaction’ of sector i ; μ_i^g and μ_i^h are the location quotients of region g and h for sector i ; n is the number of regions. As can be seen from Equation (13), if $g \neq h$, $c_i^{gh} > 1$, the bigger the value is, the higher the degree of interaction is. If $g = h$, $c_i^{gh} = 1$, it means there is no interaction. For a specific industry in both region g and region h , a high location quotient in both regions implies a greater degree of interaction between these two regions for that industry, whereas small location quotients in both regions imply few interactions between these two regions for that industry.

In order to calibrate the intra-sector relationship we introduce the ‘intra-sector trade condition’ to capture the impact of sector i in region g on sector i in region h . This condition

assumes that if a large proportion of an industry's total output is used as intermediate inputs to itself, then the industry shows an intra-industry cooperative relationship. To compute the intra-sector trade condition we use the ratio of intra-sector input as shown in Equation (14)

$$\theta_i = \bar{\delta} - \delta_i \quad (14)$$

where θ_i is a ratio of intra-sector input of sector i . δ_i measures the amount of intermediate inputs in sector i out of total input of sector i . $\bar{\delta}$ is the average ratio of intermediate inputs for all production sectors. If θ_i is positive, it reveals that there is a competitive relationship in sector i across regions. By contrast, if θ_i is negative, it indicates that a cooperative relationship in sector i across regions. For example, in the 2007 input-output table, the average ratio of intermediate input among all production sectors ($\bar{\delta}$) is 0.1628; whilst the amount of intermediate inputs in agriculture out of total input of agriculture (δ) is 0.1407, giving an intra-sector influence coefficient (θ) for agriculture of 0.0222. It suggests that agriculture reveals a competitive relationship.

The estimation of trade flows between regions is adjusted by the degree of interaction and the intra-sector trade condition as given in Equation (15)

$$Y' = \frac{\hat{Y}}{(c_i^{gh})^{\theta_i}} \quad (15)$$

where Y' is trade flow adjusted by the degree of interaction and the intra-sector trade condition; $\hat{Y}$ is the trade flow estimated by the GWR; and $(c_i^{gh})^{\theta_i}$ represents the sector-specific coefficient.

Equation (15) can be used to calculate the inter-regional trade matrix, in which elements on diagonal are set to zero (a region does not have trade flows to itself). Then, we incorporate the total inflows and total outflows from all provincial input-output tables to balance the matrix by using the RAS method.

3 Applications

In order to explain the model in details, the trade flow matrix of three economic sectors, including food and tobacco, metal smelting and processing, and electrical equipment for 30 provinces in China is examined. We also compare the characteristics of different sectors in this section.

3.1 Data source

Data on total inflows and outflows (domestic) as well as total import and export (from other countries) in each sector for each province are from the 2007 provincial input-output tables (IO tables, hereafter). The 30 provincial IO tables are compiled and published by National Bureau of Statistics (2011). The official IO tables are available at the format of 42×42 sectors. The final demand category in Chinese IO tables consists of rural and urban households' consumption, government expenditure, capital formation and exports. To reduce the computing workload, we aggregate the provincial tables into 30×30 production sectors. We selected from provincial IO tables those sectors that have significant inflows and outflows (there are four columns of aggregated data on inflows from and outflows to other provinces

as well as imports from and exports to other countries in provincial IO tables), and combine other sectors having limited inflows and outflows into one sector. All the sectors in the primary and secondary industries are kept while some sectors in the tertiary industry are combined.

Distances between provinces are estimated by using the average value of railway distances and highway distances between capital cities. The location quotient of a sector in a province is computed as the ratio of the proportion of the sector's output in the provincial total output to that of the sector in national total output. Data on national and provincial total outputs are from the 2007 national and provincial IO tables, respectively. Sample data on inter-provincial trade in food and tobacco, metal smelting and processing and electrical equipment sectors are from statistics on bulk commodities by the Ministry of Transport and the Ministry of Railways of China. Grain data, steel and non-ferrous metal data are used to estimate trade flows in the food and tobacco sector, metal smelting and processing sector, and electrical equipment sector, respectively. To match the IO tables in value terms, sample trade flows are converted from physical units (e.g. tons) into monetary units (e.g. RMB yuan), based on the price of each product. For example, the grain price is calculated using the weighted average price of four agricultural products, including rice, wheat, corn and soybeans. Figure 1 shows the inter-provincial trade flows of metal smelting and processing sector. According to the characteristics of the sample data, the bandwidth of the weighted geographical function (b) is set to 500 in this study.

Figure 1 Sketch map of metal sample trade flows (>4 billion yuan)

If we take the logarithm of Equation (7) in the model, sample number cannot be zero or negative. Thus, we replace zeros with a very small value; for instance, if each element of the sample is larger than 1000, we replace those 0s with 1s. Then, in the final result, for values less than 1 we replace them with 0s. For distances between regions, we assume that each region has a small distance to itself, e.g. 10 km. For the intra-sector trade condition, we assume that a sector does not compete with itself in the same region. Thus regional intra-sector trade coefficient is 1. Values on the diagonal of the initial trade flow matrix are set to zero.

3.2 Results

Since the spatial weighted matrix W is known, by introducing X_1 , X_2 and X_3 into the above equations and assuming the expected error $\varepsilon = 0$, we can calculate Y . Then we get the initial values of inter-provincial trade flows of the three sectors. X_3 represents distances and its value is negative, indicating that inter-provincial trade flows are inversely proportional to distance. Table 1 shows the results generated from Equation (11).

Table 1 Geographically weighted regression equation of inter-regional product flows for selected sectors

Sector	GWR equations	R^2	F test
Food and tobacco	$Y = 0.725W_0Y + 0.907W_dY - 0.662W_wY - 0.699L_n + 0.019X_1 + 0.081X_2 - 0.076X_3$	0.719	381.5
Metal processing	$Y = 0.730W_0Y + 0.839W_dY - 0.652W_wY - 0.975L_n + 0.093X_1 + 0.113X_2 - 0.167X_3$	0.731	404.3
Electrical equipment	$Y = 0.909W_0Y + 0.943W_dY - 0.896W_wY + 2.220L_n - 0.005X_1 + 0.017X_2 - 0.252X_3$	0.713	368.9

Using the 2007 national and provincial IO tables, we estimate the intra-sector trade conditions of food and tobacco, metal smelting and processing, hh and electrical equipment sectors between provinces, according to Equations (13) and (14). The computed coefficients are 0.0197, -0.1738 and -0.02758 for food and tobacco, metal smelting and electrical equipment sector, respectively. The result shows that there is more competition in food and tobacco between provinces and more intra-sector co-operation between provinces in such sectors as metal smelting and electrical equipment. By introducing an initial value of Y , intra-sector trade coefficients and processing into Equation (15), we can obtain an initial trade flow matrix at sector level between provinces.

Then, we replace the diagonals of the initial trade flow matrix with zero and use the RAS method to balance the matrix, and we obtain the final trade flow matrix of the food and tobacco, metal smelting and processing, and electrical equipment sectors between provinces. Figure 2 shows the computed inter-provincial trade flows of the metal smelting and processing sector.

Figure 2 Estimation of trade flows for metal smelting and processing sector between regions (>4 billion yuan)

3.3 Discussion

3.3.1 Reliability of geographically weighted regression (GWR) model

Table 2 compares the results generated from the GWR model and the normal regression model. It shows that the application of GWR model has improved the value of R^2 and the significance of the F statistic substantially. Comparing the value of the t-test, the significance level of the spatial variable is higher in the GWR model, which means that spatial interaction has a significant impact on inter-regional trade flows. Thus, the application of GWR model can improve the credibility of estimations of inter-regional trade flows significantly.

Table 2 Comparison between the estimations of geographically weighted regression model and basic gravity model

Sectors	Food and tobacco				Metal smelting				Electrical equipment			
	GWR		Basic GM		GWR		Basic GM		GWR		Basic GM	
R^2	0.719		0.065		0.731		0.406		0.713		0.232	
F test	381.48		20.83		404.3		203.87		368.9		90.04	
	t-test	P-value	t-test	P-value	t-test	P-value	t-test	P-value	t-test	P-value	t-test	P-value
Constant	-0.659	0.145	1.370	0.065	-0.929	0.515	-3.10	0.034	1.58	0.086	6.92	0.000
ρ_0	17.726	0.000	—	—	17.972	0.000	—	—	20.95	0.000	—	—
ρ_d	37.833	0.000	—	—	24.97	0.000	—	—	31.57	0.000	—	—
ρ_w	12.383	0.000	—	—	11.267	0.000	—	—	-8.72	0.000	—	—
β_1	0.607	0.408	3.440	0.000	1.978	0.019	14.673	0.000	0.297	0.561	-1.74	0.001
β_2	1.621	0.056	3.927	0.000	2.959	0.002	11.804	0.000	0.31	0.400	8.03	0.000
β_3	-1.756	0.068	4.795	0.000	-3.245	0.037	12.899	0.000	4.876	0.002	13.46	0.000

3.3.2 Sectoral features of inter-regional trade flows

First, major trade destinations of the food and tobacco sector are Shanghai, Jiangsu, Shandong, Jilin, Guangdong, Shaanxi and Beijing, which are either more advanced or more populated than the other regions. The origins of trade are from Henan, Shandong, Guangdong, Hebei, Jilin and Inner Mongolia, which are more advanced in food manufacturing than the other areas. Regions such as Shandong and Guangdong are both major exporters and importers of food products, which implies the existence of diverse needs of consumers for food. Intensive trade flows take place between Hebei, Jilin, Shandong, Zhejiang and Shanghai; between Hunan, Hubei and Guangdong; and between Shandong and Jilin. It shows that there are major trade flows in the food and tobacco sector from Central China to east coastal regions, which are closely related in terms of geographical location.

Second, major destinations of the metal smelting and processing sector are Zhejiang, Guangdong, Jiangsu, Shanghai, Liaoning and Hebei, which are located in East and South China. Total imports amount to 600 billion yuan. These regions have more advanced technology in metal processing and manufacturing industry. Major origins of trade flows are Central and Western regions, including Henan, Inner Mongolia, Yunnan, Hunan, Gansu, Guangxi and Eastern regions, such as Hebei, Liaoning, Guangdong, Jiangsu, which have competitive advantages in the metal smelting industry. Two major routes of trade flows are

1) from Central and Western China to Yangtze River Delta (YRD) and Pearl River Delta (PRD) and 2) from North China, such as Hebei and Tianjin to YRD and Northeast China. In general, trade flows of metal processing and smelting products are from Central and Western regions to the more developed coastal regions with characteristics of geographical proximity. For example, Yunnan, Guangxi and Guizhou export products to the PRD regions; YRD regions import products mainly from Hebei, Shandong, Henan and Anhui.

Third, destinations of electrical equipment are Guangdong, Shanghai, Beijing, Zhejiang, Fujian, Henan and Jiangsu, which are either well populated or more developed than the other regions. Regions, such as Guangdong, Jiangsu, Zhejiang, Anhui, Shanghai, Hebei, Henan and Liaoning, are major exporters of electrical equipment. The products flows from Guangdong, Zhejiang and Jiangsu to Eastern regions such as Shanghai, Beijing and Fujian and Central and Western regions, such as Sichuan, Guizhou, Yunnan and Shaanxi. They also flow within the Eastern regions themselves, such as between Zhejiang and Guangdong. It implies the existence of a cooperative relationship between regions for the electrical equipment sector. Table 3 shows the estimations of major trade flows for food and tobacco, metal smelting and processing and electrical equipment sectors between regions in details.

Table 3 Top 15 origin-destination pairs of trade flows in selected sectors (billion yuan)

Rank	Food and tobacco		Metal smelting		Electrical equipment	
1	Shandong-Jilin	20.6	Hebei-Zhejiang	85.1	Zhejiang-Guangdong	24.2
2	Guangdong-Shaanxi	17.6	Hebei-Jiangsu	57.6	Guangdong-Fujian	15.8
3	Hebei-Shanghai	12.4	Henan-Zhejiang	53.7	Guangdong-Shanghai	14.5
4	Jilin-Shanghai	12.1	Yunnan-Guangdong	45.7	Jiangsu-Shanghai	13.5
5	Zhejiang-Shanghai	12.0	Guangxi-Guangdong	35.4	Guangdong-Yunnan	12.6
6	Hubei-Guangdong	11.3	Guangdong-Zhejiang	33.2	Zhejiang-Shanghai	12.1
7	Guangdong-Jiangsu	11.3	Henan-Jiangsu	27.9	Guangdong-Guangxi	11.6
8	Hunan-Guangdong	10.5	Liaoning-Jilin	27.4	Jiangsu-Beijing	11.4
9	Jilin-Shandong	10.1	Hunan-Guangdong	26.1	Guangdong-Zhejiang	11.3
10	Guangdong-Beijing	10.1	Jiangsu-Zhejiang	25.4	Guangdong-Sichuan	9.4
11	Hebei-Jilin	9.9	Hebei-Shandong	24.0	Guangdong-Guizhou	8.7
12	Heilongjiang-Liaoning	9.8	Hebei-Liaoning	23.7	Anhui-Shanghai	8.0
13	Jilin-Jiangsu	9.5	Liaoning-Heilongjiang	19.9	Jiangsu-Shaanxi	7.9
14	Shandong-Shanghai	8.9	Hebei-Tianjin	19.7	Shanghai-Heilongjiang	7.7
15	Yunnan-Chongqing	8.6	Gansu-Zhejiang	18.5	Jiangsu-Inner Mongolia	7.4

3.3.3 Sensitivity analysis of weighted function and bandwidth

As mentioned above, in the weighted spatial regression, the choice of bandwidths has a significant impact on regression results. Thus, we compare the regression results of food trade flows using two bandwidths, namely $b = 1000$ and $b = 500$. Results show that the direction of trade flows in the sector does not change much, but the value of trade declines significantly. For example, trade flows of food products from Shandong to Jilin decline from 20.6 billion yuan to 15.8 billion yuan with the increase of bandwidth from 500 to 1000. For trade flows from Guangdong to Shanxi, the value declines from 17.6 billion yuan to 13.9 billion yuan (see Table 4). Values of trade flows also increase accordingly in other regions, indicat-

ing trade flows are more concentrated as bandwidths decrease. Thus, bandwidths reflect the level of geographical concentration of economic activities. The higher the concentration, the smaller the bandwidth should be; the lower the concentration, the bigger the bandwidth should be. This is consistent with the previously mentioned principle of bandwidth selection.

Table 4 Top 15 origin-destination pairs of trade flows in the food and tobacco sector with two bandwidths (billion yuan)

Rank		$b=500$		$b=1000$
1	Shandong-Jilin	20.6	Shandong-Jilin	15.8
2	Guangdong-Shaanxi	17.6	Guangdong-Shaanxi	13.9
3	Hebei-Shanghai	12.4	Zhejiang-Shanghai	10.4
4	Jilin-Shanghai	12.1	Henan-Shandong	10.3
5	Zhejiang-Shanghai	12.0	Yunnan-Guangdong	9.6
6	Hubei-Guangdong	11.3	Hebei-Shanghai	9.5
7	Guangdong-Jiangsu	11.3	Heilongjiang-Jilin	9.1
8	Hunan-Guangdong	10.5	Jilin-Shanghai	9.0
9	Jilin-Shandong	10.1	Hebei-Shandong	8.2
10	Guangdong-Beijing	10.1	Shandong-Shanghai	8.2
11	Hebei-Jilin	9.9	Inner Mongolia-Jilin	8.1
12	Heilongjiang-Liaoning	9.8	Jilin-Shandong	8.0
13	Jilin-Jiangsu	9.5	Henan-Shanghai	7.9
14	Shandong-Shanghai	8.9	Henan-Shaanxi	7.6
15	Yunnan-Chongqing	8.6	Guangxi-Guangdong	7.4

4 Conclusions

In this study, we built upon gravity model and location quotients techniques to develop a sector-specific spatial dependence input-output model, which is used to estimate inter-provincial trade flows. The results show that sector-specific spatial dependence model can increase the credibility of inter-regional trade flows estimation. Apart from analysing physical flows of product, spatial dependence models also could be used to analyse population, capital and information flows between regions.

The research also shows that trade flows of food products and metal smelting products tend to flow from the Central and Western regions to the Eastern regions. For electrical equipment, trade flows exist within the Eastern regions and also from the Eastern regions to the Central and Western regions. By comparing the results of our analysis, we find that some industries in different regions tend to cooperate as in the case of textile machinery equipment manufacturing; whilst other industries, such as coal mining and services, tend to compete between regions.

Last but not least, selection of bandwidth has a significant impact on the assessment of inter-regional trade. Trade flows are more scattered as bandwidths increase. As a result, bandwidth reflects the spatial concentration of geographical activities, which should be distinguishable for different industries. Future research could therefore focus on approaches to select appropriate bandwidths for different industries.

References

- Bergstrand J, 1985. The gravity equation in international trade: Some microeconomic foundations and empirical evidence. *The Review of Economics and Statistics*, 67: 474–481.
- Brunsdon C, Fotheringham A S, Charlton M E, 1996. Geographically Weighted Regression: A method for exploring spatial nonstationarity. *Geographical Analysis*, 28: 281–298.
- Casetti E, 1972. Generating models by the expansion method: Applications to geographical research. *Geographical Analysis*, 4: 81–91.
- Charlton M E, Fotheringham A S, 2009. Geographically Weighted Regression: White Paper. National Centre for Geocomputation, National University of Ireland Maynooth.
- Chenery H B, 1953. Regional analysis. In: Chenery H B, Clark P G, Pinna V C (eds.). *The Structure and Growth of the Italian Economy*. Rome: U.S. Mutual Security Agency.
- Fan C C, 1997. Uneven development and beyond: Regional development theory in post-Mao China. *International Journal of Urban and Regional Research*, 21: 620–639.
- Foster S A, Gorr W L, 1986. An adaptive filter for estimating spatially-varying parameters: Application to modeling police hours spent in response to calls for service. *Management Science*, 32: 878–889.
- Guan D, Hubacek K, 2007. Assessment of regional trade and virtual water flows in China. *Ecological Economics*, 61: 159–170.
- Hayami H, Nakamura M, 2007. Greenhouse gas emissions in Canada and Japan: Sector-specific estimates and managerial and economic implications. *Journal of Environmental Management*, 85: 371–392.
- He L, Duchin F, 2009. Regional development in China: interregional transportation infrastructure and regional comparative advantage. *Economic Systems Research*, 21: 3–22.
- Idaha T, 2005. How to utilize interregional input-output analysis in China. In: Okamoto N, Idaha T (eds.). *Spatial Structure and Regional Development in China*. New York: Palgrave Macmillan.
- Krugman P, Maurice O, 1997. *International Economics Theory and Policy*. 4th ed. Reading: Mass Addison-Wesley.
- Lenzen M, Pade L L, Munksgaard J, 2004. CO₂ multipliers in multi-region input output models. *Economic Systems Research*, 16: 391–412.
- LeSage J P, 1999. The theory and practice of spatial econometrics. Department of Economics, University of Toledo.
- LeSage J P, Pace R K, 2008. Spatial econometric modeling of origin-destination flows. *Journal of Regional Science*, 48: 941–967.
- Liang Q M, Fan Y, Wei Y M, 2007. Multi-regional input-output model for regional energy requirements and CO₂ emissions in China. *Energy Policy*, 35: 1685–1700.
- McGregor P G, Swales J K, Turner K, 2008. The CO₂ 'trade balance' between Scotland and the rest of the UK: Performing a multi-region environmental input-output analysis with limited data. *Ecological Economics*, 66: 662–673.
- Miller R E, Blair P D, 2009. *Input-Output Analysis: Foundations and Extensions*. New York, Cambridge University Press.
- Minx J, Baiocchi G, Wiedmann T *et al.*, 2009. Understanding changes in UK CO₂ emissions 1992–2004: A structural decomposition analysis. UK: Stockholm Environment Institute.
- Moses L N, 1955. The stability of interregional trading patterns and input-output analysis. *The American Economic Review*, 45: 803–822.
- Munksgaard J, Pedersen K A, 2001. CO₂ accounts for open economies: Producer or consumer responsibility? *Energy Policy*, 29: 327–334.
- Munksgaard J, Pedersen K A, Wien M, 2000. Impact of household consumption on CO₂ emissions. *Energy Economics*, 22: 423–440.
- National Bureau of Statistics (NBS), 2011. *Chinese Regional Input-Output Tables*. Beijing: China Statistics Press.
- National Bureau of Statistics (NBS), 2012. *China Statistical Yearbook*. Beijing: China Statistics Press.

- Norman J, Charpentier A D, MacLean H L, 2007. Economic input-output life-cycle assessment of trade between Canada and the United States. *Environmental Science & Technology*, 41: 1523–1532.
- Pan J, Phillips J, Chen Y, 2008. China's balance of emissions embodied in trade: Approaches to measurement and allocating international responsibility. *Oxford Review of Economic Policy*, 24: 354–376.
- Peters G P, Hertwich E G, 2006. Pollution embodied in trade: The Norwegian case. *Global Environmental Change*, 16: 379–387.
- Polenske K R, 1980. The U.S. Multiregional Input-output Accounts and Model. Lexington, MA, Lexington Books.
- Sargento A L M, 2007. Empirical examination of the gravity model in two different contexts: Estimation and explanation. *Jahrbuch für Regionalwissenschaft*, 27: 103–127.
- Sargento A L M, Ramos P N, Hewings G J D, 2012. Inter-regional trade flow estimation through non-survey models: An empirical assessment. *Economic Systems Research*, 24: 173–193.
- Shimoda T, Watanabe T, Ye Z *et al.*, 2008. An empirical study on interdependency of environmental load and international IO structure in the Asia-Pacific region. In: International Input-Output Meeting on Managing the Environment. Seville, Spain.
- Shui B, Harriss R C, 2006. The role of CO₂ embodiment in US-China trade. *Energy Policy*, 34: 4063–4068.
- Tobler W R. 1970. A computer movie simulating urban growth in the detroit region. *Economic Geography*, 46: 234–240.
- Turner K, Lenzen M, Wiedmann T *et al.*, 2007. Examining the global environmental impact of regional consumption activities (Part 1): A technical note on combining input-output and ecological footprint analysis. *Ecological Economics*, 62: 37–44.
- Walter I, 1973. The pollution content of American trade. *Western Economic Journal*, XI: 61–70.
- Wang J, Li L, Ge Y, 2000. A theoretic framework for spatial analysis. *Acta Geographica Sinica*, 55: 92–103.
- Weber C L, 2008. Uncertainties in constructing environmental multiregional input-output models. In: International Input-Output Meeting on Managing the Environment. Seville, Spain.
- Weber C L, Peters G P, Guan D *et al.*, 2008. The contribution of Chinese exports to climate change. *Energy Policy*, 36: 3572–3577.
- Wiedmann T, 2009. A review of recent multi-region input-output models used for consumption-based emission and resource accounting. *Ecological Economics*, 69: 211–222.
- Wiedmann T, Lenzen M, Turner K *et al.*, 2007. Examining the global environmental impact of regional consumption activities (Part 2): Review of input-output models for the assessment of environmental impacts embodied in trade. *Ecological Economics*, 61: 15–26.
- Wiedmann T, Wood R, Minx J *et al.*, 2008. Emissions embedded in UK trade - UK-MRIO Model results and error estimate. In: International Input-Output Meeting on Managing the Environment. Seville, Spain.
- Wyckoff A W, Roop J M, 1994. The embodiment of carbon in imports of manufactured products: Implications for international agreements on greenhouse gas emissions. *Energy Policy*, 22: 187–194.
- Yang Z, Cai J, 2010. Progress of spatial statistics and its application in economic geography. *Progress in Geography*, 29: 757–768.