Analysis of Interregional Commodity Flows

Wirach Hirun and Sompong Sirisoponsilp
Department of Civil Engineering, Chulalongkorn University Phayathai,
Pathumwan, Bangkok, 10330, Thailand

Abstract: Problem statement: Commodity Flow Survey (CFS) was launched to collect comprehensive freight flow data throughout the kingdom of Thailand. The survey's database is the most complete collection of commodity flow data in Thailand. The need to reveal interregional freight characteristics using available data from the CFS led to the objectives of this research. **Approach:** An origin destination matrix based on province was calibrated using a flexible Box-Cox function form. It used maximum likelihood and the backward method for calibration and Root Mean Square Error (RMSE) and Mean Relative Error (MRE) to verify the model's performance. Independent variables were classified into three groups: origin variable, destination variable and geographic variable. The origin variable represented the behavior of the trip as generated at the place of origin. Some consumption occurred at the origin. The employment and the average plant size variables were selected for potential productivity while personal income per capita and total populations were included to explain consumption behavior at the origin. Personal income per capita and total populations were selected for destination variables which act as proxy for final demand at the destination. The third category, distance, was the most conventional friction variable for geographical variables. **Results:** The calibrated model revealed that origin income, origin average plant size and origin population performed poorly. Therefore these variables were eliminated. The best developed model included four strongly significant variables at a 5% level: origin employment, destination population, destination income per capita and distance. **Conclusion:** The results showed that the selected variables and the Box-Cox functional form were successful in explaining behavior of interregional freight transportation in Thailand. The developed model was the first interregional freight transportation model to be calibrated against Thailand commodity flow survey data.

Key words: Commodity flow survey, Box-Cox transformation, freight distribution, freight data survey, freight modeling

INTRODUCTION

The paucity of available data has limited freight flow studies in Thailand. Past freight distribution models have been calibrated against a small sample size or by using an indirect method such as the input-output model. Wongoppadoldecha and Takano (2010) used the input-output model to develop a model for estimating interregional freight flow in Thailand. They verified their developed model with interregional freight flow data published by the Thailand Ministry of Transportation (MOT). Although the developed model showed impressive performance, the interregional freight flow which was used for verification was not clear.

In developed countries such as US and Japan, freight distribution models are calibrated using Commodity Flow Survey (CFS) because its performance is superior to other survey methods. Data collected by CFS includes commodity type, origin and destination, weight, product value, mode of transport and more from a large sample throughout the country. Since knowing such interregional freight flow characteristics would likewise be useful for transport and logistics planning in Thailand, the Office of National Statistics of Thailand launched the Commodity Flow Survey (CFS) in 2007 to collect comprehensive freight flow data from a large sample of shippers throughout the kingdom of Thailand. The survey's database is the most complete collection of freight flow data in Thailand.

To reveal interregional freight characteristics, the most important data is the freight origin destination matrix. Well known calibration methods are the gravity model and the regression based model. Of the two

Corresponding Author: Wirach Hirun, Department of Civil Engineering, Chulalongkorn University Phayathai, Pathumwan, Bangkok, 10330, Thailand

methods, the regression based model is superior because it can unveil the causative relationship of the flow with the set of policy variables. This research attempts to calibrate the origin destination matrix using available data from the CFS and a regression based model using a flexible Box-Cox function form.

Literature review: In the context of freight distribution study, a well known model, the gravity model, has always been employed to produce a freight distribution matrix. The fundamental concept of the gravity model is that the attraction between two objects is proportional to their mass and inversely proportional to their respective distance. Consequently, the general formulation of spatial interactions can be adapted to reflect this basic assumption to form the elementary formulation of the gravity model using impedance variables (e.g., distance, travel time, generalized cost) to describe the way in which freight demand decreases (Ogden, 1978; Toosi *et al.*, 2009; Ardakani *et al.*, 2009). The gravity model calibration process attempts to explore the impedance variable function as closely as possible using the observed trip length distribution (Ortuzar and Willumsen, 1994). The weaknesses of this method are its inability to explain the relationship of explanatory factors and socio economic variables of the study area and that it requires a large amount of data for calibration. Moreover, previous works which employ the gravity model focus on best fit with little theoretical foundation (Celik, 2004).

Another way to depict behavior of freight distribution, the so-called regression based model, was proposed by Ashtakala and Murthy (1988). They applied the Box-Cox transformation to the distance between a pair of communities in the gravity model. They used six commodity groups' data from a 1977 survey of commodity flows between communities within Alberta, Canada, for calibration. Only three independent variables were included in the model: production (origin), consumption (destination) of a commodity and the distance between a pair of communities. Although the R2 of the model varies between 0.71 and 0.88 for the six commodity groups, which is slightly high, the model's variables are unable to explain the relationship of explanatory factors and socio economic variables of the study area. The concept of freight distribution modeling proposed by Ashtakala and Murthy (1988) was revised by Celik and Guldmann (2007). They used a regression based model to determine freight distribution in US by using a Box-Cox functional form as a transformation function. The flow of 16 commodity groups for the 48 continental states of the US using the 1993 Commodity Flow Survey was explored. A set of explanatory variables that characterized the economic structure of the origins and destinations was used as input variables. They succeeded in proving the significance of the input variables in explaining interregional commodity flow.

In Thailand, Khamphoem and Sirisoponsilp (2009) attempted to use a Box-Cox functional form to analyze intercity commodity flows in Thailand. They were successful in explaining the distribution behavior of food and commodity categories. However, they calibrated the model using the percentage of total movement instead of tonnage units. For this reason, the developed model is unable to determine the tonnage of freight movement in the whole kingdom of Thailand.

MATERIALS AND METHODS

Theory background: The regression based model was derived from the concept of spatial price equilibria of interregional trade. Each point which supplies commodities does so in the form of the firm's production, while at the demand point there are firms and households demanding certain quantities from that supply point. From this concept the model was developed using four equations.

The supply function at the supply point is defined as:

$$S_i = \sigma_i(p_i, s_i)$$

Where:
S_i = The supply quantity at supply point i
p_i = The f.o.b. (free on board) price at i
s_i = A vector of other variables

The relation at demand point j is defined as:

$$y_j = \delta_j(q_j, w, d_j)$$

Where:
y_j = The O-D flow terminating at j
q_j = The c.i.f. (cost, insurance, freight) price vector
w = A vector of parameters that measure the supply characteristics influencing purchase choices
d_j = A vector measuring demand characteristics

Commodity prices at the supply point and the demand point are quite different. The firms at the supply point face f.o.b. while the firms and households at the demand point face c.i.f. price. The relation between c.i.f. prices and f.o.b. prices is defined as:

$$q_{ij} = p_i + c_{ij}$$

where, c_{ij} is the transportation cost between i and j.

At equilibrium condition the relationship between y_{ij} and Si can be written as:

$$\sum_j y_{ij} = S_i \text{ for } \forall_i$$

Eliminating price leads to a reduced form of the model in which the equilibrium flow is directly assigned to the vector of exogenous variables (s, w, d, c). The function of the flow can be written as:

$$Y^* = \xi(s, w, d, c)$$

An approach to analyze the flow from this equation is to view forms such as gravity as empirically functional:

$$\xi_{ij} (s, w, d, c) = a_i (s, w, d, c)f(c_{ij})b_j (s, w, d, c)$$

Where:
a_i = The supply point factor
b_j = The demand point factor
c_{ij} = The interaction factor

Celik and Guldmann (2007) suggested the Box-Cox transformation as the transform function for model parameter estimation. The Box and Cox (1964) power transformation is widely used to achieve a normalizing transformation on a positive-valued response variable as shown below:

$$w = \begin{cases} (Y^2 - 1) & \text{for } \lambda \neq 0 \\ \text{In } Y & \text{for } \lambda = 0 \end{cases}$$

There are a number of variable transformed methods such as left-hand-side only, right-hand-side only, transform independent and dependent variable with the same transform variable and transform independent and dependent variable with different transform variables and no transform of some variables. Among verity of transformation method, transform independent and dependent variables with the same transform variables are considered. The general formulation of transformation model would be:

$$y_j^{(\lambda)} = \beta_1 x_{1j}^{(\lambda)} + \beta_2 x_{2j}^{(\lambda)} + ... + \beta_k x_{kj}^{(\lambda)} + \varepsilon_j$$

Where:
ε_j = Assumed to be a normally distributed error
λ = The transformation parameter

The λ value was estimated by Maximum Log-likelihood function as:

$$LL = -\frac{n}{2}\ln\left[\sum_{i=1}^{n} \frac{(y_i^{(\lambda)} - \overline{y}^{(\lambda)})^2}{n}\right] + (\lambda - 1)\sum_{i=1}^{n} \ln y_i$$

Where:
LL = Log-likelihood function
λ = Maximum likelihood of Box-Cox parameter
N = Number of data
$y_i^{(\lambda)}$ = Number of transformed data
$\overline{y}^{(\lambda)}$ = The arithmetic mean of transformed data calculated by:

$$\overline{y}^{(\lambda)} = \frac{1}{n}\sum_{i=1}^{n} \ln y_i^{(\lambda)}$$

The chi-square test was employed to test whether the estimated transformation parameter is significant compared to the optimal one.

Thailand commodity flow survey data: The office of National Statistics of Thailand (NSO) started a Commodity Flow Data survey on January 2007 which ended on February 2008. The survey collected comprehensive quarterly data of freight transportation in Thailand including commodity type, origin and destination, weight, product value, mode of transport and more from a large sample of shippers in Thailand. The sample was selected according to the ISIC (International Standard Industrial Classification of All Economic Activities: ISIC Rev. 3.0) classification. Shippers with 11+ workers were the population of the survey. A total of 17,149 shippers were included in the survey. Industrial types included in the sample size were: mining and quarrying (1.9%), manufacturing, wholesale and retail trade (69.6%), repair of motor vehicles, motorcycles and personal and household goods (28.1%) and transport, storage and communications (0.4%). The survey divided the commodities into 20 categories using the Harmonized System. Since food and commodity are the only two types of freight which were distributed across the country, only combined food and commodity freight data was used in this research.

Model and variable:
Variable: Origin variables explain the behavior of the trip as generated at the place of origin. The origin acts as the supply point where freight is manufactured for transport to customers in either the zone of origin or in other zones. Therefore activity at the origin consists of

both supply and demand, thus the variables should be proxies for both. Origin variables consist of:

- Employment data presents potential productivity. Since labor is an input for production, origins with greater employment should generate more freight. This research uses wholesale and retail sartorial employment datasets as input variables
- Population and personal income per capita are proxy variables which reflect consumption at the origin
- The average plant size presents scale or diversification effects in the industry. It is estimated by dividing employment by the number of establishments in a sector. The assumption of the variable is that large establishments should produce and export more freight than small establishments

Destination variables serve as proxies for commodity demands, both intermediate and final. Intermediate demand is the demand of manufacturing which uses freight as raw material in its production processes, while final demand represents the demand of customers. Since this research focuses on food and consumer goods, final demand is more appropriate than intermediate demand. Personal income per capita and total populations are proxies for final demand.

Geographical variables are variables that represent impedance of transportation between origin and destination. Distance is the most conventional friction variable used in all spatial interaction models and is measured by highway distance.

The impact of described variables is shown in Table 1. The impact of employment data is expected to be positive while population and personal income per capita at origin are expected to be negative. The other variables, origin and average plant size (Opz), may have either a positive impact, pointing to dominant scale effects, or a negative impact, pointing to dominant diversification effects in the industry. The destination variable is expected the impact to be positive while distance serving as proxy for geographic variable is expected to be negative.

Table 1: Input variable signs-initial expectations

Variable	Description	Expected sign
Dist	Distance	-
Oemp	Wholesale and retail sectoral employment at origin	+
Opop	Total population at origin	-
Oinc	Personal income per capita at origin	-
Opz	Average plant size at origin	+/-
Dpop	Total population at destination	+
Dinc	Personal income per capita at destination	+

Model evaluation: To evaluate the model's performance, Root Mean Square Error (RMSE) and Mean Relative Error (MRE) are employed. The root mean square error is calculated by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(e_i - p_i)^2} \quad For \ \ i = 1 \ to \ n$$

Where:
e_i = The actual value from survey
p_i = The predicted value by models
n = The number of data points

The other performance measure, the mean relative error, MRE, is calculated by:

$$MRE(\%) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left|\frac{p_i - e_i}{e_i} x100\right|} \quad For \ \ i = 1 \ to \ n$$

Where:
e_i = Actual value from experiments
p_i = The predicted value by models
N = The number of data points

RESULTS

The variable as described was input into the model using the backward method. The chi square test was employed to verify the significance of the variables. The results show that almost all variables agree with the expected sign, except origin income (Oinc). Variable origin average plant size (Opz) and origin population (Opop) perform even more poorly. The unsatisfactory variables were excluded and a new model was calibrated.

The results of the final model are shown in Table 2. All variables have high chi square values, especially distance. The distance variable of the final model is highly significant and agrees with the expected sign; when the distance increases, shipments decrease. Since distance is a proxy for transportation cost, for commodities such as cement or industrial products which aren't specific to a certain province, nearby provinces may be preferable to those that are distant.

Only one origin variable, employment at origin (Oemp), has a positive sign and significant impact. Employment is a proxy of manufacture output which also facilitates consumption of the commodity by the final demand. Within the destination variables, Dpop and Dinc are significant and have the expected sign in the model.

Table 2: Parameters of final model

Variable	Parameter	Chi square
Oemp	0.668*	124.317
Dpop	0.523*	77.437
Dinc	0.509*	67.517
Dist	-1.483*	382.651
Const	-7.511*	-

The demand for commodity goods comes from people and the variable which shows the power of personal buying is income. The performance of the model was verified against two performance measures, Root Mean Square Error (RMSE) and Mean Relative Error (MRE). The model has an RMSE value of 94,216.32 and an MRE of 30.56.

DISCUSSION

The results indicated that the model includes four strongly significant variables, origin employment, destination population, destination income per capita and distance. The signs of all variables agree with the expected signs, with three variables being positive while distance is negative. On the other hand, other variables, origin income, origin average plant size and origin population perform even more poorly. The parameter of origin employment (0.668) is close to the value of 0.12- 3.79 reported by Celik and Guldmann (2007). The distance, which is a proxy of geographical variables representing transportation cost, has a negative effect and is highly significant. The parameter value is -1.483 which is close to that reported by Celik and Guldmann (2007), which varied from -1.07 to -2.20. For the destination variable, total population and personal income per capita, which represent final demand at the destination, have coefficients of 0.523 and 0.509 respectively; Celik and Guldmann (2007) reported values which varied from 0.24-3.22 and 0.26-4.82 respectively. Two excluded variables, origin income and origin population, increasing consumption of a local commodity. These results show that most commodities manufactured in Thailand are exported to other regions with minimal local consumption.

CONCLUSION

The need to reveal interregional freight characteristics using available data from Thailand Commodity Flow Survey led to this research. An origin destination matrix based on province to province was calibrated using a flexible Box-Cox function form. The independent variable was classified into three groups: origin variable, destination variable and geographic variable. The origin variable represents the behavior of the trip as generated at the place of origin. Some consumption occurs at the origin point. The employment and the average plant size variables were selected for potential productivity while personal income per capita and total populations were included to explain consumption behavior at origin. Personal income per capita and total populations were selected for destination variable which acts as a proxy for final demand at the destination. The third category, distance, is the most conventional friction variable for geographical variables. Using maximum likelihood and the backward method, the calibrated model reveals that origin income, origin average plant size and origin population perform poorly; therefore these variables were eliminated from the final model. The best model includes four strongly significant variables, as shown by the chi square test. They are origin employment, destination population, destination income per capita and distance. The signs of all variables agree with the expected signs, with three variables being positive while distance is negative. These results demonstrate that the model is reliable and therefore valid. The performance of the model was verified using Root Mean Square Error (RMSE) and Mean Relative Error (MRE).

The developed model explains the behavior of inter-regional freight distribution in Thailand. Among 37 possible freight categories, combined food and consumer goods was chosen to calibrate and verify the method because it was generated and distributed in all provinces. However, to fully develop freight flow data for modeling in Thailand, all commodities must be modeled. This is an important task for future research.

REFERENCES

Ardakani, Z., S. Yazdani and O. Gilanpour, 2009 Studying the effects of non-tariff barriers on the export of the main agricultural products of Iran. Am. J. Applied Sci., 6: 1321-1326. DOI: 10.3844/.2009.1321.1326

Ashtakala, B. and A.S.N. Murthy, 1988. Optimized gravity models for commodity transportation. J. Trans. Eng., 114: 393-408. DOI: 10.1061/(ASCE)0733-947X(1988)114:4(393)

Box, G.E.P. and D.R. Cox, 1964. An analysis of transformations. J. R. Stat. Soc., 26: 211-252. http://www.jstor.org/stable/2984418

Celik, H.M., 2004. Modeling freight distribution using artificial neural networks. J. Transport Geogr., 12: 141-148. DOI: 10.1016/j.jtrangeo.2003.12.003

Celik, H.M. and J.M. Guldmann, 2007. Spatial interaction modeling of interregional commodity flows. Soc.-Econ. Plann. Sci., 41: 147-162. DOI: 10.1016/j.seps.2005.10.003

Khamphoem, J. and S. Sirisoponsilp, 2009. Freight distribution modeling of intercity commodity flows in Thailand using regression models. Proceeding of the 4th Graduate Seminar of Civil Engineering Department, Mar. 2009, Chulalongkorn University, Bangkok, pp: 56-65.

Ogden, K.W., 1978. The distribution of truck trips and commodity flow in urban areas: A gravity model analysis. Transport. Res., 12: 131-137. DOI: 10.1016/0041-1647(78)90052-7

Ortuzar, J.D. and L.G. Willumsen, 1994. Modeling Transportation. 2nd Edn., John Wiley and Sons, Chichester, England, ISBN: 047194193X, pp: 454.

Toosi, M., R. Moghaddasi, S. Yazdani and M. Ahmadian, 2009. Regionalism and its effects on Iranian agricultural exports: The case of economic cooperation organization. Am. J. Applied Sci., 6: 1380-1384. DOI: 10.3844/.2009.1380.1384

Wongoppadoldecha, S. and S. Takano, 2010. Thailand's interregional freight flow estimation. J. Soc. Trans. Traffic Stud., 1: 108-116.