

## **A Spatial Multivariate Count Model for Firm Location Decisions**

### **Chandra R. Bhat\***

The University of Texas at Austin

Dept of Civil, Architectural and Environmental Engineering

301 E. Dean Keeton St. Stop C1761, Austin TX 78712-1172

Phone: 512-471-4535 ; Fax: 512-475-8744 ; Email: [bhat@mail.utexas.edu](mailto:bhat@mail.utexas.edu)

and

King Abdulaziz University, Jeddah 21589, Saudi Arabia

### **Rajesh Paleti**

Parsons Brinckerhoff

One Penn Plaza, Suite 200

New York, NY 10119

Phone: 512-751-5341; Fax: 212-465-5096; Email: [paletir@pbworld.com](mailto:paletir@pbworld.com)

### **Palvinder Singh**

Parsons Brinckerhoff

400 SW Sixth Avenue, Suite 802

Portland, OR 97204

Phone: 503-478-2873; Fax: 503-274-1412; Email: [singhp@pbworld.com](mailto:singhp@pbworld.com)

\*corresponding author

### **ABSTRACT**

This paper proposes a new spatial multivariate model to predict the count of new businesses at a county level in the State of Texas. Several important factors including agglomeration economies/diseconomies, industrial specialization indices, human capital, fiscal conditions, transportation infrastructure and land development characteristics are considered. The results highlight the need to use a multivariate modeling system for the analysis of business counts by sector type, while also accommodating spatial dependence effects in business counts. (C31, C35, C51)

*Keywords:* Multivariate analysis, spatial econometrics, business counts, composite marginal likelihood.

### **ACKNOWLEDGEMENTS**

The authors are grateful to Lisa Macias for her help in formatting this document. Three referees provided valuable comments on an earlier version of the paper.

## 1. INTRODUCTION

The choice of a location to start a new business or to expand into new locations for an existing business is critical to the success of the entity making such decisions (we will refer to this decision-making entity broadly as the “firm” in this paper). After all, firms incur high fixed capital and time costs in locating their businesses, and have to consider such cost-related factors as tax incentives offered by local jurisdictions, transportation infrastructure in the region, the availability and cost of human capital, and real-estate costs (see Alañón-Pardo and Arauzo-Carod, 2011; Hanson and Rohlin, 2011). At the same time, firms also have to estimate the potential gains (both in the short-term as well as in the long-term) from locating in specific jurisdictions, based on the demand for their product and the price levels that can be set for the product (Strotmann, 2007; Alamá-Sabater et al., 2011). On the other side of the decision-making process, local jurisdictions also have costs and benefits to having businesses locate in their areas. The costs can include congestion effects, environmental quality degradation, and excess commuting (Arauzo-Carod, 2008; Fullerton et al., 2008), while the benefits can include high economic productivity, high employment rates, and an overall better quality of life (Basile et al., 2010; Alañón-Pardo and Arauzo-Carod, 2011; Alamá-Sabater et al., 2011). Thus, business location choice is an important area of interest for both firms as well as local and regional political jurisdictions.

In addition to firms and political jurisdictions, business location choice is also of interest to transportation and urban planning agencies. From a transportation perspective, and as already alluded to in the previous paragraph, increased employment opportunities result in more commuting trips as well as more non-work trips during the traditional peak commuting periods in the day (the latter is triggered by workers chaining activities and pursuing non-work activities during the work commute; see, for example, Bhat and Sardesai, 2006 and Van Acker and Witlox, 2011). Further, a high activity intensity in a region, coupled with good economic conditions, can also result in higher levels of trip-making of residents of the region as well as of neighboring regions (see Chen et al., 2011). So, quite understandably, predicting the employment patterns in the region for future years constitutes an important preliminary step of a travel demand forecasting exercise (Pendyala et al., 2012). From an urban planning perspective, the land use intensity and composition (*i.e.*, the fraction of land acreage under residential, retail, and commercial use) in a region has a significant impact on many long-to-medium term decisions of

households, including residential location and auto ownership, which can in turn impact day-to-day short term mobility decisions related to travel (such as commute mode choice, use of non-motorized modes of transportation for non-work activities, and the decision to telecommute; see Pinjari et al., 2011 and Singh et al., 2013). Indeed, many local and regional jurisdictions have developed visions and plans for land-use in their urban areas to promote sustainable growth. For instance, the City of Austin recently drafted a vision to develop several mixed land-use corridors with housing, retail, and recreation to curb urban sprawl and promote sustainable travel patterns over the next 30 years (City of Austin, 2012). The intent is to achieve the urban vision through such policy instruments as fiscal incentives and disincentives, planning controls, and public transportation and bicycle infrastructure investments. Overall, it is important for transportation and urban planners to be able to predict the expected number of new firms of different sectors in each spatial pocket within a region as a function of relevant covariates, both for forecasting purposes as well as to inform policy making to achieve desired end-states.

To be sure, the empirical analysis of business location decisions has been a fertile area of research in several fields, but particularly in regional science. In this context, increased availability and accessibility to urban and region business location data, coupled with advancements in the specification and estimation of econometric models, has led to important progress in recent years. This earlier research has been dominated by one of two modeling approaches. The first, discrete choice modeling, approach considers the firm as the unit of analysis, and investigates business location choices of firms as a function of firm characteristics (such as firm size and industry sector) and alternative territorial location characteristics (such as population, human capital measures, and transportation infrastructure) (see Alamá-Sabater et al., 2011; Basile et al., 2009; Barrios et al., 2006). The central idea of the discrete choice approach is that a firm makes a rational decision based on the theory of profit maximization and cost minimization so that the accrued benefits exceed the initial capital investments as well as subsequent organizational expenses. In almost all of these studies, the unit of territorial analysis used to define the alternatives in a firm's location choice set is a municipality or a county.<sup>1</sup> The

---

<sup>1</sup> The reader is referred to Arauzo-Carod (2008) and Arauzo-Carod and Manjón-Antolín (2012) for extensive discussions of the territorial unit of analysis. Since this issue is not itself of direct relevance to the current study, we do not discuss it in any great detail in the current paper. However, we should mention that, while much of the earlier research in the field until about a decade back used a coarse territorial level such as states, metropolitan areas, or countries (see Disdier and Mayer, 2004, and Basile et al., 2004), recent studies have moved toward more disaggregate territorial levels such as municipalities, counties, and districts (see Alamá-Sabater et al., 2011, Arauzo-

second, count modeling, approach considers the territory as the unit of analysis, and investigates how location attributes can influence business location decisions in the form of the count of businesses in each territorial unit. The fundamental assumption underlying the count approach is that the number of new establishments that start in a territory over a time period is determined by an equilibrium condition between a stochastic supply function representing the desire of firms to start a business in the territory, and a stochastic demand function for new firms in the territory. This equilibrium condition can be represented by a reduced form stochastic distribution for the count of new businesses (Becker and Henderson, 2000). As in the first approach, the dominant territorial unit of analysis in this second approach is also the municipality or the county.

The discrete choice and count modeling approaches have their own advantages and limitations (see Arauzo-Carod et al., 2010 for a detailed discussion). The discrete choice approach can be derived as a structural process of firm location decisions and can accommodate both firm level and territory characteristics, while the count approach can only be derived as an aggregate-level reduced form equilibrium process and can accommodate only territory characteristics. Thus, the discrete choice approach has behavioral foundation advantages. However, most discrete choice models of business location use few firm-level characteristics anyway because of the difficulty in obtaining such data, and become unwieldy when the number of territorial units (alternatives) is high. The common way of dealing with the latter issue is by either moving toward aggregate territorial units or using the restrictive multinomial logit/nested logit structures so that one can sample alternatives. But such methods effectively undo the structural behavioral benefits of the approach. Further, another limitation of the discrete choice approach is that, during estimation, it does not use the location characteristics of those spatial alternatives that are never chosen by any firm. On the other hand, the count modeling approach is appealing when the number of territorial units is high (indeed, doing so contributes more observations in the count approach, so that what is a problem in the discrete choice approach becomes a statistical efficiency gain in the count approach). It also uses the characteristics of all

---

Carod and Manjón-Antolín, 2012, and Kim et al., 2008). The choice of municipalities and counties in particular as the territorial unit has been driven by (a) less heterogeneity in location characteristics at these lower levels of geography, (b) an implied belief that municipalities and counties are the “true” territorial units of analysis used by most domestic firms in their decision-making, (c) the relatively easy availability today of accurate data on location characteristics at the municipality or county level data, which reduces “error measurement” in the exogenous variables, and (d) the recognition that the use of administrative units (rather than arbitrary functional geographical units) is more relevant to policy-making, since municipality and county planning boards can develop and implement policy mechanisms to shape business location decisions in their respective jurisdictions.

territorial units in analyzing business location choice. The net result is that most recent studies in the business location choice field have adopted the count modeling approach.

In this paper, we too consider a count modeling approach for business location decisions and use a county-level territorial unit of analysis (for ease in presentation, in the rest of this paper, we will use the term “county” generically to refer to any territorial unit of space). However, unlike earlier studies, we consider the business location decisions by industry sector, to recognize that the determinants of business location decisions are likely to be different across sectors. For example, businesses in the agricultural sector are heavily affected by the land costs in the county (which are generally represented using the population density in the county), but land costs may have little to no effect on new businesses in the manufacturing sector. Similarly, a good roadway network is extremely important for businesses in the manufacturing sector for unhindered delivery of raw materials from other regions to the business locations and finished products from business locations to the markets. In comparison, businesses in the agriculture sector are not so heavily dependent on the roadway infrastructure in the county.

The multivariate count model proposed in this paper for modeling industry sector-specific business location decisions recognizes many econometric issues at once: (a) It conveniently accommodates over-dispersion and excess zero problems in the county-level count of new businesses by sector type, (b) it considers the presence of common county-level unobserved factors that simultaneously influence the county-level count of new businesses in different sectors, and (c) it considers spatial dependence effects across counties that are likely to be present because of the spatial nature of the analysis. In this regard, we see the current paper as a methodological contribution to the econometrics and spatial econometrics fields, motivated by characteristics that are specific to business location analysis (though our spatial multivariate count model may also be applicable to a wide variety of other fields too). In particular, to our knowledge, this is the first formulation and application of a multivariate spatial count model. However, from an empirical standpoint, this study also extends extant business firm location models by modeling the birth of new businesses in multiple industry sectors all at once as well by providing a mechanism to comprehensively account for spatial dependency effects in business location choice. Thus, the emphasis of the paper is on developing a new spatial econometric method that is appropriate for business location choice, and demonstrating its application to business location choice. We are embracing Arauzo-Carod et al.’s (2010) call here when they

lamented that “the scarce use of spatial econometric techniques may be due to the lack of appropriate tools, while future developments in spatial econometrics should shortly be followed by applications to industrial location”. In addition, the modeling approach offers a nice interpretative device for disentangling the effects of exogenous determinants on the demand for businesses of each sector within each county and the supply of businesses of each sector within each county, which we hope will be exploited in future business location empirical studies. More generally, we hope that the methodology developed in this paper will open up a whole new direction of intense empirical exploration using appropriate econometric tools for business location analysis.

The remainder of the paper is structured as follows. Section 2 discusses the relevant earlier literature and positions the current study. Section 3 describes the methodology and estimation procedure used in our analysis. Section 4 provides an overview of the data used and some key descriptive statistics. Section 5 presents the empirical findings. Section 6 concludes the study by summarizing important findings and identifying policy implications.

## 2. ECONOMETRIC CONSIDERATIONS AND THE CURRENT STUDY

In formulating a multivariate model for the county-level counts of new businesses by industry sector type (in the rest of this paper, we will refer to “industry sector” simply as “sector” and “counts of new businesses” as “business counts”), three econometric considerations are important to recognize, as discussed in turn below.

### *Over-Dispersion and Excess Zeros*

Several types of discrete probability distributions may be considered in modeling count data, though the workhorse discrete distributions are the Poisson and the negative binomial (NB) distributions. The NB distribution is a generalization of the Poisson, where the variance of the distribution is allowed to be higher than the mean (unlike the Poisson distribution, where the variance is equal to the mean). In the business location literature, examples of the use of a Poisson distribution include Arauzo-Carod and Manjón-Antolín (2004), Guimaraes et al. (2004), Arauzo-Carod (2005, 2008), Jofre-Monseny and Solé-Ollé (2010), and Jofre-Monseny et al. (2011), while those that use a NB distribution include Mota and Brandão (2013), Alañón-Pardo and Arauzo-Carod (2011), Arauzo-Carod and Viladecans-Marsal (2009), and Gabe and Bell

(2004). More generally, in the traditional count model, overdispersion can be accommodated by introducing an additional multiplicative continuous mixture error term in the conditional mean parameter (the NB model is a specific case where the continuous mixing error term has a gamma distribution).

A related consideration in business count models is that there are typically a large number of counties with zero values for one or more sectors. The most commonly used approach in business location count models to accommodate this issue is the zero-inflated approach. The approach identifies two separate states for the count generating process – one that corresponds to a “zero” state in which the expected value of counts is so close to zero as being indistinguishable from zero, and another “normal” state in which a typical count model (with either a Poisson or NB distribution) operates (see, for example, Gabe, 2003, Arauzo-Carod, 2008, and Manjón-Antolín and Arauzo-Carod, 2011). Effectively, the zero-inflated approach is a discrete-mixture model involving a discrete error distribution that modifies the probability of the zero outcome. Another similar approach to account for excess zeros is the hurdle-count approach (in which a binary outcome process of the count being below or above a hurdle (zero) is combined with a truncated discrete distribution for the count process being above the hurdle (zero) point. Interestingly, the hurdle approach has not seen use in the business location modeling literature, with the exception of Liviano and Arauzo-Carod (2013) who found that the hurdle approach fits their industrial sector location data better than the zero-inflated approach.

Clearly, the business count literature has seen the use of count models that accommodate both over-dispersion and excess zeros. However, all of these studies have been in the context of a single univariate count of businesses (focused either on the pooled count of businesses across sectors or on sector-specific count of businesses). Unfortunately, extending the continuous and discrete mixing approaches just discussed to a multivariate business count model becomes cumbersome to impractical (see Lee et al., 2006; Alfò and Maruotti, 2010). Other means to accommodate over-dispersion and excess zeros are desirable in multivariate count settings, as we propose in the current paper.

### *Multivariate Business Count Model*

Almost all earlier business count studies have focused on a single count in the form of a single sector such as the manufacturing sector or the industrial sector (see Arauzo-Carod, 2005,

2008; Arauzo-Carod and Manjón-Antolín, 2004; Guimaraes et al., 2004; Gabe and Bell, 2004; Holl, 2004a,b). A handful of studies have disaggregated the manufacturing sector count into specific technology types (such as high technology, intermediate technology, and low technology; see Arauzo-Carod and Viladecans-Marsal, 2009) or specific specializations (such as food, drinks, tobacco, clothes and leather, and wood and furniture). But these studies develop independent count models for each technology type or specialization. Similarly, a few studies have examined the determinants of business counts for multiple sectors such as manufacturing, FIRE (financial, retail, and real estate), wholesale trade, construction, and transportation (see Hanson and Rohlin, 2011; Kim et al., 2008; Holl, 2004c; Gabe, 2003; and Blonigen, 1997). Again, however, these are all independent models for each sector. While these business count models by sector highlight the marked differences in the location determinants of new firms, thus underscoring the need to estimate sector-specific models, they fundamentally ignore the presence of county-level unobserved factors that may either increase or decrease the county-level counts of specific sectors. To the extent that there will inevitably be some unobserved determining factors that will have similar or opposite effects on the business counts of different sectors, it is only reasonable to expect error correlations across the sector-level counts. For instance, a county may have zoning regulations that place restrictions on the land available for locating businesses. If such zoning regulations are not included in a model of business counts, it will generate an unobserved correlation across the counts of multiple sectors. To our knowledge, no such multivariate model has been developed for business counts of multiple sectors.

In the broader count literature, the most common approach to analyze multivariate counts is based on a mixing approach in which one or more (typically normally distributed) random terms are introduced in the parameterization of the mean (so that the mean is not only a function of exogenous variables, but also includes one or more random terms within the exponentiation). The correlation (positive or negative) between counts in different categories is generated by common random error terms in the mean count specification of these categories. However, unless one is willing to settle for restrictive covariance patterns (that also have to be specified *a priori*), the estimation of such mixed multivariate count data models can be cumbersome and time-consuming despite advances in frequentist and Bayesian simulation techniques (see Müller and Czado, 2005; Agüero-Valverde and Jovanis, 2006; Ver Hoef and Jansen, 2007; and Herriges et al., 2008 for discussions). Further, as indicated earlier, it is anything but straightforward to



modify these multivariate count mixing approaches to get their zero inflated and hurdle variants. Finally, the approaches discussed above to accommodate multivariate counts are already so computationally difficult that extending the approaches to accommodate spatial dependency structures becomes impractical, if not literally infeasible.

### *Spatial Dependency*

A key attribute of business location data is that they are, by construction, geo-referenced. To the extent that the location characteristics of a particular county are likely to be affected by the realizations of the location characteristics of proximally located counties (see Arauzo-Carod and Manjón-Antolín, 2012), there will be an attenuating spatial dependence in the business counts of proximally located counties. Indeed, it is a well established fact now that there tends to be positive spatial dependency effects (across territorial units) in many demographic and social activities due to the presence of external economies and inter-county spillovers (see Parent and LeSage, 2008; Alama-Sabater et al., 2011). This spatial dependence arising due to observed economic and social activity of proximally located counties is typically referred to as *spatial spillover* effects, and has been well documented in the theoretical as well as empirical business location literature (see, for example, Arauzo-Carod et al., 2010; Fujita and Thisse, 2002). As observed by Alamá-Sabater et al. (2011) and Guimaraes et al. (2004), spatial spillover effects are likely to be particularly relevant in count models that use disaggregate territorial units (such as municipalities and counties) because site characteristics can more easily extend their influence into neighboring territorial units (than if entire regions or states were used as the territorial unit of analyses). But spatial dependency may also arise due to overall (unobserved to the analyst) perceptions about gains and profitability of locating in one county impacting the perceptions about the gain and profitability of locating in proximally located counties. This can happen through a variety of effects associated with the networking of employees of firms in the same sector, and/or unobserved pecuniary, knowledge, and technological spillovers in location decision-making across firms of the same sector (see, for example, Autant-Bernard and LeSage, 2011, who also provide an exhaustive discussion of the many reasons for spatial effects in general and knowledge-related spatial effects in particular). Of course, there may also be unobserved (to the analyst) determinants of business location decisions such as high land prices, county zoning preferences and public policies, and even public attitudes that get transmitted over

space. All of these spatial dependencies arising due to unobserved factors of proximally located neighborhoods will be referred to in this paper as the *spatial error* effect. Ignoring these spatial effects will, in general, result in inconsistent and inefficient parameter estimates in non-linear models (see Arauzo-Carod and Manjón-Antolín, 2012; LeSage and Pace, 2009; Franzese and Hays, 2008).<sup>2</sup> Note also that the spatial error effect can be positive because of the various positive effects just discussed, or negative because of negative unobserved externalities, or zero because of the absence of unobserved effects or more likely due to the positive and negative effects cancelling out one another (see Griffith and Arbia, 2010 for further discussions along these lines).

The most common approach for dealing with spatial dependencies in business count data has been to use spatially lagged explanatory variables in the specification of the mean count parameter (see Arauzo-Carod and Manjón-Antolín, 2012; Alañón-Pardo et al., 2007). The underlying assumption here is that spatial dependency is generated only through observed factors of proximally located neighborhoods (that is, only through spatial spillover effects and not through spatial error effects). However, this assumption is difficult to justify, since unobserved factors of proximal counties should affect business counts in a county just as observed factors do. Another alternative approach uses a conditional autoregressive (CAR) or a joint prior on a spatial random effect term that is introduced in the parameterization of the expected value of the discrete distribution for the count variable. Liviano and Arauzo-Carod (2012) used this approach to analyze the determinants of the number of new manufacturing firms. Unfortunately, this approach (which is essentially a mixing approach of the type discussed in the previous section, except with the mixing undertaken over space) becomes difficult as the number of spatial units increases. Besides, this approach considers spatial error effects, but not spatial spillover effects. A third approach to spatial dependency in count models was recently proposed by Basile et al. (2010), who developed a semi parametric negative binomial model with a spatially lagged dependent variable in the mean count specification, and estimated the resulting model using a

---

<sup>2</sup> Many papers in the business location literature suggest that the typical aspatial count model avoids the independence from irrelevant alternatives (IIA) property of the multinomial logit discrete choice model. However, this is rather misleading. After all, observed and unobserved factors that increase the utility of proximally located counties in a discrete choice model (leading to the violation of IIA) are precisely what will cause spatial spillover and spatial error effects across proximally located counties in a count model. In effect, using an aspatial count model is tantamount to using a multinomial logit model in discrete choice.

control function approach. However, the nature of spatial dependence is difficult to interpret in this approach, and it is difficult to extend the approach to a multivariate count case.

In the broader spatial econometric literature, spatial dependency specifications have been used extensively in the case of a continuous dependent variable, but the application of such a specification to count models has only been achieved very recently. In particular, Castro, Paleti, and Bhat (2012) (CPB for short in the rest of this paper) showed that even count models can be recast in the form of an underlying latent continuous variable framework (so that the spatial dependency specifications can then be applied to the latent continuous propensity variables characterizing the observed count data).

### *Current Study in Context*

In this paper, we develop a multivariate count model for analyzing the county-level count of new businesses by sector that at once accommodates over-dispersion and excess zero considerations, recognizes the presence of county-level unobserved factors that may influence the counts across sectors, and recognizes spatial dependency effects across counties. Our approach is based on recasting the basic count model as a special case of a generalized ordered-response (GOR) model, as proposed by CPB. The resulting model has a likelihood function that is analytically intractable. Traditional simulation approaches are all but infeasible for the typical sample sizes used for estimation in business count models. Instead, we use a composite marginal likelihood (CML) inference approach that is simple to implement and is based on evaluating lower-dimensional marginal probability expressions. To our knowledge, this is the first spatially dependent multivariate count model formulation in the business location literature and more generally in the econometric literature.

The model is applied to estimate the count of new businesses in each of 254 counties in Texas and for each of 11 different sectors. A wide range of factors related to agglomeration economies/diseconomies, fiscal policy, human capital, transportation infrastructure, geographical position, and demographics of the county are considered in the analysis.

### 3. METHODOLOGY

#### *Basic Count Model Recasting*

Let  $q$  ( $q = 1, 2, \dots, Q$ ) be the index for the territorial unit of analysis (a “county” in the current paper) and let  $s$  ( $s = 1, 2, \dots, S$ ) be the index for sector type. Let  $y_{qs}$  be the index for the count of new businesses in sector  $s$  in county  $q$ , and let  $m_{qs}$  be the actual observed count of new businesses in sector  $s$  in the county  $q$ . Next, consider that there is a county-specific demand function that represents the intensity of consumption need for the services and products offered by businesses in sector  $s$ . This demand function represents the prevailing consumption needs but is not directly observed, and so may be represented by a latent (unobserved to the analyst) demand intensity  $y_{qs}^*$  for new businesses in sector  $s$  in county  $q$ . Then, in the generalized ordered response (GOR) notation, the latent demand intensity  $y_{qs}^*$  is written as a function of a  $(K \times 1)$ -vector of observed covariates  $\mathbf{x}_q$  (excluding the constant) as:

$$(1) \quad y_{qs}^* = \boldsymbol{\beta}'_s \mathbf{x}_q + \eta_{qs}, \quad y_{qs} = m_{qs} \Rightarrow \psi_{qs, m_{qs}-1} < y_{qs}^* < \psi_{qs, m_{qs}}.$$

In the above specification,  $\boldsymbol{\beta}_s$  is a  $(K \times 1)$ -vector whose elements capture the effects of the elements in the  $\mathbf{x}_q$  variable vector on latent demand propensity  $y_{qs}^*$ .<sup>3</sup> Finally,  $\eta_{qs}$  captures county-specific unobserved factors that increase or decrease the latent demand propensity for new businesses in sector  $s$ .

The thresholds in Equation (1) take the following form:

$$(2) \quad \psi_{qs, m_{qs}} = \Phi^{-1} \left( e^{-\lambda_{qs}} \sum_{l=0}^{m_{qs}} \frac{\lambda_{qs}^l}{l!} \right) + \alpha_{s, m_{qs}}, \quad \lambda_{qs} = e^{\boldsymbol{\gamma}'_s \mathbf{h}_q}, \alpha_{s, 0} = 0, \alpha_{s, m_{qs}} = \alpha_{s, L_s} \text{ if } m_{qs} > L_s,$$

where  $\Phi^{-1}$  is the inverse function of the univariate cumulative standard normal,  $\psi_{qs, -1} = -\infty \forall q, s$ ,  $\alpha_{s, 0} = 0 \forall s$  (this restriction is needed for identification, given the parameterization of the thresholds),  $\mathbf{h}_q$  is a vector of exogenous variables (including a constant) associated with county  $q$  (there can be common variables in  $\mathbf{h}_q$  and  $\mathbf{x}_q$ ),  $\boldsymbol{\gamma}_s$  is a corresponding

---

<sup>3</sup> The set of explanatory variables can vary across sectors; that is, there may be some variables in  $\mathbf{x}_q$  that may be important in determining the latent demand propensity for businesses in some sectors but not others. This situation is accommodated within our notation system by letting the corresponding elements in the vector  $\boldsymbol{\beta}_s$  be equal to zero.

coefficient vector to be estimated for sector  $s$ , and  $L_s$  is a pre-defined count level determined based on the empirical context under consideration. As in the typical ordered-response framework, the values of  $\alpha_{s,m_{qs}}$  should be such that the ordering condition on the thresholds  $(-\infty < \psi_{qs,0} < \psi_{qs,1} < \psi_{qs,2} < \dots)$  is satisfied. The presence of the  $\alpha_{s,m_{qs}}$  terms provides flexibility to accommodate high or low probability masses for specific count outcomes without the need for cumbersome treatment using hurdle or zero-inflated mechanisms (see CPB). Also, if these terms are set to zero, and all elements of the vector  $\beta_s$  are set to zero, the result is the traditional Poisson count model (CPB).

As mentioned earlier, the parameters in the count modeling framework correspond to an equilibrium condition between the demand for businesses of each sector within each county and the supply of businesses of each sector within each county. In traditional count models, the demand and supply functions get co-mingled, and so it is not possible to disentangle the demand and supply side effects of county variables. The GOR framework that we adopt, on the other hand, offers a nice interpretive device to separate out these two effects in a more structural fashion. In particular,  $y_{qs}^*$  in Equation (1) may be viewed as the latent prevailing demand intensity for new businesses in sector  $s$  in county  $q$  that is impacted by county-specific variables included in the  $x_q$  vector (such as the availability of human capital as a proxy for individuals seeking employment, residential population, and existing counts of businesses). On the other hand, the threshold values  $\psi_{qs,l}$  ( $l = 0, 1, 2, \dots, \infty$ ) may be viewed as supply “tipping points” that determine, given prevailing demand, the number of new businesses in sector  $s$  that ultimately decide to enter into county  $q$ . These thresholds are impacted by localized county-specific characteristics embedded in the  $h_q$  vector (such as transportation infrastructure, population change trends indicative of potential future demand, the availability of human capital as a proxy for skill availability at reasonable cost, and local government tax incentives) that determine the profitability margin calculations for businesses in sector  $s$ . Of course, the same county characteristics can enter into both the latent demand intensity  $y_{qs}^*$  and the supply tipping points  $\psi_{qs,l}$ , though the expectation is that many more county characteristics will enter the demand intensity rather than the supply tipping points (see Arauzo-Carod et al., 2010 for a discussion).

Overall, two counties may have the same latent demand intensity  $y_{qs}^*$ , but may have different observed counts of new businesses  $m_{qs}$  because of different supply tipping points defining the  $y_{qs}^*$  - to -  $m_{qs}$  mapping.

### *Motivation for Spatial Model*

Following Autant-Bernard and LeSage (2011), we now consider spatial dependence (see section entitled ‘‘Spatial Dependency’’) across counties  $q$  in the  $\mathbf{x}_q$  vector as well as in the unmeasurable terms  $\eta_{qs}$ . For ease in presentation in this motivation section, we assume a single exogenous variable in the  $\mathbf{x}_q$  vector (the motivation applies also to the realistic case of multiple exogenous variables, but unnecessary cumbersome in the context of motivating the spatial formulation). Assume the following spatial autoregressive processes:  $x_q = \tilde{\psi} \sum_{q'=1}^Q w_{qq'} x_{q'} + u_q$  and

$$\eta_{qs} = \rho_s \sum_{q'=1}^Q w_{qq'} \eta_{q's} + v_{qs}. \quad u_q \text{ and } v_{qs} \text{ are shocks to the observable and unobservable inputs, and}$$

$w_{qq'}$  is the usual distance-based spatial weight corresponding to counties  $q$  and  $q'$  (with  $w_{qq} = 0$  and  $\sum_{q'} w_{qq'} = 1$ ) for each (and all)  $q$ . To allow correlation in the explanatory variables and the error term  $\eta_{qs}$  within a county, we assume a linear dependence structure between  $u_q$  and  $v_{qs}$ :

$$v_{qs} = \gamma u_q + \varepsilon_{qs}, \text{ where } \varepsilon_{qs} \text{ is independent of } u_q. \text{ Next, vertically stack the elements } y_{qs}^* \text{ into a}$$

vector  $\mathbf{y}_s^*$ , the elements  $x_q$  into a vector  $\tilde{\mathbf{x}}$ , the elements  $v_{qs}$  into a vector  $\mathbf{v}_s$  and the elements

$\eta_{qs}$  into a vector  $\boldsymbol{\eta}_s$ . Collect all the weights  $w_{qq'}$  into a row-normalized spatial weight matrix  $\mathbf{W}$ .

Then, we have the vector equations  $\mathbf{y}_s^* = \boldsymbol{\beta}'_s \tilde{\mathbf{x}} + \boldsymbol{\eta}_s$ ,  $\tilde{\mathbf{x}} = \tilde{\psi} \mathbf{W} \tilde{\mathbf{x}} + \mathbf{u}$ ,  $\boldsymbol{\eta}_s = \delta_s \mathbf{W} \boldsymbol{\eta}_s + \mathbf{v}_s$ , and

$$\mathbf{v}_s = \gamma \mathbf{u} + \boldsymbol{\varepsilon}_s. \quad \text{We also may write } \mathbf{u} = (\mathbf{I}_Q - \tilde{\psi} \mathbf{W}) \tilde{\mathbf{x}}, \text{ and}$$

$$\boldsymbol{\eta}_s = (\mathbf{I}_Q - \delta_s \mathbf{W})^{-1} \mathbf{v}_s = (\mathbf{I}_Q - \delta_s \mathbf{W})^{-1} (\gamma \mathbf{u} + \boldsymbol{\varepsilon}_s). \text{ Next, substituting this expression for } \boldsymbol{\eta}_s \text{ in the}$$

vector equation for the latent demand propensity, we obtain  $\mathbf{y}_s^* = \boldsymbol{\beta}'_s \tilde{\mathbf{x}} + (\mathbf{I}_Q - \delta_s \mathbf{W})^{-1} (\gamma \mathbf{u} + \boldsymbol{\varepsilon}_s)$ ,

Equivalently, we may write:

$$\begin{aligned}
(\mathbf{I}_Q - \delta_s \mathbf{W}) \mathbf{y}_s^* &= (\mathbf{I}_Q - \delta_s \mathbf{W}) \boldsymbol{\beta}'_s \tilde{\mathbf{x}} + (\gamma \mathbf{u} + \boldsymbol{\varepsilon}_s) \\
(3) \quad &= \boldsymbol{\beta}'_s \mathbf{x}_q - (\delta_s \mathbf{W}) \boldsymbol{\beta}'_s \tilde{\mathbf{x}} + \gamma (\mathbf{I}_Q - \tilde{\psi} \mathbf{W}) \tilde{\mathbf{x}} + \boldsymbol{\varepsilon}_s \\
&= (\boldsymbol{\beta}'_s + \gamma) \tilde{\mathbf{x}} + (-\delta_s \boldsymbol{\beta}'_s - \tilde{\psi} \gamma) \mathbf{W} \tilde{\mathbf{x}} + \boldsymbol{\varepsilon}_s.
\end{aligned}$$

The equation above can be rewritten as:

$$(4) \quad \mathbf{y}_s^* = \delta_s \mathbf{W} \mathbf{y}_s^* + \mathbf{b}'_s \tilde{\mathbf{x}} + \pi_s' \mathbf{W} \tilde{\mathbf{x}} + \boldsymbol{\varepsilon}_s, \quad \mathbf{b}_s = (\boldsymbol{\beta}_s + \gamma) \quad \text{and} \quad \pi_s = (-\delta_s \boldsymbol{\beta}_s - \tilde{\psi} \gamma).$$

Thus, for each sector, the demand propensity driving the count of the number of new firms takes a spatial Durbin (SD) model specification. It is also useful to note that the spatial autoregressive coefficient  $\delta_s$  in the above equation is generated as a result of the spatial error effect and the correlation in the explanatory variables and the error term  $\eta_{qs}$ .  $\delta_s$  may be positive or negative ( $-1 < \delta_s < 1 \forall s$ ). The net result, in Manski's (1993) classification, is that there is an endogenous interaction effect represented by  $\delta_s$  as well as an exogenous interaction effect represented by  $\pi_s$ .<sup>4</sup>

In the econometric formulation, we also consider the joint nature of the demand propensities across sectors for each county  $q$ . We do so by allowing the elements  $\varepsilon_{qs}$  of the

---

<sup>4</sup> Note that we are able to identify both the endogenous and exogenous interaction effects in non-linear models, such as in the latent variable reformulation of count models in Equation (4) above. This is different from the case of linear models (that is, when observation is made directly on  $\mathbf{y}_s^*$  in Equation (4)). In such a linear case, Manski (1993) identifies an identification problem that, in general, will not allow the analyst to disentangle the endogenous and exogenous interaction effects. This is because a change in any element of  $\tilde{\mathbf{x}}$  not belonging to county  $q$  in Equation (4) gets manifested in the variable  $\mathbf{W} \tilde{\mathbf{x}}$  and shifts all elements of  $\mathbf{y}_s^*$  not belonging to county  $q$ . But these shifts in  $\mathbf{y}_s^*$  for other counties affect the element of  $\mathbf{y}_s^*$  for county  $q$  through the  $\mathbf{W} \mathbf{y}_s^*$  effect. Effectively, then, in linear models,  $\mathbf{W} \tilde{\mathbf{x}}$  and  $\mathbf{W} \mathbf{y}_s^*$  become linearly dependent, causing an inability to disentangle the exogenous  $\pi_s$  effect from the endogenous  $\delta_s$  effect. However, in our non-linear count model, the relationship in Equation (4) is on the latent variables that then get translated in a non-linear fashion through thresholds to the observed count variable. In such models, the effect of  $\mathbf{y}_s^*$  for county  $q$  (which includes the  $\mathbf{W} \mathbf{y}_s^*$  influence) on the observed count is non-linear, while the effect of  $\mathbf{W} \tilde{\mathbf{x}}$  on  $\mathbf{y}_s^*$  is still linear. This fact allows the exogenous and endogenous interaction effects to be separately identifiable. The reasoning here is somewhat similar to, but different from, the discussion in Brock and Durlauf, 2005. In particular, the endogenous interaction specification in Blume and Durlauf, 2005 is not the same as the one used here; in Blume and Durlauf, the non-linear binary choice model specifies, in our notation, the latent variable vector  $\mathbf{y}_s^*$  as a function of a weighted average of the choice probabilities of other observation units (for the endogenous interaction effect), as opposed to our specification of  $\mathbf{y}_s^*$  as a weighted average of the latent variable vector  $\mathbf{y}_s^*$  (through the  $\mathbf{W} \mathbf{y}_s^*$  effect).

vector  $\boldsymbol{\varepsilon}_s$  to be correlated across sectors for each county  $q$ . To develop this multivariate count model, we start from the scalar version of the SD model of Equation (4), but now allow the vector  $\mathbf{x}_q$  to again be composed of  $K$  elements:

$$(5) \quad y_{qs}^* = \delta_s \sum_{q'=1}^Q w_{qq'} y_{q's}^* + \mathbf{b}'_s \mathbf{x}_q + \sum_{k=1}^K \pi_{sk} \sum_{q'=1}^Q w_{qq'} x_{q'k} + \varepsilon_{qs}, \quad y_{qs} = m_{qs} \Rightarrow \psi_{qs, m_{qs}-1} < y_{qs}^* < \psi_{qs, m_{qs}}.$$

An important point to be noted here is that the spatial dependency in counts is generated in the equation above through spatial “spillover” effects and spatial error correlation effects in the latent demand propensity (thus capturing inter-relationships in prevailing demand intensities), not through the localized county-specific characteristics that impact the supply tipping points.

### *Spatial Model Formulation and Estimation*

To proceed with the model formulation, we assume that the error terms  $\varepsilon_{qs}$  in Equation (5) are realizations of a standard normal error term uncorrelated across counties  $q$  (note that the use of a standard normal error term is innocuous, and is needed for identification). The  $\varepsilon_{qs}$  terms may be correlated across different sectors for the same county because of county-level unobserved factors that influence the demand for new businesses across sectors (see section entitled “Multivariate Business Count Model”). Formally, define  $\boldsymbol{\varepsilon}_q = (\varepsilon_{q1}, \varepsilon_{q2}, \varepsilon_{q3}, \dots, \varepsilon_{qs})'$ . Then,  $\boldsymbol{\varepsilon}_q$  is multivariate normally distributed with a mean vector of zeros and a correlation matrix as follows:

$$(6) \quad \boldsymbol{\varepsilon}_q = N \left[ \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_{12} & \rho_{13} & \dots & \rho_{1s} \\ \rho_{21} & 1 & \rho_{23} & \cdot & \cdot \\ \rho_{31} & \rho_{32} & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{s1} & \rho_{s2} & \cdot & \cdot & 1 \end{pmatrix} \right],$$

or  $\boldsymbol{\varepsilon}_q \sim N[\mathbf{0}, \boldsymbol{\Sigma}]$ .

The model framework we propose nests several other count modeling structures. Specifically, if all the off-diagonal elements of  $\boldsymbol{\Sigma}$  in Equation (6) are set to zero, and if  $\pi_{sk} = 0 \forall s, k$  and  $\delta_s = 0 \forall s$  in Equation (5), the result is independent flexible count (IFC) models for each sector (if, in addition,  $\alpha_{s, y_{qs}} = 0$  for all values of  $y_{qs}$  in Equation (2), the result is independent Poisson count models for different sectors). If the restriction that all the off-diagonal



elements are zero is relaxed from the IFC model, one gets the joint flexible count (JFC) model. If the restriction that  $\pi_{sk} = 0 \forall s, k$  is lifted from the JFC, the result is a spatially lagged joint flexible count model. Finally, if the restriction that  $\delta_s = 0 \forall s$  is also lifted, the result is the most general spatial Durbin joint flexible (SDJFC) model.

To write the demand propensity part of Equation (5) compactly, we define several vectors and matrices. Let  $\mathbf{y}_q^* = (y_{q1}^*, y_{q2}^*, \dots, y_{qS}^*)'$ ,  $\mathbf{y}_q = (y_{q1}, y_{q2}, \dots, y_{qS})'$ , and  $\mathbf{m}_q = (m_{q1}, m_{q2}, \dots, m_{qS})'$  be  $(S \times 1)$  vectors of vertically stacked propensities and observed count outcomes, respectively. Also, define additional vectors and matrices:  $\mathbf{y}^* = [(y_1^*)', (y_2^*)', (y_3^*)', \dots, (y_Q^*)']'$  ( $QS \times 1$  vector),  $\mathbf{y} = [(y_1)'], (y_2)'], (y_3)'], \dots, (y_Q)']'$  ( $QS \times 1$  vector),  $\mathbf{m} = (m_1', m_2', m_3', \dots, m_Q')'$  ( $QS \times 1$  vector),  $\boldsymbol{\varepsilon} = (\varepsilon_1', \varepsilon_2', \dots, \varepsilon_Q')'$  ( $QS \times 1$  vector),  $\mathbf{b} = (b_1', b_2', b_3', \dots, b_S')'$  ( $SK \times 1$  vector),  $\boldsymbol{\pi}_s = (\pi_{s1}, \pi_{s2}, \pi_{s3}, \dots, \pi_{sK})'$  ( $K \times 1$  vector),  $\boldsymbol{\pi} = (\boldsymbol{\pi}_1', \boldsymbol{\pi}_2', \boldsymbol{\pi}_3', \dots, \boldsymbol{\pi}_S')'$  ( $SK \times 1$  vector),  $\boldsymbol{\delta} = (\delta_1, \delta_2, \delta_3, \dots, \delta_S)$  ( $S \times 1$  vector),  $\tilde{\boldsymbol{\delta}} = \mathbf{1}_Q \otimes \boldsymbol{\delta}$  ( $QS \times 1$  vector;  $\mathbf{1}_Q$  is a vector of size  $Q$  with all elements equal to 1),  $\tilde{\mathbf{x}}_q = \mathbf{I}_S \otimes \mathbf{x}'_q$  ( $S \times SK$  matrix;  $\mathbf{I}_S$  is an identity matrix of size  $S$ ),  $\tilde{\mathbf{x}} = (\tilde{\mathbf{x}}_1', \tilde{\mathbf{x}}_2', \tilde{\mathbf{x}}_3', \dots, \tilde{\mathbf{x}}_Q')'$  ( $QS \times SK$  matrix),  $z_{qk} = \sum_{q'=1}^Q w_{qq'} x_{q'k}$  (scalar),  $\mathbf{z}_q = (z_{q1}, z_{q2}, z_{q3}, \dots, z_{qK})'$  ( $K \times 1$  vector),  $\tilde{\mathbf{z}}_q = \mathbf{I}_S \otimes \mathbf{z}'_q$  ( $S \times SK$  matrix), and  $\tilde{\mathbf{z}} = (\tilde{\mathbf{z}}_1', \tilde{\mathbf{z}}_2', \tilde{\mathbf{z}}_3', \dots, \tilde{\mathbf{z}}_Q')'$  ( $QS \times SK$  matrix). With these definitions, the demand propensity part of Equation (5) may be re-written as:

$$\mathbf{y}^* = [\tilde{\boldsymbol{\delta}} \cdot * (\mathbf{W} \otimes \mathbf{I}_S)] \mathbf{y}^* + \tilde{\mathbf{x}} \mathbf{b} + \tilde{\mathbf{z}} \boldsymbol{\pi} + \boldsymbol{\varepsilon},$$

where the operation  $\mathbf{H} = \mathbf{M} \cdot * \mathbf{N}$  in the equation above is used to refer to the element by element product of a vector  $\mathbf{M}$  and a matrix  $\mathbf{N}$ , i.e.,  $\mathbf{H}_{i,j} = \mathbf{M}_i \cdot * \mathbf{N}_{i,j}$ . After further matrix manipulation, we obtain:

$$(7) \quad \mathbf{y}^* = \mathbf{C} \tilde{\mathbf{x}} \mathbf{b} + \mathbf{C} \tilde{\mathbf{z}} \boldsymbol{\pi} + \mathbf{C} \boldsymbol{\varepsilon}, \text{ where } \mathbf{C} = [\mathbf{I}_{QS} - \tilde{\boldsymbol{\delta}} \cdot * (\mathbf{W} \otimes \mathbf{I}_S)]^{-1}.$$

The expected value and variance of  $\mathbf{y}^*$  may be obtained from the above equation after developing the covariance matrix for the error vector  $\mathbf{C} \boldsymbol{\varepsilon}$ . To do so, note that the error vector  $\boldsymbol{\varepsilon}$  is distributed multivariate normal with a mean vector of zero and covariance matrix  $\mathbf{I}_Q \otimes \boldsymbol{\Sigma}$  (of size  $QS \times QS$ ). Then, we obtain  $\mathbf{y}^* \sim MVN_{QS}(\mathbf{B}, \boldsymbol{\Psi})$ , where

$$\mathbf{B} = \mathbf{C}\tilde{\mathbf{x}}\mathbf{b} + \mathbf{C}\tilde{\mathbf{z}}\boldsymbol{\pi} \text{ and } \boldsymbol{\Psi} = \mathbf{C}[\mathbf{I}_Q \otimes \boldsymbol{\Sigma}] \mathbf{C}'.$$

The parameter vector to be estimated in the SDJFC model is  $\boldsymbol{\theta} = (\mathbf{b}', \boldsymbol{\pi}', \boldsymbol{\delta}', \boldsymbol{\gamma}', \boldsymbol{\alpha}', \boldsymbol{\Omega}')$ , where  $\boldsymbol{\alpha}$  is a column vector obtained by vertically stacking the  $\alpha_{s,r}$  ( $r = 0, 1, 2, \dots, L_s$ ) parameters across all sectors and  $\boldsymbol{\Omega}$  is a column vector obtained by vertically stacking all the correlation parameters (*i.e.*, off-diagonal elements of  $\boldsymbol{\Sigma}$ ). The likelihood function for the model is:

$$(8) \quad L(\boldsymbol{\theta}) = P(\mathbf{y} = \mathbf{m}) = \int_{D_{\mathbf{y}^*}} \phi_{QS}(\mathbf{y}^* | \mathbf{B}, \boldsymbol{\Psi}) d\mathbf{y}^*,$$

where  $D_{\mathbf{y}^*} = \{\mathbf{y}^* : \psi_{(qs, m_{qs}-1)} < y_{qs}^* < \psi_{qs, m_{qs}}, \forall q = 1, 2, \dots, Q, s = 1, 2, \dots, S\}$  and  $\phi_{QS}(\cdot)$  is the multivariate normal density function of dimension  $QS$ . The integration domain  $D_{\mathbf{y}^*}$  is simply the multivariate region of the elements of the  $\mathbf{y}^*$  vector determined by the observed vector of count outcomes. The dimensionality of the rectangular integral in the likelihood function is  $QS$ . Existing estimation methods including the Maximum Simulated Likelihood (MSL) method and the Bayesian Inference method become cumbersome and encounter convergence problems even for moderately sized  $Q$  and  $S$  (Bhat et al., 2010). The alternative is to use the composite marginal likelihood (CML) approach. In the current study, we use the pairwise composite marginal likelihood method based on the product of the likelihood contributions from pairs of counties across all sectors. To write this function, define threshold vectors as follows:

$$\boldsymbol{\varphi}_q = (\psi_{q1, m_{q1}-1}, \psi_{q2, m_{q2}-1}, \dots, \psi_{qS, m_{qS}-1}) \text{ and } \boldsymbol{\vartheta}_q = (\psi_{q1, m_{q1}}, \psi_{q2, m_{q2}}, \dots, \psi_{qS, m_{qS}}) \text{ (} S \times 1 \text{ vectors)}$$

$$\boldsymbol{\varphi} = (\boldsymbol{\varphi}'_1, \boldsymbol{\varphi}'_2, \dots, \boldsymbol{\varphi}'_Q)' \text{ and } \boldsymbol{\vartheta} = (\boldsymbol{\vartheta}'_1, \boldsymbol{\vartheta}'_2, \dots, \boldsymbol{\vartheta}'_Q)' \text{ (} QS \times 1 \text{ vectors)}$$

Let  $g$  be an index that can takes the values from 1 to  $QS$ . Then,

$$(9) \quad L_{CML}(\boldsymbol{\theta}) = \left( \prod_{g=1}^{QS-1} \prod_{g'=g+1}^{QS} P([\mathbf{y}]_g = [\mathbf{m}]_g, [\mathbf{y}]_{g'} = [\mathbf{m}]_{g'}) \right) \\ = \left( \prod_{g=1}^{QS-1} \prod_{g'=g+1}^{QS} \left[ \begin{array}{c} \Phi_2(\tilde{\boldsymbol{\varphi}}_g, \tilde{\boldsymbol{\varphi}}_{g'}, \mathbf{v}_{gg'}) - \Phi_2(\tilde{\boldsymbol{\varphi}}_g, \tilde{\boldsymbol{\vartheta}}_{g'}, \mathbf{v}_{gg'}) \\ -\Phi_2(\tilde{\boldsymbol{\vartheta}}_g, \tilde{\boldsymbol{\varphi}}_{g'}, \mathbf{v}_{gg'}) + \Phi_2(\tilde{\boldsymbol{\vartheta}}_g, \tilde{\boldsymbol{\vartheta}}_{g'}, \mathbf{v}_{gg'}) \end{array} \right] \right),$$

$$\text{where } \tilde{\boldsymbol{\varphi}}_g = \frac{[\boldsymbol{\varphi}]_g - [\mathbf{B}]_g}{\sqrt{[\boldsymbol{\Psi}]_{gg}}}, \tilde{\boldsymbol{\vartheta}}_g = \frac{[\boldsymbol{\vartheta}]_g - [\mathbf{B}]_g}{\sqrt{[\boldsymbol{\Psi}]_{gg}}}, \mathbf{v}_{gg'} = \frac{[\boldsymbol{\Psi}]_{gg'}}{\sqrt{[\boldsymbol{\Psi}]_{gg}} \sqrt{[\boldsymbol{\Psi}]_{g'g'}}}.$$

In the above expression,  $[\mathcal{G}]_g$  represents the  $g^{th}$  element of the column vector  $\mathcal{G}$ , and similarly for other vectors.  $[\Psi]_{gg'}$  represents the  $gg'th$  element of the matrix  $\Psi$ . The CML estimator is obtained by maximizing the logarithm of the function in Equation (9).

Under usual regularity assumptions, the CML estimator of  $\theta$  is consistent and asymptotically normal distributed with asymptotic mean  $\theta$  and covariance matrix given by the inverse of Godambe's (1960) sandwich information matrix (see Zhao and Joe, 2005):

$$(10) \quad \mathbf{V}_{CML}(\hat{\theta}) = [\mathbf{G}(\theta)]^{-1} = [\mathbf{H}(\theta)]^{-1} \mathbf{J}(\theta) [\mathbf{H}(\theta)]^{-1},$$

$$\text{where } \mathbf{H}(\theta) = E \left[ - \frac{\partial^2 \log L_{CML}(\theta)}{\partial \theta \partial \theta'} \right] \text{ and } \mathbf{J}(\theta) = E \left[ \left( \frac{\partial \log L_{CML}(\theta)}{\partial \theta} \right) \left( \frac{\partial \log L_{CML}(\theta)}{\partial \theta'} \right) \right].$$

The reader is referred to CPB for complete details regarding the estimation of the matrices  $\mathbf{H}(\theta)$  and  $\mathbf{J}(\theta)$  in Equation (10) above. To ensure the constraints on the autoregressive terms  $\delta_s$ , we parameterize each of these terms as  $\delta_s = 1/[1 + \exp(\tilde{\delta}_s)]$ . Once estimated, the  $\tilde{\delta}_s$  estimate can be translated back to estimate of  $\delta_s$ .

### *Model Selection*

For the purpose of comparing two nested models estimated using the CML approach, the analyst can use the adjusted composite likelihood ratio test (*ADCLRT*) statistic, which is asymptotically chi-squared distributed similar to the likelihood ratio test statistic for the maximum likelihood approach. The reader is referred to Bhat (2011) for details regarding the *ADCLRT* test statistic.

## 4. DATA DESCRIPTION AND VARIABLE SPECIFICATION

In the current study, we model the count of new businesses in 2007-2008 by sector in each of the 254 counties in Texas. The data for the analysis is drawn from several sources. Information on the count of new businesses by sector in each county (which is the dependent variable in our analysis) is primarily extracted from the U.S. Census Bureau website that hosts county-level business and employment (BE) data (see <http://www.census.gov/econ/susb/data/susb2008.html>). The variable forming the basis for constructing the dependent variable in the current study is the number of "Establishment Births"

in 20 major sectors (defined based on the first two digits of the North American Industry Classification System (NAICS) codes) in each county during the period 2007-2008. The Statistic of U.S. Businesses program, which forms the basis for the BE data, defines an “Establishment” as *“a single physical location where business is conducted or where services or industrial operations are performed”* and “Establishment births” as *“establishments that have zero employment in the first quarter of the initial year and positive employment in the first quarter of the subsequent year”*.

For the current analysis, we translated the 20-sector classification into an 11-sector classification because some sectors had zero births for almost all counties, leading to a situation where the variation in births was not adequate to tease out the effects of county-level explanatory variables. Table 1 presents the mapping between the first two digits of the NAICS codes in the original data and the new merged 11-sector sector classification, including the label we will use in the rest of the paper for each of the 11 sectors. The final two columns of Table 1 show the percentage distribution of births in different sectors within Texas and the US as a whole, respectively, for 2007-2008 (the percentages add up to 100 percent within each column). Overall, the births in each sector as a percentage of total new births are remarkably similar for Texas and the US as a whole. The largest disparities are for the construction and services sectors (where Texas shows a lower percentage of births than in the country), and the agriculture and finance sectors (where Texas shows a higher percentage of births than in the country). Across all sectors, the highest percentages of births are in the trade, services, finance, and administration sectors.

The total number of new businesses per county varies between 0 and 1432, with an average of about 18 new businesses per county (in the rest of the paper, we will consistently use the term “new businesses” to refer to the number of establishment births during the period 2007-2008 year, and the term “existing businesses” to refer to the number of establishments prior to the period 2007-2008). Figure 1 provides the percentage distribution of new businesses across all Texas counties for the four sectors with the highest percentage of new businesses. The large number of counties with zero count values is clearly discernible for all sectors, though there are also secondary spikes at other count values (such as the higher number of counties with seven new businesses than six new businesses for the service and administration sectors). Such count accumulations (or inflations) in discrete probability mass can be accommodated in our proposed model using the threshold parameters  $\alpha$ . The figure also indicates substantial variation in the

range and distribution profiles for the different sectors. For example, nearly 26 percent of the counties have zero new businesses in the finance sector, while the corresponding figure is only 14 percent for the trade sector. These variations point to the need to analyze new business counts by sector rather than pooling businesses of all sectors together.

### *Variable Specification*

Table 2 presents the county-specific sample statistics of the explanatory variables considered in our analysis. The second column describes the variable, while the third column provides information on the data sources used for construction of the variable. Almost all variables have been constructed for the year 2006 or slightly before (depending on data availability), to recognize that new business decisions in an upcoming year will depend on circumstances in the earlier year(s). The variables were introduced as such as well as in a spatially lagged form to accommodate possible exogenous spatial interaction effects originating from inter-county spillover effects in the explanatory variables (see discussion in sections entitled “Spatial Dependency” and “Basic Count Model Recasting”). Also, we consider the variables both as determinants of the latent demand intensity  $y_{qs}^*$  for new businesses as well as the supply of new businesses through the tipping points  $\psi_{qs,l}$  (see discussion in section entitled “Basic Count Model Recasting”). As can be observed from Table 2, many data sources, as listed at the bottom of the table, were used to assemble the county-specific explanatory variables. The explanatory variables may be grouped into six categories, as identified in the first column of Table 2. The theoretical and conceptual basis for inclusion of these explanatory variables is discussed below.

*Agglomeration Economies/Diseconomies.* The costs and benefits in firm location decisions are oftentimes evaluated in the context of agglomeration economies and agglomeration diseconomies. “Agglomeration economies” refer to the benefits that firms obtain when locating near each other due to factors such as availability of raw products, skilled labor, and readily available markets for manufactured products. In contrast, “agglomeration diseconomies” refer to the negative effects that firms experience when locating near each other because of such factors as increased competition and congestion effects on the infrastructure. Agglomeration economies are further distinguished into localization economies (concentration of similar activities as the

firm making the location decision in a region) and urbanization economies (general clustering of economic activity in a region across all types of firms). There is now a vast body of empirical literature that shows that agglomeration effects are an important cause of the uneven distribution of economic activities in a region. Most of this body of literature originates in the field of regional science that explores firm locations at an aggregate territorial level, as does our current study (see Rosenthal and Strange, 2004; Groot et al., 2009). However, the presence of agglomeration effects also has support from a more basic microeconomic increasing returns/imperfect competition-based theory and economies of scale-related considerations at the individual firm level, originating from what has now come to be referred to as the New Economic Geography theory (see Fujita and Thisse, 2003; Krugman, 1995). In this theory, Marshallian-type (Marshall, 1980) external localization economies arise not only from convenient access to specialized raw product suppliers (and therefore to lower per unit input production costs) and skilled labor (leading to lower search costs) at a location, but also due to knowledge and other spillovers by way of diffusion of ideas, innovations, products, and technology through firm contacts and/or employee interactions across firms. Similarly, Isard-type (Isard, 1956) external urbanization economies arise because of an association with a well-diversified labor pool, good public infrastructure, knowledge-generating education/research institutions that foster innovation, and greater overall stability and lower transaction costs due to large internal markets. The reader is referred to Fujita and Thisse (2003), Arauzo-Carod et al. (2010), and Mota and Brandão (2013) for extensive theoretical and conceptual discussions on agglomeration effects (see also Jacobian-type externalities discussed in the next section).

The variables in the agglomeration category in our aggregate county-level analysis include population density (residential population per square mile), population change from the beginning of 2000 to the end of 2006, total number of existing businesses per square mile, number of existing trade businesses per square mile, number of existing trade sector employees per square mile (as a proxy for employment density), and sector-specific number of existing businesses per square mile.<sup>5,6</sup> In addition to the linear effects, we also considered squared terms

---

<sup>5</sup> The number of existing businesses per square mile varies by sector; but, due to space considerations, we do not provide the descriptive statistics for each sector separately. Rather, in Table 2, the descriptive statistics corresponding to the “Number of existing businesses per square mile by sector” represent overall statistics across all sectors.

of these variables to capture potential inverted U-shaped effects of spatial concentration on demand for new businesses (see Viladecans-Marsal, 2004; Arauzo-Carod et al., 2010). The conceptual basis for this is that while agglomeration economies are likely to be at play initially (as discussed under localization and urbanization economies), a substantial level of clustering of firms can lead to negative externalities and agglomeration diseconomies.

*Specialization Indices and Firm Size.* Two sets of indices; the Location Quotient (*LQ*) index (see Gabe and Bell, 2004) and the Herfindahl-Hirschman Index (*HHI*) (see Duranton and Puga, 2000; Holl, 2004a); are developed to capture specialization effects. The Location Quotient (defined as the percentage of businesses in a given sector for each county normalized by the percentage of businesses in the same sector for the entire nation) is a sector-specific measure of specialization of a county relative to the specialization of the entire nation in the sector.

Mathematically, it can be written as:  $LQ_{qs} = \frac{EST_{qs} / TotEst_q}{Est_{Ns} / TotEst_N}$ , where  $Est_{qs}$  represents the number

of businesses in sector  $s$  in county  $q$  and  $TotEst_q$  refers to the total number of businesses in county  $q$ .  $Est_{Ns}$  and  $TotEst_N$  are corresponding numbers for the US. An  $LQ_{qs}$  (or relative specialization) value greater than one implies a higher degree of specialization in sector  $s$  in county  $q$  compared to the nation, while a value between zero and one implies a lower degree of specialization in sector  $s$  in county  $q$ .<sup>7</sup> The *LQ* index is a relative measure of sector specialization that is distinctly different from agglomeration-based localization economies that correspond to the absolute density of businesses in a sector. Thus, it is possible for a county to have a high level of the *LQ* index in a particular sector even if the number of businesses in that sector in the county is low. This can happen if the total number of businesses in the county is so low that the relative specialization in a particular sector is still higher than the national-level specialization in that sector. The *LQ* effect captures similar effects as the localization economies, because a higher concentration of businesses in a sector will imply a higher human capital in that county that is skilled in, and looking for employment in, that sector. In addition, the economies related to

---

<sup>6</sup> The trade-related variables are included because earlier studies have suggested that trade sector activity sometimes is a better predictor of urbanization economies/diseconomies than total economic activity, at least for some business sectors (see Kim et al., 2008).

<sup>7</sup> The *LQ* values vary across sectors. However, to conserve on space, we present descriptive statistics (in Table 2) across all sectors for this variable.

inputs and other services (such as raw materials and well established market pipelines) may make it beneficial to firms (and counties) to invest more in a sector that the county is already invested substantially in (see Gabe and Bell, 2004; Alañón-Pardo and Arauzo-Carod, 2011; Alamá-Sabater et al., 2011 for similar results).

The second index, *HHI*, is an overall (sector-independent) measure of specialization computed as  $HHI_q = \sum_{s=1}^S \left( \frac{Est_{qs}}{TotEst_q} \right)^2$ . The value of *HHI* ranges between zero and one. A higher

value for *HHI* indicates more specialization, with *HHI*=1 representing the case of full specialization in one sector. Note also that several earlier studies have used the *HHI* index as a measure of agglomeration (urbanization) economies. However, as with the difference between the *LQ* index and localization economies, there is also a difference between the *HHI* index and the measures used in our study for urbanization economies. In particular, while the *HHI* index is a sector-independent measure of specialization index, it does not capture a measure of the absolute levels of existing businesses. Thus, a high *HHI* index can be achieved even with low values for the measures used in this paper for urbanization economies. In this regard, the *HHI* index captures Jacobian-type (Jacobs, 1969) external specialization effects; that is, firms may find areas with a high level of current specialization (that is, existing businesses focused in few industry sector types) less appealing to enter into and/or counties already specialized in specific sectors may want to continue doing so because that is where their competitive advantage lies. As discussed by Frenken et al. (2007) and Bok and Oort (2011), and based particularly on New Economic Geography concepts, greater specialization dampens “interaction, generation, replication, modification, and recombination of ideas and applications across industries” (Bok and Oort, 2011), making both a firm less likely to locate in such counties as well as counties less likely to be welcoming of new business entrants.

In addition to the above two indices, the average size of existing businesses in each sector (defined as the average number of employees across existing businesses within the sector) is used to capture the effect of existing firm size within each sector.

*Human Capital.* Counties with substantial human capital will have a high demand for businesses to locate there (so their work force can be gainfully employed), while businesses need



skilled labor to carry out their production and other related activities. Thus, human capital should impact the number of new businesses, which is accommodated in our analysis through variables representing education achievement, average wage per job, and unemployment rate in the county.<sup>8</sup> Human capital effects have been considered in earlier studies, with the general conceptual notion that higher education (Mota and Brandão, 2013) and better wages (see, for example, Basile et al., 2009) capture both market size and its accessibility, while unemployment rate is a negative indicator of market size and accessibility.

*Fiscal Conditions.* The fiscal conditions of local governments can determine both the demand for new businesses as well as the intensity of efforts to woo new businesses. In this study, we employ property tax rates in each county and total government expenditure in each county to study the effects of fiscal conditions on new business decisions.<sup>9</sup> These fiscal conditions serve as indicators for the potential market for the final demand, access to customers, as well as suppliers of intermediate inputs. High property tax rates have been found, in general, to lead to a negative effect on the count of new businesses, a possible result of firms choosing not to enter into counties with high property tax rates (see Jofre-Monseny and Solé-Ollé, 2010; Guimaraes et al., 2004). On the other hand, government expenditure is typically viewed as potential for growth and upkeep by firms and may be expected to have a positive impact on the birth of new businesses (Holl, 2004c).

*Transportation Infrastructure and Land Development Characteristics.* Transportation infrastructure and land development intensity in the county can substantially influence the profit margin projections of firms, because the movement of raw materials (from suppliers) and products (to markets) is primarily based on the infrastructure and location of the region. Thus, our strong expectation is that these variables will affect the supply tipping points  $\psi_{qs,l}$  much more so than the demand intensity  $y_{qs}^*$ . The transportation infrastructure attributes include the length of roadway network, number of airports, and dummy variables for the presence of interstate highways and sea-shorelines (*i.e.*, coast areas). In addition, the extent of built density in

---

<sup>8</sup> We used 2006-2010 five year estimates for the education variables, since specific 2006 data were not available.

<sup>9</sup> We considered government expenditure in specific categories such as public assistance, highway construction, and highway maintenance. However, these variables did not come out to be statistically significant in our analysis.

each county is captured by designating each of the 254 counties of Texas in one of the following three categories: (a) metropolitan, (b) non-metropolitan but adjacent to a metropolitan county, or (c) non-metropolitan and non-adjacent to any metropolitan county. This was accomplished using Beale's codes (also known as Rural-Urban Continuum Codes; see USDA, 2003). Further, we also considered land area in our analysis. All these variables provide measures of proximity to transportation infrastructure access points and markets (Bok and Oort, 2011). As indicated by Hakimi (1964) and Louveaux et al. (1982), transportation costs have a substantial bearing on firm locations. In general, if there are no fixed transportation costs (that is, there are no costs independent of the length of haul), then an extension of Hakimi's theorem applies and states that firms will gravitate toward transportation network nodes, input and output market places, and intermodal hubs. In our analysis, the length of the roadway network in a county may be viewed as a proxy for the number of transportation network nodes, while the number of airports and the presence of interstate highways and sea-shorelines capture intermodal hub presence. Similarly, the characterization of counties based on built density and land area may be viewed as indicators of the intensity of input and output market places.

*County-Specific Factors.* In addition to the variables discussed above, we also introduced indicator variables for five counties: Harris, Dallas, Tarrant, Bexar, and Travis County during model estimations. This is because we observed that, during 2007-2008, over 50 percent of new businesses located in one of these five counties. These counties were so different and unique from the remaining 254 counties in terms of the dependent variable values and exogenous variables (such as population and population changes, and human capital) that we had to introduce county-specific factors to accommodate the unique "brand" location these counties seemed to hold in the perception space of firms.

## 5. EMPIRICAL ANALYSIS

We considered all the variables presented in Table 2 for our analysis. Several variable specifications and functional forms (including a logarithmic transformation and higher order effects of continuous variables) were tested to arrive at the final specification. Dummy variables created from the continuous variables were also considered to capture non-linear effects. All variables were introduced both in the latent demand propensity and threshold specifications. The

model estimation process was guided by prior research, intuitiveness, and parsimony considerations. A few variables that were only marginally statistically significant (*i.e.*, not significant at the 0.05 level of significance) were retained in the final model specification because of their intuitive effects and potential to guide future research efforts.

We estimated three different model formulations, as discussed in the methodology section: IFC, JFC, and SDJFC models. However, in this section, we present and discuss the estimation results corresponding only to the SDJFC model to focus the presentation and conserve on space. However, we will discuss data fit considerations from all three models.<sup>10</sup>

Tables 3a and 3b presents the estimation results corresponding to the final specification of the SDJFC model. In the following discussion, we first discuss variable effects on the demand intensity  $y_{qs}^*$ , and then on the supply tipping points  $\psi_{qs,l}$ .

#### *Demand Intensity Effects*

*Agglomeration Economies/Diseconomies.* The effects of agglomeration economies and diseconomies have been studied extensively in the business location modeling literature. We include variables to accommodate both urbanization and localization economies/diseconomies. Urbanization economies relate to the benefits a business can accrue because of the overall density of economic activity and cross-sectoral agglomeration effects, while urbanization diseconomies reflect infrastructure congestion effects and high competitive forces. Localization economies relate to benefits that a business in a specific sector can obtain through clustering of business in the same sector, attributable to information/knowledge spillovers and specialized labor availability, while localization diseconomies can be related to high competition of human capital.

The results in Table 3a are consistent with our initial hypothesis that the variables in this category will primarily affect demand intensity. Among the many variables, the effect of residential population per square mile and resident population change influenced the number of new businesses through the supply tipping points, but not the demand intensity (these effects are discussed later). The other urbanization economies/diseconomies variables representing the total and trade numbers of existing businesses, and the number of existing trade employees per square

---

<sup>10</sup> The detailed estimation results for the IFC and JFC models are posted as a supplementary document by Bhat et al., 2013 at <http://www.caee.utexas.edu/prof/bhat/ABSTRACTS/Firm/SuppNote.pdf>.

mile, have the expected positive sign, supporting the urbanization economies hypothesis. The positive coefficient for the arts sector on the “square of the total number of existing businesses per square mile” variable (see the last variable under “agglomeration economies/diseconomies” in Table 3a) reveals a relatively high urbanization economies effect for the arts sector. The coefficient on the corresponding spatially lagged variable is also positive indicating strong inter-county urbanization economies for the arts sector. Furthermore, the coefficient on the spatially lagged variable corresponding to the number of existing trade businesses is positive for the construction sector, suggesting the presence of significant inter-county spillover effects because of which trade businesses in the neighboring Counties increase the latent demand intensity for the construction sector.<sup>11</sup> Also, pure localization economies are at work for the trade and administration sectors (see the positive effects on the “sector-specific number of existing businesses per square mile” variable for the trade and administration columns). In contrast, both urban and local economies are not present for the agricultural sector, perhaps because this sector may be viewed as a relatively stand-alone sector that is not very sensitive to inter-sectoral or intra-sectoral activity. Interestingly, however, there also is a highly statistically significant negative effect on the “square of the total number of existing businesses per square mile” for all sectors except the arts sector, suggesting the presence of urbanization diseconomies for all non-art sectors beyond a certain threshold agglomeration density. In fact, for the agricultural sector, the results suggest only agglomeration diseconomies. Among the other sectors, one can estimate the threshold agglomeration density at which urbanization economies turn into diseconomies for five of the remaining nine sectors. These values are: manufacturing (16.3 businesses per square mile), transportation (19.2 businesses per square mile), finance (8.5 businesses per square mile), health (8.6 businesses per square mile), and hospitality (8.1 businesses per square mile).<sup>12</sup> In terms of the distribution of the total number of existing businesses across counties that existed in

---

<sup>11</sup> Note here that while the spatially lagged variables just discussed have larger coefficients than the corresponding non-lagged variables, this does not mean that the (inter-county) spillover effects are higher than the corresponding intra-county explanatory variable effects. This is because a unit change in an explanatory variable within a county  $q$  implies a change in the latent demand intensity for businesses in that county that is equal to the non-lagged coefficient. On the other hand, a unit change in an explanatory variable of another neighboring county  $q'$  will have very little change on the spatially lagged variable value for county  $q$  (because this variable is a weighted average of the explanatory variable across many counties). Thus, the effects of changes in neighboring counties will still be much smaller than the intra-county effect.

<sup>12</sup> For the construction, trade, and services sectors, the urbanization economies are reflected through the “number of existing trade sector businesses per square mile” variable, while the diseconomies are manifested in the “square of the number of existing businesses per square mile” variable; thus, one cannot compute a threshold agglomeration density.

2007, the results suggest that agglomeration diseconomies are at work in 0.8 percent of the counties for the manufacturing sector, 0.8 percent of the counties for the transportation sector, 2.0 percent of the counties for the finance sector, 2.0 percent of the counties for the health sector, and 2.0 percent of the counties for the hospitality sector. These figures indicate that, given the existing number of businesses, most counties are still operating at the agglomeration economies range and have not yet reached the threshold point where agglomeration diseconomies have set in.

To summarize, the results of the agglomeration economies/diseconomies variables provide four important insights: (1) The arts sector is associated with a relatively high positive sensitivity to urbanization economies, (2) There is an inverted U-shaped profile for the agglomeration effect for many sectors, (3) Urbanization economies are, in general, much more dominant than localization economies (see Holl, 2004c and Arauzo-Carod, 2005 for a similar result), and (4) Agglomeration benefits accrue longer for the manufacturing and transportation sectors than for the finance, health, and hospitality sectors.

*Specialization Indices and Firm Size.* The two sector specialization indices (location quotient (*LQ*) and Herfindahl-Hirschman index (*HHI*)) and firm size have statistically significant effects on the demand intensity for most sectors. The positive coefficients on the *LQ* variable for all sectors except the hospitality and administration sectors imply that counties that are more specialized in a specific sector relative to national norms have a higher demand intensity for new businesses in that sector, as hypothesized earlier and consistent with earlier studies. The coefficient on the *HHI* index is consistently negative and statistically significant for all sectors except the manufacturing and administration sectors. This indicates a lower (higher) demand intensity for new businesses among counties with higher overall specialization (diversity), presumably a reflection of greater diversity providing improved access to resources, rounded human capital, and markets. Also, the coefficient on the corresponding spatially lagged *HHI* index is negative and statistically significant for trade, hospitality, and agriculture sectors. So, diverse business environment in the neighboring counties is conducive for starting new businesses in a county. It is also possible that new businesses in a county benefit from the

competitive environment created by a dynamic business environment in that county as well as neighboring counties (Arauzo-Carod, 2005; Alamá-Sabater et al., 2011).<sup>13</sup>

Finally, counties with larger-sized existing trade businesses have higher demand intensity for new trade-oriented businesses than counties with smaller-sized existing trade businesses. This result should be of particular interest to small trade businesses, since it suggests that they can locate and position themselves in counties with large trade employers, and still potentially benefit from an established retail customer base.

*Human Capital.* All the three types of variables (education achievement within the population group over 25 years of age, unemployment rates, and average wage per job) used to capture the availability of human capital in a county emerged as being significant determinants of the demand intensity for new businesses. However, the effects are sector-specific. In this regard, three observations may be made from the effects of the education variables: (a) For the agricultural, and transportation sectors, a more highly educated population leads to a lower demand for new businesses, which is not entirely surprising because these sectors tend to employ a good pool of individuals who are not necessarily highly educated (see Liviano and Arauzo-Carod, 2013; Cieslik, 2005; Arauzo-Carod and Manjón-Antolín, 2004; Tadesse and Ryan, 2004), (b) For the manufacturing, health, and arts sectors, the results indicate a U-shaped effect of education (that is, counties with a high fraction of individuals with a high school degree, which is the intermediate education category in Table 3a, have a lower demand intensity for new businesses than counties with populations having an overall lower or higher education level, and (c) For the hospitality, arts, and administration sectors, counties with high fractions of very well educated (bachelor's degree or higher) adult populations have the highest demand intensity for new businesses. These human capital-based effects need further exploration in future studies.

High unemployment rate in a county may be viewed as a proxy for the availability of labor force in the county seeking employment and thus may lead to higher demand intensity for new businesses. On the other hand, the difficult economic conditions in the county may also result in rigid labor markets discouraging new establishment of new businesses in that county (see Basile, 2004; Basile et al., 2004, 2010; Cieslik, 2005). While the latter effect should be

---

<sup>13</sup> Alamá-Sabater et al., 2011 considered a diversification index =  $1-HHI$  and found a positive sign on the diversification index, consistent with the results here.

better reflected in the supply tipping points  $\psi_{qs,l}$  than in the demand intensity, we did not find any statistically significant effect of unemployment rate on the supply thresholds. But the results clearly indicate the negative effect of high unemployment rate on the demand intensity for new businesses. The average wage per job in each county positively impacts the demand intensity for new businesses in the agriculture and transportation sectors. In addition, higher average wages per job in neighboring counties increases the demand intensity for new businesses in these two sectors, as indicated by the positive coefficients on the corresponding spatially lagged wage variable.

*Fiscal Conditions* . Of the two variables considered in the group of fiscal policy related variables – property tax rates and total government expenditure – the first variable has an impact through the demand intensity (discussed here), while the second variable has an impact through the supply tipping points (discussed later).<sup>14</sup> Specifically, higher property tax rates decrease the demand intensity for new businesses in ten of the eleven sectors (the hold out being the arts sector). Also, the parameter estimate on the corresponding spatially lagged tax variable is negative and statistically significant for four of the ten sectors whose demand intensity is affected by tax rates. This is perhaps a reflection of a reticence to bring in additional firms when a county and its neighboring counties already have high property taxes, given the uncertainty of what that may do to the already high property tax rates. Of course, this effect may also be the result of firms choosing not to enter into counties and clusters of counties with high property tax rates, though such an effect should be manifested through the supply tipping points (see Jofre-Monseny and Solé-Ollé, 2010; Guimaraes et al., 2004). Thus, while our GOR-based count model provides more of a structural basis than traditional count models, it still can be difficult to extricate demand and supply effects from count observations that reflect equilibrium between the two.

*Transportation Infrastructure and Land Development Characteristics*. As expected, the transportation infrastructure and land development characteristics affect the count of new businesses almost exclusively through the supply tipping points, suggesting that, while not

---

<sup>14</sup> Other forms of taxes such as income tax and corporate tax usually do not vary across counties within Texas. So, we considered only county level property tax rates in this study.

always being able to extricate demand and supply effects, our GOR-based count model is able to do so for many variables.

*County-Specific Factors.* The statistically significant parameter estimates on the five indicator variables for the five counties in Table 3a indicate the uniqueness of these counties relative to other Texas counties. These coefficients do not have any substantive interpretations, but capture the mean effects of unobserved factors in these counties. In general, the coefficients are positive, suggesting higher demand intensity for new businesses in these counties, except for a couple of specific sector-county combinations.

#### *The Supply Tipping Points Effects (or Threshold Effects)*

The supply tipping point parameters include the elements of the  $\alpha$  vector and the  $\gamma$  vector (see Equation 1 in Section 3). The elements of the  $\alpha$  vector do not have any substantive interpretations, but play the very important role of accommodating high or low probability masses for specific outcomes (after controlling for the effects of other exogenous variables). As indicated in Section 4, identification is achieved by specifying  $\alpha_{s,0} = 0$ ,  $\alpha_{s,m_{qs}} = \alpha_{s,L_s}$  if  $m_{qs} > L_s$ . In the present specification, we initially set  $L_s = 20$  for each sector  $s$  and progressively reduced  $L_s$  based on statistical significance considerations and general data fit. These  $\alpha$  vector elements obviate the need for cumbersome zero-inflated and hurdle treatments. Further, they not only accommodate excess zero effects, but also probability mass clusterings on other outcome values. The results in Table 3a illustrate the importance of incorporating such flexibility in multivariate count models, with the  $\alpha$  vector elements turning out to be statistically significant for many sector-outcome combinations.

The elements in the  $\gamma$  vector are presented next in Table 3a. The constants within the  $\gamma$  vector do not have any particular interpretation, given the inclusion of several continuous variables in the  $\mathbf{h}_q$  vector. For the other variables, a positive coefficient shifts all the thresholds toward the left of the demand intensity scale (see CPB for a detailed discussion), which has the effect of reducing the probability of the zero count for new businesses as the variable value increases (or, equivalently, increasing the attractiveness for firms to make positive business location decisions). A negative coefficient, on the other hand, shifts all thresholds toward the



right of the demand intensity scale, which has the effect of increasing the probability of the zero count for new businesses (or, equivalently, decreasing the attractiveness for firms to make positive business location decisions).

The variable effects on the supply tipping points are discussed next. In the following presentation, we do not discuss the effects by variable group (as done for the demand intensity function) because there are not too many variables affecting the supply tipping points (consistent with our expectation in the section entitled “Motivation for Spatial Model” that many more county characteristics will enter the demand intensity function rather than the supply tipping points).

*Variable Effects on Supply Tipping Points.* Table 3a reveals that residential population per square mile and residential population change affect the supply tipping points. Additional functional forms for these two variables were also considered, such as squared terms and other non-linear functional forms, but the linear forms provided the best results. The parameter estimates on these variables are consistently negative for the sectors in which they are statistically significant, implying that counties with high residential population densities and population changes are less likely to be chosen by firms as their business locations. This is perhaps related to high land costs, commercial land scarcity, and infrastructure congestion in the county (see Alañón-Pardo and Arauzo-Carod, 2011; Jofre-Monseny and Solé-Ollé, 2010; Barrios et al., 2006).

The coefficient on the natural logarithm of government expenditures variable is consistently positive and statistically significant in the threshold specifications of all sectors. This shows a clear positive bias of firms toward counties that invest in improving public facilities and other supporting infrastructure, a finding also observed in earlier studies. For instance, Bolinger and Ihlanfeldt (2003) found that higher government expenditure on fire safety and parks in a Census tract increases employment in that Census tract. More generally, firms are likely to view a proactive local government as a sign of a dynamic market environment.

Several transportation infrastructure and land development variables have a significant positive impact on the supply tipping points. The positive parameter estimates on the natural logarithm of the length (in miles) of the roadway network in the county is intuitive, since firms are likely to locate their businesses in counties with good transportation infrastructure. Similar,

but more sector-specific effects are observed for the presence of an interstate roadway and the number of airports in the county. The coefficients on the indicator variables for whether the county is metropolitan or adjacent to a metropolitan county reflect the tendency of firms in the manufacturing, trade, and finance sectors to locate their businesses in metropolitan counties or other neighboring counties rather than in non-metropolitan (non-adjacent to Metropolitan County) counties. This suggests a tendency to locate close to resources for inputs as well as markets for finished goods and services. Finally, firms in the manufacturing, finance, health, and hospitality sectors also prefer to locate their businesses in counties with large land areas, presumably capturing a general size effect as well as a lower land cost effect.

### *Cross-Correlation and Spatial Effects*

Table 3b presents the results corresponding to the correlation matrix capturing cross-sector dependencies within any given county. As mentioned earlier in the methodology section, the off-diagonal elements in  $\Sigma$  capture the error correlations between the underlying latent demand intensities of any given county for different sectors. In the final model specification, 35 out of the 55 off-diagonal elements in the correlation matrix  $\Sigma$  are statistically significant, a clear evidence that the counts of new businesses in different sectors are correlated due to unobserved factors (and strongly confirming the need to adopt a multivariate modeling approach). This is an important finding, given that all earlier studies in the business location modeling literature have ignored unobserved correlations in the counts for different sectors. Also, all the 35 correlation parameters in Table 3b are positive and reasonably strong, reflecting a complementary relationship among the demand intensities for new businesses in different sectors.

We tested several different specifications for constructing the weight matrix in the spatial model estimations, including inverse distance, inverse of the square of the distance, and inverse of the cube of the distance between counties. Among all weight matrix specifications considered, the best model fit was obtained using the inverse of the square of distance between the counties.<sup>15</sup> The spatial autoregressive parameter  $\delta_s$  corresponding to this weight matrix specification in the final spatial Durbin formulation is highly statistically significant for the manufacturing and finance sectors, with a positive value of 0.2755 (t-statistic of 6.45) and 0.2985

---

<sup>15</sup> The spatial models that use different weight matrices are not directly comparable, since they are non-nested models. We used the composite likelihood information criterion (CLIC) introduced by Varin and Vidoni (2005) to compare models with different weight matrix specifications. Details are available on request from the authors.

(t-statistic of 14.56), respectively. This result supports the hypothesis that the number of new businesses in a county is not just a function of its characteristics, but is also influenced by the observed factors (such as number of businesses, employment, and road network characteristics) and unobserved factors (such as high land prices, county zoning preferences, and public attitudes) of spatially proximate counties. The spatial autoregressive parameter might be coming out to be statistically insignificant for the remaining nine sectors because of the cancelling out of positive and negative spatial dependencies, as suggested by Griffith and Arbia (2010), and discussed in the section entitled “Spatial Dependency”.<sup>16</sup> Also, we would like to note that for some (but not all) of the remaining nine sectors, removing the spatially lagged explanatory variables led to a statistically significant spatial autoregressive parameter, suggesting, as in Corrado and Fingleton (2012), that a model including only endogenous interaction effects can be comingling “spurious” endogenous interaction effects with that of “true” exogenous interaction effects (*i.e.*, spatially lagged explanatory variable effects). However, the reverse was also true. That is, removing the statistically significant spatial autoregressive parameter for the finance sector immediately showed up as statistically significant exogenous interaction effects (that were not statistically significant otherwise), suggesting that a model including only exogenous interaction effects can also be comingling “spurious” exogenous interaction effects with “true” endogenous interaction effects. The spatial Durbin model in the current paper offers a way to accommodate both effects and allows the empirical data to disentangle the two effects.

### *Model Selection and Data Fit*

The spatial Durbin joint flexible count model (SDJFC) is superior to both the joint flexible count model (JFC) model and the independent flexible count model (IFC), as should be evident from the statistically significant spatial autoregressive parameter and the correlation parameters in Table 3b. An alternative way to compare data fit across these nested models is through the adjusted composite likelihood ratio test (*ADCLRT*) statistic. The composite log-likelihood (CLL) values for the SDJFC, JFC, and IFC models are -739449.88 (with 219 parameters), -741396.66 (with 207 parameters), and 742172.76 (with 172 parameters), respectively. The *ADCLRT*

---

<sup>16</sup> The spatial autoregressive parameters for the nine sectors in the model specification of Table 3b were very small in magnitude and statistically insignificant. They are as follows (t-statistics in parenthesis): Agriculture 0.0002 (0.0064), Construction 0.0580 (1.48), Trade 0.0250 (0.7924), Transportation 0.0004 (0.0218), Services 0.0004 (0.028), Health 0.0126 (0.9962), Hospitality 0.0002 (0.03), Arts 0.0047 (0.6173), and Administration 0.0145 (0.4004). These are not included in Table 3b to reduce clutter.

statistic for the comparison between the SJFC and JFC models is 23.51, which is greater than the critical chi-squared value corresponding to 12 degrees of freedom at the 5 percent level of significance. Similarly, the *ADCLRT* statistic for the comparison between the JFC model and the IFC model is 222.14, which is once again higher than the critical chi-squared table value with 35 degrees of freedom at any reasonable level of significance. This demonstrates very strong evidence of cross-sector dependence within any given county and spatial dependence across counties.

### *Application Demonstration*

The estimated parameters of the multivariate count model can be used to forecast the number of new businesses in a county in response to changing county economics and demographics. It can also be used by policy makers to evaluate the effects of potential policy actions. To demonstrate the ability of our multivariate model to provide forecasts, as well as to demonstrate the differences between considering spatial dependency effects and ignoring these effects, we apply the model to examine the impacts of increasing property tax rates by 20 percent on five sectors: manufacturing, transportation, services, finance, and administration. It is important to note that the model developed in this paper can be used in many different ways to examine a change in a variable in a specific county on the number of new businesses in each sector of that county (direct effects), each other county (indirect effects), and the entire state of Texas as a whole (total effects). But, to summarize these effects, it is typical to compute average effects. Further, the effects themselves can be computed in several ways. We do so in a way that is generalizable to any explanatory variable (whether it is a continuous explanatory variable or not) and to any magnitude of change in the explanatory variable (the procedure suggested in LeSage and Pace, 2009, on the other hand, is specific to continuous explanatory variables and to an infinitesimal change in an explanatory variable). The procedure we adopted to compute the elasticity effects is as follows. We first increase the tax rate in County 1 by 20 percent (while keeping all other values fixed), and compute the effect of this change on the expected value of the number of new businesses in each sector of County 1 (as discussed in the next paragraph). Subsequently, the percentage change (from the base case) in the expected number of businesses by sector in County 1 is computed, and designated as the direct effect corresponding to County 1. Similarly, the percentage change (from the base case) in the expected number of businesses by sector in

County 1 because of a 20 percent increase in the tax rate of all other counties (but not County 1) is obtained, and designated as the indirect effect corresponding to a change in the property tax in all neighboring counties. Finally, the overall percentage change (from the base case) in the expected number of businesses by sector in County 1 because of a 20 percent increase in the tax rate of all counties (including County 1) is also obtained, and labeled as the total effect for County 1. This process is repeated in turn for each of the 254 counties. Next, the overall measure of direct, indirect, and total percentage effects are obtained as the average across the county-specific direct, indirect, and total percentage changes, respectively. Note that the total percentage effect will not be equal to the sum of the direct and indirect effects because we work with percentage changes.<sup>17</sup>

The prediction process above requires the computation of the expected value of the number of new businesses in sector  $s$  in each county  $q$ :

$$(11) \quad E(y_{qs}) = \sum_{i=0}^{\infty} P(y_{qs} = i) \cdot i,$$

where  $P(y_{qs} = i)$  is the probability that  $i$  businesses of sector  $s$  will start in county  $q$ . Although the summation in the equation above extends until infinity in our count model, we consider counts only up to  $i = 60$  in our prediction procedure (this value represents the 95 percentile value of the count of new businesses across all counties and across all sectors in the estimation sample). Beyond the count value of 60, the probabilities are very close to zero and have little impact on the predictions.

The expected value in Equation (11) is a function of all the exogenous variable vectors  $\mathbf{x}_q \forall q = 1, 2, \dots, Q$  in the latent demand intensity expressions as well as a function of all the variable vectors  $\mathbf{h}_q \forall q = 1, 2, \dots, Q$  embedded in the thresholds in Equation (2). To estimate  $P(y_{qs} = i)$  in Equation (11), we simulate the  $QS \times 1$  – vector  $\mathbf{y}^*$  in Equation (7) five hundred times using the estimated values of  $\delta$ ,  $\mathbf{b}$ , and the  $QS \times 1$  – vector  $\boldsymbol{\varepsilon}$ . Subsequently, we compare

---

<sup>17</sup> Note also that the multivariate spatial model can provide the effect of variable changes in one county on the joint distribution of any combination of counts of new businesses by sector for each county in the region. In this demonstration exercise, we are effectively taking the marginals of each county-level joint distribution to predict the count in each individual sector within that county. In doing so, we are essentially ignoring the benefit of the joint modeling of the sectors within each county. That is, the focus of our demonstration is to illustrate the difference between the spatial and aspatial models, not the multivariate nature of our model. Thus, we present the results only for the SDJFC and IFC models in this section.

each of the 500 draws of the  $q^{th}$  element of  $\mathbf{y}^*$  with the corresponding thresholds for the  $q^{th}$  element from Equation (3), and assign the count value for each of the 500 draws based on this comparison. The share of each count prediction is taken across the 500 draws to estimate  $P(y_{qs} = i)$ . We also compute the standard errors and corresponding t-statistics (against the value of zero) for the elasticity effects by using 200 bootstrap draws from the sampling distributions of the estimated parameters.<sup>18</sup>

Table 4 presents the predicted effects, and the corresponding standard errors and t-statistics, for the IFC and SDJFC models. For instance, the entry in the first row of the table under the column entitled “Direct” for the IFC model indicates that an increase in the tax rate by 20 percent in a county would, on average, result in about a 19.41 percent decrease in the number of new businesses in the manufacturing sector in that county. There is no indirect effect for the IFC model because spatial dependency is entirely ignored. For the SDJFC model, on the other hand, there is both a direct and indirect effect, as may be observed from the table. The important point to note also is that most of the indirect effects in the SDJFC model are statistically significant at well beyond a 95 percent level of confidence for a one-tailed t-test (since we expect that an increase in property tax will reduce the number of new businesses). That is, the expected number of new businesses locating in a county is impacted by both the property tax in that county as well as other counties. While it may seem that the indirect effect is stronger than the direct effect, one needs to be cautious in drawing any such conclusion. This is because the indirect effect is being summarized here as the impact on a county of a 20 percent change in property tax in each other county. If we had computed the indirect effect as the effect of a 20 percent change in one county on the expected number of businesses in a neighboring county, this would be much smaller than the direct effect. The more important point to take away is that there are statistically significant indirect effects. The presence of the indirect effects, in turn, leads to a generally higher total effect in the SDJFC model relative to the IFC model. Of course, all the figures in Table 4 are average measures and for a single variable, and our proposed model can predict the impact of changes in any variable and for specific counties on each other county.

---

<sup>18</sup> For computational ease in bootstrapping, we fix the spatial parameter  $\delta$  and the  $\Sigma$  parameters at the estimated values so that the covariance matrix of  $\mathbf{y}^*$  ( $= \Psi$ ) is fixed. Further, in computing the elasticity effects of variables that influence only the demand intensity, we fix the supply tipping points, and in computing the elasticity effects of variables that influence only the supply tipping points, we fix the demand intensity.

In addition to the difference between the spatial and aspatial models, Table 4 indicates that the magnitude of the effects of variables varies across different sectors. This supports the multivariate model developed in this paper, rather than aggregating business count data across all sectors and estimating a univariate model.

## 6. CONCLUSIONS

This paper has developed a multivariate modeling framework for analyzing business counts in different sectors that simultaneously addresses several key econometric issues: (a) It accommodates over-dispersion and excess zero problems in the county-level business count by sector type, (b) it captures cross-correlation in counts across different sectors within the same county due to the presence of common county-level unobserved factors, and (c) it explicitly models spatial dependence effects across counties due to observed as well as unobserved factors. The resulting multivariate spatial model is estimated using the composite marginal likelihood inference approach, which is simple to implement and also does not involve simulation. From a conceptual perspective, the generalized ordered-response (GOR) reformulation of the count model offers a nice interpretive device to separate out the effects of demand and supply functions in business location choice decisions relative to the reduced form formulation of traditional count models.

The proposed model is applied to analyze new county-level business counts in eleven different sectors in the state of Texas. Several important factors including agglomeration economies/diseconomies, industrial specialization indices, human capital, fiscal conditions, and transportation infrastructure and land development characteristics are considered. The empirical results provide several substantive insights. In summary, variables representing agglomeration economies/diseconomies and fiscal conditions affect both the county demand intensity for new businesses as well as the supply function representing the desire of firms to start a business in the county. On the other hand, the industrial specialization and human capital considerations, according to our analysis, influence the count of new businesses solely through the demand intensity, while the transportation infrastructure and land development characteristics impact the count of new businesses solely through the supply function. Overall, many more county characteristics enter the demand intensity function rather than the supply function. In the context

of variables impacting the demand function, urbanization economies are much more dominant than localization economies.

The multivariate spatial count model proposed in the paper fits the data better than other more restrictive models that ignore the jointness of business counts across multiple sectors and/or the spatial nature of business counts. Further, the estimated impacts of variables are generally higher from the spatial model developed in this paper relative to more restrictive aspatial models. The estimated model may be used to forecast the number of new businesses in a county in response to changing county characteristics or to evaluate the effects of policy actions.



## REFERENCES

- Aguero-Valverde, Jonathan and Paul P. Jovanis. 2006. "Spatial Analysis of Fatal Injury Crashes in Pennsylvania," *Accident Analysis and Prevention*, 38(3), 618–625.
- Alamá-Sabater, Luisa, Andres Artal-Tur, and Jose Miguel Navarro-Azorín. 2011. "Industrial Location, Spatial Discrete Choice Models and the Need to Account for Neighbourhood Effects," *The Annals of Regional Science*, 47(2), 393–418.
- Alañón-Pardo, Ángel and Josep Maria Arauzo-Carod. 2011. "Agglomeration, Accessibility and Industrial Location: Evidence from Spanish Municipalities," Working paper, Facultad de Ciencias Económicas y Empresariales, Universidad Complutense de Madrid.
- Alañón-Pardo, Ángel, Josep Maria Arauzo-Carod, and Rafael Myro-Sánchez. 2007. "Accessibility, Agglomeration and Location," in J.M. Arauzo-Carod, and M.C. Manjón-Antolín (eds.), *Entrepreneurship, Industrial Location and Economic Growth*, Chentelham: Edward Elgar Publishing.
- Alfò, Marco and Antonello Maruotti. 2010. "Two-Part Regression Models for Longitudinal Zero-Inflated Count Data," *Canadian Journal of Statistics*, 38(2), 197–216.
- Arauzo-Carod, Josep Maria. 2005. "Determinants of Industrial Location: An Application for Catalan Municipalities," *Papers in Regional Science*, 84(1), 105–120.
- \_\_\_\_\_. 2008. "Industrial Location at a Local Level: Comments on the Territorial Level of the Analysis," *Tijdschrift voor Economische en Sociale Geografie*, 99(2), 193–208.
- Arauzo-Carod, Josep Maria and Miguel C. Manjón-Antolín. 2004. "Firm Size and Geographical Aggregation: An Empirical Appraisal in Industrial Location," *Small Business Economics*, 22(3), 299–312.
- \_\_\_\_\_. 2012. "(Optimal) Spatial Aggregation in the Determinants of Industrial Location," *Small Business Economics*, 39(3), 645–658.
- Arauzo-Carod, Josep Maria and Elisabet Viladecans-Marsal. 2009. "Industrial Location at the Intra-Metropolitan Level: The Role of Agglomeration Economies," *Regional Studies*, 43(4), 545–558.
- Arauzo-Carod, Josep Maria, Daniel Liviano-Solis, and Miguel C. Manjón-Antolín. 2010. "Empirical Studies in Industrial Location: An Assessment of their Methods and Results," *Journal of Regional Science*, 50(3), 685–711.
- Autant-Bernard, Corinne and James P. LeSage. 2011. "Quantifying Knowledge Spillovers using Spatial Econometric Models," *Journal of Regional Science*, 51(3), 471–496.
- Barrios, Salvador, Holger Gorg, and Eric Strobl. 2006. "Multinationals' Location Choice, Agglomeration Economies and Public Incentives," *International Regional Science Review*, 29(1), 81–107.
- Basile, Roberto. 2004. "Acquisition versus Greenfield Investment: The Location of Foreign Manufacturers in Italy," *Regional Science and Economics*, 34(1), 3–25.
- Basile, Roberto, Davide Castellani, and Antonello Zanfei. 2004. "Location Choices of Multinational Firms in Europe: The Role of National Boundaries and EU Policy," ERSA conference papers, European Regional Science Association.

- \_\_\_\_\_. 2009. "National Boundaries and the Location of Multinational Firms in Europe," *Papers in Regional Science*, 88(4), 733–748.
- Basile, Roberto, Luigi Benfratello, and Davide Castellani. 2010. "Location Determinants of Greenfield Foreign Investments in the Enlarged Europe: Evidence from a Spatial Autoregressive Negative Binomial Additive Model," Working paper No. 10, former Department of Economics and Public Finance, University of Torino.
- Becker, Randy and Vernon Henderson. 2000. "Effects of Air Quality Regulations on Polluting Industries," *Journal of Political Economy*, 108(2), 379–421.
- Bhat, Chandra R. 2011. "The Maximum Approximate Composite Marginal Likelihood (MACML) Estimation of Multinomial Probit-based Unordered Response Choice Models," *Transportation Research Part B*, 45(7), 923–939.
- Bhat, Chandra R. and Rupali Sardesai. 2006. "The Impact of Stop-making and Travel Time Reliability on Commute Mode Choice," *Transportation Research Part B*, 40(9), 709–730.
- Bhat, Chandra R., Rajesh Paleti, and Palvinder Singh. 2013. "Estimation Results of Aspatial Models: A Supplementary Note to *A Spatial Multivariate Count Model for Firm Location Decisions*," Department of Civil, Architectural and Environmental Engineering, The University of Texas at Austin. Available at: <http://www.cae.utexas.edu/prof/bhat/ABSTRACTS/Firm/SuppNote.pdf>.
- Bhat, Chandra R., Cristiano Varin, and Nazneen Ferdous. 2010. "A Comparison of the Maximum Simulated Likelihood and Composite Marginal Likelihood Estimation Approaches in the Context of the Multivariate Ordered-response Model," in W. Greene and R.C. Hill (eds.), *Advances in Econometrics: Maximum Simulated Likelihood Methods and Applications*, Bingley, UK: Emerald Group Publishing Limited, pp. 65–106.
- Blonigen, Bruce A. 1997. "Firm-Specific Assets and the Link between Exchange Rates and Foreign Direct Investment," *American Economic Review*, 87(3), 447–465.
- Blume, Lawrence E. and Steven N. Durlauf, 2005. "Identifying Social Interactions: A Review", Working Paper 12, Social Systems Research Institute, University of Wisconsin, Madison, WI.
- de Bok, Michael and Frank van Oort. 2011. "Agglomeration Economies, Accessibility, and the Spatial Choice Behavior of Relocating Firms," *Journal of Transport and Land Use*, 4(1), 5–24.
- Bolinger, Christopher R. and Keith R. Ihlanfeldt. 2003. "The Intraurban Spatial Distribution of Employment: Which Government Interventions Make a Difference?," *Journal of Urban Economics*, 53(3), 396–412.
- Castro, Marisol, Rajesh Paleti, and Chandra R. Bhat. 2012. "A Latent Variable Representation of Count Data Models to Accommodate Spatial and Temporal Dependence: Application to Predicting Crash Frequency at Intersections," *Transportation Research Part B*, 46(1), 253–272.

- Chen, Yali, Srinath Ravulaparthi, Kathleen Deutsch, Pamela Dalal, Seo Youn Yoon, Ting Lei, Konstadinos G. Goulias, Ram M. Pendyala, Chandra R. Bhat, Hsi-Hwa Hu. 2012. "Development of Indicators of Opportunity-Based Accessibility," *Transportation Research Record: Journal of the Transportation Research Board*, 2255, 58–68.
- Cieslik, Andrzej 2005. "Location of Foreign Firms and National Border Effects: The Case of Poland," *Tijdschrift voor Economische en Sociale Geografie*, 96(3), 287–297.
- City of Austin. 2012. "Imagine Austin," Comprehensive Plan, adopted by Austin City Council June 15, 2012.
- Corrado, Luisa and Bernard Fingleton, 2012. "Where is the Economics in Spatial Econometrics?," *Journal of Regional Science*, 51(2), 1-30.
- Disdier, Anne-Celia and Thierry Mayer. 2004. "How Different is Eastern Europe? Structure and Determinants of Location Choices by French Firms in Eastern and Western Europe," *Journal of Comparative Economics*, 32(2), 280–296.
- Duranton, Gilles and Diego Puga. 2000. "Diversity and Specialization in Cities: Why, Where and When Does it Matter?," *Urban Studies*, 37(3), 533–555.
- Franzese, Robert J. and Jude C. Hays. 2008. "Empirical Models of Spatial Interdependence," in J. Box-Steffensmeier, H., Brady and D. Collier (eds), *Oxford Handbook of Political Methodology*, New York: Oxford University Press, pp. 570–604.
- Frenken, Koen, Frank Van Oort, and Thijs Verburg. 2007. "Related Variety, Unrelated Variety and Regional Economic Growth," *Regional Studies*, 41 (5), 685–697.
- Fujita, Masahisa and Jacques-Francois Thisse. 2002. *Economics of Agglomeration: Cities, Industrial Location, and Regional Growth*. Cambridge: Cambridge University Press.
- \_\_\_\_\_. 2003. "Agglomeration and Market Interaction," in *Advances in Economics and Econometrics. Theory and Applications*, Eight World Congress, 1, Cambridge: Cambridge University Press, pp. 302–338.
- Fullerton, Don, Andrew Leicester, and Stephen Smith. 2008. "Environmental Taxes. Reforming the Tax System for the 21<sup>st</sup> Century," The Institute of Fiscal Studies. Available at: <http://www.ifs.org.uk/mirrleesreview/reports/environment.pdf>
- Gabe, Todd. 2003. "Local Industry Agglomeration and New Business Activity," *Growth and Change*, 34(1), 17–39.
- Gabe, Todd and Kathleen P. Bell. 2004. "Tradeoffs Between Local Taxes and Government Spending as Determinants of Business Location," *Journal of Regional Science*, 44(1), 21–41.
- Godambe, Vidyadhar P. 1960. "An Optimum Property of Regular Maximum Likelihood Estimation," *Annals of Mathematical Statistics*, 31(4), 1208–1211.
- Griffith, Daniel A. and Giuseppe Arbia. 2010. "Detecting Negative Spatial Autocorrelation in Georeferenced Random Variables," *International Journal of Geographical Information Science* 24(3), 417–437.

- Groot, Henri L.F. de, Jacques Poot, and Martijn J. Smit. 2009. "Agglomeration Externalities, Innovation and Regional Growth: Theoretical Perspectives and Meta-Analysis," in R. Capello and P. Nijkamp (eds.), *Handbook of Regional Growth and Development Theories*, Cheltenham: Edward Elgar, pp. 256–281.
- Guimaraes, Paulo, Octávio Figueiredo, and Douglas Woodward. 2004. "Industrial Location Modeling: Extending the Random Utility Framework," *Journal of Regional Science*, 44(1), 1–20.
- Hakimi. S. Louis 1964. "Optimum Locations of Switching Centers and the Absolute Centers and Medians of a Graph," *Operations Research*, 12(3), 450–459.
- Hanson, Andrew and Shawn Rohlin. 2011. "Do Location-based Tax Incentives Attract New Business Establishments?," *Journal of Regional Science*, 51(3), 427–449.
- Herriges, Joseph A., Daniel J. Phaneuf, and Justin L. Tobias. 2008. "Estimating Demand Systems When Outcomes are Correlated Counts," *Journal of Econometrics*, 147(2), 282–298.
- Holl, Adelheid. 2004a. "Start-Ups and Relocations: Manufacturing Plant Location in Portugal," *Papers in Regional Science*, 83(4), 649–668.
- \_\_\_\_\_. 2004b. "Manufacturing Location and Impacts of Road Transport Infrastructure: Empirical Evidence from Spain," *Regional Science and Urban Economics*, 34(3), 341–363.
- \_\_\_\_\_. 2004c. "Transport Infrastructure, Agglomeration Economies, and Firm Birth: Empirical Evidence from Portugal," *Journal of Regional Science*, 44(4), 693–712.
- Isard, Walter. 1956. *Location and Space-Economy: A General Theory Relating to Industrial Location, Market Areas, Land Use, Trade, and Urban Structure*. Cambridge, MA: MIT Press.
- Jacobs, Jane. 1969. *The Economies of Cities*. New York: Random House.
- Jofre-Monseny, Jordi and Albert Solé-Ollé. 2010. "Tax Differentials in Intraregional Firm Location: Evidence from New Manufacturing Establishments in Spanish Municipalities," *Regional Studies*, 44(6), 663–677.
- Jofre-Monseny, Jordi, Raquel Marín-López, and Elisabet Viladecans-Marsal. 2011. "The Mechanisms of Agglomeration: Evidence from the Effect of Inter-Industry Relations on the Location of New Firms," *Journal of Urban Economics*, 70(2-3), 61–74.
- Kim, Hyungtai, Paul Waddell, Venkataraman N. Shankar, and Gudmundur F. Ulfarsson. 2008. "Modeling Micro-Spatial Employment Location Patterns: A Comparison of Count and Choice Approaches," *Geographical Analysis*, 40(2), 123–151.
- Krugman, Paul. 1995. *Development, Geography, and Economic Theory*. Cambridge MA: MIT Press.
- Lee, Andy H., Kui Wang, Jane A. Scott, Kelvin K.W. Yau, and Geoffrey J. McLachlan. 2006. "Multi-Level Zero-Inflated Poisson Regression Modelling of Correlated Count Data with Excess Zeros," *Statistical Methods in Medical Research*, 15(1), 47–61.
- LeSage, James P., and Robert K. Pace. 2009. *Introduction to Spatial Econometrics*. Boca Raton, FL: Chapman & Hall/CRC, Taylor & Francis Group.

- Liviano, Daniel and Josep Maria Arauzo-Carod. 2012. "Industrial Location and Spatial Dependence: An Empirical Application," *Regional Studies*, forthcoming.
- \_\_\_\_\_. 2013. "Industrial Location and Interpretation of Zero Counts," *The Annals of Regional Science*, 50(2), 515–534.
- Louveaux, Francois, Jacques-Francois Thisse, and Hubert Beguin. 1982. "Location Theory and Transportation Costs," *Regional Science and Urban Economics*, 12(4), 529–45.
- Manjón-Antolín, Miguel C., and Josep Maria Arauzo-Carod. 2011. "Locations and Relocations: Modelling, Determinants, and Interrelations," *The Annals of Regional Science*, 47(1), 131–146.
- Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem," *The Review of Economic Studies*, 60(3), 531–542.
- Marshall, Alfred. 1980. *Principles of Economics*. New York: Prometheus.
- Mota, Isabel, and António Brandão. 2013. "The Determinants of Location Choice: Single Plants versus Multi-Plants," *Papers in Regional Science*, 92(1), 31–49.
- Müller, Gernot, and Claudia Czado. 2005. "An Autoregressive Ordered Probit Model with Application to High-frequency Financial Data," *Journal of Computational and Graphical Statistics*, 14(2), 320–338.
- Parent, Olivier, and James P. LeSage. 2008. "Using the Variance Structure of the Conditional Autoregressive Spatial Specification to Model Knowledge Spillovers," *Journal of Applied Econometrics*, 23(2), 235–256.
- Pendyala, Ram M., Chandra R. Bhat, Konstadinos G. Goulias, Rajesh Paleti, Karthik C. Konduri, Raghuprasad Sidharthan, Hsi-hwa Hu, Guoxiong Huang, and Keith P. Christian. 2012. "Application of Socioeconomic Model System for Activity-Based Modeling," *Transportation Research Record: Journal of the Transportation Research Board*, 2303, 71–80.
- Pinjari, Abdul R., Ram M. Pendyala, Chanda R. Bhat, and Paul A. Waddell. 2011. "Modelling the Choice Continuum: An Integrated Model of Residential Location, Auto Ownership, Bicycle Ownership, and Commute Tour Mode Choice Decisions," *Transportation*, 38(6), 933–958.
- Rosenthal, Stuart S., and William C. Strange. 2004. "Evidence on the Nature and Sources of Agglomeration Economies," in J.V. Henderson and J-F. Thisse (eds.), *Handbook of Regional and Urban Economics, Volume 4: Cities and Geography*, Amsterdam: Elsevier B.V., pp. 2119–2171.
- Singh, Palvinder, Rajesh Paleti, Sydney Jenkins, and Chandra R. Bhat. 2013. "On Modeling Telecommuting Behavior: Option, Choice, and Frequency," *Transportation*, 40(2), 373–396.
- Strotmann, Harald. 2007. "Entrepreneurial Survival," *Small Business Economics*, 28(1), 87–104.
- Tadesse, Bedassa, and Michael Ryan. 2004. "Host Market Characteristics, FDI, and the FDI-trade Relationship," *The Journal of International Trade & Economic Development*, 13(2), 199–229.

- United States Department of Agriculture (USDA). 2003. "Rural-Urban Continuum Codes 2003," Economic Research Service, USDA. Available at <http://www.ers.usda.gov/data-products/rural-urban-continuum-codes.aspx>.
- Van Acker, Veronique, and Frank Witlox. 2011. "Commuting Trips within Tours: How is Commuting Related to Land Use?," *Transportation*, 38(3), 465–486.
- Varin, Cristiano, and Paolo Vidoni. 2005. "A Note on Composite Likelihood Inference and Model Selection," *Biometrika*, 92(3), 519–528.
- Ver Hoef, Jay M., and John K. Jansen. 2007. "Space-Time Zero-Inflated Count Models of Harbor Seals," *Environmetrics*, 18(7), 697–712.
- Viladecans-Marsal, Elisabet. 2004. "Agglomeration Economies and Industrial Location: City-Level Evidence," *Journal of Economic Geography*, 4(5), 562–582.
- Zhao, Yinshan, and Harry Joe. 2005. "Composite Likelihood Estimation in Multivariate Data Analysis," *The Canadian Journal of Statistics*, 33(3), 335–356.

## **LIST OF FIGURES**

FIGURE 1: Percentage Distribution of New Businesses across Counties

## **LIST OF TABLES**

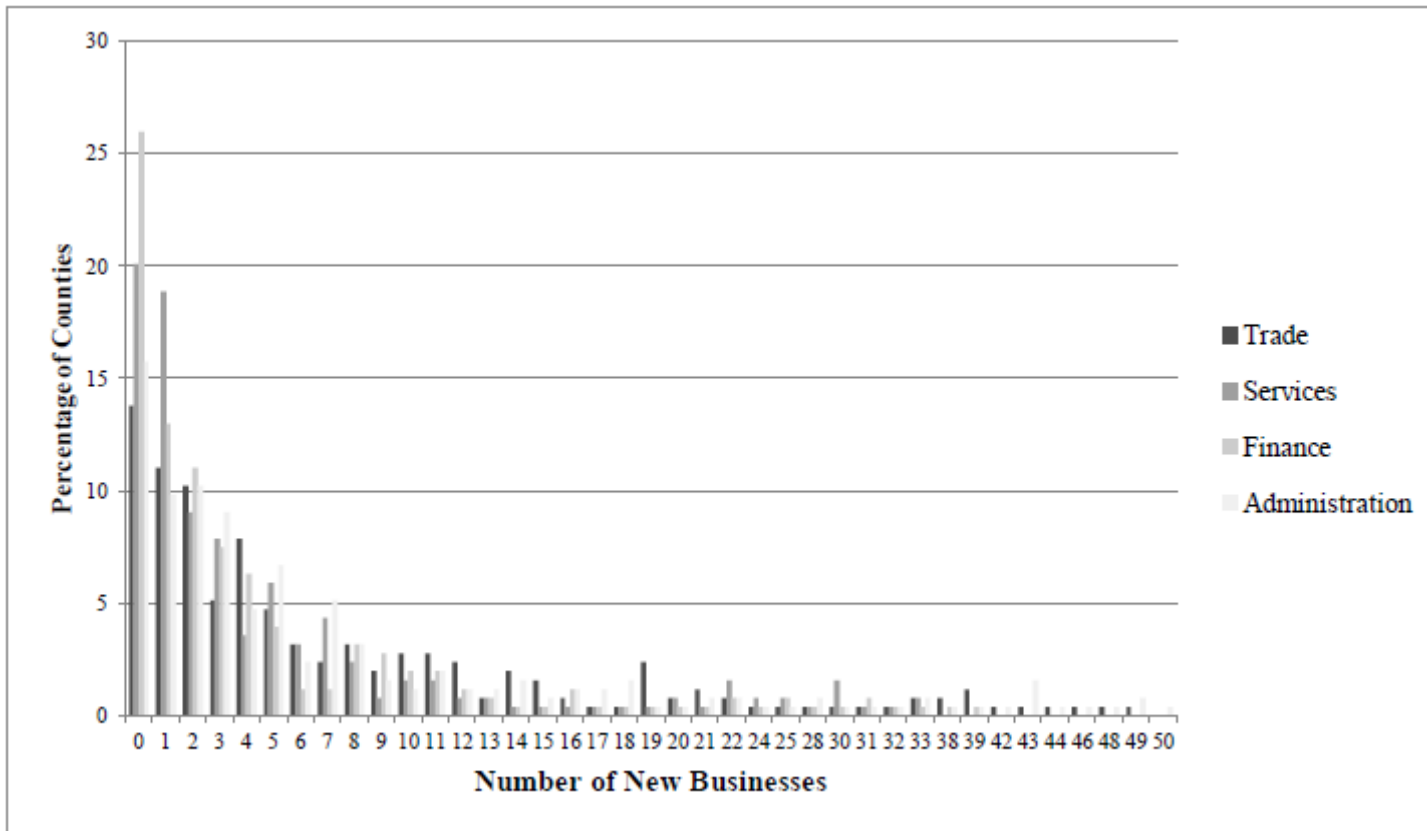
TABLE 1: Percentage Distribution of New Businesses by Sector in Texas and United States

TABLE 2: Data Sources, Variable Definitions, and Descriptive Statistics of the Sample

TABLE 3a: Estimation Results of the SJFC Model

TABLE 3b: Estimation Results of the SJFC Model: Cross-Sector Correlation Matrix and Spatial Autoregressive Parameters

TABLE 4: Elasticity Effects from the IFC and SDJFC Models



**FIGURE 1: Percentage Distribution of New Businesses across Counties**



**TABLE 1: Percentage Distribution of New Businesses by Sector in Texas and United States**

<b>Sector label</b>	<b>Sector</b>	<b>2-digit NAICS Codes</b>	<b>New Businesses in Texas (%)</b>	<b>New Businesses in U.S. (%)</b>
Agriculture	Agriculture, Forestry, Fishing and Hunting, Mining, Quarrying	11, 21	2.11	0.68
Construction	Construction	23	9.83	11.68
Manufacturing	Manufacturing	31-33	2.85	2.80
Trade	Wholesale and Retail Trade	42, 44-45	16.33	16.82
Transportation	Transportation and Warehousing	48-49	4.03	3.53
Services	Information, Professional, Scientific and Technical Services, Management, and Educational Services	51, 54, 55, 61	14.70	16.65
Finance	Finance, Insurance, Real Estate and Rental, and Leasing	52, 53	14.11	12.21
Health	Health Care and Social Assistance	62	9.33	8.49
Hospitality	Accommodation and Food Services	72	10.34	10.01
Arts	Arts, Entertainment and Recreation	71	1.37	1.82
Administration	Public Administration, Administrative and Support, and Other Services	22, 56, 81, 99	15.00	15.31

**TABLE 2: Data Sources, Variable Definitions, and Descriptive Statistics of the Sample**

Type of Variable	Definition	Sources	Descriptive Statistics			
			Min.	Max.	Mean	Std. Dev.
Agglomeration Economies/ Diseconomies	Residential population per sq. mile	Calculated using data from TxSDC <sup>a</sup>	0.09	2,714.17	99.89	291.99
	Residential population change between 2000 and 2006 (in 1000s)	TxSDC	-5.95	490.84	12.02	45.92
	Number of existing businesses per sq. mile	BITS <sup>b</sup>	0.00	13.72	1.11	1.94
	Number of existing trade businesses per sq. mile	BITS	0.00	4.09	0.27	0.47
	Number of existing trade employees per sq. mile	BITS	0.00	386.25	10.17	36.07
	Sector-specific number of existing businesses per sq. mile	BITS	0.00	3.45	0.10	0.23
Specialization Indices and Firm Size	Location Quotient (LQ) of businesses	Calculated using data from CBP <sup>c</sup>	0.00	152.64	1.61	4.25
	Herfindahl-Hirschman Index (HHI) of businesses	Calculated using data from CBP	0.00	0.56	0.15	0.05
	Average size of existing businesses (across all sectors)	Calculated using data from CBP	0.00	537.00	13.59	19.20
Human Capital	Percentage of persons 25 years or above with educational attainment less than high school	US CB <sup>d</sup>	7.20	52.10	22.67	8.29
	Percentage of persons 25 years or above with high school degree	US CB	35.90	74.50	59.92	6.85
	Percentage of persons 25 years or above with bachelor's degree or higher	US CB	7.60	48.30	17.41	6.99
	Rate of unemployment (in %)	TxWC <sup>e</sup>	2.10	1.15	4.28	1.21
	Average wage per job (in 1000s of dollars)	US BEA <sup>f</sup>	21.45	62.62	32.46	6.75

<sup>a</sup> TxSDC: Texas State Data Center

<sup>b</sup> BITS: Business Information Tracking Series, Statistics of U.S. Businesses program

<sup>c</sup> CBP: County Business Patterns

<sup>d</sup> US CB: US Census Bureau - State & County Quick Facts

<sup>e</sup> TxWC: Texas Workforce Commission, Labor Force Statistics for Texas Counties

<sup>f</sup> US BEA: U.S. Bureau of Economic Analysis

**TABLE 2 (Contd.): Data Sources, Variable Definitions, and Descriptive Statistics of the Sample**

Type of Variable	Definition	Sources	Year	Descriptive Statistics			
				Min.	Max.	Mean	Std. Dev.
Fiscal Conditions	Property tax rates	WSG-TCPA <sup>g</sup>	2007	0.20	1.07	0.51	0.16
	Total Government expenditure (in 1,000,000s of dollars)	WSG-TEC <sup>h</sup>	2007	0.57	11,077.61	309.48	1,112.16
Transportation Infrastructure and Land Development Characteristics	Length (in miles) of Roadway network	Calculated using TNRIS and TSF <sup>i</sup>	2007	18.60	9171	1134.03	798.09
	Dummy Variable, 1 if Interstate roadway passes through the county	Calculated using TNRIS and TSF	2007	0	1.00	0.42	0.49
	Dummy Variable, 1 if county has shoreline areas	US CB-NOAA <sup>j</sup>	2007	0	1.00	0.16	0.37
	Number of Airports in each county	Calculated using TNRIS and TSF	2007	0	128.00	7.26	11.33
	Dummy Variable, 1 if county is metropolitan	USDA ERS <sup>k</sup>	2003	0	1.00	0.30	0.46
	Dummy Variable, 1 if county is non-metropolitan and adjacent to a metropolitan county	USDA ERS	2003	0	1.00	0.43	0.50
	Dummy Variable, 1 if county is non-metropolitan and non-adjacent to a metropolitan county	USDA ERS	2003	0	1.00	0.27	0.44
	Land area (in sq. miles)	US CB	2007	127.00	6183.70	1028.48	657.41
County-Specific Factors	5 dummies corresponding to special counties - Harris, Dallas, Tarrant, Bexar and Travis	Our own generated variables	2007	NA	NA	NA	NA

<sup>g</sup> WSG-TCPA: Window on State Government - Texas Comptroller of Public Accounts

<sup>h</sup> WSG-TEC: Window on State Government - Texas State Expenditure by County

<sup>i</sup> TNRIS: Texas Natural Resources Information Systems; TSF: Tiger Shape Files

<sup>j</sup> US CB-NOAA: US Census Bureau - National Oceanic and Atmospheric Administration

<sup>k</sup> USDA ERS: United States Department of Agriculture Economic Research Service

**TABLE 3a: Estimation Results of the SJFC Model**

<b>Sector</b>	<b>Agriculture</b>	<b>Construction</b>	<b>Manufacturing</b>	<b>Trade</b>	<b>Transportation</b>	<b>Services</b>	<b>Finance</b>	<b>Health</b>	<b>Hospitality</b>	<b>Arts</b>	<b>Administration</b>
Variables	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)
<b>Demand Intensity Variables</b>											
<i>Agglomeration Economies/Diseconomies</i>											
Total number of existing businesses per sq. mile	--	--	1.078 (6.488)	--	0.673 (6.769)	--	2.248 (20.671)	2.143 (24.184)	1.637 (14.668)	--	--
Square of the number of existing businesses per square mile	-0.005 (-1.435)	-0.020 (-3.665)	-0.066 (-5.789)	-0.022 (-5.516)	-0.035 (-5.384)	-0.030 (-8.735)	-0.263 (-20.087)	-0.250 (-24.629)	-0.201 (-20.124)	0.007 (8.389)	-0.157 (-8.106)
Spatially lagged Square of the number of existing businesses per square mile	--	--	--	--	--	--	--	--	--	0.024 (6.524)	--
Number of existing trade businesses per sq. mile	--	1.304 (4.470)	--	--	--	2.053 (10.886)	--	--	--	--	--
Spatially lagged # of existing trade businesses per sq. mile	--	1.226 (2.732)	--	--	--	--	--	--	--	--	--
Number of existing trade employees per sq. mile/10	--	--	--	--	--	--	--	0.087 (4.77)	--	--	--
Sector-specific number of existing businesses per sq. mile	--	--	--	1.317 (5.181)	--	--	--	--	--	--	5.340 (7.647)
<i>Specialization Indices and Firm Size</i>											
Location Quotient of businesses /10	0.188 (1.921)	11.231 (7.098)	9.191 (3.441)	10.125 (4.138)	3.049 (4.253)	17.078 (7.709)	5.801 (3.622)	11.935 (12.773)	--	1.415 (4.612)	--
Herfindahl-Hirschman Index of businesses	-3.794 (-1.783)	-9.157 (-3.406)	--	--	-6.604 (-3.426)	-10.786 (-8.396)	-6.907 (-7.142)	-14.382 (-9.826)	-9.368 (-8.214)	-9.979 (-9.145)	--
Spatially lagged Herfindahl-Hirschman Index of businesses	-25.615 (-2.631)	--	--	-9.947 (-4.179)	--	--	--	--	-14.117 (-6.211)	--	--
Average size of businesses/10	--	--	--	1.111 (6.298)	--	--	--	--	--	--	--

**TABLE 3a (Contd.): Estimation Results of the SJFC Model**

<b>Sector</b>	<b>Agriculture</b>	<b>Construction</b>	<b>Manufacturing</b>	<b>Trade</b>	<b>Transportation</b>	<b>Services</b>	<b>Finance</b>	<b>Health</b>	<b>Hospitality</b>	<b>Arts</b>	<b>Administration</b>
<b>Variables</b>	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)
<i>Human Capital</i>											
Education (Base Case: % of persons 25 years or above with educational attainment less than high school)	--	--	--	--	--	--	--	--	--	--	--
% of persons 25 years or above with high school degree /100	--	--	-1.501 (-3.108)	--	-1.668 (-2.258)	--	--	-2.500 (-6.303)	--	-3.425 (013.342)	--
% of persons 25 years or above with bachelor's degree or higher /100	-2.510 (-1.907)	--	--	--	-3.054 (-3.293)	--	--	--	2.631 (4.779)	6.215 (15.056)	3.703 (2.546)
Rate of unemployment /100	-27.241 (-2.576)	-17.709 (-3.199)	--	-6.912 (-1318)	--	-17.089 (-4.296)	--	--	--	--	--
Average wage per job (in 10,000s of dollars)	0.248 (1.915)	--	--	--	0.312 (4.812)	--	--	--	--	--	--
<i>Fiscal Conditions</i>											
Property tax rates	-1.046 (-2.279)	-1.196 (-3.033)	-0.952 (-2.256)	-1.167 (-3.999)	-1.607 (-5.504)	-1.047 (-4.365)	-1.855 (-11.485)	-0.618 (-3.696)	-0.354 (-2.633)	--	-1.417 (-3.055)
Spatially lagged property tax rates	--	--	-3.632 (1.650)	--	-1.962 (-2.412)	-2.077 (-4.017)	--	--	--	--	-2.958 (-2.773)
<i>County-Specific Factors</i>											
Dummy Variable; 1 if region is Harris county, 0 otherwise	--	--	-2.557 (-4.984)	--	--	4.953 (8.429)	--	--	--	--	--
Dummy Variable; 1 if region is Dallas county, 0 otherwise	--	5.113 (6.408)	--	--	--	7.216 (10.527)	--	--	--	--	--
Dummy Variable; 1 if region is Tarrant county, 0 otherwise	--	4.516 (8.310)	--	2.780 (5.572)	2.330 (4.525)	7.185 (10.985)	--	3.619 (15.541)	2.993 (6.224)	3.333 (8.608)	--
Dummy Variable; 1 if region is Bexar county, 0 otherwise	--	--	--	--	--	--	-1.190 (-6.558)	--	--	--	--
Dummy Variable; 1 if region is Travis county, 0 otherwise	--	--	--	--	--	--	--	--	--	--	-2.046 (-3.639)

**TABLE 3a (Contd.): Estimation Results of the SJFC Model**

Sector	Agriculture	Construction	Manufacturing	Trade	Transportation	Services	Finance	Health	Hospitality	Arts	Administration
Variables	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)
<b>Supply Tipping Point (Threshold) Effects</b>											
<b><math>\alpha</math> Vector Elements</b>											
$\alpha_1$	--	--	-0.219 (-1.872)	--	--	0.341 (8.420)	--	--	--	--	-0.194 (-1.598)
$\alpha_2$	-0.151 (-1.553)	--	--	--	--	--	--	--	--	--	--
$\alpha_3$	--	--	--	-0.249 (-5.441)	--	--	--	--	--	--	--
$\alpha_4$	-0.384 (-2.672)	-0.134 (-1.756)	--	--	--	0.116 (3.057)	--	--	--	--	-0.398 (-2.524)
$\alpha_5$	--	-0.246 (-2.180)	-0.343 (-2.068)	-0.422 (-6.456)	--	--	-0.141 (-4.240)	-0.099 (-3.954)	--	--	--
$\alpha_6$	--	-0.454 (-3.350)	--	--	--	--	--	--	--	--	-0.642 (-3.594)
$\alpha_7$	-0.644 (-3.702)	--	-0.478 (-2.073)	-0.669 (-8.132)	--	--	-0.443 (-10.800)	--	--	--	--
$\alpha_8$	-0.766 (-3.288)	--	-0.616 (-2.207)	--	--	--	-0.550 (-10.392)	--	--	--	--
$\alpha_9$	-1.114 (-4.408)	-0.729 (-5.061)	--	-0.893 (-9.4711)	-0.213 (-6.100)	-0.222 (-12.282)	--	--	--	--	-0.900 (-4.632)
$\alpha_{10}$	--	-0.874 (-5.248)	--	--	--	--	--	-0.162 (-3.122)	--	--	-1.091 (-4.940)
$\alpha_{11}$	--	--	--	--	-0.431 (-8.525)	--	--	--	--	--	--
$\alpha_{12}$	--	-1.005 (-5.365)	--	--	--	-0.392 (-6.706)	--	--	--	--	-1.259 (-5.185)
$\alpha_{13}$	--	-1.210 (-6.035)	--	-1.14 (-10.031)	--	--	--	--	--	--	--
$\alpha_{14}$	-3.917 (-9.962)	--	--	--	--	-0.537 (-7.762)	--	--	--	--	--

**TABLE 3a (Contd.): Estimation Results of the SJFC Model**

Sector	Agriculture	Construction	Manufacturing	Trade	Transportation	Services	Finance	Health	Hospitality	Arts	Administration
Variables	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)	Estimate (t-stat)
<i>γ</i> Vector Elements											
Constant	-3.021 (-2.287)	-4.451 (-6.020)	-8.197 (-4.428)	-3.166 (-14.376)	-2.4291 (-5.295)	-3.111 (-14.581)	-6.728 (-12.123)	-3.742 (-7.421)	-2.062 (-3.071)	-8.265 (-16.440)	-5.234 (-4.640)
<i>Agglomeration Economies/Diseconomies</i>											
Residential population per sq. mile / 100	-0.032 (-2.442)	-0.059 (-2.896)	--	--	--	-0.074 (-8.458)	--	--	-0.025 (-1.395)	--	--
Residential population change between 2000 and 2007 (in 10,000s)	--	--	--	-0.020 (-4.643)	--	--	-0.015 (-7.418)	-0.032 (-8.674)	-0.018 (-2.928)	--	-0.033 (-5.833)
<i>Fiscal Conditions</i>											
Natural logarithmic of government expenditure (in dollars)	0.272 (5.051)	0.309 (9.073)	0.199 (2.531)	0.259 (72.88)	0.189 (5.989)	0.285 (66.532)	0.247 (12.441)	0.148 (7.774)	0.192 (5.355)	0.470 (23.073)	0.330 (8.893)
<i>Transportation Infrastructure and Land Development Characteristics</i>											
Natural logarithmic of length (in miles) of roadway network	--	0.109 (2.099)	0.145 (3.138)	0.076 (3.424)	0.142 (3.513)	0.056 (2.665)	0.093 (2.935)	0.200 (7.498)	--	0.121 (8.958)	0.186 (4.345)
Dummy Variable; 1 if Interstate roadway passes through the county, 0 otherwise	--	--	0.217 (1.830)	--	--	--	--	--	--	--	--
Number of airports in each county /10	--	0.056 (2.202)	--	0.046 (2.648)	--	--	--	0.080 (5.420)	0.101 (9.875)	--	--
Metropolitan status (Base Case: County is a non metropolitan area and also non-adjacent to metropolitan county)	--	--	--	--	--	--	--	--	--	--	--
Dummy Variable; 1 if county is metropolitan, 0 otherwise	--	--	0.307 (2.528)	--	--	--	--	--	--	--	--
Dummy Variable; 1 if county is a non-metropolitan but adjacent to a metropolitan county, 0 otherwise	--	--	0.147 (1.20)	0.080 (2.337)	--	--	0.198 (6.257)	--	--	--	--
Natural logarithmic of land area (in sq. miles)	--	--	0.626 (2.014)	--	--	--	0.420 (6.093)	0.226 (5.160)	0.139 (4.557)	--	--

**TABLE 3b: Estimation Results of the SJFC Model: Cross-Sector Correlation Matrix and Spatial Autoregressive Parameters**

Sector	Agriculture	Construction	Manufacturing	Trade	Transportation	Services	Finance	Health	Hospitality	Arts	Administration	$\delta_s$
<b>Agriculture</b>	1.000											--
<b>Construction</b>	0.137 (1.711)	1.000										--
<b>Manufacturing</b>	0.000	0.202 (2.140)	1.000									0.275 (6.452)
<b>Trade</b>	0.000	0.454 (6.351)	0.279 (2.682)	1.000								--
<b>Transportation</b>	0.000	0.169 (2.437)	0.000	0.171 (2.265)	1.000							--
<b>Services</b>	0.147 (1.663)	0.442 (6.167)	0.230 (2.709)	0.455 (7.057)	0.229 (2.928)	1.000						--
<b>Finance</b>	0.000	0.340 (4.336)	0.210 (2.774)	0.343 (5.221)	0.166 (1.917)	0.291 (3.426)	1.000					0.298 (14.561)
<b>Health</b>	0.000	0.294 (3.273)	0.165 (1.838)	0.299 (3.738)	0.000	0.335 (4.667)	0.382 (5.377)	1.000				--
<b>Hospitality</b>	0.000	0.172 (1.885)	0.298 (3.56)	0.313 (3.829)	0.000	0.181 (1.845)	0.131 (1.341)	0.131 (1.492)	1.000			--
<b>Arts</b>	0.000	0.000	0.000	0.000	0.1736 (1.9776)	0.000	0.000	0.000	0.000	1.000		--
<b>Administration</b>	0.000	0.342 (4.489)	0.223 (2.209)	0.385 (5.333)	0.000	0.373 (4.702)	0.229 (2.391)	0.282 (3.560)	0.402 (5.463)	0.000	1	--



**TABLE 4: Elasticity Effects from the IFC and SDJFC Models**

Sector	IFC						SDJFC					
	Direct		Indirect		Total		Direct		Indirect		Total	
	Mean	Tstat.	Mean	Tstat.	Mean	Tstat.	Mean	Tstat.	Mean	Tstat.	Mean	Tstat.
<b>Manufacturing</b>	-19.41	-3.21	0.00	--	-19.41	-3.21	-5.30	-1.13	-16.46	-0.60	-19.14	-0.70
<b>Transportation</b>	-16.43	-4.02	0.00	--	-16.43	-4.02	-14.22	-3.16	-16.49	-1.90	-28.15	-2.88
<b>Services</b>	-11.39	-4.56	0.00	--	-11.39	-4.56	-8.46	-3.52	-15.56	-3.12	-22.69	-3.99
<b>Finance</b>	-15.26	-8.34	0.00	--	-15.26	-8.34	-14.14	-6.57	-5.86	-6.04	-19.12	-6.90
<b>Administration</b>	-14.00	-3.64	0.00	--	-14.00	-3.64	-9.77	-2.20	-19.36	-1.92	-27.17	-2.45