Computers, Environment and Urban Systems xxx (2012) xxx–xxx

Contents lists available at SciVerse ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/compenvurbsys

Modeling interregional commodity flows with incorporating network autocorrelation in spatial interaction models: An application of the US interstate commodity flows

Yongwan Chun [a,*], Hyun Kim [b,1], Changjoo Kim [c,2]

[a] School of Economic, Political and Policy Sciences, The University of Texas at Dallas, Richardson, TX 75080, United States
[b] Department of Geography, University of Tennessee, Knoxville, TN 37996, United States
[c] Department of Geography, University of Cincinnati, Cincinnati, OH 45221, United States

ARTICLE INFO

Article history:
Received 21 July 2010
Received in revised form 21 April 2012
Accepted 22 April 2012
Available online xxxx

Keywords:
Interregional commodity flows
Network autocorrelation
Spatial autoregressive models
Spatial interaction models

ABSTRACT

Spatial interaction models are frequently used to predict and explain interregional commodity flows. Studies suggest that the effects of spatial structure significantly influence spatial interaction models, often resulting in model misspecification. Competing destinations and intervening opportunities have been used to mitigate this issue. Some recent studies also show that the effects of spatial structure can be successfully modeled by incorporating network autocorrelation among flow data. The purpose of this paper is to investigate the existence of network autocorrelation among commodity origin–destination flow data and its effect on model estimation in spatial interaction models. This approach is demonstrated using commodity origin–destination flow data for 111 regions of the United States from the 2002 Commodity Flow Survey. The results empirically show how network autocorrelation affects modeling interregional flows and can be successfully captured in spatial autoregressive model specifications.

1. Introduction

Interregional commodity flows play an essential role in maintaining economic productivity and supporting regional economics. They also have a considerable influence over national and regional transportation systems, such as road management and traffic congestion. In 2007, the total value of interregional commodity flows was more than $11.6 trillion, and their total weight exceeded 12.5 billion tons (US Bureau of Transportation Statistics & CB., 2007). The sheer size justifies the need to accurately estimate commodity flows in general; thus, understanding the determinants of interregional commodity flows is a critical step in developing a comprehensive model for them.

The purpose of this paper is to investigate the interregional commodity flows in the US, considering autocorrelation structures among the flows in the context of spatial interaction modeling. To begin, studies show the usefulness of spatial interaction models for modeling interregional commodity flows (Ashtakala & Murthy, 1988; Black, 1972; Chisholm & O'Sullivan, 1973). Spatial interaction models for interregional flows improve when the spatial structure effect is taken into consideration (e.g., Brown & Anderson, 2002), which is commonly measured with an accessibility variable (Kwan, 1998; Thill & Kim, 2005). In particular, Celik and Guldmann (2007) demonstrate that incorporating competing destination effects (Fotheringham, 1983) and intervening opportunities (Stouffer, 1960) can further expand and improve the spatial interaction model. These are called either network variables or geographic separation variables. Roy and Thill (2004) provide an extensive review of spatial interaction modeling methods and approaches.

However, spatial interaction models often fail to explicitly incorporate dependencies among flows, referred to as network autocorrelation (Black, 1992). As discussed in the literature of spatial autocorrelation (e.g., Anselin, 1988; Cliff & Ord, 1981), the presence of network autocorrelation violates the independence assumption and may result in biased and inefficient parameter estimates in spatial modeling. As a result, recent studies, including Chun (2008), Fischer and Griffith (2008), Griffith (2009), and LeSage and Pace (2008), discuss how to accommodate network autocorrelation in spatial interaction models. They show that a successfully specified model for network autocorrelation produces unbiased parameter estimates and improves the model.

In this study, following Bröcker (1989), a spatial interaction model is first specified and estimated in the context of a linear regression of the 2002 Commodity Flow Survey (CFS) data (US BTS & CB, 2007). An extended model with competing destination

* Corresponding author. Tel.: +1 972 883 4719
E-mail addresses: ywchun@utdallas.edu (Y. Chun), hyunkim@utk.edu (H. Kim), changjoo.kim@uc.edu (C. Kim).
[1] Tel.: +1 865 974 602.
[2] Tel.: +1 513 556 3424.

0198-9715/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.002

effects and intervening opportunities is also examined based on Celik and Guldmann (2007). Next, a spatial regression approach is employed to account for network autocorrelation in interregional commodity flows and is critical for the following reasons. First, it is anticipated that spatial regression that incorporates network autocorrelation produces unbiased estimates. Second, spatial regression is subsequently useful to examine statistically significant interregional commodity flow determinants. Finally, a spatial regression approach tends to produce a better model fit than a linear regression, thanks to the corrections provided by network autocorrelation.

The remainder of this paper is organized as follows. Section 2 presents a relevant literature review. Section 3 then discusses the analysis method and data. Next, Section 4 presents the results of linear and spatial regression. Finally, Section 5 presents the conclusions and discussion.

2. Literature review

Spatial interaction models are commonly used to model interregional commodity flows. As discussed by Bröcker (1989), the interregional flows of trade can be modeled with spatial interaction models. In such models, the demand-and-supply side among areas with several parameters constrains spatial equilibrium. Bröcker's work helps clarify how inconsistency forms between a spatial trade theory and empirical analyses. Importantly, he shows that the form of spatial interaction models effectively explains empirical flows, in terms of classical spatial equilibrium models (Enke, 1951; Samuelson, 1952). In terms of less-smaller geographic scales, many efforts are made to estimate interregional flows in various fields, such as commodities, telecommunications, and migration, in a given set of spatial units using the framework of spatial interaction model formulation (Black, 1972; Fotheringham & O'Kelly, 1989; Guldmann, 1999). In the context of international flows, Anderson and van Wincoop (2004) discuss that bilateral trade resistance and all trading partners also influence spatial interactions in trade flows among countries. This highlights the importance of capturing price differentials, tariff structures, and political barriers in modeling international flows.

From the geographic perspective, spatial structure effects (or spatial configuration) among commodity regions have also been specified, mainly as a form of competing destinations (CDs) and intervening opportunities (IOs); these are expected to capture the potential estimation misspecifications in interacting flows, which enhance the model fitness to estimate interregional spatial interactions. Specifically, the CD model is an approach that incorporates spatial structure to correct under- or over-estimation which often manifests in spatial interaction models (Fotheringham, 1983). In general, an accessibility-type measurement among destinations is entered in the model as a treatment. In contrast, the IO approach aims to investigate a structural factor to reduce or increase the interactions between origins and destinations (Stouffer, 1960). For example, when large IOs are observed in regions between origin and destination, the opportunities of inter-regions tend to partially absorb interactions, and therefore become less than expected. Hence, IOs can be interpreted as a measure of the spatial structure among origins (Guldmann, 1999).

Recent work by Celik and Guldmann (2007) explained interregional commodity flows in the US using spatial interaction models. Their proposed spatial interaction model was specifically applied to the 1993 US Commodity Flow Survey (CFS). By following Bröcker's framework, they captured the commodity flow characteristics as a function of the variables origin, destination, and distance. They then introduced three variables that can reflect the effects of spatial structure: CD effects, IOs, and border effects. Additionally, they used an adjacency dummy variable in the model to specify the border effect. This variable was also viewed as a proxy of a spatial structure variable to capture a (highly anticipated) heavy interaction of commodity flows between regions that share a common boundary. Their model showed these three variables (competing destination effects, intervening opportunities, and border effects) to be statistically significant for describing interregional commodity flows in the US. They concluded that specifying spatial structure in the spatial interaction models improves the models' capability to explain interregional flows.

Furthermore, spatial autocorrelation is also used to capture the effect of spatial structure in flows. Novak, Hodgdon, Guo, and Aultman-Hall (2010) propose a freight generation model based on a spatial regression model. Specifically, they suggest that accounting for spatial autocorrelation in freight data improves a linear regression model. However, as their analyses were conducted on outbound flows from each origin without considering destinations, their models did not form a spatial interaction model. For example, distance-decay, which is an important element in spatial interaction, was absent in their model.

In other recent research, the effectiveness of embedding network autocorrelation in spatial interaction models has been shown in flow data modeling (Chun 2008; Chun & Griffith 2011; Fischer & Griffith 2008; LeSage & Pace 2008). Spatial dependency generally occurs based on its geographic proximity. Likewise, it is expected to be embedded in commodity flows, which are composed of actual freight shipments in industries such as mining, manufacturing, wholesale, and retail. This is due to commodity transactions being highly dependent on the closeness of the spatial units, specifically origins and destinations of the freight movements. This spatial dependency issue in the spatial interaction model has been identified since the 1970s (e.g., Curry, 1972; Griffith & Jones, 1980; Sheppard, Griffith, & Curry, 1976). LeSage and Pace (2008) discuss how to account for network autocorrelation in a spatial autoregressive model. Fischer and Griffith (2008) also show that incorporating network autocorrelation, with an application of knowledge transmission using patent citation data from the European Union, improves spatial interaction models. Chun (2008) presents that the network dependency structure can be specified based on spatial choice behavior in the context of migration. Specifically, he discusses how a network weight matrix is defined to reflect competing destination effects and internal IOs. Further, Chun and Griffith (2011) reveal that specifying network autocorrelation for panel data, where multiple time spans are involved, is also effective treatment to improve the model fitting and provide better parameter estimates.

Of critical concern in these efforts in the literature is capturing spatial arrangement or dependency in commodity flow patterns. These concerns have been reflected in the models using specific treatments. Likewise, selecting appropriate variables is highlighted to improve the model's applicability in empirical analysis. Overall, recent studies show that network autocorrelation has an influence on spatial interaction models.

3. Analysis framework

3.1. Modeling interregional commodity flows

In this paper, gravity type spatial interaction models are used to model interregional commodity flows in the United States. A simple gravity type spatial interaction model can be written as:

$$F_{ij} = k \cdot P_i^{\beta_O} \cdot P_j^{\beta_D} \cdot exp(\beta_{dist} \cdot d_{ij}), \quad i,j = 1, \ldots, n \tag{1}$$

where F_{ij} is a flow from origin i to destination j, P_i and P_j are population at i and j, respectively, d_{ij} is the distance between i

and j, and (β_O, β_D, β_{dist}, k) are parameters to be estimated. A linearized gravity model is expressed by taking the natural logarithm on the both sides of the equation with an error term, ε_{ij}:

$$ln(F_{ij}) = ln(k) + \beta_O \cdot ln(P_i) + \beta_D \cdot ln(P_j) + \beta_{dist} \cdot d_{ij} + \varepsilon_{ij} \quad (2)$$

The natural logarithm is generally used for a transformation of the dependent variable. However, in this study, the Box–Cox transformation is used to make the dependent variable at least close to a normal distribution. The Box–Cox is conducted with the equation below:

$$F_{ij}^{\lambda} = \begin{cases} \frac{F_{ij}^{\lambda}-1}{\lambda}, & \lambda \neq 0 \\ ln(F_{ij}) & \lambda = 0 \end{cases} \quad (3)$$

Hence, the natural logarithm is a special case of the Box–Cox transformation when $\lambda = 0$. The estimated λ is 0.0296 for the US interregional commodity flows. The independent variables are used in natural logarithm form. Such a model has been frequently estimated with linear regression, assuming the error term independently and identically follows a normal distribution. This gravity model can be further extended by introducing more independent variables, the details of which are provided in the next section.

The extended gravity type spatial interaction models are also estimated within a spatial regression framework. This paper uses spatial lag models. When unobserved latent variables such as regional price information lead to spatial dependence, a spatially lagged dependent variable can be used to account for the unobserved variables. The matrix form below expresses a spatial lag model specification to account for network autocorrelation:

$$(\mathbf{I} - \rho\mathbf{W})\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4a)$$

$$\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (4b)$$

where $\mathbf{I}$ is an identity matrix, $\mathbf{Y}$ is a dependent variable, $\mathbf{X}$ is a design matrix for independent variables, β is a vector of parameters, $\mathbf{W}$ is a network weight matrix, ρ is a parameter for network autocorrelation, and ε is a vector of errors. In conjunction with Eqs. (2) and (3), $\mathbf{Y}$ is expressed as a vector of F_{ij}^{λ} s with $n^2 \times 1$ dimension given n regions. That is, $\mathbf{Y} = (F_{11}^{\lambda}F_{12}^{\lambda}, \cdots, F_{1n}^{\lambda}, F_{21}^{\lambda}, F_{22}^{\lambda}, \cdots, F_{2n}^{\lambda}, \cdots, F_{n1}^{\lambda}, F_{n2}^{\lambda}, \cdots, F_{nn}^{\lambda})$. Similarly, $\mathbf{X}$ has $n^2 \times k$ dimension, where k is the number of independent variables and ε has $n^2 \times 1$ dimension. It is assumed that ε follows a normal distribution, $N(0, \sigma^2\mathbf{I})$. The network weight matrix, $\mathbf{W}$, is correspondingly $n^2 \times n^2$ dimensional. The structure of a network weight matrix is discussed later in this section.

Additionally, the model specification is further extended considering large internal flows within a region. In a flow dataset, internal flows within a region generally dominate interregional flows. That is, among n^2 flows, n internal flows have much larger values than interregional flows. While a dummy variable is often used to mitigate this issue, LeSage and Fischer (2010) suggest a procedure to modify independent variables by replacing values of independent variables for internal flows with zero values. These modified independent variables are denoted as $\mathbf{X}_{inter}$. Then, the model introduces additional independent variables containing non-zero values for n internal flows from the original independent variables and zero values for the other n^2–n flows, which are denoted as $\mathbf{X}_{intra}$. This modification is also similarly applied to intercept. The spatial lag model with this modification becomes

$$\mathbf{Y} = \rho\mathbf{W}\mathbf{Y} + \mathbf{X}_{inter}\boldsymbol{\beta}_{inter} + \mathbf{X}_{intra}\boldsymbol{\beta}_{intra} + \boldsymbol{\varepsilon} \quad (5)$$

Here, the model anticipates that $\mathbf{X}_{intra}$ accounts for internal flows and, hence, interregional flows are modeled avoiding the impact of dominating internal flows.

Accordingly, linear regression and spatial lag models estimate US interregional freight flows, with and without consideration of large internal flows. The linear regression is estimated with the ordinary least square method, which is the same as maximum likelihood estimation (MLE) in linear regression. The spatial lag models are estimated with MLE, which is available in *spdep* package in R (see Bivand, Pebesma, & Gómez-Rubio, 2008).

3.2. Network weight matrix specifications

A network weight matrix reflects a dependence structure among flows. Chun (2008), Fischer and Griffith (2008), and LeSage and Pace (2008) basically present four different specifications of network dependence structure. The first specification is based on the equation below:

$$b_{ij,kl}^N = \begin{cases} 1, & \text{if } i = k \text{ and } b_{jl}^S = \mathbf{1} \\ \mathbf{0}, & \text{otherwise} \end{cases} \quad (6)$$

where $b_{ij,kl}^N$ is an element of a binary type weight matrix, $\mathbf{B}^N$, with $n^2 \times n^2$ dimension and b_{jl}^S is an element of spatial weight matrix, $\mathbf{B}^S$, with $n \times n$ dimension. The non-zero value of $b_{ij,kl}^N$ indicates F_{ij} and F_{kl} are network neighbors in the specification. That is, flows that have a same origin and spatially neighbored destinations are considered network neighbors. LeSage and Pace (2008) discuss this specification to capture spatial autocorrelation among destinations. Chun (2008) describes this specification to capture competing destination effects.

The second specification is defined with same destination and spatially neighbored origins. Each element of a network matrix has an assigned value as below:

$$b_{ij,kl}^N = \begin{cases} 1, & \text{if } j = l \text{ and } b_{ik}^S = \mathbf{1} \\ \mathbf{0}, & \text{otherwise} \end{cases} \quad (7)$$

This can be interpreted as a network dependence structure to reflect spatial autocorrelation among origins. Chun (2008) argues that this dependence structure can be understood to capture intervening opportunity effects. The third specification can be defined as a combination of the first and second specifications. That is, the addition of two network weight matrices based on Eqs. (6) and (7) constructs another network weight matrix to reflect origin- and destination-based autocorrelation structures together. This can be expressed as

$$b_{ij,kl}^N = \begin{cases} 1, & \text{if } i = k \text{ and } b_{jl}^S = 1, or \text{ if } j = l \text{ and } b_{ik}^S = \mathbf{1} \\ \mathbf{0}, & \text{otherwise} \end{cases} \quad (8)$$

Under the fourth specification, flows from spatially neighbored origins to spatially neighbored destinations are considered as network neighbors to each other. It can be expressed as

$$b_{ij,kl}^N = \begin{cases} 1, & \text{if } b_{ik}^S = 1 \text{ and } b_{jl}^S = 1 \\ 0, & \text{otherwise} \end{cases} \quad (9)$$

Using Kronecker product and Kronecker sum can easily create a network weight matrix based on the above four specifications. Chun and Griffith (2011) show that corresponding to the order of the above four specifications, a binary type network matrix can be defined as

$$\mathbf{B}^N = \mathbf{B}^S \otimes \mathbf{I} \quad (10a)$$

$$\mathbf{B}^N = \mathbf{I} \otimes \mathbf{B}^S \quad (10b)$$

$$\mathbf{B}^N = \mathbf{B}^S \otimes \mathbf{B}^S = \mathbf{B}^S \otimes \mathbf{I} + \mathbf{I} \otimes \mathbf{B}^S \quad (10c)$$

$$\mathbf{B}^N = \mathbf{B}^S \otimes \mathbf{B}^S \quad (10d)$$

where $\otimes$ and $\oplus$ denote Kronecker product and Kronecker sum respectively, and $\mathbf{I}$ is an $n \times n$ identity matrix.

In this paper, network neighbors used for measurement are defined with Eqs. (8), (10c). The network weight matrix, $\mathbf{W}$ in Eqs. (4a) and (4b), is prepared through the row-standardization of $\mathbf{B}^N$. Fischer and Griffith (2008) discuss the utility of this network neighbor specification by showing the improvement of their spatial interaction models for patent citation. Chun (2008) also shows improvement of interstate migration models with this network neighbor specification.

3.3. The effects of spatial structure in spatial interaction models

Different forms in spatial interaction modeling show the effects of spatial structure. Conventionally, distance decay is used as a primary variable to reflect a geographical effect. This paper uses an exponential type of distance decay (see Fotheringham and O'Kelly (1989) for details). In addition to distance decay effect, the effects of spatial structure have been introduced following two modeling frameworks: competing destination and intervening opportunity. Previous research (for example, Almeida & Goncalves, 2001; Celik & Guldmann, 2007; Fik & Mulligan, 1990; Raphael, 1998; Roy, 1993) shows the improvement of spatial interaction models incorporating CD and IO effects.

These two effects are commonly specified with accessibility type variables. For example, Celik and Guldmann (2007) include two variables in their spatial interaction models. First, a variable to reflect CD effects is defined as

$$CD_{ij} = \sum_{k} TE_k / d_{kj}, \quad k \neq (i,j) \tag{11}$$

where TE_k is total employment at k and d_{kj} is the distance between k and j. Considering the spatial distribution of destinations, the CD variable is calculated with other potential destinations among the destination j. While a positive sign indicates an agglomeration effect, a negative sign implies competition among destinations (Fotheringham, 1983). Second, another variable is formulated to capture IOs. As the IOs between an origin and a destination dyad increase, flows between the pair decrease (Stouffer, 1960). Considering IOs on the path from an origin to a destination, the IO variable is calculated with potential opportunities around the origin. The IO variable is calculated as

$$IO_{ij} = \sum_{k} TE_k / d_{ik}, \quad k \neq (i,j) \tag{12}$$

Hence, this effect reflects spatial configuration around origins, similar to destination cases (Guldmann, 1999). Celik and Guldmann (2007) found that these two variables are highly significant in their models.

4. An analysis of the US interstate commodity flows

4.1. Data

As mentioned in the previous section, a major concern in the freight analysis is the construction of a set of variables that appropriately characterize origin, destination, and geographical separation among regions. In this paper, we used the 2002 Commodity Flow Survey (CFS) data to construct interregional commodity flow models. The Census Bureau and the Bureau of Transportation Statistics of the United States determines and publishes this data, and more importantly, the data include the Origin–Destination (hereafter O–D) data of commodity flows in the United States in terms of Freight Analysis Framework (FAF) zones. As illustrated in Fig. 1, the FAF regions are basically delineated based on

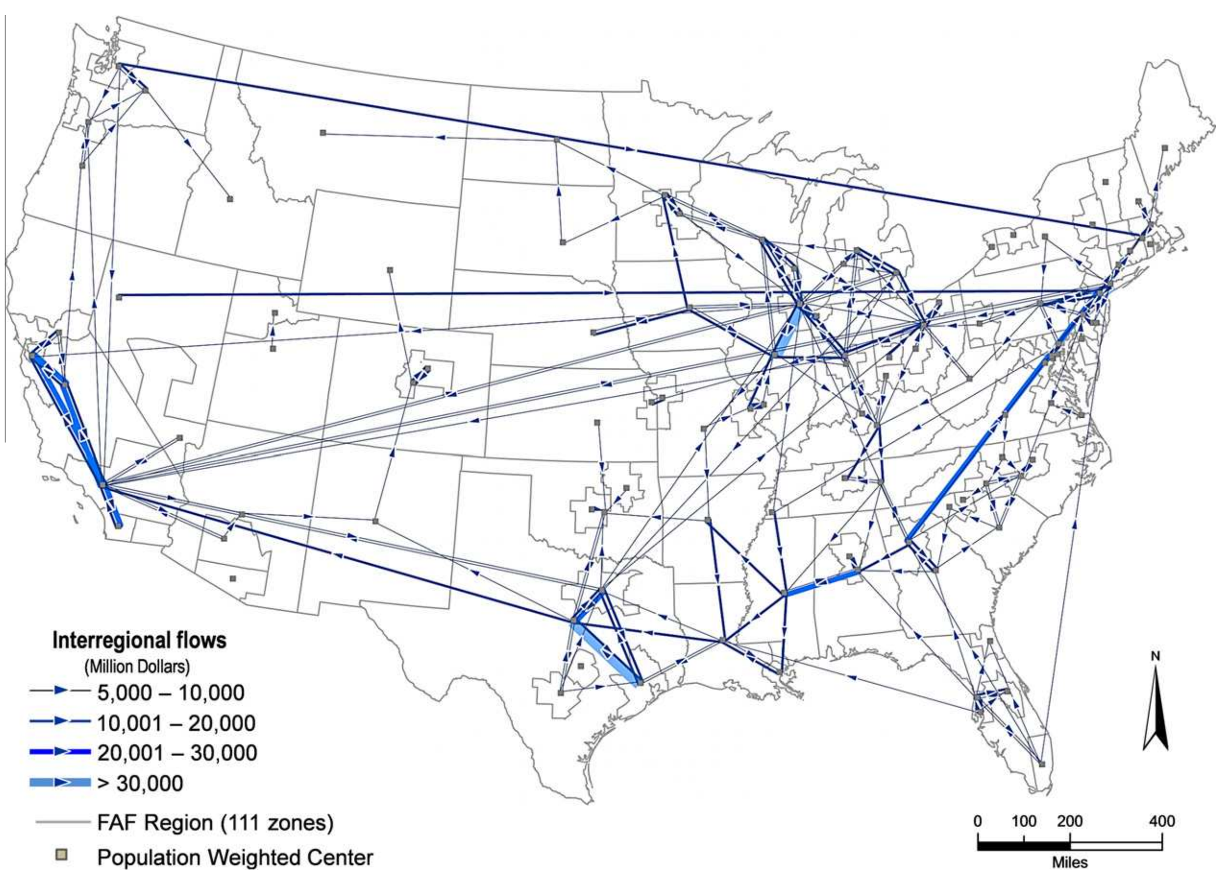

Fig. 1. 2002 CFS interregional flows with more than 5 billion dollars among 111 FAF regions.

Please cite this article in press as: Chun, Y., et al. Modeling interregional commodity flows with incorporating network autocorrelation in spatial interaction models: An application of the US interstate commodity flows. *Computers, Environment and Urban Systems* (2012), http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.002

Metropolitan Statistical Areas (MSAs), Consolidated Statistical Areas (CSAs), or balances of States. Our models consider 111 FAF regions which cover the contiguous United States. In other words, our models do not involve other FAF regions outside of the inland United States in the analysis because dependency among O–D flows can be specified in the model when they are geographically structured in a contiguous space and specified as a spatial weights matrix to measure network autocorrelation. Note that we considered all types of commodity flows to construct the O–D flows among FAF regions for a generalized model. Although it is difficult to display the amount of interregional flows for 111 FAF regions, Fig. 1 displays the dominant interregional flows with more than 5 billion dollars among the 111 FAF regions. For effective flow pattern visualization, flows are treated as directional line symbols, which connect the population weighted centroid of each region.

This paper specifies spatial interaction models to accommodate underlying spatial structure effects as well as exogenous variables. To do this, the explanatory variables of origin, destination, and geographical separation are prepared, along with the framework of Bröcker (1989) and Celik and Guldmann (2007). First, the 2000 US Census provides the population size, whose county-level population is aggregated according to the 111 FAF regions. Second, this paper uses the 2002 Economic Census by US Census Bureau to construct variables that reflect regional economic status, which includes the number of employees, average production value, average plant size, and manufacturing. Third, the model includes family income per capita to represent a proxy for purchase demand using data from the 2002 Bureau of Economic Analysis (BEA), which is used as a benchmark of each region's economic status. Finally, the model calculates a population weighted centroid of each region to measure the geographic distances among the 111 regions. In detail, the geographic distances are measured as the spherical distance between a dyad of population weighted centroids.

In modeling commodity flows among interregional regions using spatial interaction models, it is primarily important to specify variables for the demand and supply sides. The origin variables include income per capita (*oinc*), the number of employees (*oemp*), average value of production (*oprod*), and average plant size (*oplant*). Two variables among them, *oemp* and *oprod*, are employed to represent production at the origin. While the income variable, *oinc*, is considered a proxy for demand conditions at the origin, the income level also reflects origin economic conditions. The average plant size, *oplant*, is expected to reflect the economy scale in terms of production. The three destination variables, population (*dpop*), income per capita (*dinc*), and manufacturing (*dmanuf*), represent demands at each destination. While *dpop* and *dinc* are considered final demand, *dmanuf* is a proxy for intermediate demand.

5. Results

Four different linear regression models are specified with the adjustment of large intra flows and two of the above geographical variables (i.e., CD and IO). The base linear regression model (LM-1u) contains only origin and destination variables, as well as a distance-decay effect variable. Adjusting large intra flows, as specified in Eq. (5) (LM-1a), has further extended this model. Then, the CD and IO variables, which Eqs. (11) and (12) prepare, are introduced additionally for the base linear models, labeled as LM-2u and LM-2a, respectively. These four models are estimated in spatial lag model specifications, called similarly SLag-1u, SLag-1a, SLag-2u, and SLag-2a.

Table 1 reports the results of the four linear regression models. First, the inclusion of the two geographic variables CD and IO is statistically significant. The estimates for these variables are significant at the 1% level in LM-2 models (their p-values <0.001). Adding the two variables also improved the model fit. The AIC values of LM-2 models decreased from those of LM-1 models, decreasing from 43,731.66 to 43,159.61 for the model without the adjustment of intra flows and 42,180.73–41,661.67 for the model with the adjustment. Adding these variables also triggers the change of statistical significance for variables. While *oinc* and *dinc* variables are not significant in LM-1 models, they are significant at the 1% level in LM-2 models. In contrast, the *oprod* variable is significant in LM-1 models at the 1% level but not in LM-2 models.

Table 1
The results of linear regression models (OLS).

	Base model				CD and IO model			
	Unadjusted (LM-1u)		Adjusted (LM-1a)		Unadjusted (LM-2u)		Adjusted (LM-2a)	
	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. error*
intercept	−29.0420	1.3825***	−29.8409	1.3076***	−22.3749	1.3823***	−23.8293	1.3104***
intra_intercept	–	–	2.7627	12.4002	–	–	4.3228	12.1410***
ln(*oinc*)	0.1067	0.1262	0.1524	0.1190	0.8024	0.1305***	0.7784	0.1232***
ln(*oemp*)	1.3430	0.0191***	1.3551	0.0180***	1.2961	0.0188***	1.3126	0.0178***
ln(*oprod*)	0.0899	0.0172***	0.0682	0.0162***	0.0009	0.0179	−0.0104	0.0170
ln(*oplant*)	−0.2963	0.0920**	−0.1538	0.0868	0.5534	0.1061***	0.6015	0.1001***
ln(*dpop*)	0.3862	0.0325***	0.3576	0.0307***	0.2523	0.0328***	0.2404	0.0310***
ln(*dinc*)	−0.0190	0.0810	−0.0220	0.0764	0.5265	0.0843***	0.4663	0.0797***
ln(*dmanuf*)	0.8971	0.0269***	0.9313	0.0254***	0.9827	0.0269***	1.0054	0.0254***
ln(*intra_inc*)	–	–	−1.1514	1.4186	–	–	0.3241	1.3904
ln(*intra_emp*)	–	–	2.5197	1.4295	–	–	2.1603	1.3997
ln(*intra_prod*)	–	–	0.1057	0.1739	–	–	−0.0976	0.1705
ln(*intra_plant*)	–	–	−1.3185	0.9568	–	–	0.4303	0.9398
ln(*intra_pop*)	–	–	−1.0506	1.1597	–	–	−0.8945	1.1354
ln(*intra_manuf*)	–	–	0.2522	0.3267	–	–	0.3561	0.3199
dist	−0.0015	0.0000***	−0.0014	0.0000***	−0.0017	0.0000***	−0.0016	0.0000***
CD	–	–	–	–	−0.7134	0.0382***	−0.6374	0.0361***
IO	–	–	–	–	−0.7575	0.0440***	−0.6850	0.0415***
z-Score of Moran's I (p value)	97.36 (0.0000)		82.40 (0.0000)		92.25 (0.0000)		78.31 (0.0000)	
AIC	43731.66		42180.73		43159.61		41661.67	

Significance codes:
* 0.05.
** 0.01.
*** 0.001.

Table 2
The results of spatial lag models.

	Base model				CD and IO model			
	Unadjusted (SLag-1u)		Adjusted (SLag-1a)		Unadjusted (SLag-2u)		Adjusted (SLag-2a)	
	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. error*	Coeff.	*Std. error*
intercept	−19.4079	1.0405***	−20.6682	2.6349***	−16.1960	1.1654***	−17.7440	1.0480***
intra_intercept	–		3.7747	21.6137	–		4.5549	3.0949
ln(*oinc*)	0.1236	0.0734	0.1554	0.1328	0.4501	0.1082***	0.4601	0.1008***
ln(*oemp*)	0.8302	0.0171***	0.8721	0.0187***	0.8233	0.0165***	0.8683	0.0161***
ln(*oprod*)	−0.0037	0.0075	−0.0124	0.0080	−0.0416	0.0160**	−0.0468	0.0124***
ln(*oplant*)	−0.1495	0.0576**	−0.0546	0.1275	0.2404	0.0964*	0.3048	0.0752***
ln(*dpop*)	0.3833	0.0266***	0.3624	0.0288***	0.3094	0.0265***	0.2957	0.0256***
ln(*dinc*)	−0.1315	0.0802	−0.1206	0.1452	0.1682	0.0669**	0.1555	0.0671*
ln(*dmanuf*)	0.4188	0.0226***	0.4750	0.0241***	0.4815	0.0228***	0.5343	0.0221***
ln(*intra_inc*)	–	–	−0.9492	2.0246	–	–	−0.1838	0.2942
ln(*intra_emp*)	–	–	1.5934	1.0314	–	–	1.4388	0.6422*
ln(*intra_prod*)	–	–	0.0062	0.1638	–	–	−0.0967	0.1000
ln(*intra_plant*)	–	–	−1.5886	0.7193*	–	–	−0.6630	0.6570
ln(*intra_pop*)	–	–	−0.4310	1.1975	–	–	−0.3718	0.5630
ln(*intra_manuf*)	–	–	−0.0345	0.2872	–	–	0.0303	0.0670
dist	−0.0004	0.0000***	−0.0004	0.0000***	−0.0006	0.0000***	−0.0006	0.0000***
CD	–	–	–	–	−0.3877	0.0312***	−0.3559	0.0306***
IO	–	–	–	–	−0.3593	0.0382***	−0.3366	0.0338***
rho (*p* value)	0.7206 (0.0000)		0.6726 (0.0000)		0.6983 (0.0000)		0.6482 (0.0000)	
AIC	39200.2		38119.51		38976.44		37932.67	

Significance codes:
* 0.05.
** 0.01.
*** 0.001.

Second, adjusting large intra flows improved the model fit and caused the change of statistical significance. The LM models with the adjustment have smaller AIC values than their counterpart models without the adjustment. This triggered the change of statistical significance of the *oplant* variable in LM-1 models. This *oplant* variable is significant in LM-1u at the 1% level but becomes insignificant in LM-1a. The estimates for the distance decay parameter find one significant change from this adjustment. Although their estimates are not much different (i.e., from −0.0015 to −0.0014 in LM-1 models and from −0.0017 to −0.0016 in LM-2 models), their 99% confidence intervals do not overlap with each other; the 99% confidence intervals are (−0.00156, −0.00145) for LM-1u, (−0.00142, −0.00132) for LM-1a, (−0.00178, −0.00166) for LM-2u, and (−0.00163, −0.00152) for LM-2a. Although all of the independent variables corresponding to only intra flows are not statistically significant in both LM-1a and LM-2a models, it may not be of primary interest, as LeSage and Fischer (2010) point out.

Third, the linear regression models have significant positive network autocorrelation. The *z*-scores of Moran's *I* for LM-1u, LM-1a, LM-2u, and LM-2a are very high: 97.36, 82.40, 92.25, and 78.31, respectively. These Moran's *I* values were calculated considering the estimation effect by the independent variables. Their high *z*-scores indicate that the residuals have strong positive network autocorrelation. The smaller Moran's *I* values of LM-2u and LM-2a, respectively, compared to those of LM-1u and LM-1a may imply that CD and IO are able to capture a part of network autocorrelation. Also, the adjustment of large intra flows noticeably reduced the network autocorrelation level. LM-1a and LM-2a have smaller *z*-scores for Moran's *I* than LM-1u and LM-2u, respectively. Their high *z*-scores of Moran's *I* indicate that network autocorrelation needs to be incorporated in their model specifications.

Table 3
Likelihood ratio test to SLag-2a model.

	Log likelihood	*df*	LR stat	*p* Value
LM-1u	−21855.83	10	5818.993	0.0000
LM-1a	−21073.37	3	4254.064	0.0000
LM-2u	−21567.80	8	5242.939	0.0000
LM-2a	−20811.83	1	3731.002	0.0000
SLag-1u	−19589.10	9	1285.537	0.0000
SLag-1a	−19041.75	2	190.840	0.0000
SLag-2u	−19475.22	7	1057.774	0.0000
SLag-2a	−18946.33	–	–	–

Table 2 reports the results of the spatial lag models to incorporate network autocorrelation. The spatial lag models are preferred over the linear regression models with their smaller AIC values. Incorporating network autocorrelation dramatically decreased the AIC values from those of their counterpart linear regression models. Among the four spatial lag models, SLag-2a has the best model fit with the smallest AIC value, and SLag-1u has the largest AIC values. Table 3 reports the results of the Likelihood ratio test between each model and SLag-2a which has the smallest AIC value. The results statistically confirm that SLag-2a has a better model fit than the other seven models and, hence, is preferred among the eight models.

The interpretation of the spatial lag models has been done with estimated impacts rather than the estimated coefficients, as suggested by LeSage and Pace (2009). Tables 4 and 5 report the estimated direct, indirect, and total impacts for the spatial lag models. Regarding the statistical significance, each variable has a same statistical significance level for direct, indirect, and total impacts. In the results of the SLag-1u model, while *oemp*, *oplant*, *dpop*, *dmanuf*, and *dist* variables are significant at the 1% level for all of direct, indirect, and total impacts, the other variables are not significant for any of those impacts. This significance pattern is consistent for the impact estimates of all of the spatial lag models.

Compared with the results of the linear regression models, the spatial lag models produce different statistical inferences for two variables. While the *oprod* variable is statistically significant with a positive sign at the 1% level in LM-1 models, its impact estimates

Table 4
Impact estimates for the base models.

	Unadjusted (SLag-1u)						Adjusted (SLag-1a)					
	Direct effects		Indirect effects		Total effects		Direct effects		Indirect effects		Total effects	
	Mean estimate	*z*-Values	Mean estimate	*z*-Values	Mean estimate	*z*-Values	Mean estimate	*z*-Values	Mean estimate	*z*-Values	Mean estimate	*z*-Values
ln(*oinc*)	0.1347	1.7173	0.3076	1.7177	0.4422	1.7187	0.1668	1.1485	0.3078	1.1456	0.4746	1.1470
ln(*oemp*)	0.9039	51.1239***	2.0668	24.9110***	2.9708	33.4988***	0.9365	44.8845***	1.7275	24.8354***	2.6640	33.8190***
ln(*oprod*)	−0.0041	−0.5038	−0.0093	−0.5044	−0.0134	−0.5044	−0.0133	−1.5875	−0.0245	−1.5845	−0.0377	−1.5864
ln(*oplant*)	−0.1624	−2.6104**	−0.3722	−2.5865**	−0.5350	−2.5972**	−0.0586	−0.4521	−0.1081	−0.4535	−0.1666	−0.4531
ln(*dpop*)	0.4175	14.7073***	0.9543	12.4100***	1.3716	13.3909***	0.3892	12.4121***	0.7179	10.8181***	1.1071	11.5710***
ln(*dinc*)	−0.1417	−1.6112	−0.3275	−1.6051	−0.4707	−1.6080	−0.1295	−0.8358	−0.2388	−0.8338	−0.3683	−0.8347
ln(*dmanuf*)	0.4559	19.1743***	1.0425	16.9099***	1.4984	18.5221***	0.5100	19.8294***	0.9408	17.9562***	1.4509	19.7827***
ln(*intra_inc*)	–	–	–	–	–	–	−1.0193	−0.4124	−1.8802	−0.4086	−2.8995	−0.4100
ln(*intra_emp*)	–	–	–	–	–	–	1.7109	1.4881	3.1561	1.4904	4.8670	1.4905
ln(*intra_prod*)	–	–	–	–	–	–	0.0067	−0.0057	0.0123	−0.0089	0.0190	−0.0078
ln(*intra_plant*)	–	–	–	–	–	–	−1.7058	−2.0978	−3.1467	−2.0901*	−4.8525	−2.0948*
ln(*intra_pop*)	–	–	–	–	–	–	−0.4628	−0.3100	−0.8537	−0.3060	−1.3166	−0.3074
ln(*intra_manuf*)	–	–	–	–	–	–	−0.0371	−0.1530	−0.0683	−0.1556	−0.1054	−0.1547
dist	−0.0005	−21.0581***	−0.0011	−24.2060***	−0.0015	−25.5781***	−0.0004	−19.2956***	−0.0008	−22.3332***	−0.0013	−22.9887***

Significance codes:
* 0.05.
** 0.01.
*** 0.001.

Table 5
Impact estimates for CD and IO models.

	Unadjusted (SLag-2u)						Adjusted (SLag-2a)					
	Direct effects		Indirect effects		Total effects		Direct effects		Indirect effects		Total effects	
	Mean estimate	*z*-Values	Mean estimate	*z*-Values	Mean estimate	*z*-Values	Mean estimate	*z*-Values	Mean estimate	*z*-Values	Mean estimate	*z*-Values
ln(*oinc*)	0.4867	4.3377***	1.0050	4.2573***	1.4917	4.2982***	0.4910	4.6099***	0.8168	4.5871***	1.3078	4.6121***
ln(*oemp*)	0.8904	50.6209***	1.8383	24.4513***	2.7287	33.3816***	0.9265	55.4143***	1.5415	26.9588***	2.4680	39.4768***
ln(*oprod*)	−0.0450	−2.6555**	−0.0930	−2.6190**	−0.1380	−2.6345**	−0.0499	−3.7226***	−0.0830	−3.6824***	−0.1330	−3.7066***
ln(*oplant*)	0.2599	2.6090**	0.5367	2.6119**	0.7966	2.6147**	0.3252	4.0523***	0.5411	4.0424***	0.8663	4.0581***
ln(*dpop*)	0.3346	11.8915***	0.6909	10.4074***	1.0255	11.0641***	0.3155	11.5993***	0.5249	10.4371***	0.8404	11.0584***
ln(*dinc*)	0.1818	2.5700*	0.3755	2.5635*	0.5573	2.5690*	0.1659	2.3427*	0.2760	2.3424*	0.4418	2.3449*
ln(*dmanuf*)	0.5208	21.4590***	1.0752	18.9041***	1.5959	21.1405***	0.5702	25.1489***	0.9485	21.0200***	1.5187	24.4741***
ln(*intra_inc*)	–	–	–	–	–	–	−0.1962	−0.6152	−0.3263	−0.6143	−0.5225	−0.6148
ln(*intra_emp*)	–	–	–	–	–	–	1.5352	2.2502*	2.5541	2.2477*	4.0894	2.2509*
ln(*intra_prod*)	–	–	–	–	–	–	−0.1032	−0.9443	−0.1716	−0.9410	−0.2748	−0.9425
ln(*intra_plant*)	–	–	–	–	–	–	−0.7074	−1.0030	−1.1769	−0.9980	−1.8843	−1.0002
ln(*intra_pop*)	–	–	–	–	–	–	−0.3968	−0.6532	−0.6601	−0.6540	−1.0568	−0.6538
ln(*intra_manuf*)	–	–	–	–	–	–	0.0324	0.4519	0.0538	0.4533	0.0862	0.4528
dist	−0.0006	−25.5832***	−0.0013	−27.0043***	−0.0019	−30.6152***	−0.0006	−25.6742***	−0.0010	−27.7361***	−0.0016	−30.9336***
CD	−0.4193	−12.5181***	−0.8657	−11.4292***	−1.2850	−12.0554***	−0.3798	−12.0263***	−0.6319	−11.2549***	−1.0117	−11.7923***
IO	−0.3885	−9.5557***	−0.8022	−9.3906***	−1.1907	−9.6010***	−0.3592	−9.8447***	−0.5975	−9.5032***	−0.9567	−9.7825***

Significance codes:
* 0.05.
** 0.01.
*** 0.001.

Please cite this article in press as: Chun, Y., et al. Modeling interregional commodity flows with incorporating network autocorrelation in spatial interaction models: An application of the US interstate commodity flows. *Computers, Environment and Urban Systems* (2012), http://dx.doi.org/10.1016/j.compenvurbsys.2012.04.002

have a negative sign in SLag-2 models and are significant at the same level. This *oprod* variable is not significant in LM-2 and SLag-1 models. That is, considering both network autocorrelation and the geographical variables CD and IO, the sign of its estimate became the opposite. When only one of them is considered, its estimate is not significant. However, the destination income variable, *dinc*, has a different significance level. While, the coefficient of *dinc* is significant at the 1% level in LM-1 models, its impact estimates in SLag-2 models are not significant at the 1% level. Nevertheless, these impact estimates are still significant at the 5% level with a same positive sign. Although the variables correspond only to intra flows and are not a primary interest, *intra_plant* and *intra_emp* variables are significant at the 5% level in SLag-1a and SLag-2a models, respectively.

The influence of CD and IO variables can also be observed from the impact estimates of the spatial lag models, as with the results of the linear regression models. Significantly, the addition of CD and IO variables increases the magnitude of the distance-decay effect toward to the negative direction. While the total impact estimates for the distance-decay are −0.0015 and −0.0013 in SLag-1 models, the total impacts estimates in SLag-2 models are −0.0019 and −0.0016. The magnitude increase is statistically significant with their 99% confidence intervals not overlapping with each other: their 99% confidence intervals are (−0.00169, −0.00138) for SLag-1u, (−0.00140, −0.00112) for SLag-1a, (−0.00207, −0.00175) for SLag-2u, and (−0.00171, −0.00145) for SLag-2a. It is worth noting that these total impact estimates for the spatial lag models' distance-decay are not significantly different from the coefficients of their corresponding variables in the linear regression models' results. For the CD variable, the 99% confidence intervals are (−0.8118, −0.6150) for LM-2u, (−0.7302, −0.5445) for LM-2a, (−1.5610, −1.0091) for SLag-2u, and (−1.2320, −0.7913) for SLag-2a. For the IO variable, the 99% confidence intervals are (−0.8709, −0.6440) for LM-2u, (−0.7918, −0.5781) for LM-2a, (−1.5106, −0.8708) for SLag-2u, and (−1.2088, −0.7046) for SLag-2a.

Based on SLag-2a with the best model fit, the independent variables at origins and destinations are significant at the 1% level, except the *dinc* variable, which is significant at the 5% level. Only *oprod* has a negative relationship to interregional flows among the origin variables. Three destination variables also show a positive relationship. It appears that we have observed the distance decay effect. With negative signs for CD and IO, competing effects among destinations appear, and IOs negatively impact interregional flows. The independent variables also have significant indirect (or neighborhood) impacts on the interregional commodity flows. Hence, the interregional flows can be described as follows: (1) the distance-decay is significantly negative, meaning that the shorter the distance, the more the commodity flows; (2) competitive effects are present among destinations and origins; (3) the *oinc*, *oemp*, and *oplant* variables, which can represent the economic size at origins, are positive and significant. However, the negative sign of *oprod* may need further investigation; finally, (4) the three variables representing demands at destinations are significant with a positive sign.

The results have concerning implications in commodity flow analysis, the structure of commodity. The model can incorporate the structure of commodity flow patterns. Structural characteristic reflects the tendency for sets of origins, destinations, or dyads to have either similar or different commodity movement characteristics. Moreover, this spatial structure reveals significant interstate or regional connections. Although this approach is applied to commodity flows, it can also be applied to passenger flows. In addition, the proposed model suggests that the model's results can be used to diagnose commodity volume flows over individual routes to identify outliers.

The existence of network autocorrelation on interstate flows shows that an origin–destination cost for each pair of places is geographically dependent. The results also imply that network autocorrelation affects the specification and solutions of spatial interaction models with flow dependent costs. One weakness of the spatial interaction model is that routes (costs) for each pair of places are not necessarily the same throughout the network. This implies that forecasting spatial interactions is less reliable when interdependence among network segments exists. Results support that segment-specific considerations are important in interregional commodity flow models.

Although different model specification results also support that models estimated by different approaches are not necessarily due to any fundamental difference in principle, several variables (origin employees, destination population and manufacturing, distance, competing effects and intervening opportunities) powerfully explain variations in interregional commodity flows. First, as the number of employees at the origin increases, the interregional commodity outflows from these nodes also increase. Second, as population and the number of manufacturing at the destination increases, the interregional commodity inflows to these nodes also increase. Third, distance working as it relates to transportation costs is a deterrent to interregional commodity trade. Last, competing destination effects and intervening opportunities play important roles in interregional commodity flows. Specifically, as other destinations are geographically clustered to a specific destination, the commodity flows to destinations decrease. Similarly, destination nodes clustered around the origin serve as alternative destinations. These results from interregional commodity flow are consistent with findings of other spatial interaction applications, and thus imply that the ability to obtain better origin and destination information about supplies and demands are important factors in explaining the interstate commodity flows.

6. Summary and conclusions

This paper analyzed interregional commodity flows with spatial interaction models, using the 2002 CFS commodity flows among 111 regions in the contiguous United States. In spatial interaction models, an exponential type of distance decay was configured, as were origin and destination variables. In addition, two geographical variables were considered in order to reflect the effect of spatial structure: competing destination (CD) effects and intervening opportunities (IOs). These spatial interaction models were estimated in linear regression and spatial lag model specifications to incorporate network autocorrelation.

The Moran's *I* tests for the residuals of the linear regression models report that the network autocorrelation in the linear regression models is highly significant. Thus, the estimates of the linear regression may be statistically biased. The estimates for the network autocorrelation parameter, ρ, in the spatial lag models have high magnitudes ($\rho \geqslant 0.6482$). The magnitudes of ρ parameter estimates reflect the level of network autocorrelation in the linear models. SLag-1u, which corresponds to LM-1u with the largest *z*-score of Moran's *I*, has the largest ρ estimate, and SLag-2a, which corresponds to LM-2a with the smallest *z*-score, has the smallest ρ estimate.

The geographical variables are statistically significant and adding the variables increased model fits. When the geographical variables are added in model specifications, the *z*-scores of Moran's *I* are smaller in LM models, and the ρ estimates are smaller in the spatial lag models. This indicates that the geographical variables partially account for network autocorrelation. This may also imply that the magnitudes of the geographic variables could be overestimated (or underestimated) unless network autocorrelation is appropriately accounted for. Chun (2008) discusses that the network weight matrix specification is defined based on CD and IOs. Hence, conducting further investigation may find how the effects of CD and IO variables behave when a network autocorrelation is incorporated in a model specification.

However, there are some limitations in this analysis. First, since this study uses the total amounts of the interregional commodity flows in the US; different commodity types may lead to different patterns in their interregional flows (for example, Ham, Kim, & Boyce, 2005). The impact of commodity types needs further investigation. Second, this paper only applies network autocorrelation framework to domestic commodity flows in the US. The spatial interaction models do not include some other factors which are important in international commodity flows, such as tariff structure and political barriers (see Anderson & van Wincoop, 2004). Hence, these factors need to be considered when this modeling framework is applied to interactional trades.

Conclusively, the network autocorrelation framework to US interregional commodity flows proves useful as the results confirm that network autocorrelation needs to be explicitly accounted for in modeling interregional commodity flows. Geographical variables based on spatial behavioral theories, including CD and IOs, significantly improve spatial interaction models and capture a level of network autocorrelation. Nevertheless, network autocorrelation can still present itself in spatial interaction models and impact parameter estimation and standard errors. Thus, spatial interaction models should be specified to explicitly explain network autocorrelation. It is obvious that the frameworks presented in this paper are applicable in other areas, including passenger flows estimation in air transportation and predicting information flows in telecommunications, which requires capturing the potential misspecification in estimating flows distribution at various geographic scales with higher reliability (Sherali & Park, 2001).

Acknowledgments

We thank three anonymous reviewers on their valuable comments.

References

Almeida, L. M. W., & Goncalves, M. B. (2001). A methodology to incorporate behavioral aspects in trip-distribution models with an application to estimate student flow. *Environment and Planning A, 33*, 1125–1138.
Anderson, J. E., & van Wincoop, E. (2004). Trade costs. *Journal of Economic Literature, 42*(3), 691–751.
Anselin, L. (1988). *Spatial econometrics: Methods and models*. London: Kluwer.
Ashtakala, B., & Murthy, A. (1988). Optimized gravity models for commodity transportation. *Journal of Transportation Engineering (ASCE), 114*, 393–408.
Black, W. R. (1972). Interregional commodity flows: Some experiments with the gravity model. *Journal of Regional Science, 12*(1), 107–118.
Black, W. R. (1992). Network autocorrelation in transport network and flow systems. *Geographical Analysis, 24*(3), 207–222.
Bivand, R. S., Pebesma, E. J., & Gómez-Rubio, V. (2008). *Applied spatial data analysis with R*. New York: Springer.
Bröcker, J. (1989). Partial equilibrium theory of interregional trade and the gravity model. *Papers of the Regional Science Association, 66*, 7–18.
Brown, W. M., & Anderson, W. P. (2002). Spatial markets and the potential for economic integration between Canadian and US regions. *Papers in Regional Science, 81*(1), 99–120.
Bureau of Economic Analysis (BEA). (2002). *Regional economic accounts*. <http://www.bea.gov/regional/reis/crius.cfm>.
Celik, H. M., & Guldmann, J.-M. (2007). Spatial interaction modeling of interregional commodity flows. *Socio-Economic Planning Sciences, 41*, 147–162.
Chisholm, M., & O'Sullivan, P. (1973). *Freight flows and spatial aspects of the British economy*. New York: Cambridge University Press.
Chun, Y. (2008). Modeling network autocorrelation within migration flows by eigenvector spatial filtering. *Journal of Geographical Systems, 10*(4), 317–344.
Chun, Y., & Griffith, D. A. (2011). Modeling network autocorrelation in space–time migration flow data: An eigenvector spatial filtering approach. *Annals of the Association of American Geographers, 101*(3), 523–536.
Cliff, A. D., & Ord, J. K. (1981). *Spatial processes: Models and applications*. London: Pion Limited.
Curry, L. (1972). A spatial analysis of gravity flows. *Regional Studies, 6*, 131–147.
Enke, S. (1951). Equilibrium among spatially separated markets: Solution by electric analogue. *Econometrica, 19*(1), 40–47.
Fik, T. J., & Mulligan, G. F. (1990). Spatial flows and competing central places: Towards a general theory of hierarchical interaction. *Environment and Planning A, 22*, 527–549.
Fischer, M. M., & Griffith, D. A. (2008). Modeling spatial autocorrelation in spatial interaction data: An application to patent citation data in the European Union. *Journal of Regional Science, 48*, 969–989.
Fotheringham, A. S. (1983). A new set of spatial-interaction models: The theory of competing destinations. *Environment and Planning A, 15*, 15–36.
Fotheringham, A. S., & O'Kelly, M. E. (1989). *Spatial interaction models: Formulations and applications*. Amsterdam: Kluwer Academic.
Griffith, D. A. (2009). Modeling spatial autocorrelation in spatial interaction data: Empirical evidence from 2002 Germany journey-to-work flows. *Journal of Geographical Systems, 11*, 117–140.
Griffith, D. A., & Jones, K. G. (1980). Explorations into the relationship between spatial structure and spatial interaction. *Environment and Planning A, 12*, 187–201.
Guldmann, J.-M. (1999). Competing destinations and intervening opportunities interaction models of inter-city telecommunication. *Papers in Regional Science, 78*, 179–194.
Ham, H., Kim, T. J., & Boyce, D. (2005). Assessment of economic impacts from unexpected events with an interregional commodity flow and multimodal transportation network model. *Transportation Research Part A: Policy and Practice, 39*(10), 849–860.
Kwan, M.-P. (1998). Space–time and integral measures of individual accessibility: A comparative analysis using a point-based framework. *Geographical Analysis, 30*(3), 191–216.
LeSage, J. P., & Pace, R. K. (2008). Spatial econometric modeling or origin–destination flows. *Journal of Regional Science, 48*(5), 941–967.
LeSage, J. P., & Pace, R. K. (2009). *Introduction to spatial econometrics*. Roca Raton: CRC Press.
LeSage, J. P., & Fischer, M. M. (2010). Spatial econometric methods for modeling origin–destination flows. In M. M. Fischer & A. Getis (Eds.), *Handbook of applied spatial analysis* (pp. 409–433). Heidelberg: Springer.
Novak, D. C., Hodgdon, C., Guo, F., Aultman-Hall, L. (2010). Nationwide freight generation models: A spatial regression approach. *Network and Spatial Economics* (in press) doi: 10.1007/s11067-008-9079-2. <http://www.springerlink.com/content/an2mx454g47u8155/fulltext.pdf>.
Raphael, S. (1998). Intervening opportunities, competing searchers, and the intrametropolitan flow of male youth labor. *Journal of Regional Science, 38*(1), 43–59.
Roy, J. R. (1993). Comment on "The development of a new gravity-opportunity model for trip distribution". *Environment and Planning A, 25*, 1689–1691.
Roy, J. R., & Thill, J.-C. (2004). Spatial interaction modeling. *Papers in Regional Science, 83*, 339–361.
Samuelson, P. A. (1952). Spatial price equilibrium and linear programming. *American Economic Review, 42*, 283–303.
Sheppard, E. S., Griffith, D. A., & Curry, L. (1976). A final comment on misspecification and autocorrelation in those gravity parameters. *Regional Studies, 10*(3), 337–339.
Sherali, H. D., & Park, T. (2001). Estimation of dynamic origin–destination trip tables for a general network. *Transportation Research B, 35*(3), 217–235.
Stouffer, S. A. (1960). Intervening opportunities and competing migrants. *Journal of Regional Science, 2*(1), 1–26.
Thill, J.-C., & Kim, M. (2005). Trip making, induced travel demand, and accessibility. *Journal of Geographical Systems, 7*(2), 229–248.
US Bureau of Transportation Statistics and Census Bureau (US BTS & CB). (2007). *2002 Commodity flow survey (CFS)*, June 2005. <http://www.bts.gov/publications/commodity_flow_survey/>.