Spatial interaction modeling of interregional commodity flows

H. Murat Celik[a], Jean-Michel Guldmann[b],*

[a]*Department of Civil Engineering, Izmir Institute of Technology, Urla, Izmir 35430, Turkey*
[b]*Department of City and Regional Planning, The Ohio State University, 298 Knowlton Hall, 275 West Woodruff Avenue, Columbus, OH 43210, USA*

Abstract

Drawing from both the spatial price equilibrium theoretical framework and the empirical literature on spatial interaction modeling, this paper expands models of interregional commodity flows (CFs) by incorporating new variables and using a flexible Box–Cox functional form. The 1993 US commodity flows survey provides the empirical basis for estimating state-to-state flow models for 16 commodity groups over the 48 continental US states. The optimized Box–Cox specification proves to be superior to the multiplicative one in all cases, while selected variables provide new insights into the determinants of state-to-state CFs.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Spatial interaction models; Interregional commodity flows; Freight transportation; Box–Cox estimation

1. Introduction

Understanding the determinants of interregional commodity flows (CFs) is critical for both national transportation infrastructure planning (highways, railroad tracks, river/port facilities) and regional development policies (location of activities, reducing regional disparities). Unfortunately, limited data availability has, in the past, hindered empirical research in this area. Prior to the 1993 US commodity flows survey (CFS), the 1977 survey was the most recent one. There has also been a dearth of similar data in other countries (see Section 2). However, the US Bureau of Transportation Statistics has released the results of the 1993 CFS, making them widely available. The structure of these flow data is very suitable for empirical analyses.

Using Bröcker's [1] theoretical framework, this paper expands past empirical research on interregional CFs. It specifies a spatial interaction model that incorporates: (1) variables similar to those used in past CF studies, (2) variables used in international trade models, and (3) a set of completely new variables. The selection of the variables is consistent with Bröcker's framework and with inter-industry transactions considerations. For instance, the origins and destinations are characterized by proxy variables representing final and intermediate demands. Adjacency and customs district dummy, distance, competing destination (Fotheringham [2]), and intervening opportunities (Guldmann [3]) variables are also considered.

Instead of the multiplicative functional form used in the past, a flexible Box–Cox transformation specification is estimated with 1993 CFS data in the current paper. The geographical coverage is the 48 US continental states, while the industry coverage is 16 two-digit manufacturing groups. Interregional service flows, which take primarily the form of information flows, are not part of the CFS and are thus not included in our analysis. These flows are indeed becoming more important in the new information economy, but comprehensive data on them are not available. Therefore, the results obtained here for commodity flows cannot be transferred to service flows, which are likely determined by other factors.

The remainder of the paper is organized as follows. Section 2 consists of a literature review. The modeling methodology is presented in Section 3. Data are described in Section 4. The results are discussed in Section 5. Section 6 is devoted to an elasticity analysis, and Section 7 concludes the paper.

2. Literature review

2.1. Input–output models

Input–output analysis, originally developed to model inter-industry relationships, has been extended to analyze interregional relationships, including the "pure inter-regional input–output model" by Isard [4], the "column coefficient input–output model" by Moses [5], and the "multi-regional input–output model" by Leontief and Strout [6]. In these models, regional (e.g., state) input–output tables are linked via interregional flows, the structure of which is assumed *a priori* (e.g., the percentage distribution of outgoing flows remains fixed across all possible destinations) and is not derived from empirical analyses. Some of these assumptions lead to inconsistent forecasts (e.g., negative ouputs). In any case, input–output models are data-hungry, and building a multi-region and multi-industry structure is prohibitively expensive. Further, because of ever-changing production technologies, these models are not dependable for long-term forecasting.

2.2. Spatial price equilibrium models

Spatial price equilibrium (SPE) models are primarily theoretical, and represent the processes of production and consumption of various commodities in several regions, as well as the resulting trade among these regions. Some of these models have been made computable. Two streams can be distinguished: (1) traditional SPE models, where each region is characterized, for any product, by supply and demand functions, and where the interactions between these functions and interregional transportation costs determine the equilibrium in each region and the pattern of trade; and (2) "new economic geography" (NEG) models, based on the micro-behavior of individual consumers and producing firms, and on assumptions of internal plant economies of scale and monopolistic competition.

The SPE model developed by Samuelson [7] is the early prototype of the traditional stream. It is formulated as an optimization model, where the objective function is equal to the sum of all regional consumers' and producers' surpluses, net of transportation costs. The solution represents an equilibrium, where commodities flow from high-price regions to low-price ones, with price differentials between regions equal to transportation costs. In this model, opportunities for trade are created by differences in production and consumption structures across regions (e.g., Ricardian comparative advantage). Autarky, with no trade at all, would result if the regional supply and demand functions were all the same.

Takayama and Judge [8] propose a quadratic programming approach to solving Samuelson's SPE model, using linear regional demand and supply functions. Another SPE approach is illustrated by the Tinbergen–Bos models developed by Paelinck et al. [9]. These models assume multiple sectors, multiple production centers, and multiple market areas. Centers are ranked according to the types of goods they produce, the lowest producing only agricultural goods, and the highest producing all types of industrial goods. The models consist of four types of relations: (1) trade equations, where imports are set equal to exports for each type of center; (2) product equations, with equilibrium between supply and demand for each sector; (3) income equations; and (4) relationships between the numbers of firms in the various centers.

These constructs clearly combine standard equilibrium modeling with hierarchical concepts borrowed from central place theory. The minimization of transportation costs leads to the concentration of industrial

activities in the highest-ranked centers. They explicitly use distances and the numbers of goods produced (i.e., rank) to explain relations between centers (i.e., regions).

More recently, Brocker [1] has shown that the gravity model is consistent with the implications of traditional SPE models (see Section 3.1). Despite its sound theoretical basis, the SPE model faces problems related to (1) estimating model parameters with available data, and (2) designing efficient algorithms for numerical approximation of the equilibrium (Bröcker [1]). In a multi-regional and multi-commodity framework, it would be very difficult to obtain enough observations for empirical estimation. In addition, the SPE model assumes that any commodity is perfectly homogenous. However, the available commodity flow data are too aggregated to allow for homogeneity, and always involve cross-hauling.

The NEG model was initially developed by Krugman [10], combining, in a spatial framework, the monopolistic competition model proposed by Dixit and Stiglitz [11] with economies of scale at the plant level (through fixed costs) and interregional transportation costs for manufactured goods only. The NEG approach reconciles economies of scale in production with competition through monopolistic competition, considering a variety of products. The basic core–periphery (CP) model involves two regions and two goods (agriculture and manufacturing). The agriculture sector is perfectly competitive and produces a single homogeneous good F that can be transported at no cost. The manufacturing sector provides a large variety of differentiated goods. An aggregate manufacturing production variable M is linked to these differentiated goods through a CES function. Manufacturing labor is mobile, but agricultural labor is not. Consumers have the same Cobb–Douglas utility function, which is optimized in a two-step process: (1) finding the optimal combination of M and F, and (2) finding the optimal variety of differentiated manufacturing goods. Producers maximize profits, and zero-profit conditions prevail at equilibrium. Both sectors employ a single input, labor. Finally, transportation is modeled as an iceberg technology. Instantaneous equilibrium leads to three equations per region: (1) income in region r as a function of manufacturing wage in region r; (2) manufacturing price index in region r as a function of manufacturing wages in all regions; and (3) manufacturing wage in region r as a function of incomes and wages in all regions. These equations are highly non-linear, do not have closed-form solutions, and must be solved numerically.

A particular focus of numerical sensitivity analyses has been on assessing the conditions under which manufacturing is concentrated in one region and agriculture in the other. Some empirical evidence supporting the NEG model is reported in Brackman et al. [12], particularly regarding the spatial wage structure and the home-market effect. Ricci [13] has extended the basic NEG model by including the effects of differences in production costs across countries (Ricardian comparative advantage in traditional trade theory). Ricci shows that the agglomeration of industries at a specific location may reduce their specialization, and that economic integration may reduce regional polarization because of reduced trade costs.

Finally, it is worth noting the attempt by Brackman et al. [12] to show that the gravity model is consistent with the NEG model. A model with 24 regions and congestion costs is solved numerically for different transportation cost parameters. The resulting pseudo data on interregional shipments are regressed on incomes at the origin and destination regions, and the distance. However, in some cases the coefficient of the origin region income is negative.

In summary, both the SPE and NEG models are primarily theoretical, and have not been subjected to rigorous empirical testing. However, they suggest that the characteristics of the regional production and consumption structures, including diversification, specialization, agglomeration, and competition, are important determinants of interregional trade, thus providing a framework for the selection of explanatory variables in any empirical analysis.

2.3. Spatial interaction models

2.3.1. Interregional commodity flows

The problems discussed in Section 2.2 regarding the utilization of SPE/NEG models for empirical analyses of CFs have led to the development and estimation of various forms of spatial interaction models to better understand and forecast CFs. Reed [14], in one of the first studies of CFs, analyzes the interactions of the Bengal Bihar area with the rest of India. Railroad shipments are related to population, employment, and distance. Two separate models, for outflows and inflows, are used. Reed accounts for the effects of intervening

and otherwise competing supplies and demands by introducing potential variables, as well as variables capturing redistribution and concentration effects.

Chisholm and O'Sullivan [15], using UK 1962 and 1964 CF data, covering 78 zones and 13 commodity groups, estimate a production-constrained gravity model (GM). The R^2 obtained by comparing actual and estimated flows varies between 0.24 for steel and 0.62 for food. They also use a linear programming (LP) transportation model to predict flows, and conclude that the R^2s between observed and LP-estimated flows are high for completely homogenous products only. Black [16,17] analyzes the properties and determinants of the distance exponent in the GM using the 1967 US CFS for 24 major shipper groups. He concludes that "(1) the greater the proportion of total shipments from the largest producer (or shipper), the lower the exponent, and (2) the greater the proportion of total flow which is local, the higher the exponent."

Ashtakala and Murthy [18], using data from a 1977 survey of commodity flows between communities within Alberta, estimate a production-constrained GM, where the independent variables are the production (origin) and consumption (destination) of a commodity, and the distance between a pair of communities. The Box–Cox transformation is applied to the distance, and the focus of the estimation is on the transformation parameter through iterative least squares. The R^2 varies between 0.71 and 0.88 for six commodity groups.

Although spatial interaction models have been extensively used in such areas as migration, commuting, shopping, and telecommunication, their application to CF modeling has remained limited. The few studies reviewed thus far focus on best-fitting, very simple models, with little theoretical foundation. Black [16,17], Chisholm and O'Sullivan [15], and Ashtakala and Murthy [18] employ a basic GM (with two mass and one friction variables), where their focus is on estimating the exponent of the distance variable.

In addition to these simple models, Reed [14] adds two more variables: supply/demand potentials, and redistribution/concentration effects. Except for Reed, no one has attempted to account for the effects of the spatial structure on interregional CFs (e.g., Fotheringham [2]). Interestingly, the existence of these effects was first mentioned by Ullman [19], but in a completely qualitative way.

2.3.2. International trade flows

There is a long tradition, starting with Linneman [20], of using GMs in empirical analyses of international (country-to-country) trade flows, which are, of course, closely connected to interregional CF analyses. A good recent example is provided by Frankel and Wei [21], who assess the effects of regional trade arrangements on world trading patterns. Their data set covers 63 countries (3906 exporter–importer pairs). In addition to the standard explanatory variables, such as GNP and GNP per capita in both origin and destination countries, the distance between countries, and dummy variables related to adjacency, language, and trading block affiliation, they create "overall distance" variables, which measure how far an exporting or importing country is from all other countries. Use of these remoteness variables is linked to the idea that the remoteness of an exporter from the rest of the world has a positive effect on trade volume. Thus, Frankel and Wei appear to be the first to introduce measures of the spatial structure in international trade GMs.

A related, more recent and hybrid stream of research has focused on trade flows between Canadian provinces and US states, focusing on the effect of the US–Canada border. McCallum [22] uses a simple GM and 1988 data on trade between each of the 10 Canadian provinces, and between these provinces and the 30 US states that account for 90% of US–Canadian trade. In addition to the GDPs of the origin and destination regions, and the distance between them, a dummy variable is used to differentiate inter-provincial from province-to-state trade. McCallum finds that provinces trade 20 times more with other provinces than with states of similar size and distance, pointing to a significant border effect.

This result is further confirmed by Helliwell [23] with additional data for 1989–1990, and a specific focus on Quebec. Finally, Brown and Anderson [24] estimate an attraction-constrained GM with 1993 US CFS and 1996 Canada TIERS data, to measure the border effect on trade patterns at the two-digit commodity level. Most of their independent variables characterize the origin region, and include output, value added per employee, wage rate, and location quotient. Other variables include distance, dummy variables characterizing the origin–destination (O–D) pair, and a spatial structure variable measuring the origin's accessibility to potential markets. Dummy variables for all destinations are used to ensure that the predicted demand equals the actual demand. Brown and Anderson find that eliminating the border effect would increase Canadian trade with the US by 130%, and that distance remains a primary determinant of trade.

2.3.3. Summary

The empirical interregional CFs studies using spatial interaction models often lack of a strong theoretical basis, being best-fitting heuristics with different specifications and variables. A common point among these studies is that distance is negatively related to CFs. Although the effect of spatial configuration is considered in many other applications of spatial interaction modeling (e.g., Fotheringham [2]; Guldmann [3]), it is only considered by Reed [14] and Brown and Anderson [24] in the case of interregional CFs.

The current research analyzes these flows within a more consistent theoretical framework, using proxy variables for intermediate and final demand conditions, new geographic variables, diversification, specialization, agglomeration, and competition effects, as well as a flexible Cox–Box functional form with endogenously-determined transformation parameters.

3. Modeling methodology

3.1. Theoretical background

Bröcker [1] shows that all forms of the GM are reduced forms of spatial price equilibria of interregional trade, using a modified version of Samuelson's SPE model. At each supply point i, there are firms supplying the commodities, while at each demand point j, there are firms and households demanding certain quantities $(y_{1j}, y_{2j}, \ldots, y_{ij}, \ldots y_{Ij})$ from the supply points ($i = 1 \rightarrow I$). The supply firms are faced with f.o.b. (free on board) prices, and the buying firms and households with c.i.f. (cost, insurance, freight) prices. This model consists of four equations.

The supply functions are defined as

$$S_i = \sigma_i(p_i, \boldsymbol{s}_i) \tag{1}$$

where S_i is the supply quantity at supply point i, p_i is the f.o.b. price at i, and $\boldsymbol{s}_i$ is a vector of other variables, such as prices of other commodities. σ_i is monotone, non-decreasing in p_i. For each demand point j, there is a demand correspondence, a point-to-set mapping which assigns the vector of O–D flows terminating at j, $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{ij}, \ldots y_{Ij})$, to the c.i.f. price vector, $\boldsymbol{q}_j = (q_{1j}, \ldots, q_{ij}, \ldots, q_{Ij})$, so that

$$\boldsymbol{y}_j \varepsilon \delta_j(\boldsymbol{q}_j, \boldsymbol{w}, \boldsymbol{d}_j), \tag{2}$$

where $\boldsymbol{w}$ is a vector of parameters that measures the supply characteristics influencing purchase choices, and $\boldsymbol{d}_j$ is a vector measuring demand characteristics, including income, prices of other commodities, etc.

The third equation defines c.i.f. prices:

$$q_{ij} = p_i + c_{ij}, \tag{3}$$

where c_{ij} is the transportation cost between i and j. The fourth equation states the equilibrium conditions:

$$\sum_j y_{ij} = S_i \ \forall i. \tag{4}$$

A spatial price equilibrium is characterized by prices and quantities satisfying (1)–(4), which represent the *explicit* (or *structural*) form of the trade model, with both prices and quantities as endogenous variables. Eliminating prices leads to the reduced form of the model, where equilibrium flows are directly assigned to the vector of exogenous variables, $(\boldsymbol{s}, \boldsymbol{w}, \boldsymbol{d}, \boldsymbol{c}) = (s_1, \ldots, s_I, w_1, \ldots, w_I, d_1, \ldots d_J, c_{11}, \ldots, c_{IJ})$. The reduced form is denoted by $\zeta = (\zeta_{11}, \ldots, \zeta_{ij}, \ldots, \zeta_{IJ})$, so that, for any equilibrium flow matrix

$$\boldsymbol{Y}^* = (\boldsymbol{y}_I^*, \ldots, \boldsymbol{y}_J^*),$$
$$\boldsymbol{Y}^* = \zeta(\boldsymbol{s}, \boldsymbol{w}, \boldsymbol{d}, \boldsymbol{c}). \tag{5}$$

Of course, there is no closed mathematical formulation of ζ. One way to think about this function is to solve the equilibrium problem for a wide range of combinations of values for the input parameters $(\boldsymbol{s}, \boldsymbol{w}, \boldsymbol{d}, \boldsymbol{c})$, for instance over a grid. The resulting flow values Y_{ij}^* could then be regressed over the input parameters, providing an approximation of the function ζ. An alternative approach is to view (5) as a general guide for the selection of simpler, and empirically estimable functional forms. Brocker [1]

shows that the generalized GM form

$$\zeta_{ij}(\boldsymbol{s}, \boldsymbol{w}, \boldsymbol{d}, \boldsymbol{c}) = \boldsymbol{a}_i(\boldsymbol{s}, \boldsymbol{w}, \boldsymbol{d}, \boldsymbol{c}) f(c_{ij}) \boldsymbol{b}_j(\boldsymbol{s}, \boldsymbol{w}, \boldsymbol{d}, \boldsymbol{c}) \tag{6}$$

is consistent with (5). Eq. (6) suggests that the origin and destination factors, $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$, may be functions of the whole vectors $(\boldsymbol{s}, \boldsymbol{w}, \boldsymbol{d}, \boldsymbol{c})$, and not only of the components of these vectors associated with i or j, exclusively. In the standard GM,

$$\boldsymbol{a}_i = \boldsymbol{a}_i(s_i, w_i), \tag{7}$$

$$\boldsymbol{b}_j = \boldsymbol{b}_j(d_j), \tag{8}$$

the origin mass factor is only a function of origin supply variables, and the destination mass factor is only a function of the destination demand variables. Eq. (6) clearly suggests that supply and demand variables for other locations $k(\neq i, j)$ may be included in $\boldsymbol{a}_i$ and $\boldsymbol{b}_j$. The following section presents the adaptation of (6) to an empirically estimable model.

3.2. The empirical commodity flow model

3.2.1. Variables
3.2.1.1. Origin variables.   The origins serve as supply points but also consume part of this supply. The origin variables should therefore be proxies for both supply and demand conditions.

Sectoral employment, oemp, and sectoral value-added, ovlad, are used as proxies for sectoral production at the origin. Their impacts are expected to be positive. Wholesale employment, owsem, measures the expected positive effects of redistributive activities on out-shipments. Wholesale activities may also facilitate consumption of the commodity by the final demand sector at the origin. Thus, the expected impact of owsem is positive.

Total population, opop, and personal income per-capita, opipc, are proxies for final demand conditions at the origin. Their impacts are expected to be negative. As local final demand increases, commodity out-shipments decrease due to increased local consumption. Note also that the total population variable provides an implicit ranking of the origin regions, a factor considered important in the Tinbergen–Bos models [9].

The average plant size, oaps, captures scale or diversification effects in the industry. It is estimated by dividing employment by number of establishments in a sector. As plant scale increases, total production and, thus, total out-shipments, can be expected to increase due to higher production efficiency. However, the two-digit level aggregation may not reflect this effect properly. Total out-shipments by small firms may exceed out-shipments by larger firms, because smaller firms may feature more product diversity, and thus be more attractive to export markets. Therefore, oaps may either (1) have a positive impact, pointing to dominant scale effects, or (2) have a negative impact, pointing to dominant diversification effects in the industry. Note that this factor is prominent in the NEG theoretical models (Section 2.2).

3.2.1.2. Destination variables.   The destinations serve as demand points, with the destination variables serving as proxies for commodity demands, both intermediate and final. Manufacturing employment, dmnem, is a proxy for intermediate demand. Personal income per capita, dpipc, and total population, dpop, are proxies for final demand conditions. Note that dpop also provides an implicit ranking of the destination regions [9]. Finally, wholesale employment, dwsem, measures redistributions effects at the destination. Their impacts are expected to be positive.

3.2.1.3. Geographical variables.   Distance is the most conventional friction variable used in all spatial interaction models, and is measured by highway distance, great circle distance, etc. The current study uses the average distance of hauled commodities. Distance is taken as a proxy for transportation costs. This may be inappropriate when the supply of transportation infrastructure is inadequate, which could be the case in underdeveloped/developing countries or regions, where a time/cost measure would be more representative. However, this is clearly not the case of the US, with its highly developed highway, railroad, and water

transportation systems. In any case, systematic interregional transportation time/cost data are not available. The impact of distance is expected to be negative.

Two variables are employed to capture the effects of the spatial configuration of states: competing destinations, cd, and intervening opportunities, io. These variables may be viewed as integrating into the model the demand/supply effects at locations other than the origin (i) and destination (j). With reference to (5) and (6), these variables represent at least a portion of the vectors (s, w, d). The cd variable measures the accessibility of a specific destination to all other destinations. It is estimated using total employment (TE_k) and the distance between two destinations (d_{kj}), with

$$CD_{ij} = \sum_k TE_k/d_{kj}, \quad k \neq (i,j). \tag{9}$$

There is no presumption about the sign of this variable. A negative sign indicates that there is competition among destinations, wherein as other destinations k get closer to destination j, the shipments from i to j decrease. A positive sign implies agglomeration effects: flows increase as other destinations get closer to destination j, and thus make it more attractive.

The io variable is defined by a formula similar to (9), but the distance is d_{ki} instead of d_{kj}. According to the intervening opportunities concept, flows to a destination decrease when the opportunities between the origin and the destination increase. Just like clusters at destinations, the io variable may be used to describe the spatial configuration of the clusters around origins.

According to this approach, a positive sign indicates that, when other origins get closer (implying an economic concentration around the origin), the flow to destination increases. This suggests agglomeration effects at the supply level. However, a negative sign implies that destinations around the origin may act as competing destinations, thus reducing the flow to destination. An alternative interpretation of a negative sign could be linked to agglomeration diseconomies. The larger the cluster, the larger the negative effects (e.g., congestion); hence, the lesser the demand and the flow to destination. The cd and io variables may also be viewed as representing the competitive and agglomerative factors that play a dominant role in the NEG models (see Section 2.2).

Three dummy variables are also used in the proposed model. First, the adjacency dummy, adjncy, measures whether having a common physical border has an effect on commodity flows between states. The sign is expected to be positive: trade flows between neighboring states increase because of better business information, regional cultural commonalities, etc.

Imports and exports are included in the 1993 CFS, from, and up to, the custom districts where the commodity enters or leaves the US. For this reason, two customs district dummy variables, ocddmy for the origin and dcddmy for the destination, measure the effects of foreign trade at either the origin or the destination, on commodity flows. The magnitudes of their coefficients depend on the foreign trade share of interregional commodity flows.

These variables may have either a negative or a positive sign. A positive ocddmy indicates that the sector may have a significant foreign import of the commodity, while a negative sign implies a significant foreign export. A positive dcddmy, on the other hand, implies that the sector may have a significant foreign export, while a negative sign would point to imports. States with custom districts are coastal (Oceans or Great Lakes) and along the borders with Canada and Mexico (Montana, North Dakota, and Arizona). Expectations about the signs of all the selected variables are summarized in Table 1.

3.2.2. Functional form

The commodity flow F_{ij} between two points (i,j) can be formulated as a function of the variables described in Section 3.2.1, using the framework of Eq. (6):

$$F_{ij} = a_i(\text{ovlad, oemp, opop, opipc, oaps, owsem, io, ocddmy}) \cdot g_{ij}(\text{dist, adjncy})$$

$$b_j(\text{dwsem, dmnem, dpop, dpipc, dcddmy, cd}), \tag{10}$$

where a_i is the supply point factor, b_j the demand point factor, and g_{ij} the interaction factor. Eq. (10) could be, in line with past empirical research, represented by a multiplicative functional form. However, other

Table 1
Initial expectations about variable signs

Variable	Sign
adjncy	+
ocddmy	+/−
dcddmy	+/−
cd	+/−
io	+/−
dist	−
opipc	−
opop	−
oemp	+
owsem	+
ovlad	+
oaps	+/−
dmnem	+
dwsem	+
dpipc	+
dpop	+

functional forms may be acceptable, and there are no strong theoretical reasons to prefer one functional form to another. It is therefore reasonable to allow for the endogenous selection of the functional form.

The Box–Cox transformation, wherein the variable X is transformed into the variable $X^{(\lambda)}$ according to

$$X^{(\lambda)} = (X^\lambda - 1)/\lambda, \tag{11}$$

is ideally suited to this purpose (Box and Cox [25]). Two different transformation parameters are considered: one for all the independent variables (λ) and one for the dependent variable (θ). Dummy variables, however, are not transformed. A general formulation of the Box–Cox model could be

$$\frac{Y^\theta - 1}{\theta} = a_0 + a_1 X_1 + a_2 \frac{X_2^\lambda - 1}{\lambda} + \cdots + a_n \frac{X_n^\lambda - 1}{\lambda} + \varepsilon, \tag{12}$$

where X_1 is a dummy variable and ε is assumed to be a normally distributed error, with $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \sigma^2 I$.

The Box–Cox transformation (11) is continuous at $\lambda = 0$ because $X^{(\lambda)}$ tends toward $\ln X$ when $\lambda \to 0$. Thus, the linear and multiplicative functional forms are simply specific points ($\lambda = 1$ and 0) on a continuum of forms allowing for different degrees of independence and interaction among the variables. The fundamental criterion for comparing the infinite number of *a priori* possible models is how well they are able to explain the data, with the best model maximizing the likelihood of the original observations. Once the optimal functional form (λ^*, θ^*) has been determined, it is possible to test whether an alternate form (λ, θ) is significantly different from the optimal one, using a chi-square test.

4. Data sources and processing

Four main data sources are used in the current study: the 1993 CFS, the 1993 county business patterns (CBP), the 1992 Census of Manufacturing (CM), and the annual state personal income (ASPI) database. The CFS provides data for the dependent flow and distance variables, while the other sources provide data for the independent variables. References for accessing these data are provided in the Appendix.

4.1. Dependent variable

The dependent variable, flow, is drawn from File 9 of the 1993 CFS, and represents the value (Million $) of out-shipments from each origin state to every other state, for each of 16 commodity groups (see Table 2),

Table 2
Commodity groups codes and definitions

Code	Definition
20	Food and kindred products
24	Lumber or wood products
25	Furniture or fixture
26	Pulp, paper or allied products
28	Chemicals or allied products
29	Petroleum or coal products
30	Rubber or plastics products
32	Clay, concrete, glass and stone products
33	Primary metal products
34	Fabricated metal products
35	Machinery, excluding electrical, products
36	Electrical machinery products
37	Transportation equipment
38	Precision instruments
39	Miscellaneous freight shipment
75	Textile, apparel and leather products

primarily defined at the two-digit SIC level (the highest level of disaggregation of O–D flows in the CFS). The file includes 83,232 flow observations, with 22,476 missing (27%) because of data disclosure and sampling problems; but these missing flows represent 12% of the total flow. The missing rate varies from a high of 36% for leather products to a low of 3% for food and kindred products. Missing observations are eliminated from the database. Geographical coverage is the 48 US continental states. Imported products shipments are included after they leave the importer's domestic location for another location. Export shipments are also included until they reach the port of exit from the US Shipments through a foreign country (e.g., Canada), with both the origin and destination in the US, are included.

4.2. Independent variables

The adjacency dummy variable, adjncy, is equal to 1 if the origin and destination states have a common border, and 0 otherwise. The customs district variables, ocddmy and dcddmy, are equal to 1 if the state contains at least one customs district, and 0 otherwise. All the employment variables are drawn from the CBP, and include (1) origin sectoral employment, oemp; (2) origin wholesale employment, owsem; (3) destination manufacturing employment, dmnem; and (4) destination wholesale employment, dwsem.

The origin average establishment size, oaps, is estimated by dividing the origin sectoral employment by the number of establishments in that sector. The numbers of establishments are drawn from the CBP. The value-added variable, ovlad, is drawn from the 1992 CM. The state personal incomes per-capita, opipc and dpipc, and the state populations, opop and dpop, are drawn from the ASPI database.

The distance, dist, is directly derived from the 1993 CFS, which has O–D data on both tons and ton-miles for each commodity. Dividing the flow of commodity measured in ton-miles by the same flow measured in tons provides an estimate of the average hauled distance for each commodity and O–D pair. This measure accounts for the multiplicity of departure points within the origin state, and of arrival points within the destination state. It represents a weighted average of these individual distances, thus accounting for the structure of the transportation networks involved. This is certainly more precise than computing the distance between state capitals, as is often done.

4.3. Statistics overview

Table 3 presents statistics for all variables across all commodities. The total value of the 16 commodities traded in the US within and across state lines was $ 5160 billions in 1993. The largest share characterizes food

Table 3
Descriptive statistics for all commodities combined

Variable	N	Mean	St. Dev.	Minimum	Maximum
Flow ($M)	26023	92	280	0	7800
Cd	26023	201995	178751	16323	771064
io	26023	197308	179558	16323	771064
dist (miles)	26016	1241	757	40	3519
opipc ($)	26023	20660	2841	15468	29602
opop (#)	26023	5473805	5785073	460000	30380000
oemp (#)	26023	22100	32709	1	284042
owsem (#)	26023	143180	153831	7807	783658
ovlad ($M)	26023	1730	2889	0	21698
oaps (#)	26023	71	94	2	1715
dmnem (#)	26023	417532	390355	11285	1898885
dwsem (#)	26023	146218	153495	7807	783658
dpipc ($)	26023	20736	2881	15468	29602
dpop (#)	26023	5585155	5766049	460000	30380000

(16.5%), followed by transportation equipment (12.4%), and chemicals (10.2%). Other significant products are non-electrical machinery (8.6%), and textile, leather, and apparel (8.6%).

The highest value product is precision instruments ($2,566/ton), and the cheapest (bulkiest) is clay, concrete, glass, and stone ($11/ton). Other high-value products include electrical ($1363/ton) and non-electrical ($1295/ton) machineries, transportation equipment ($748/ton), and textile, leather, and apparel (($1050/ton). Lumber and wood ($19/ton), coal and petroleum ($19/ton), and primary metals ($86/ton) belong to the low value/bulky products group. Low-value commodities are expected to have short hauls, with high-value ones requiring long hauls, reflecting the impact of transportation costs on total production costs.

With approximately 11.5% of shipments, California displays the largest manufacturing concentration. The second largest concentration is in Texas, with 8% of shipments. Illinois, Ohio, New Jersey, Michigan, New York, Pennsylvania, North Carolina, Georgia, Florida, and Indiana export/import approximately 30–40% of all shipments.

Foreign trade is included in CFS flows to and from US customs states. Non-electrical machinery had the highest share of foreign export (19.4%), followed by clay, concrete, glass, and stone (17.7%), electrical machinery (14.6%), transportation equipment (11.6%), and precision instruments (11%). The products with high shares of foreign imports include clay, concrete, glass, and stone (24.2%), non-electrical machinery (22.3%), electrical machinery (20.7%), transportation equipment (16.2%), coal and petroleum (16.3%), textile, leather, and apparel (14.9%), furniture (13.1%,), and precision instruments (11.1%).

5. Results

Results of the estimation are presented in Tables 4 and 5. The selected variables are, in most cases, significant in explaining interregional commodity flows, with the expected signs. The optimized Box–Cox specification is always statistically superior to the linear-in-logarithms specification, the mainstay of past empirical work.

A common physical border significantly increases commodity exchanges between contiguous states. For 15 of the 16 commodity groups, adjncy is thus significant at the 5% level. This result is consistent with the international trade GMs. While an increasing distance guarantees a declining interaction, this decline is attenuated among contiguous states. The ability to obtain better business information about suppliers and/or consumers, as well as possible cultural commonalities, are likely factors. A business trying to expand its market beyond state boundaries is likely to first focus on neighboring states before expanding beyond, thus ensuring a differential advantage to these states. It is also possible that short-haul transportation between contiguous states may be different and less expensive than for greater distances.

Table 4
Box–Cox models of interregional commodity flows—groups (20–32)

	20		24		25		26		28		29		30		32	
adjncy	1.62	*	1.82	*	0.50	*	0.51	*	1.84	*	8.25	*	1.14	*	1.37	*
ocddmy	−0.27	*	−0.20	**	−0.10		0.27	*	0.01		1.84	*	−0.17	**	−0.12	
dcddmy	0.21	**	0.07		0.20	**	−0.01		0.16		0.33		−0.05		0.08	
cd	−0.85	*	−1.84	*	−1.13	*	−0.60	*	−0.23	*	−0.41	**	−0.42	*	−0.77	*
io	−0.67	*	−1.28	*	−0.58	*	−0.50	*	−0.20	*	−0.18		−0.09		0.26	*
dist	−2.20	*	−2.22	*	−1.64	*	−1.62	*	−1.29	*	−3.00	*	−1.07	*	−1.41	*
opipc	−0.09		2.22	*	0.09		−0.37		0.00		−2.19	*	−0.69	*	0.14	
opop	−0.57	*	0.09		0.07		0.22	**	−0.10		−0.45	**	−0.33	*	−0.29	
oemp	0.87	*	−0.23		1.38	*	0.52	*	0.45	*	1.95	*	0.12		0.96	*
owsem	0.90	*	0.42		0.68	*	0.15		0.23		0.78	*	0.75	*	0.07	
ovlad	1.05	*	2.37	*	−0.06		0.49	*	0.60	*	−0.12		0.68	*	0.36	**
oaps	−0.88	*	0.33		0.45	*	−0.16		−0.71	*	−2.85	*	−0.01		−0.66	*
dmnem	0.19	*	1.02	*	−0.09		0.37	*	0.10	*	0.03		0.24	*	0.36	*
dwsem	1.02	*	0.76	*	0.71	*	0.67	*	0.73	*	0.28		0.42	*	0.35	*
dpipc	1.05	*	3.40	*	2.42	*	0.84	*	−0.12		−0.10		0.44	*	1.26	*
dpop	0.47	*	1.12	*	1.25	*	0.17		0.05		0.31		0.24	*	0.65	*
λ	0.08	*	0.02		0.04	*	0.09	*	0.11	*	0.19	*	0.10	*	0.07	*
θ	0.19	*	0.17	*	0.12	*	0.18	*	0.18	*	−0.24	*	0.19	*	0.13	*

*Significant at 5 % level. **Significant at 10 % level.

Table 5
Box–Cox models of interregional commodity flows—groups (33–75)

	33		34		35		36		37		38		39		75	
adjncy	0.85	*	1.22	*	1.52	*	1.30	*	0.06		1.51	*	1.47	*	1.05	*
ocddmy	−0.22	**	−0.22	*	0.04		−0.18	**	−0.60	*	0.18		0.01		−0.21	
dcddmy	−0.36	*	0.11		0.17	**	−0.08		0.07		0.23	*	0.13		−0.12	
cd	−0.48	*	−0.54	*	−0.42	*	−0.67	*	−0.56	*	−0.26	*	−0.58	*	−2.61	*
io	−0.41	*	−0.21	*	−0.15	*	−0.02		0.29	*	0.09		−0.16	*	−1.12	**
dist	−1.57	*	−1.26	*	−0.86	*	−0.96	*	−1.24	*	−0.70	*	−0.93	*	−2.04	*
opipc	0.04		−0.30		−0.26	**	−1.16	*	−0.79	*	−1.01	*	0.69	*	1.59	
opop	0.20		−0.31	*	−0.32	*	−0.45	*	−0.30	*	−0.98	*	−0.13		−3.10	*
oemp	1.23	*	−0.04		0.16		1.54	*	0.15		0.92	*	0.78	*	3.79	*
owsem	0.30	**	0.90	*	0.88	*	0.81	*	0.41	*	1.42	*	0.57	*	3.46	*
ovlad	−0.21	**	0.86	*	0.49	*	−0.36	*	0.59	*	0.07		0.26	*	−0.36	**
oaps	−0.11		0.25		0.07		−1.03	*	−0.80	*	−0.85	*	0.04		−1.35	*
dmnem	0.49	*	0.24	*	0.09	*	0.05		0.16	*	−0.11	**	0.05		1.48	*
dwsem	0.29	*	0.35	*	0.47	*	0.56	*	0.18	**	0.49	*	0.32	*	0.64	
dpipc	−0.45		0.37	*	0.26	**	1.52	*	0.76	*	0.98	*	1.05	*	4.82	*
dpop	0.57	*	0.43	*	0.33	*	0.93	*	0.42	*	0.58	*	0.82	*	3.22	*
λ	0.07	*	0.10	*	0.11	*	0.07	*	0.13	*	0.08	*	0.07	**	−0.03	
θ	0.18	*	0.21	*	0.20	*	0.18	*	0.14	*	0.16	*	0.19	*	0.13	*

*Significant at 5 % level. **Significant at 10 % level.

The foreign trade variables, ocddmy and dcddmy, do not display the same level of consistency as adjncy. Only five commodity groups have a significant occdmy at the 5% level, and four groups at the 10% level. The other foreign trade variable, dcddmy, performs even more poorly: it is significant at the 5% level for three commodity groups, and at the 10% level for only two groups. Focusing on ocddmy, note that seven out of nine significant coefficients are negative (groups 20, 24, 30, 33, 34, 36, 37), which suggests that foreign exports taking place at the origin node reduce the interregional commodity outflows from these nodes. These results are consistent with the

significant foreign export volumes of sectors 36 (non-electrical machinery) and 37 (transportation equipment), and, to a lesser extent, of sector 20 (food), 30 (rubber and plastics), and 33 (primary metals).

In contrast, the coefficient is positive for groups 26 and 29, which suggests that when the origin node imports pulp and paper, and petroleum and coal, these imports stimulate interregional flows out of these origin nodes. In the case of petroleum, this result is very much consistent with the importance of foreign oil in the US economy.

The competing destination variable, cd, is uniformly negative and significant in all groups: in 15 at the 5% level, and in 1 at the 10% level. These results suggest that competition effects at destinations are strong determinants of interregional commodity flows. As other destinations are physically closer (clustered) to a specific destination, the flow of commodities reaching this destination decreases. Every other factor remaining constant, this clustering absorbs part of the flow that would have ended at this destination under a less clustered configuration. This result is consistent with similar effects uncovered in the case of other spatial interactions.

The intervening opportunities variable, io, does not have the same highly consistent effects as cd. Ten groups have negative signs at the 5% level, one group has a negative sign at the 10% level, and two groups have a positive and significant coefficient. Overall, competition effects at the supply level appear dominant. Destination nodes clustered around the origin serve as alternative destinations for the commodity, and absorb part of the flow that would have ended at the selected destination.

The distance variable, dist, is always negative and highly significant for all groups. Distance can be viewed as a proxy for transportation cost, where increasing transportation costs are an obvious deterrent to trade. Also, from an information viewpoint, the farther away, the lesser is the information about business opportunities, and, hence, the fewer the interactions.

The distance coefficients for sectors 35–39 are generally lower (in absolute terms) than those for the other sectors, which indicates that the shipping distances for these goods are greater. This is consistent with the value per weight of these commodities, which ranges from $748 to $ 5566/ton. The latter characterizes precision instruments (group 38), with the lowest coefficient (−0.70). The other groups, with values per weight ranging from $11 to $427/ton, have distance coefficients varying from −1.07 to −2.20.

The origin state personal income per capita, opipc, is presumed to have a negative sign: when the local consumption of a commodity increases, its out-of-state shipments decrease. This is generally verified for the commodities involved in final consumption (e.g., petroleum, rubber, machinery, transportation equipment, precision instruments). Six groups have negative signs at the 5% level, and two groups at the 10% level. The origin state total population, opop, is similar to opipc: both opop and opipc are used as surrogates for local final consumption. Eight groups have negative and significant signs at the 5% level, and two groups at the 10% level. When significant, opop and opipc generally have the same sign, which supports their use as measures of final consumption at the origin.

The origin sectoral employment, oemp, has a generally positive and significant impact (13 groups at the 5% level). The other sectoral production variable, ovlad, also has generally positive coefficients: nine groups at the 5% level, and one group at the 10% level. Scale or diversification effects have been tested using average establishment size, oaps. Eight of the 16 group have significant negative signs, implying that these sectors are characterized by diversification effects. Only one group, furniture and fixture, has a positive and significant coefficient, pointing to economies of scale effects.

Another important result is that wholesale employment, representing redistributive activities at both origin and destination, does indeed facilitate interregional commodity flows, by buying commodities from producers and reselling them to the intermediate and final demand sectors: 13 groups thus have a significant positive sign for owsem (12 at the 5% level), and 14 groups for dwsem (13 at the 5% level).

Manufacturing employment at destination, dmnem, is a proxy for the effects of intermediate demand, and is presumed to have a positive sign. This is verified for 11 groups at the 5% level. The variables representing final demand at destinations, dpipc and dpop, are also expected to have positive effects. This is the case for 13 groups for both variables.

Focusing on their likely demand structure, it is possible to assign commodities to three main groups: (1) primarily inputs to final demand, (2) primarily inputs to intermediate demand, and (3) inputs to both demands.

(1) Product 25 (furniture and fixture), product 29 (coal and petroleum), product 36 (electrical machinery), product 38 (precision instruments), and products 39 (miscellaneous manufacturing), are primarily inputs to

final demand. The intermediate demand variable, dmnem, is not significant for any of these products, confirming the hypothesis. The origin local demand for furniture and fixture is not significant, probably because the production of this commodity is spatially concentrated, with economies of scale (oaps is positive and significant for this group only). Product 29, however, is very sensitive to origin local demand conditions but not to destination local demand, probably because it is highly sensitive to shipping distance. The other groups, however, are sensitive to both origin and destination final demands.

(2) Product 24 (lumber and wood), product 28 (chemicals), product 32 (clay, concrete, glass, and stone), and product 33 (primary metals), are the sectors that produce primarily for intermediate demand. The variable dmnem is very significant for all four groups, whereas the final demand variables are not significant.

(3) Product 20 (food), product 26 (pulp and paper), product 30 (rubber and plastic), product 34 (fabricated metals), product 35 (non-electrical machinery), product 37 (transportation equipment), and product 75 (textile, apparel, and leather), on the other hand, are supplied to both intermediate and final demands. The variables representing both demands are highly significant for these products.

6. Elasticity analysis

While the estimation results reported in Section 5 point to the significance and direction of impact of the independent variables, they do not allow for a straightforward comparative analysis of these effects. In order to assess and compare the sensitivities of shipments to variations in the non-dummy independent variables, an elasticity analysis was conducted using the estimated equations. The elasticity of Y with respect to X_i measures the percentage increase in Y resulting from a one percent increase in X_i. In the case of the Box–Cox model, the elasticity is

$$\varepsilon_{X_i} = a_i \frac{X_i^\lambda}{Y^\theta}. \tag{13}$$

When $\lambda \to 0$ and $\theta \to 0$, Eq. (13) becomes a log-linear specification, and the elasticity is a_i. Elasticities were computed over all sample observations with non-zero flows ($Y > 0$). The means of these computed elasticities are presented in Table 6.

The mean elasticities of cd are all negative, indicating a competition effect at the destination that is highest (−1.63) for transportation equipment, and lowest (−0.48) for precision instruments. Lumber and wood (−1.60), furniture or fixture (−1.49), clay, concrete, glass, and stone (−1.48), and textile, apparel, and leather (−1.35) are also high-competition-effect products.

Most io elasticities are negative, indicating a competition effect at the origin that is highest for lumber and wood (−1.10), and lowest for electrical machinery products (−0.03). Three products have positive elasticities, indicating agglomeration effects at the origin: transportation equipments (0.86), clay, concrete, glass, and stone (0.50), and precision instruments (0.16). Bulkier or cheaper products have high distance elasticities: clay, concrete, glass, and stone (−1.88), pulp and paper (−1.79), furniture and fixture (−1.75), lumber and wood (−1.77), and food (−1.73). The lowest elasticity product is precision instruments (−0.81).

Transportation equipment, precision instruments, and petroleum/coal are most sensitive to final demand conditions at the origin. The mean elasticities of opipc and opop are −2.44 and −1.36 for petroleum/coal, −1.78 and −1.35 for transportation equipment, and −1.50 and −2.24 for precision instruments. The mean elasticity of oemp varies between 2.08 (textile, apparel, and leather) and 0.19 (rubber or plastic). The mean elasticity of ovlad varies between 1.88 (lumber and wood) and 0.09 (precision instruments). The variable oaps has both positive and negative mean elasticities. The highest positive elasticity, indicating scale effects, is 0.43 (furniture and fixture), and the lowest is 0.03 (miscellaneous). The highest negative elasticity is −1.49 (petroleum and coal).

Wholesale employment at both origins and destinations is a strong determinant of interregional CFs. The mean elasticity of owsem varies between 1.81 (apparel, textile, and leather) and 0.36 (lumber and wood). The mean elasticity of dwsem varies between 1.46 (chemicals) and 0.34 (apparel, textile, and leather). The mean elasticity of the intermediate demand variable, dmnem, varies between 0.72 (clay, concrete, glass, and stone) and 0.07 (petroleum and coal, electrical machinery). The final demand variables at the destinations, dpipc and dpop, have the highest elasticities: 2.94 and 1.89 for furniture and fixture, 2.85 and 1.02 for lumber and wood,

Table 6
Mean elasticities of variables by commodity groups

Variable	20	24	25	26	28	29	30	32	33	34	35	36	37	38	39	75
cd	−1.00	−1.60	−1.49	−1.05	−0.48	−0.71	−0.86	−1.48	−0.67	−0.91	−0.86	−0.82	−1.63	−0.45	−0.86	−1.35
io	−0.78	−1.10	−0.77	−0.89	−0.41	−0.29	−0.19	0.50	−0.58	−0.37	−0.29	−0.03	0.86	0.16	−0.24	−0.58
dist	−1.73	−1.77	−1.75	−1.79	−1.54	−1.09	−1.31	−1.88	−1.52	−1.33	−1.00	−0.83	−1.85	−0.81	−0.97	−1.20
opipc	−0.09	***	***	−0.56	***	−2.44	−1.16	***	***	−0.42	−0.42	−1.24	−1.78	−1.50	***	***
opop	−0.86	***	***	***	−0.30	−1.36	−0.94	−0.70	***	−0.73	−0.95	−0.69	−1.35	−2.24	−0.25	−1.47
oemp	0.86	***	1.62	0.72	0.70	1.36	0.19	1.51	1.40	***	0.27	1.65	0.34	1.31	0.93	2.08
ovlad	0.88	1.88	***	0.53	0.75	***	0.81	0.47	***	0.91	0.59	***	0.95	0.09	0.25	***
owsem	1.02	0.36	0.88	0.25	0.44	1.19	1.49	0.13	0.40	1.47	1.71	0.96	1.15	1.41	0.82	1.81
oaps	−0.57	0.25	0.43	−0.14	−0.62	−1.04	−0.01	−0.70	−0.09	0.19	0.05	−0.76	−0.92	−0.80	0.03	−0.86
dmmem	0.24	0.89	***	0.68	0.23	0.07	0.53	0.72	0.71	0.42	0.20	0.07	0.50	***	0.08	0.76
dwsem	1.16	0.65	0.92	1.13	1.46	0.44	0.83	0.65	0.39	0.57	0.91	0.67	0.50	0.83	0.45	0.34
dpipc	1.05	2.85	2.94	1.21	***	***	0.74	2.09	***	0.52	0.43	1.63	1.71	1.46	1.37	1.63
dpop	0.72	1.02	1.89	0.40	0.15	0.94	0.69	1.56	1.00	1.00	0.98	1.44	1.92	1.32	1.53	1.53

***Elasticities not computed.

and 2.09 and 1.56 for clay, concrete, glass, and stone. Table 6 shows that, in facilitating interregional CFs, the final demand variables, opipc, opop, dpipc, and dpop, have the highest elasticities. Distance also has high elasticities. Wholesale employment has lower elasticities, but with a more stable and consistent pattern.

7. Conclusions

This paper has expanded past models of interregional commodity flows by incorporating new explanatory variables and using a flexible Box–Cox specification. Based on input–output concepts of intermediate and final demands, the new variables include more detailed descriptions of the economies of the origin and destination states, such as employment and value added for the commodity sector at the origin state, wholesale employment at both ends, manufacturing employment at the destination state, and population and per-capita income at both ends. In addition, the average establishment size for the commodity at the origin is intended to measure scale or diversification effects. The competitive or agglomerative effects of the economic spatial structure are captured with competing destination and intervening opportunities variables. In addition to the average hauling distance between states, the model includes dummy variables measuring the effects of a common physical border, and the role of customs districts states.

Overall, the results show that the selected variables and Box–Cox functional form are successful in explaining shipments variations: (1) the distance effect is negative and highly significant, with bulkier products hauled over shorter distances; (2) the adjacency effect is significant, with neighboring states trading more with one another, even after accounting for distance; (3) the impact of the spatial structure is of a competitive nature in most cases; (4) the effects of imports and exports are significant for specific commodities; (5) wholesale activities at both origins and destinations are important facilitators of commodity flows; (6) except in one case (furniture), flows increase with product diversification; and (7) the role of intermediate and final demands for the commodities are clearly reflected by the selected employment, population and per-capita income variables. Importantly, the Box-Cox specification, with endogenous determination of the transformation parameters, is superior, in all cases, to the gravity-type log-linear specification used in past research.

This research could be extended in several directions. First, the US database could be expanded into a multi-year framework. Complete CF data are available for 1993 and 1997, and as are preliminary data for 2002. Capturing the effects of differential growth among regions on the pattern of commodity flows should then become possible. Second, new economic data and detailed input–output information, based on the national US input–output table, could be used to better capture final and intermediate demand effects. Third, expanding the Box–Cox specification with more than two transformation parameters would allow for a more precise estimation of elasticities. Finally, modeling the relationship between commodity flows and modes of transportation could be used for assessing the need and planning for improvements in transportation infrastructure.

Appendix A. Data sources

Commodity Flow Survey, 1993 on CD-ROM CD-CFS-93-1, US Department of Commerce, Bureau of the Census, Washington, DC.

County Business Patterns, 1992–1993 on CD-ROM, US Department of Commerce, Bureau of the Census, Washington, DC.

Economic Census, Census of Manufacturing, 1992 on CD-ROM CD-EC92-1i, US Department of Commerce, Bureau of the Census, Washington, DC.

Annual State Personal Income Database, US Department of Commerce, Bureau of Economic Analysis, Washington, DC. www.bea.doc.gov/bea/regional/spi.

References

[1] Bröcker J. Partial equilibrium theory of interregional trade and the gravity model. Papers of the Regional Science Association 1989;66:7–18.

[2] Fotheringham AS. A new set of spatial-interaction models: the theory of competing destinations. Environment and Planning A 1983;15:15–36.

[3] Guldmann J-M. Competing destinations and intervening opportunities interaction models of inter-city telecommunication flows. Papers in Regional Science 1999;78:179–94.

[4] Isard W. Interregional and regional input–output analysis: a model of a space-economy. The Review of Economics and Statistics 1951;33:157–69.

[5] Moses LN. The stability of interregional trading patterns and input–output analysis. American Economic Review 1955;45:803–26.

[6] Leontief W, Strout A. Multiregional input–output analysis. In: Barna T, editor. Structural dependence and economic development. New York: St. Martin's Press; 1963. p. 119–50.

[7] Samuelson PA. Spatial price equilibrium arid linear programming. American Economic Review 1952;42:283–303.

[8] Takayama T, Judge GG. Spatial equilibrium and quadratic programming. Journal of Farm Economics 1964;46:67–93.

[9] Paelinck JHP, Ancot JP, Kuiper JH. Formal spatial economic analysis. Aldershot, UK: Gower Publishing; 1983.

[10] Krugman PR. Increasing returns and economic geography. Journal of Political Economy 1991;99:483–99.

[11] Dixit AK, Stiglitz JE. Monopolistic competition and optimal product diversity. American Economic Review 1977;67:297–308.

[12] Brackman S, Garretsen H, Marrewijk CV. An introduction to geographical economics. New York and London: Cambridge University Press; 2001.

[13] Ricci LA. Economic geography and comparative advantage: agglomeration versus specialization. European Economic Review 1999;43:357–77.

[14] Reed WE. Areal interaction in India: commodity flows in the Bengal–Bihar industrial area. Research Papers Series, No. 110. Department of Geography. Chicago: The University of Chicago; 1967.

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

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

[17] Black WR. Interregional commodity flows: some experiments with the gravity model. Journal of Regional Science 1972;12:107–18.

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

[19] Ullman EL. American commodity flow. Seattle and London: University of Washington Press; 1967.

[20] Linneman H. An econometric study of international trade flows. Amsterdam: North-Holland Publishing Company; 1966.

[21] Frankel JA, Wei SJ. Regionalization of world trade and currencies: economics and politics. In: Frankel JA, editor. The regionalization of the world economy. Chicago: The University of Chicago Press; 1998. p. 189–219.

[22] McCallum J. National borders matter: Canada–US regional trade patterns. American Economic Review 1995;85:615–23.

[23] Helliwell JF. Do national borders matter for Quebec's trade? The Canadian Journal of Economics 1996;29:507–22.

[24] Brown WM, Anderson WP. Spatial markets and the potential for economic integration between Canadian and US regions. Papers in Regional Science 2002;81:99–120.

[25] Box GEP, Cox DR. An analysis of transformations. Journal of the Royal Statistics Society 1964;26:211–43.

H. Murat Celik is an Assistant Professor, Department of Civil Engineering and Department of City and Regional Planning, Izmir Institute of Technology, Izmir, Turkey. He earned Bachelors and Masters degrees in city planning from Mimar Sinan University, Istanbul, Turkey, a masters in urban transportation planning from The University of Kansas, Lawrence, and a Ph.D. in city and regional planning from The Ohio State University, Columbus. Dr. Celik's research interests focus on transportation planning, economic planning, quantitative modeling, energy planning, remote sensing, and urban GIS. He teaches several graduate and undergraduate courses in these areas. He has published in the *Journal of Transport Geography* and *Transportation Planning and Technology*.

Jean-Michel Guldmann is Professor, Department of City and Regional Planning, The Ohio State University, Columbus. He earned an M.S. in industrial and systems engineering from the Ecole des Mines, Nancy, France, and a Ph.D. in urban and regional planning from the Technion-Israel Institute of Technology, Haifa, Israel. Dr. Guldmann's current research focuses on air quality modeling, natural gas policy, telecommunications costs and demand, and urban models. He has co-authored Industrial Location and Air Quality Control: A Planning Approach (John Wiley and Sons, 1980) and has extensively published in such journals as *Socio-Economic Planning Sciences, Environment and Planning A and B, Geographical Analysis, Papers in Regional Science, Regional Science and Urban Economics, Operations Research, Management Science, Energy, and Resources* and *Energy*. Dr. Guldmann's professional affiliations include the Regional Science Association International, the American Economic Association, INFORMS, and the International Association of Energy Economists.