Public school consolidation: a partial observability spatial bivariate probit approach

David M. Brasington*
Department of Economics
University of Cincinnati
Carl H. Lindner Hall
2925 Campus Green Drive
PO Box 0371
Cincinnati, OH 45221-0371
Phone:513-556-2616
Fax:513-556-2669
david.brasington@uc.edu
and
Olivier Parent
Department of Economics
University of Cincinnati
olivier.parent@uc.edu

Abstract

A pair of municipalities may consolidate services if they are contiguous. Traditional estimation methods assume that each voting process is independent. Instead we propose a new estimation procedure that allows the probability of consolidation to be influenced by neighboring decisions. We extend a model of local interaction by allowing consolidation effort of neighbors to be either strategic substitutes or strategic complements. We disentangle direct effects arising from a change in one's own characteristics from indirect or spillover effects associated with a change in the other municipalities' characteristics. Results reveal that the endogenous peer effect coming from neighbors is a primary determinant of willingness to consolidate.

KEYWORDS: social interactions, discrete choice modeling, spatial econometrics, school district consolidation, land annexation.

JEL: C11, C21, H41, H7, I2

*We would like thank the Taft Research Center, the University Research Council and the research committee of the Carl H. Lindner College of Business of the University of Cincinnati for providing generous research support. The authors acknowledge comments from Chunrong Ai, James P. LeSage, David MacArthur and other participants at the Midwest Econometric Group 2012, NARSC 2011, 2013 and WRSA 2013 conferences.

1 Introduction

In recent years, economists have paid considerable attention to political integration. One key challenge for the empirical literature has been to model the interaction between decision makers.

We specify a model where the probability of consolidation for individual pairs of contiguous municipalities is influenced by neighboring pairs. For an example at the nation level, when West Germany integrated with East Germany in 1990, West Germany became contiguous to Poland, which mattered for West Germany's trade and defense purposes, because Poland was a member of the Warsaw Pact and (unlike East Germany) had fought against Germany in World War II. The new border between a unified Germany and Poland also allowed for Poland to potentially become a member of NATO in the future. If Poland's proximity represents peer effects, and these effects are ignored, a model of consolidation will yield biased parameter estimates. The case we study is school boards. If our school board is considering consolidating with an adjacent school board, and the adjacent school board's neighbors affect our decision, ignoring the peer effects of those neighbors will lead to biased estimates of the determinants of agreeing to consolidate.

We propose a new estimation procedure based on the spatial Durbin bivariate probit model to address the influence of neighboring jurisdictions in a consolidation decision. Traditional attempts to model consolidation involve regressing the number of political units in an urban area (like school districts) as a function of population heterogeneity (e.g., Nelson, 1990). Recent studies use probit analysis to model the decision of Connecticut municipalities to consolidate public health services (Bates, Lafrancois and Santerre, 2011), and logit analysis to model the desire of Norwegian town officials to consolidate their municipality with one or more neighbors (Sorensen, 2006). Brasington (1999) argues that such approaches fail to model the true decision-making involved. The decision to consolidate should be modeled as a joint decision by two specific, contiguous neighboring municipalities. This bivariate probit approach allows each municipality to decide whether to consolidate with its neighbor, with each neighbor having veto power. Brasington (2003a, 2003b) extends this approach and lets researchers compare the decision-making of the smaller and larger member of a pair.

Gordon and Knight (2009) introduce a matching algorithm to model consolidation. While they do not allow for multiple member consolidation, they relax the assumption of independence in merger decisions of Brasington (1999). The current study extends the approaches proposed by Brasington (1999) and Gordon and Knight (2009) by allowing the probability of consolidation between pairs of municipalities to depend on the probability of consolidation with their respective neighbors while retaining many of the other attractive features of their models: mutual veto power, multiple member consolidation, and the ability to separately analyze the decisions of the larger and smaller members of a consolidation. We extend the model of local interaction of Glaeser and Scheinkman (2001) by allowing the effort of neighboring members of a potential consolidation pair to be either strategic substitutes or strategic complements. We analyze a unique data set of 1,596 pairs of municipalities that could potentially consolidate schooling services across 8 urban areas in Ohio and Texas.

The main contribution relies on a simple extension of discrete choice models that would allow latent utilities to contain strategic interactions similar to the spatial autoregressive formulation of the linear-in-means models (as defined in Lee, 2007 and Bramoulle et al.

2014 among others) while observing the same stability conditions. This avoids complex contraction mapping properties to define the rational expectation equilibrium. Estimates of group interaction might also suffer from the presence of unobservable factors that could affect all individual pairs within the same urban area. People typically use traditional fixed effects approaches, but these require transformations that cannot be implemented in discrete choice models. Instead we invoke the correlated random effects approach (Chamberlin, 1984) to model the unobservable factors while avoiding the incidental parameters problem. Another advantage of this approach is that it makes it easy to evaluate partial effects for discrete choice models. With spatial dependence, changes in one explanatory variable associated with one municipality directly affect the willingness to consolidate of the other member of the pair but are also reflected in the decisions of all other municipalities located within the same urban area. Our approach makes it feasible to properly interpret the partial derivative impacts from changes in the explanatory variables. Results reveal that local as well as global interaction effects correspond to a complementarity in consolidation efforts.

The outline of this paper is as follows. Section 2 presents the theoretical framework of the linear-in-means model. Extension to discrete choice models is provided in Section 3 using Poirier's (1980) approach to specify a bivariate model with partial observability. Section 4 presents the estimation procedure along with Monte Carlo experiments. We describe the data in Section 5 while empirical results are discussed in Section 6. Section 7 concludes.

2 Theoretical framework

For the theoretical model, we extend to the bivariate case the traditional social network model recently surveyed in Liu, Patacchini and Zenou (2011), who use network analysis to study the impact of peer pressure on adolescents' sports, education, and video game-playing activities. Our objective is to probabilistically describe the choice of consolidation of public schooling provision for a pair $i = (1, \ldots, n)$ of neighboring municipalities $j = (1, 2)$.

Let n_r be the number of pairs of neighboring municipalities in an urban area $r = (1, \ldots, \bar{r})$ that could potentially consolidate. Across all urban areas $\bar{r}$, the total number of pairs is equal to $n = \sum_{r=1}^{\bar{r}} n_r$. Municipalities are laid out across urban areas, and urban areas are assumed to be disconnected from each other.[1] We describe connections between pairs of adjacent municipalities as: $G_r = [g_{il,r}]$, where $[g_{il,r}] = 1$ if pairs i and l have municipalities that are direct neighbors, and $[g_{il,r}] = 0$ otherwise. Because municipalities are either adjacent or not, $g_{il,r} = g_{li,r}$, which satisfies reciprocity in the networking literature, and, as in the spatial econometric literature, we assume $g_{ii,r} = 0$, that a municipality cannot be its own neighbor. The data set for our empirical section will be a map of neighboring municipalities, some of whom have independent school districts, others of whom have formed consolidated school districts with one or more neighbors. Taking this map as an equilibrium, we model how the map came to be in this equilibrium as a simultaneous set of decisions by contiguous pairs of municipalities. We work at the pair level to be able to draw simple analogies between the linear-in-means model and the bivariate discrete choice model.

Within each pair i a municipality $j = (1, 2)$ tries to consolidate schooling services with an

[1] Urban areas like Cincinnati and Dayton are distant from each other and do not share any common boundary.

adjacent municipality $(-j) = (1, 2)$. The two municipalities of that pair i must expend some effort studying the costs and benefits of consolidating with each other. In our static game, each municipality $j = (1, 2)$ of the pair i in urban area r expends $y^{\star}_{ij,r} \geq 0$ in consolidation effort. To better understand how the model is set up, consider the illustration in Appendix A. It shows five political jurisdictions in a metropolitan area. Each jurisdiction can maintain its own school district, or it can consolidate with a willing neighbor. Neighboring pairs are defined as dyads having one municipality in common. Outside the direct consolidation process taking place between the two players of that pair i, each of municipality j's neighbors belonging to other pairs l also expend a certain amount of effort consolidating with j. We let $\bar{y}_{ij,r}$ be the average effort of j's neighbors:

$$\bar{y}^{\star}_{ij,r} = \frac{1}{g_{i,r}} \sum_{l=1}^{n_r} g_{il,r} y^{\star}_{lj,r}, \tag{1}$$

where $g_{i,r} = \sum_{l=1}^{n_r} g_{il,r}$ is the number of neighboring pairs for individual pair i. Note that as of now the neighboring influence is coming from different pairs but the same order $j = (1, 2)$ of municipality. We will later see that each pair i is composed of a small municipality $(j = 1)$ and large municipality $(j = 2)$ that are contiguous. This equation says that municipality j's expectations of the average effort in urban area r equals the average effort of adjacent municipalities.

To introduce the bivariate setting, we extend the model of local interaction developed by Glaeser and Scheinkman (2001). Local dependence is based on the interaction between both municipalities j in each pair i and global dependence is based on $\bar{y}^{\star}_{ij,r}$, the neighbors of municipality j.

Municipality j of pair i chooses its own effort level to maximize utility:

$$\begin{aligned} U_{ij,r}(Y^{\star}_r, G_r) &= (a_{ij,r} + \lambda_r + \epsilon_{ij,r}) y^{\star}_{ij,r} - \frac{(1 - \psi - \varphi)}{2} y^{\star 2}_{ij,r} - \\ &\quad \frac{\psi}{2}(y^{\star}_{ij,r} - y^{\star}_{i(-j),r})^2 - \frac{\varphi}{2}(y^{\star}_{ij,r} - \bar{y}^{\star}_{ij,r})^2 \end{aligned} \tag{2}$$

Certain symbols in Equation (2) require explanation. We let $Y^{\star}_r = (y^{\star\prime}_{1,r}, \ldots, y^{\star\prime}_{n_r,r})'$ represent the population effort profile in each urban area r with $y^{\star}_{1,r} = (y^{\star}_{i1,r}, y^{\star}_{i2,r})'$. As in Liu et al. (2011), $a_{ij,r}$ represents the idiosyncratic heterogeneity, which is perfectly observable by all municipalities. This term captures all the observable characteristics of the municipality j (like e.g. enrollment, racial composition, education levels, and property values etc.) and the average observable characteristics of the neighboring municipalities (contextual effects). In fact, $a_{ij,r}$ can be written as:

$$a_{ij,r} = x'_{ij,r}\beta_2 + (1/g_{i,r}) \sum_{l=1}^{n_r} g_{il,r} x'_{lj,r} \gamma_2 \tag{3}$$

where $x_{ij,r}$ is the Q-dimensional vector of individual-specific characteristics, and β_2 and γ_2 are parameters of interest.

The parameter λ_r represents unobservable network or urban area characteristics, and $\epsilon_{ij,r}$ is an error term, unobservable to the modeler but known to the municipality, whose inclusion indicates that there is some uncertainty about the benefit of consolidation. The first term of (2) captures the intrinsic utility and reveals that utility depends on a municipality's own characteristics and the characteristics of its neighbors.

For the pair i, municipality $j = (1, 2)$ must decide whether to consolidate with the other municipality $(-j) = (1, 2)$ of that same pair. As in the univariate case intrinsic utility is defined by a municipality's own independent effort toward consolidation $\frac{(1-\psi-\varphi)}{2} y^{\star 2}_{ij,r}$. The utility function is strictly concave in own effort as long as $\psi + \varphi < 1$. In the bivariate case a municipality's effort is also a function of the effort of its neighbor. As detailed in Ballester et al. (2006), neighboring influences are captured by the cross-derivatives that are pair-dependent. Following Glaeser and Scheinkman (2001), local interdependence is captured by the third term $\frac{\psi}{2}(y^{\star}_{ij,r} - y^{\star}_{i(-j),r})^2$ and depends on the difference between both municipalities' effort within each pair i. Negative values for the cross-derivatives (positive ψ) would reveal that consolidation efforts for both members of pair i are strategic substitutes. If ψ is negative, an increase in municipality $(-j)$'s consolidation effort would trigger a positive shift in j's reaction, allowing effort for both municipalities to be strategic complements.

Global dependence is represented by the last term in (2). It corresponds to the disutility for deviating from the group or urban area norms between pairs of potential consolidations. Similar to local dependence, the effort of consolidation with neighboring municipalities is measured by the parameter φ which also represents a taste for conformity. A negative value for φ would deter a municipality from merging when its neighbors have a high willingness to consolidate. A high, positive value would indicate a high taste for conformity in effort, while a low positive value suggests a municipality doesn't care much about conforming to its neighbors' consolidation effort. Whether local and global strategic interaction are substitutes or complements is theoretically unknown; our empirical section will estimate the values of those parameters.

The idea is that a municipality cares about its residents and the education of its students. It works to undertake a consolidation with a neighbor if the consolidation benefits residents, but there is a difference between a hostile takeover and a friendly merger. In Ohio and Texas, mutual consent is required for consolidation, but having the two parties agree equally is likely to provide a more peaceful transition from independent to consolidated school districts. Newspaper articles discuss the poisoned atmosphere of a school district consolidation where one party was more eager than the other to consolidate.

In addition, the $\bar{y}^{\star}_{ij,r}$ term in (2) implies that the more neighbors a municipality has, the more influences the utility will be averaged over. In general, residents sorting into an urban area with a large number of municipalities will be able to find a closer match between preferences over taxes, public services, and demographic characteristics, and thus achieve a higher utility, than residents who sort into an urban area with few municipalities. Similarly, the more neighbors a municipality has, the better the chances of a good fit with a neighbor for purposes of school district consolidation. Consolidation is more attractive between neighbors with similar size, property values, demographic characteristics and other unobserved attributes.

Sorting the members of each pair of potential consolidation by size will allow us to identify separate impacts for the larger and smaller municipality's neighbors, instead of estimating an average impact. Previous studies (e.g. Brasington, 2003a) show that there are differences in how the larger and smaller member of a pair makes its consolidation decisions, so there might be differences in how the neighbors of the larger and the neighbors of the smaller member of a pair influence consolidation, too.

More importantly, within a given urban area, it is hard to disentangle endogenous peer effects coming from interactions between pairs of municipalities from correlated unobservable effects λ_r associated with similar individual preferences (Moffitt, 2001). In fact, Lee (2007) shows that interaction effects cannot be identified if there is no variation in size between groups or urban areas. He proposes to model unobserved heterogeneity via fixed effects and discusses efficient estimators to overcome the incidental parameter problem. Because of our interest in identifying marginal effects, we implement the popular correlated random effects model (Chamberlain, 1984). Dependence between the unobserved effects and explanatory variables is restricted through assumptions on the conditional distribution of heterogeneity given the explanatory variables. Therefore, we do not need to assume independence like a traditional random effects model does, and we avoid the incidental parameters problem as well.

When all pairs of municipalities choose effort level $y^{\star}_{ij,r}$ simultaneously to maximize (2), the following first-order conditions result:

$$\frac{\delta U_{ij,r}(Y^{\star}_r, G_r)}{\delta y^{\star}_{ij,r}} = a_{ij,r} + \lambda_r + \epsilon_{ij,r} - y^{\star}_{ij,r} + \psi y^{\star}_{i(-j),r} + \varphi \bar{y}^{\star}_{ij,r} \tag{4}$$

Using the definition of $\bar{y}^{\star}_{ij,r}$ from (1) along with (4) yields the best-response function:

$$\begin{aligned} y^{\star}_{ij,r} &= \varphi \frac{1}{g_{i,r}} \sum_{l=1}^{n_r} g_{il,r} y^{\star}_{lj,r} + (a_{ij,r} + \lambda_r + \epsilon_{ij,r} + \psi y^{\star}_{i(-j),r}) \\ y^{\star}_{ij,r} &= \psi y^{\star}_{i(-j),r} + \varphi \frac{1}{g_{i,r}} \sum_{l=1}^{n_r} g_{il,r} y^{\star}_{lj,r} + \\ &\quad x'_{ij,r}\beta_2 + (1/g_{i,r}) \sum_{l=1}^{n_r} g_{il,r} x'_{lj,r}\gamma_2 + \lambda_r + \epsilon_{ij,r} \end{aligned} \tag{5}$$

For the univariate case, when $\psi = 0$, Liu, Patacchini and Zenou (2011) show that if $|\varphi| < 1$ then the game has a unique Nash equilibrium in pure strategies given by (5). By setting specific constraints on ψ and φ, Glaeser and Scheinkman (2001) force strategic interaction to be characterized by efforts that are either complements or substitutes. This simplifies tremendously the stability condition of the Nash equilibrium. As discussed in Bramoule et al. (2014) in great detail, introducing substitutability into a model of local interaction effects with complementarity in effort greatly complicates the stability conditions for a unique Nash equilibrium. Depending on the magnitude of the minimum eigenvalue of the entire network, multiple equilibria could arise. One crucial difference with Bramoule et al. (2014) is that we allow for both matrices of global and local interactions to be row-normalized. The standard dominance diagonal conditions will ensure the inferiority and uniqueness of the equilibrium. Stationary regions are defined by the maximum and minimum eigenvalues of the entire network. Maximum eigenvalues for both matrices of local and global dependence are equal to one and the largest characteristic root is thus equal to $\psi + \varphi$. If both parameters are positive, the equilibrium conditions will be satisfied if $\psi + \varphi < 1$. When one or both parameters are negative, minimum eigenvalues have to be calculated to define stationary conditions. We will use the complete set of stationary conditions defined by Elhorst et al. (2012) to ensure the system is stable.

In Ohio and Texas, school boards must each vote in favor for a consolidation to occur, so consolidation involves effort on the part of each school board. The effort of each school board depends on the other's effort, according to (4) and (5), and these are related decisions. And while the theoretical model has an observed, continuous effort $y^\star$, we do not observe $y^\star$. Instead we observe whether effort is high enough on the part of both parties to result in a consolidation or not. The measure of effort $y^\star$, or willingness to consolidate, that we use is the utility differential before and after a potential consolidation. Because the consolidation process takes place between two municipalities of a same pair i, it argues for a bivariate model. Because we observe consolidation or not, it argues for a probit model. And because we only observe consolidation if both parties vote for it, but we do not observe actual consolidation votes, we have a bivariate probit with partial observability. The spatial component of the bivariate probit model arises from the influence of neighboring pairs - and the neighbors of the other member of our pair - in the consolidation decision.

Traditionally with discrete choice models, the econometrician only observes the choice $y_{ij,r} = m$, $m \in 0,1$ of municipality j that generates the highest utility level. If latent utility were observed by the econometrician, estimating parameters would reduce to linear regression. Using the Bayesian approach, the Data Augmentation step (Tanner and Wong, 1987) allows the econometrician to simulate the latent consolidation effort. The idea behind data augmentation is simply to expand the parameter space with a set of ancillary variables in order to simplify the estimation of the model.

3 Discrete Choice Modeling

For discrete choice models, the observed choice $y_{ij,r}$ depends on the latent utility $U_{ij,r}$ in a non-linear way. As described by Brock and Durlauf (2001) and Tamer (2003) the main challenge is that in general, multiple equilibria arise when we try to estimate latent utility models. Lee, Li and Lin (2014) have recently shown that subjective data on expectations provides important information in the modeling of social interactions. Using logit and probit models, they prove that the existence of an equilibrium is guaranteed only by the fixed point theorem, whether the expected probability is constant or heterogeneous across individuals in a group. We model interactions with other individuals in the same group through their latent utilities and therefore equilibrium can be achieved with simple properties of stability as in the linear-in-means framework, without relying on the fixed point theorem. The main purpose of this section is to show that the latent utility difference derived from the bivariate probit model is equivalent to the Nash equilibrium defined in (5).

Suppose that we no longer observe actual consolidation effort $y^\star_{ij,r}$ but only whether or not two municipalities have consolidated $y_{ij,r}$. Following Brock and Durlauf (2001), we set the observed discrete outcome $y_{ij,r} = m$ where $m \in \{-1, 1\}$. The utility function $U_{ijm,r}$ developed in (2) is now defined with the index $m \in \{-1, 1\}$ and is equivalent to:

$$\begin{aligned} U_{ijm,r}(Y_r, G_r) &= y_{ij,r}(a_{ij,r} + \lambda_r) - \frac{\psi}{2} E(y_{ij,r} - y^\star_{i(-j),r})^2 - \\ &\quad \frac{\varphi}{2} E(y_{ij,r} - \mu_{ij,r})^2 + \epsilon_{ij,r}(y_{ij,r}), \end{aligned}$$

where $\mu_{ij,r} = \sum_{l=1}^{n_r} g_{il,r} y^\star_{lj,r}$ denotes the expectation municipality j places on neighboring consolidation effort. Similar to the linear-in-means model, the penalty terms are expressed as the expected square deviation of j's consolidation choice from the expected mean effort of others. Then, the willingness to consolidate for municipality j is equivalent to:

$$U_{ij1,r}(Y_r, G_r) - U_{ij(-1),r}(Y_r, G_r) = 2(a_{ij,r} + \lambda_r) + 2\psi E(y^\star_{i(-j),r}) + 2\varphi \sum_{l=1}^{n_r} g_{il,r} E(y^\star_{lj,r}) + \varepsilon_{ij,r}, \quad (6)$$

with $\varepsilon_{ij,r} = \epsilon_{ij,r}(1) - \epsilon_{ij,r}(-1)$. In this case, equilibrium could be characterized using the logistic density and maintaining the i.i.d. error assumptions. Blume et al. (2011) draw an interesting analogy between this binary choice model and the quadratic utility function previously described in the linear-in-means model. Because choice models cannot be identified, the scalar 2 in front of each term of the willingness to consolidate defined in (6) could be removed. A detailed discussion on identification will be discussed in this section. We will now show that a similar expression can be obtained using the traditional setting $y_{ij,r} = m$ where $m \in \{0, 1\}$, following Poirier (1980).

We consider in an urban area r, a pair $i = (1, \ldots, n)$ of neighboring municipalities $(j = 1, 2)$ each faced with the binary choice of consolidation $y_{ij,r} = m$ where $m \in \{0, 1\}$. As in Brasington (2003a), we examine the effect of size in the consolidation decision by imposing $j = 1$ to be the smaller municipality and $j = 2$ the larger one. The framework could also be used to split the data by income, race or even to randomly assign pairs. The common formulation of the merger decision between two communities relies on the bivariate probit model. For a given pair i of neighboring communities, $v_{ijm,r}$ represents the vector of individual characteristics for municipality j where $m \in \{0, 1\}$ for an urban area r. The bivariate probit model proposed by Poirier (1980) relies on the fact that the two municipalities $j = 1, 2$ have the following utility functions for $m \in \{0, 1\}$:

$$U_{i1m,r} = h_{i1m}(v_{i1m,r}, y^\star_{i2,r}) + \omega_{1m,r} + \eta_{i1m,r}$$
$$U_{i2m,r} = h_{i2m}(v_{i2m,r}, y^\star_{i1,r}) + \omega_{2m,r} + \eta_{i2m,r},$$

where h_{ijm} is a scale function, $\omega_{jm,r}$ and $\eta_{ijm,r}$ represent the unobserved attributes.[2] Even if we only observe a single binary outcome of consolidation, the utility maximization framework reflects the binary joint choices of two decision-makers' choices. However, as emphasized by Gordon and Knight (2009), this model assumes that the consolidation decision process for each pair of neighboring municipalities is independent from all the other potential partners. We extend the bivariate probit model proposed by Poirier (1980) by assuming that utility for each member in the pair is a function of the member's willingness to potentially consolidate with each of its neighbors. In fact, we want to measure to what extent the decision process for the consolidation of two municipalities depends on the potential consolidations neighboring municipalities could undertake. As for the linear-in-means model, we separate the direct influence of the other member of the pair (local effect) from the influence of neighboring pairs (global effect). We will show the new latent utility formulation is similar to the best reply function defined in (5) and stability conditions remain

[2] For identification perspective, the parameter expansion technique requires that within each urban area r, small $(j = 1)$ and large $(j = 2)$ municipalities have different fixed effects $\omega_{jm,r}$.

identical. For each utility function, we define a new vector $Y^{\star}_{\delta_{ij}}$ measuring the willingness to consolidate of all the neighboring communities around a municipality $j = (1, 2)$ of pair i.[3] Utility functions are now defined for $m \in 0, 1$ as:

$$U_{i1m,r} = h_{i1m}(v_{i1m,r}, y^{\star}_{i2,r}, Y^{\star}_{\delta_{i1,r}}) + \omega_{1m,r} + \eta_{i1m}$$
$$U_{i2m,r} = h_{i2m}(v_{i2m,r}, y^{\star}_{i1,r}, Y^{\star}_{\delta_{i2,r}}) + \omega_{2m,r} + \eta_{i2m}.$$

For each pair i, we measure the willingness to consolidate for a municipality $j = (1, 2)$, with the utility differential:

$$y^{\star}_{ij,r} = U_{ij1,r} - U_{ij0,r}.$$

One way to represent the effect of neighboring decisions is to assume that for a potential consolidation pair i, the willingness to consolidate of the smaller community $y^{\star}_{i1,r}$ (resp. the larger, $y^{\star}_{i2,r}$) is first influenced by a set of neighbors δ_{i1} (resp. δ_{i2}). The influence of neighbors on the decision process between two municipalities is modeled using the spatial weight matrix W_j. The elements of the matrix show whether or not two municipalities that belong to different pairs are themselves contiguous.[4] The coefficients $w_{iljj'}$ of the $(n \times n)$ dimensional matrix $W_{jj'}$, $j, j' = (1, 2)$, are equal to one if community j' of pair i is a neighbor of community j belonging to pair l (see Appendix A for further details). The spatial weight is based on the number of neighboring pairs $i = (1 \ldots, n)$ that could potentially consolidate.

Thus, we obtain for each municipality $j = (1, 2)$ of the pair $i = (1, \ldots, n)$:

$$\begin{aligned} h_{i11}(v_{i11,r}, y^{\star}_{i2,r}, Y^{\star}_{\delta_{i1}}) - h_{i10}(v_{i10,r}, y^{\star}_{i2,r}, Y^{\star}_{\delta_{i1}}) &= \psi_1 y^{\star}_{2,r} + \varphi \sum_{l=1}^{n_r} w_{il11} y^{\star}_{l1,r} + \\ &\quad \varphi \sum_{l=1}^{n_r} w_{il21} y^{\star}_{l2,r} + \tilde{x}_{i1,r}\delta_1 \end{aligned} \quad (7)$$

$$\begin{aligned} h_{i21}(v_{i21,r}, y^{\star}_{i1,r}, Y^{\star}_{\delta_{i2}}) - h_{i20}(v_{i20,r}, y^{\star}_{i1,r}, Y^{\star}_{\delta_{i2}}) &= \psi_2 y^{\star}_{i1,r} + \varphi \sum_{l=1}^{n_r} w_{il22} y^{\star}_{l2,r} + \\ &\quad \varphi \sum_{l=1}^{n_r} w_{il12} y^{\star}_{l1,r} + \tilde{x}_{i2,r}\delta_2, \end{aligned} \quad (8)$$

where we extended Poirier's (1980) model for the individual characteristics $\tilde{x}_{ij,r}$ to include the following variables:

$$\tilde{x}_{ij,r} = \left[x_{ij,r} \quad \sum_{l=1}^{n_r} w_{il11} x_{l1,r} + \sum_{l=1}^{n_r} w_{il21} x_{l2,r} \right]. \quad (9)$$

The first term $x_{ij,r}$ of the concatenation (9) is a k_j- raw vector of individual characteristics and the second term is also a k_j- raw vector containing the contextual effects obtained from averaging characteristics (with row-normalized spatial weight matrix) of the small $(x_{l1,r})$ and large $(x_{l2,r})$ municipalities in neighboring pair l. The $2k_j-$dimensional vector δ_j represents parameters of interest similar to the marginal effects β_2 and γ_2 defined in the

[3] Note that this vector excludes the other municipality $(-j)$ of pair i.

[4] The same municipality belonging to two different pairs is not defined as a neighbor.

best-reply function (5) of the linear-in-means model. However, unlike the linear-in-means model, Poirier (1980) allows for local interaction to be different between small and the large municipalities of each pair. In fact, the effect (ψ_1) of the large municipality on the willingness to consolidate of the small municipality is set to be different from the effect (ψ_2) of the small on the large municipality's willingness to consolidate. Note that this assumption will be challenged when we investigate identification issues. The parameter φ measures the endogenous effects or the strength of the spatial dependence between neighboring municipalities. It corresponds to the same effect measuring the global dependence in the linear-in-means model.

Note that each municipality can be influenced either by a smaller municipality $y_{l1,r}$ of the neighboring pair l or by a larger municipality $y_{l2,r}$. Even though this separation was not made in the linear-in-means model, it will help us to identify endogenous effects coming from large and small neighboring municipalities. This is a necessary step to analytically solve the reduced form developed in Appendix B. This separation forces the unobserved group or urban area effects λ_r to be split between small municipalities ($\lambda_{1,r}$) and large municipalities ($\lambda_{2,r}$)) where $\omega_{11,r} - \omega_{10,r} = \lambda_{1,r}$, $\omega_{21,r} - \omega_{20,r} = \lambda_{2,r}$. We define the new random term as $\eta_{i11,r} - \eta_{i10,r} = \epsilon_{i1,r}$ and $\eta_{i21,r} - \eta_{i20,r} = \epsilon_{i2,r}$.

Let $w_{1ij} = (w_{ij11}, w_{ij21})$, $w_{2ij} = (w_{ij12}, w_{ij22})$ and $Y^{\star}_{i,r} = (y^{\star}_{i1,r}, y^{\star}_{i2,r})'$. Thus for each pair of municipalities that could consolidate, the spatial bivariate probit model with partial observability can be rewritten as:

$$
\begin{aligned}
y^{\star}_{i1,r} &= \tilde{x}_{i,r}\beta_1 + \rho_1 \sum_{l=1}^{n_r} w_{1il} Y^{\star}_{l,r} + \rho_2 \sum_{l=1}^{n_r} w_{2il} Y^{\star}_{l,r} + \mu_{1,r} + \nu_{i1,r}, \\
y^{\star}_{i2,r} &= \tilde{x}_{i,r}\beta_2 + \rho_1 \sum_{l=1}^{n_r} w_{2il} Y^{\star}_{l,r} + \rho_2 \sum_{l=1}^{n_r} w_{1il} Y^{\star}_{l,r} + \mu_{2,r} + \nu_{i2,r}
\end{aligned}
\tag{10}
$$

where $\rho_1 = \frac{\varphi}{(1-\psi^2)}$ and $\rho_2 = \frac{\psi\varphi}{(1-\psi^2)}$, by setting $\psi_1 = \psi_2 = \psi$. This restriction does not affect the identification of the marginal effects but simplifies the effect of the neighboring municipalities. In fact, the parameter ψ represents the effect of local dependence defined in the linear-in-means model. Because the reduced form involves solving the system with respect to the local dependence, the global dependence is now characterized by two spatial weight matrices w_{1ij} and w_{2ij}. The decomposition of other marginal effects and unobserved random terms are defined in Appendix B. For each pair of potential consolidation the error term can be rewritten as:

$$
\begin{pmatrix} \nu_{i1,r} \\ \nu_{i2,r} \end{pmatrix} \overset{iid}{\sim} N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, R \right],
$$

where the correlation matrix R is defined as

$$
R = \begin{pmatrix} 1 & \varrho \\ \varrho & 1 \end{pmatrix}.
$$

It is important to emphasize that the willingness to consolidate for each community $j = (1, 2)$ depends on the willingness of its neighbors but also on the neighbors of the other municipality under direct consideration for consolidation. For identification purposes, $var(\nu_{1,r})$ and $var(\nu_{2,r})$ are usually set to one. However we will propose a more flexible specification for the variance matrix that relaxes this constraint and allows for a simpler interpretation of the estimation results. A discussion about this variance constraint is provided in Appendix C. Identification conditions also require the set of explanatory variables of the small municipality to be distinct from the large municipality for each pair of potential consolidation (see Poirier 1980).[5]

Note that consolidation only happens if each municipality's board of education votes in favor of it. The consolidation between two contiguous municipalities fails if one or both of the boards votes against it. Because we observe consolidation or no consolidation, instead of actual consolidation votes by each municipality, we have a situation of partial observability. Partial observability for each potential consolidation $i = 1, \dots, n$ can be modeled using the following random variable:

$$z_{i,r} = y_{i1,r} y_{i2,r}.$$

Thus, the likelihood function is:

$$\begin{aligned} \ln(\beta_1, \beta_2, \varrho, \rho_1, \rho_2, \mu_1, \mu_2) \quad = \quad & \sum_i^n \{ z_{i,r} \ln F(\tilde{x}_{i,r}\beta_1, \tilde{x}_{i,r}\beta_2, \rho_1, \rho_2, \mu_1, \mu_2; \varrho) + \\ & (1 - z_{i,r}) \ln \left[1 - F(\tilde{x}_{i,r}\beta_1, \tilde{x}_{i,r}\beta_2, \rho_1, \rho_2, \mu_1, \mu_2; \varrho)\right] \} , \end{aligned}$$

where F denotes the bivariate standard normal distribution function with correlation coefficient ϱ for each of the potential consolidation $i = 1, 2, \dots, n$.

For estimation purposes, we can rewrite the bivariate expression in matrix form. In fact, for each potential consolidation pair i, we estimate the extent to which the willingness of municipality $j = 1$ to consolidate with its partner $j = 2$ is influenced by 1's own neighbors via the matrix $\tilde{W}_1$ and by the neighbors of its partner municipality $j = 2$ via the matrix $\tilde{W}_2$. Each spatial weight $\tilde{W}_o$, $o = (1, 2)$ is a $(2n \times 2n)$-dimensional matrix, where n is the number of potential consolidation pairs. The coefficients of $\tilde{W}_1$ and $\tilde{W}_2$ are defined as:

$$\tilde{w}_{1ij} = \begin{pmatrix} w_{ij11} & w_{ij21} \\ w_{ij12} & w_{ij22} \end{pmatrix} = \begin{pmatrix} w_{1ij} \\ w_{2ij} \end{pmatrix}, \tag{11}$$

$$\tilde{w}_{2ij} = \begin{pmatrix} w_{ij12} & w_{ij22} \\ w_{ij11} & w_{ij21} \end{pmatrix} = \begin{pmatrix} w_{2ij} \\ w_{1ij} \end{pmatrix}. \tag{12}$$

For matrix form, we set the dependent variable $Y^\star = (Y_1^{\star\prime}, \dots, Y_n^{\star\prime})'$ as a vector of dimension $(2n \times 1)$. The matrix of explanatory variables $X = (X_1', \dots, X_n')'$ is of dimension $(2n \times 4K)$ from which each pair i is defined with the $2K-$dimensional vector $\tilde{x}_i$ as $X_i = (I_2 \otimes \tilde{x}_i)$. The parameters of interest $b = (\beta_1', \beta_2')'$ are of dimension $(4K \times 1)$ and the $2n-$dimensional vector of error terms is defined as $V = (V_1', \dots, V_n')'$ with $V_i = (\nu_{i1}, \nu_{i2})'$. For each urban area r, the unobserved correlated random effects $M = (m_1', \dots, m_{\bar{r}}')'$ are of dimension $(2\bar{r} \times 1)$, with $m_r = (m_{1,r}, m_{2,r})'$. The $(2n \times 2\bar{r})$ matrix Δ serves to impose a restriction that unobserved heterogeneity is the same for all pairs of municipalities within the same urban area. The $(2n \times 2\bar{r})$ block diagonal matrix corresponds to $\Delta = diag(D_1, \dots, D_{\bar{r}})$

[5] The concatenation $\hat{X}$ with the correlated random effects defined in Appendix D ensures that small and large municipalities of the same pair have different values allowing for ψ and φ to be identified.

where each block is defined as $D_r = (\iota_{n_r} \otimes I_2)$ with ι_{n_r} the n_r-dimensional vector of ones. The spatial bivariate probit defined in (10) can now be rewritten in matrix form:

$$Y^\star = Xb + \rho_1 \hat{W}_1 Y^\star + \rho_2 \hat{W}_2 Y^\star + \Delta M + V. \tag{13}$$

Analyzing stationary conditions, Elhorst et al. (2012) show that $|\rho_1| + |\rho_2| < 1$ is too restrictive. We implement their algorithm to make sure spatial models are always stationary. It is also important to note that by construction, the two spatial weight matrices $\hat{W}_1$ and $\hat{W}_2$ are row-normalized and do not have common elements. Even if two municipalities have the same neighbor, this common neighbor will never be part of the same pair. It is important to note that the stability conditions remain identical to the linear-in-means model. In fact, if we assume both parameters ρ_1 and ρ_2 to be positive, then the constraint $\rho_1 + \rho_2 < 1$ is equivalent to $\frac{\varphi}{(1-\psi^2)} + \frac{\psi\varphi}{(1-\psi^2)} < 1$. Those constraints are actually the same as the stability condition $\psi + \varphi < 1$ of the Nash equilibrium of the linear-in-means model assuming that ψ and φ are positive.[6] If either ψ and φ are negative, the real minimum eigenvalues of the matrices $W1 + (\tilde{\rho}_2/\tilde{\rho}_1)W2$ for each coordinate point located at the border of the stationary region $(\tilde{\rho}_2, \tilde{\rho}_1)$ will have to be determined to establish the system of stationary conditions (see Elhorst et al. 2012).

4 Estimation Method and Monte Carlo Simulations

The estimation is achieved via Markov Chain Monte Carlo (MCMC) methods. Letting $B = (I_{2n} - \rho_1 \hat{W}_1 - \rho_2 \hat{W}_2)$, the simultaneous spatial autoregressive process can be expressed as:

$$BY^\star = Xb + \Delta M + V. \tag{14}$$

For Bayesian analysis of this bivariate probit form, we note that the posterior density is a function of the likelihood and the prior distribution as follows:

$$p(b, M, \rho_1, \rho_2, \Sigma | Y^\star, X) \propto p(Y^\star | b, M, \rho_1, \rho_2, \Sigma, X) p(b, M, \rho_1, \rho_2, \Sigma).$$

Because direct evaluation of the posterior density is computationally intensive, we use Metropolis-Hastings algorithms to produce a sequence of MCMC samples from the conditional posterior distributions for each parameter in the model. The sample of draws converges in distribution to a joint posterior for the parameters of our model.

We briefly note some features of our MCMC sampler, with details provided in Appendix D.

To investigate the finite sample performance of our estimator we simulate the following data generating process assuming full observability:

$$\begin{aligned} Y^\star &= Xb + \rho_1 \hat{W}_1 Y^\star + \rho_2 \hat{W}_2 Y^\star + \Delta M + V. \\ V &\sim N(0, I_n \otimes \Sigma) \end{aligned} \tag{15}$$

[6] It is easy to note that $\frac{\varphi}{(1-\psi^2)} + \frac{\psi\varphi}{(1-\psi^2)} < 1 \Leftrightarrow \frac{\varphi}{(1-\psi)} < 1$.

As detailed in Appendix D, unobserved correlated effects $M = (m'_1, \ldots, m'_{\bar{r}})'$ are defined for the small and large municipalities as $m_r = (m_{1,r}, m_{2,r})$ with $m_{j,r} = \alpha_j + \bar{\tilde{x}}_{j,r}\kappa_j + \mu_{j,r}$ where $\bar{\tilde{x}}_{j,r} = \frac{1}{n_r}\sum_{i=1}^{n_r}(\tilde{x}_{ij,r})$ represents the average of individual characteristics over each group r and $\mu_r \sim N(0, \Sigma_\mu)$. When running the algorithms, we use prior distributions defined in Appendix D with $b_0 = 1.01$. Pairs of potential consolidation are created by randomly placing 250 municipalities. Each municipality faces potential mergers with its neighbors. Pairs are sorted so that the first municipality is the smaller one. We define neighboring municipalities using the 5 nearest neighbors. For each dyad, or pair of neighboring municipalities, the weight matrices $\hat{W}_1$ and $\hat{W}_2$ initially have a weight of one if two pairs have one municipality in common. The $2n \times 2n$ weight matrix $\hat{W}_1$ represents the neighboring dyad for the smaller municipalities whereas $\hat{W}_2$ represents the neighboring dyad for the larger municipalities. Both spatial weight matrices are then row normalized.

The sample data are generated with one regressor ($k_1 = k_2 = 1$) for each municipality i, $x_{ijr} \sim N(0, 4)$. For the urban area r, x_{i1r} represents the characteristics for the small municipality and x_{i2r} corresponds to the same variables for the large municipality. In the bivariate setting, the matrix of idiosyncratic heterogeneity for each pair i corresponds to $X_i = I_2 \otimes [x_{i1r} \quad \sum_{l=1}^{n_r} w_{1il}x_{l,r} \quad x_{i2r} \quad \sum_{l=1}^{n_r} w_{2il}x_{l,r}]$. The matrix $X = (X'_1, \ldots, X'_n)'$ is of dimension $2n \times 8$. The true parameters are $b = \iota_2 \otimes (0.5, -0.5, 0.5, -0.5)'$, $\alpha = (1, 1)'$ $\kappa = (2, 2)'$, $\beta = (b', \alpha', \kappa')'$, $\rho = (0.35, 0.35)'$, and

$$\Sigma = \Sigma_\mu = \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}. \tag{16}$$

As in Lee (2007), we consider two situations with 50 groups or urban areas: a case where the average size for the urban area is small, and another where it is relatively large. In the first case, the sample size for urban areas varies from 3 to 8 with an average of 5. In the large case, the number of urban areas ranges from 15 to 25 with an average of 20 and a sample size of 950. This allows us to analyze the effect of increasing the average group size.[7] For both cases, we generate 20 replications of 5,000 iterations with a burn-in period of 4,000 to calculate the posterior means reported in Table (2). Unlike Lee (2007), we do not observe increased bias when groups are larger. In fact, the within group estimator that removes the unobserved group fixed effects in Lee (2007) makes the identification of ρ difficult when groups are large. Here we do not suffer from this problem since unobserved effects are estimated. We also present numerical standard errors (NSE) which give insight into the accuracy of the Monte Carlo approximation (see Geweke 1991 for a discussion).[8]

5 Data Description

There are important advantages of our data set over that used in Brasington (1999, 2003a, 2003b). First, we add observations from Texas, which not only contributes half our sample: it also allows us to test whether there are differences in how the consolidation decision

[7]These are the two extreme cases, any situations in between will provide more efficient estimates.

[8]Based on n iterations of the Gibbs sampler, Geweke (1991) proposes estimating the NSE of the sample mean of the function of interest $g(\Theta, y)$ by the square root of its asymptotic variance $S_g(O)/N$, where $S_g(O)$ is an estimate of the spectral density of the series $g(\Theta^{(i)}, y^{(i)})$ $(i = l, \ldots, S)$ at frequency zero, and $\Theta^{(i)}$ and $y^{(i)}$ are the sampled values.

is made in two parts of the U.S.A. that differ in culture and consolidation law. Second, the papers of Brasington (1999, 2003a, 2003b) omit central cities. The argument was that central cities contained heterogeneous populations, while the theoretical model contained internally homogeneous municipalities. The previous papers also omitted any municipality that sent a substantial portion of its population to more than one school district. Our current study includes central cities and municipalities with split school district assignments because it takes a much closer look at the data. For example, the City of Houston sends its children to sixteen school districts! While it would be easier to omit Houston as having not made a unified decision about where to send its students, we instead treat each of the sixteen pieces of the City of Houston as a separate municipality. We examine maps to see which Census tracts and block groups belong to the portion of the City of Houston that sends its students to Pasadena School District, and we add up the data from these Census tracts and block groups to find the population, income, racial composition, and all the other variables of interest for this new Houston City-Pasadena School District "municipality". The same is true for smaller municipalities like Texas City, whose students attend Texas City, Dickinson, and Lamarque School Districts. While the data work is much more time-consuming, it allows us to include a larger number of observations. For example, we cover the same urban areas of Ohio as Brasington (1999) did, but while he had 298 observations for Ohio, we have 821.

We draw our sample from the largest urban areas in Ohio and Texas in 2000. The legal and geo-political environment for school consolidation across both states is detailed in Appendix E. The Ohio urban areas are Toledo, Cincinnati, Cleveland, Columbus and Dayton, which range in population from 670,000 to 2.2 million. In Texas, our sample covers Houston (5.9 million), Dallas (6.4 million), and San Antonio (2.1 million).[9] The sample contains 1596 observations, where an observation is a pair of municipalities that could potentially consolidate schooling services or belong to different school districts. Of these 1596, Ohio contributes 821 observations and Texas 775. These 1596 observation pairs come from 655 municipalities (291 in Texas and 364 in Ohio). There are 227 school districts in our sample: 101 in Texas and 126 in Ohio. Texas school districts are fewer in number and larger in size than Ohio's, probably because Texas has a minimum size of nine square miles. This difference in consolidation decisions across both states forced us to separate the determinants of consolidation between the states of Ohio and Texas. Parameters capturing endogenous effects stay the same across states.

As explained in the previous section, we examine to what extent the larger community's consolidation decisions are different from the smaller community's decisions. As in Brasington (2003a), we split the sample into larger and smaller municipalities. Sources and summary statistics are presented in Table 3 and Table 4.

6 Estimation Results

To determine the robustness of the main socioeconomic factors influencing the willingness to consolidate, we decided to present two models capturing different proxy variables for

[9] Austin was excluded because it has few municipalities and school districts and would have added few observations.

racial composition, education, income and property value. For Model 1, we implement the following proxy variables: the percent of Hispanic population, percent of students enrolled in grades 1-12 who attend private schools, percent of occupied housing units that are occupied by renters rather than owners, percent of households that are married with own child 0-17 years and property value per pupil. For Model 2, the following proxy variables are considered: an index measuring ethnic heterogeneity, percent of persons in municipality aged 5-17 years who speak English less than "very well", percentage of persons 25 years or older in municipality whose highest educational attainment is no more than a high school degree, and average income of households. For both models, we generate a single chain of length 50,000 along with a burn-in period of 10,000 iterations. We use prior distributions defined in Appendix D with $b_0 = 1.01$.

Estimation results presented in Table 5 and Table 10 show how the larger and smaller members of each potential consolidation pair decide whether to consolidate with each other. We observe a significant role for spatial dependence. We consider this one of the most important findings of the paper. As previously detailed, there are two parameters measuring spatial dependence. The parameter estimates for $\rho_1 = \frac{\varphi}{(1-\psi^2)}$, is about 0.39 for Model 1 and 0.32 for Model 2. Recall that there are two members for each pair of municipalities that could consolidate. The parameter ρ_1 suggests that the willingness to consolidate for one member of the pair is influenced by its own neighbors. The theoretical model suggests also that each municipality in a potential consolidation pair is affected by the neighbors of the other municipality in the pair. This effect is captured by the parameter ρ_2. The spatial dependence parameter $\rho_2 = \frac{\psi\varphi}{(1-\psi^2)}$ for both models is around 0.2. It suggests that the neighbors of the other member of any pair affect the decision to consolidate. For each pair of municipalities that could consolidate, the neighbors of both municipalities directly affect the willingness to consolidate of the two municipalities in the pair. More importantly the theoretical model allows us to separate the direct influence of the other member of the pair (local effect) from the influence of neighboring pairs (global effect).

More importantly, from the reduced form of the theoretical model defined in (10) we find the endogenous peer effects coming from neighboring pairs ($\psi = 0.54$ for Model 1 $\psi = 0.63$ for Model 2) to be greater than the direct effect coming from the other municipality in each pair of potential consolidation ($\varphi = 0.18$ for Model 1 and $\varphi = 0.12$ for Model 2). This result confirms the importance of modeling the influence of neighbors of the other member of our pair. Traditional bivariate probit models that omit endogenous peer effects will obtain biased estimates. The other statistic of interest, Σ_{12}, measures the correlation between the two decision makers' decisions. The estimate is to 0.35 for both models suggesting that the decision process cannot be estimated separately between the larger and smaller member of each consolidation pair.

With the presence of local and global interactions, any change of a particular municipality's characteristics will have multiple impacts. Before discussing those impacts in greater detail, marginal effects obtained from the reduced form (13) are presented in Table 5 and Table 10. Even though we cannot interpret them directly, it is possible to retrieve their initial values by observing that $b = (\beta_1', \beta_2')'$ and $\beta_1 = [\delta_1 \; \psi\delta_2] / (1 - \psi^2)$, $\beta_2 = [\psi\delta_1 \; \delta_2] / (1 - \psi^2)$. Since ψ has been previously identified, the effects of individual characteristics as well as contextual effects can now be retrieved from δ_1 and δ_2 as defined in (9). For instance, in Model 1, the contextual effect of the property values per pupil of the small municipality on its willingness to consolidate is equal to 0.114 ($= 0.081/(1 - 0.54^2)$).

There are many ways to calculate marginal effects. We decide to interpret the model estimates of the spatial Durbin bivariate probit model in light of the approach proposed by Greene (1996) that is based on the derivatives of $Prob\,[y_1 = 1, y_2 = 1|x_1, x_2]$. Details are provided in Appendix F. With the introduction of spatial dependence, changes in one explanatory variable associated with one municipality will directly affect the willingness to consolidate of the other member of the pair (bivariate model) but will also be reflected in the decisions of all other municipalities located within the same urban area. As described in LeSage et al. (2011) the indirect effects, evaluated through cross-partial derivatives, cumulate the spillover effects falling on all other observations, but the magnitude will be greater for neighboring municipalities. The concept of direct and indirect effect is similar to the one detailed in Bramoulle et al. (2014) and Ballester et al. (2006) where municipalities will have a direct effect on some other municipalities through the spatial connectivity matrix W and the indirect effects will depend on higher orders of the matrix W^k.

The direct effects capture only the own partial derivatives including the feedback effects. The sum of the direct and indirect effects represents the total effects and reflects the cumulative change in probability of consolidation arising from a change in an explanatory variable in a typical municipality. For each urban area r, the cumulation process occurs over the n_r potential pairs of consolidation. The spatial weight matrix being block diagonal, there will be no spillover impact between urban areas. Since changes are observed for the explanatory variables of each municipality, each change gives rise to $2n_r$ responses including our own municipality, the other municipality of the pair as well as all other municipalities of the other pairs. We obtain a $(2n \times 2n)$ block diagonal matrix of partial derivatives reflecting all responses. To overcome the difficulty of presenting an $(2n \times 2n)$ matrix of partial derivatives for each explanatory variable, LeSage and Pace (2009) propose the use of average scalar summary measures. LeSage et al. (2011) underline that if explanatory variables vary substantially over space, scalar summary measures might not be useful for global inferences. Since we only work with urban areas, the variables only vary across a limited geography. Therefore, summary measures for the marginal effects are more globally representative. Table 6, Table 7 and Table 8 show the posterior means of the direct, indirect, and total marginal effects for Model 1, and Table 11, Table 12 and Table 13 show the same for Model 2. These scalar summary estimates required for proper inference are calculated during each iteration of the MCMC sampling.

We focus first on the impact of the size of the municipality, as measured by the number of pupils, on the willingness to consolidate. Because of its importance, *pup* is the only variable appearing in both models. Across specifications we find consistent results. Smaller municipalities are more reluctant to consolidate. Analyzing the direct effects in Table 6 and Table 11, we observe that in Ohio, the smaller a municipality gets the less tempting are the economies of scale gains of consolidation; the city instead prefers to enjoy complete political control over the provision of public schooling. The negative estimates of (-0.002) of *pupL* for Model 1 and (-0.003) for Model 2 reveal that the small municipality in a pair of potential consolidation is more reluctant to consolidate when the other municipality gets larger. Unlike Brasington (2003a), the effects of the number of pupils on school consolidation is different between larger and smaller municipalities. Table 6 shows that in both Ohio and Texas the large municipality is more willing to consolidate if it has more pupils. The positive estimates of *pupL* (+0.003) in Ohio and (+0.001) in Texas confirm that the big municipality can gain some scale economies without sacrificing political control. This is consistent with the prediction of Ellingsen's (1998) theory that large jurisdictions favor consolidating with smaller cities. A large municipality is more encouraged to consolidate with the smaller municipality, and the effect is stronger the smaller the small municipality is. Moreover, indirect effects summarized in Table 7 and Table 12 show that the large municipality for each pair in Ohio will have greater positive spatial spillover impact (+0.007 for Model 1 and +0.002 for Model 2) on its larger neighboring municipalities.

The literature consistently shows that property value is a strong driver of the probability of consolidation. The current study finds that, the richer the municipality's school district is in property value, the less willing it is to consolidate. Any consolidation would dilute the municipality's property tax base. Similarly, large Ohio municipalities are less willing to consolidate the more property value they have; and large Texas municipalities are more willing to consolidate when the smaller municipality of the consolidation has a higher per-pupil property value. For both states, the large municipality has a positive spillover impact on the neighbors of the smaller municipality *valpup_allS* with a value of +0.018 for Ohio and +0.003 for Texas. Spillover effects are important particularly for the smaller municipality in Ohio for which the total effect of +0.020 in Table 8 is mainly explained by the large +0.016 impact in Table 7 of property value of the neighbors of the larger municipality for each potential consolidation pair.

Similar arguments can be made analyzing the impact of average household income where the small municipality is more willing to consolidate if the large municipality has higher income (+0.047 for the variable *avincL* in Texas in Table 11).

Like Alesina et al. (2000), which focused on the impact of population heterogeneity on jurisdictional consolidation, we find that racial composition does affect school consolidation in Ohio and Texas. There is a significant role for Hispanic in influencing consolidation in Ohio and Texas. However Ohio seems more influenced by the presence of Hispanics, despite a much larger presence of Hispanics in Texas than Ohio. For each pair of potential consolidation in Ohio, the smaller municipality is less likely to consolidate if the larger municipality has a larger percentage of Hispanics (−0.011). While there is no significant direct effect of population heterogeneity *leikrace* on the consolidation decision we do find negative spillover effects in Ohio in Table 12. In fact, neighbors of large municipalities in Ohio with higher levels of racial heterogeneity decrease the probability of consolidation for both municipalities of each pair. Indirect effects on smaller municipalities are equal to −0.644 and reach −0.815 on larger municipalities.

Consistently significant in Ohio, the percentage of school-aged children who do not speak English well (*pctbadenglish*) does influence the willingness to consolidate. Table 11 shows that for large jurisdictions, having a higher percentage of children with poor English language skills makes them more willing to consolidate (+0.056). Larger municipalities are less likely to consolidate (-0.060) than smaller (-0.027) when the other municipality of the pair has a higher percentage of school-aged children who do not speak English well.

Having a greater private school market share reduces consolidation effort. Among the most important drivers of private school attendance is the quality of public education available and the ability of residents to afford private schooling. For smaller municipalities in Ohio and Texas, having larger enrollment in private schools reduces their effort toward consolidation (-0.003 in Ohio and -0.002 in Texas, from Table 6). For larger Texas municipalities, private school attendance is positively related to the likelihood of consolidation. For each pair of potential consolidation in Texas, we observe positive spillover effects ($+0.008$ for *privL* in Table 7) between the smaller municipality and the neighbors of the larger municipality.

The percentage of residents who have no more than a high school degree sometimes influences consolidation. Similar to the results for *pctbadenglish*, municipalities with low education levels are more willing to consolidate with their neighbors. This is true for small municipalities in Ohio and Texas, as well as for large Texas municipalities.

Having a high percentage of renters affects the likelihood of consolidation, especially for the larger Ohio and Texas municipalities. The more renters there are in the small Ohio and Texas municipalities, the less larger municipalities are willing to consolidate schooling with them.

Family composition is also an important determinant of consolidation. For both Ohio and Texas, the willingness to consolidate will be higher if the potential candidate has a higher proportion of families married with children and lower if the municipality itself already has a high proportion of families married with children.

Finally, it is important to analyze the unobserved effects associated with each urban area. Predicted values of correlated random effects defined in Appendix D are reported in Table 9 for Model 1 and Table 14 for Model 2. Estimation results show that despite the higher number of consolidations we observe in Texas, the smaller municipalities are less likely to consolidate. Dallas and Houston in Texas and Toledo in Ohio are the only urban areas for which both models show a negative willingness to consolidate. The decision to consolidate schooling services with a neighbor varies across urban areas and further reveals the reluctance for smaller jurisdictions to do so. Analyzing municipality characteristics and their effect on neighboring decisions provides insightful explanations about the main factors related to school district consolidation. However, unobserved heterogeneity is only based on municipality characteristics averaged over urban areas, and the variation in the willingness to consolidate reveals that urban area characteristics might also provide further insights into school consolidation. The literature has mainly focused on explaining differences in public goods provision and school quality between municipalities but little is known about common factors existing at the metropolitan area level. A more coherent regional strategy, operating at the urban area level, could increase efficiency in public goods provision. For instance different degrees of racial segregation across metropolitan areas could be associated with a different willingness to spend on local public goods. Hoxby (2000) analyzes the effects of school choice across metropolitan area and shows that public schools are more productive with greater Tiebout choice among districts.

7 Conclusion

This paper proposes an approach to modeling public school consolidation using a social interaction framework allowing for both correlations between alternative consolidation choices as well as between individuals. We also help model unobserved heterogeneity in the urban area in which each municipality is located.

To illustrate this new model, we analyze how municipalities interact in making decisions about consolidation of schooling services. Strategic interaction is key to better understanding why some pairs of municipalities have a low probability of consolidation.

To reach these conclusions, we try to clearly identify the peer effects from both the contextual and unobserved correlated effects. We provide necessary and sufficient conditions for identification and propose a specification ensuring the existence of a unique equilibrium. To this end, we extend the most current estimation procedure for bivariate probit models with partial observability by introducing a spatial Durbin model with correlated random effects. The identification of these effects is of paramount importance for policy purposes. We derive analytical expressions related to partial derivatives measuring how changes in values of an explanatory variable impacts its own and other municipalities. It allows us to disentangle direct effects coming from a change in one's own characteristics from indirect or spillover effects coming from a change in the other municipalities' characteristics or willingness to consolidate.

Based on our theoretical framework in which both municipalities must agree to consolidate public good provision, differences in size discourage consolidation for the smaller municipality. In fact, only larger differences in size make big municipalities more willing to consolidate. Since smaller municipalities have veto power, larger differences in size hinder consolidation. By analyzing the willingness to consolidate between pairs of contiguous municipalities, results reveal that the merger decision is directly influenced by neighboring pairs that could consolidate. However the direct influence of the other municipality of the pair is still more important than neighboring merger decisions.

Similar to Alesina et al. (2004) we also find evidence that municipalities avoid heterogeneity because they do not want to interact with different people, but we also find evidence that people avoid heterogeneity because of the difficulty of cooperating in the provision of local public goods.

References

[1] Alesina, A., Baqir R. and Hoxby C. (2004). "Political jurisdictions in heterogeneous communities",*Journal of Political Economy*, 112, 348-396.

[2] Anderson, Simon P., Jacob K. Goeree, and Charles A. Hol, (1998). "A Theoretical Analysis of Altruism and Decision Error in Public Goods Games." *Journal of Public Economics* 70, 297-323.

[3] Ballester, Coralio, Antoni Calv-Armengol, and Yves Zenou. 2006. "Who's Who in Networks: Wanted: The Key Player." *Econometrica* 74 (5): 140317.

[4] Bates, Laurie J., Lafrancois, Becky A. and Santerre, Rexford E. 2011. An empirical study of the consolidation of local public health services in Connecticut. *Public Choice* 147(1-2) p. 107-121.

[5] Barnard, J., McCulloch, R., and Meng, X. (2000), "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations, with Application to Shrinkage," *Statistica Sinica*, 10, 1281-1311.

[6] Blume, L, W. Brock, S. Durlauf, and Y. Ioannides (2011), "Identification of Social Interactions," in J. Benhabib, A. Bisin, and M. Jackson (eds.), *Handbook of Social Economics*, Amsterdam: North-Holland.

[7] Bramoulle, Y., Kranton, R. and M. D'Amours (2014). "Strategic Interactions and Networks," *American Economic Review* (104)3, 898-930.

[8] Brasington, D.M. (1999). "Joint provision of public goods: The consolidation of school districts", *Journal of Public Economics*, **73**, 373-93.

[9] Brasington, D.M. (2003a). "Size and School District Consolidation: Do Opposites Attract?", *Economica*, 70, 673-690.

[10] Brasington, D. (2003b) Snobbery, Racism, or Mutual Distaste: What Promotes and Hinders Cooperation in Local Public-Good Provision? *Review of Economics and Statistics* 85, 874 883.

[11] Brock, William A and Durlauf, Steven N, (2001). "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68(2), 235-60

[12] Chamberlain, G. (1984), "Panel Data," in *Handbook of Econometrics*, Volume 2, ed. Z. Griliches and M.D. Intriligator. Amsterdam: North Holland, 1248-1318.

[13] Chib, S. and Greenberg, E. (1998), "Analysis of Multivariate Probit Models," *Biometrika*, 85, 347-361.

[14] Chopin, N. (2011) "Fast simulation of truncated Gaussian distributions", *Statistics and Computing*, 21(2), 275-288

[15] Elhorst, J. P., Lacombe, D. J. and Piras, G. (2012), "On model specification and parameter space definitions in higher order spatial econometric models", *Regional Science and Urban Economics* 42(1), 211-220.

[16] Elligsen T. (1998). "Externalities vs internalities: A model of political integration". *Journal of Public Economics*, **68**, 251-68.

[17] Geweke, J. (1991) Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments, in Bernardo, J., Berger, J., Dawid, A. and Smith, A. (eds.), *Bayesian Statistics 4*, pp. 641-649. Oxford: Clarendon Press.

[18] Glaeser, E.L., Scheinkman, J.A., 2001. Measuring social interactions. In: Durlauf, S.N., Young, H.P. (Eds.), *Social Dynamics.* Brooking Institute, Washington, DC, pp. 83131.

[19] Gordon, N. and B. Knight. (2009). "A Spatial Merger Estimator with an Application to School District Consolidation," *Journal of Public Economics*, **93**, 752-765.

[20] Greene W. (1996) "Marginal Effects in the Bivariate Probit Model." Working Paper No. 96-11, Department of Economics, Stern School of Business, New York University,

[21] Hoxby, C.M. (2000) "Does Competition among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, 90(5), 1209-38.

[22] Imai, K. and van Dyk, D. A. (2005). "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation", *Journal of Econometrics* 124(2): 311-334.

[23] Lee, L.F. (2007) "Identification and Estimation of Econometric Models with Group Interactions, Contextual Factors and Fixed Effects," *Journal of Econometrics* 140,2, 333-374.

[24] Lee, L.F., Li, J. and X. Lin (2014), "Binary Choice Models with Social Network under Heterogeneous Rational Expectations" *Review of Economics and Statistics*, forthcoming.

[25] LeSage, J. and R. K. Pace (2009). *An Introduction to Spatial Econometrics*, Chapman Hall/CRC Press.

[26] LeSage JP, Pace RK, Lam NSN, Campanella R, Liu X (2011) "New Orleans business recovery in the aftermath of Hurricane Katrina". *Journal of Royal Statistics Society A* 174(4), 10071027

[27] Liu Xiaodong, Eleonora Patacchini and Yves Zenou (2011) "Peer Effects in Education, Sport, and Screen Activities: Local Aggregate or Local Average?", working paper.

[28] Liu, J. and Wu, Y. (1999), "Parameter Expansion for Data Augmentation," *Journal of the American Statistical Association*, 94, 1264-1274.

[29] Manski, Charles. 2004. "Measuring Expectations." Econometrica, 72(5): 1329-1376.

[30] McCulloch, R., Polson, N., and Rossi, P. (2000), "A Bayesian Analysis of the Multinomial Probit Model with Fully Identified Parameters," *Journal of Econometrics*, 99, 173-193.

[31] McCulloch Robert E. and Rossi Peter E. (1994), "An Exact Likelihood Analysis of the Multinomial Probit Model.", *Journal of Econometrics*, 64(1-2), 207-40.

[32] Moffitt, R. A. (2001). Policy interventions, low-level equilibria, and social interactions, in S. N. Durlauf and H. P. Young (eds), *Social Dynamics*, MIT press, Cambridge.

[33] Nelson, M.A., (1990). Decentralization of the Subnational Public Sector: An Empirical Analysis of the Determinants of Local Government Structure in Metropolitan Areas in the U.S. *Southern Economic Journal* 57(2), 443-457.

[34] Nobile, A. (2000), "Comment: Bayesian Multinomial Probit Models with a Normalization Constraint," *Journal of Econometrics*, 99, 335-345.

[35] Poirier, D. (1980). Partial observability in bivariate probit models. *Journal of Econometrics*, **12**, 209-17.

[36] Sorensen, Rune J. (2006). Local government consolidations: the impact of political transaction costs. *Public Choice* 127(1-2) p. 75-95.

[37] Talhouk A., A. Doucet, K. Murphy (2012) "Efficient Bayesian Inference for Multivariate Probit Models with Sparse Inverse Correlation Matrices". *Journal of Computational and Graphical Statistics*, 21, 739-757;

[38] Tamer, E. (2003), "Incomplete simultaneous discrete response models with multiple equilibria", *Review of Economic Studies* 70, 147165.

[39] Tanner,M. A., and Wong,W. H. (1987), "The Calculation of Posterior Distributions by Data Augmentation" (with discussion), *Journal of the American Statistical Association*, 82, 528550.

[40] Texas Education Agency (2013). "Consolidations, Annexations, and Name changes For Texas Public Schools", updated 8/13/2013.

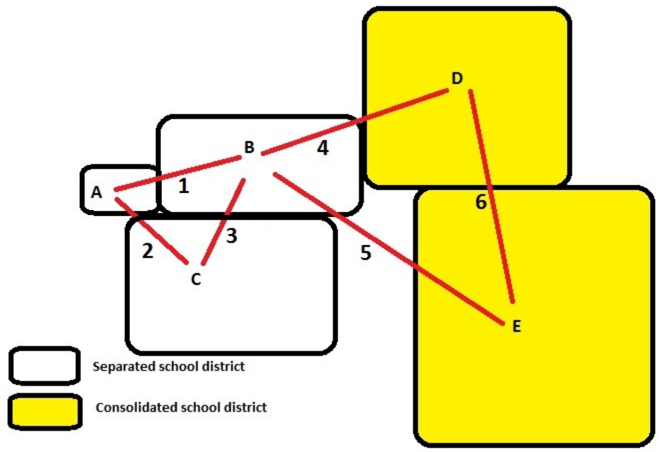
Appendix A - Illustration of Consolidation

Figure 1 illustrates a situation of an urban area containing five municipalities. In this study, the arrangement of the data is similar to Brasington (2003a, 2003b). Each municipality can maintain its own school district or it can consolidate with a contiguous neighbor. For this illustration only two municipalities decide to consolidate. Therefore, we observe four independent school districts. All pairs of potential consolidation are observed at the municipality level. Thus the dependent variable would take the value one only for pair six connecting municipalities D and E and zero for all other pairs.

The first municipality of each potential matching pair always has the lower number of pupils. We arrange the data to analyze to what extent smaller and larger communities differ in their willingness to consolidate. In Figure A.1, the physical size of each municipality reflects the relative number of pupils.

In Figure 1 only municipalities D and E consolidate. Municipalities B and E are not adjacent. They cannot legally form a consolidated school district with each other; therefore, there should be no potential consolidation pair observation between B and E directly. However, because D and E consolidate their schools, B may try to join the consolidated school district D&E. Municipality B may merge with E, then, but only through the actual consolidated school district. Therefore, potential consolidations are observed between (1) adjacent municipalities or (2) municipalities and adjacent consolidated school districts. Neighboring pairs are defined as dyads having one municipality in common. For instance, for the fourth pair of potential consolidation between municipalities B and D, B is influenced by the willingness to consolidate of two smaller municipalities A and C (via the first and third pairs, respectively) and one larger E (via the fifth pair) whereas D is only influenced by the larger municipality E (via the sixth pair). Note that the second pair of potential consolidation has no direct influence on the fourth pair of consolidation. It is important to note that each municipality will appear in several pairs, each time revealing a different willingness to consolidate.

Figure 1: School District Consolidation



Appendix B -Reduced form of Discrete Choice Modeling

The error terms $\epsilon_{i,r} = (\epsilon_{i1,r}, \epsilon_{i2,r})$ are assumed to follow a standard bivariate normal distribution, with correlation parameter ϱ. The correlation between the two disturbances allows for interdependence between the utility functions of the two municipalities. Each vote is a function of the sentiment toward consolidation with other jurisdiction. Thus, (7) and (8) are equivalent to:

$$\begin{aligned} y^{\star}_{i1,r} &= \psi_1 y^{\star}_{i2,r} + \varphi \sum_{l=1}^{n_r} w_{il11} y^{\star}_{l1,r} + \\ &\quad \varphi \sum_{l=1}^{n_r} w_{il21} y^{\star}_{l2,r} + \tilde{x}_{i1,r}\delta_1 + \lambda_{1,r} + \epsilon_{i1,r} \\ y^{\star}_{i2,r} &= \psi_2 y^{\star}_{i1,r} + \varphi \sum_{l=1}^{n_r} w_{il22} y^{\star}_{l2,r} + \\ &\quad \varphi \sum_{l=1}^{n_r} w_{il12} y^{\star}_{l1,r} + \tilde{x}_{i2,r}\delta_2 + \lambda_{2,r} + \epsilon_{i2,r}. \end{aligned}$$

These latent utilities are identical to the best-response functions defined in (5). From random utility maximization theory, municipality j of pair i will vote in favor of consolidation $m = 1$ if:

$$y_{ij,r} = 1 \ \ iff \ \ y^{\star}_{ij,r} > 0, \ \ i.e., \ \ U_{ij1,r} > U_{ij0,r}.$$

The reduced form is equivalent to:

$$
\begin{aligned}
y^{\star}_{i1,r} &= \tilde{x}_{i,r}\beta_1 + \frac{\varphi}{(1-\psi_1\psi_2)}\left(\sum_{l=1}^{n_r} w_{il11}y^{\star}_{l1,r} + \sum_{l=1}^{n_r} w_{il21}y^{\star}_{l2,r}\right) \\
&\quad + \frac{\psi_1\varphi}{(1-\psi_1\psi_2)}\left(\sum_{l=1}^{n_r} w_{il22}y^{\star}_{l2,r} + \sum_{l=1}^{n_r} w_{il12}y^{\star}_{l1,r}\right) + \mu_{1,r} + \nu_{i1,r} \\
y^{\star}_{i2,r} &= \tilde{x}_{i,r}\beta_2 + \frac{\varphi}{(1-\psi_1\psi_2)}\left(\sum_{l=1}^{n_r} w_{il22}y^{\star}_{l2,r} + \sum_{l=1}^{n_r} w_{il12}y^{\star}_{l1,r}\right) + \\
&\quad \frac{\psi_2\phi}{(1-\psi_1\psi_2)}\left(\sum_{l=1}^{n_r} w_{il11}y^{\star}_{l1,r} + \sum_{l=1}^{n_r} w_{il21}y^{\star}_{l2,r}\right) + \mu_{2,r} + \nu_{i2,r}
\end{aligned} \tag{17}
$$

The explanatory variables are based on the concatenation $\tilde{x}_{i,r} = [\tilde{x}_{i1,r}\ \tilde{x}_{i2,r}]$ of dimension $(1 \times 2K)$, $K = k_1 + k_2$ and the parameters of interest β_j are of dimension $(2K \times 1)$. In fact, $\beta_1 = [\delta_1\ \psi_1\delta_2]/(1-\psi_1\psi_2)$, $\beta_2 = [\psi_2\delta_1\ \delta_2]/(1-\psi_1\psi_2)$, the group or urban area effects are now defined as $\mu_{1,r} = (\lambda_{1,r}+\psi_1\lambda_{2,r})/(1-\psi_1\psi_2)$ and $\mu_{2,r} = (\lambda_{2,r}+\psi_2\lambda_{1,r})/(1-\psi_1\psi_2)$ and the error terms are equal to $\nu_{1,r} = (\epsilon_{1,r}+\psi_1\epsilon_{2,r})/(1-\psi_1\psi_2)$ and $\nu_{2,r} = (\epsilon_{2,r}+\psi_2\epsilon_{1,r})/(1-\psi_1\psi_2)$.

Appendix C - Estimation Procedure

Identifiability

The traditional multivariate probit model assumes that each subject $i = (1,\ldots,n)$ has J distinct binary responses. Let $Y_i^{\star} = (Y_{i1}^{\star},\ldots,Y_{iJ}^{\star})'$ denote the $J-$dimensional vector of latent utilities received by subject i and X_{ij} represent $1 \times K$ vectors of explanatory variables associated with each response. Assuming that K is the same for all responses, $X_i = diag(X_{i1}\ldots,X_{iJ})$. The parameters of interest $\beta = (\beta_1',\ldots,\beta_J')'$ form a $JK \times 1$ vector of unknown regression coefficients and ϵ_i is a $J \times 1$ error vector, normally distributed with a 0 mean and a $J \times J$ variance matrix Σ. The parameters of interest (β,Σ) are not identifiable from the observed data (Chib and Greenberg, 1998). There are an infinite number of values for those parameters which yield exactly the same model. In fact, by multiplying both sides by a positive scalar parameter υ, we obtain

$$\upsilon Y^{\star} = \upsilon(X\beta + \epsilon) = X(\upsilon\beta) + \upsilon\epsilon$$

where $Y^{\star} = (Y_1^{\star\prime},\ldots,Y_n^{\star\prime})'$ is the $nJ \times 1$ vector of latent utilities and $X = (X_1',\ldots,X_n')'$ the $nJ \times kJ$ matrix of explanatory variables. The $nJ \times 1$ vector of observed values Y will not be affected by this scalar parameter and the likelihood of $Y|X,\beta,\Sigma$ will be the same as that of $Y|X,\upsilon\beta,\upsilon^2\Sigma$. The probit model cannot distinguish β and Σ separately. To ensure identifiability of the model parameters, restrictions need to be imposed on the covariance matrix. Following Imai and Van Dyk (2005), we use the symbols $\tilde{\Sigma}$, $\tilde{\beta}$ and $\tilde{\mu}$ to represent the unidentified parameters. Removing the $\tilde{\Sigma}$ symbol signifies the identified parameters Σ, β and μ.

In the univariate case, the standard solution to this problem is to set the variance to one. However, defining such a restriction in the bivariate case is more complex. An alternative solution proposed by McCulloch and Rossi (1994) is to ignore the identifiability issue and analyze the unidentified model by scaling with the sampling covariance through the reduction function $R = d^{-1}\tilde{\Sigma}d^{-1}$, where d is a diagonal matrix with diagonal elements $d_i = \sqrt{\tilde{\Sigma}_{ii}}$. But extra care must be given to the choice of prior distributions for unidentified parameters. McCulloch et al. (2000) propose to work with an identified model by setting the first diagonal element of the covariance matrix $\sigma_{11} = 1$. Traditional Wishart distributions for covariance can no longer be used: they partition the covariance matrix and impose a specific prior on the identified parameters, but their algorithm is slower to converge. For that reason, Nobile (2000) suggests direct simulation from Wishart distributions conditional on one element of the diagonal. However, for all these approaches, imposing a normalization constraint increases the complexity of interpretation for the parameters and priors.

Alternatively, Chib and Greenberg (1998) restrict the entire identified covariance matrix Σ to be a correlation matrix R. An efficient computational approach has been proposed by Talhouk et al. (2012) to overcome the fact there is no conjugate prior for the correlation matrices. Following Bernard et al. (2000), they impose a prior on the identified matrix R whose elements follow marginal uniform distributions over the interval $[-1, 1]$. This marginally uniform prior distribution for R is given by:

$$p(R) \propto |R|^{\frac{J(J-1)}{2}-1} \left(\prod_i |R_{ii}|^{-(J+1)/2} \right) \quad (18)$$

where R_{ii} is the principal submatrix of R. Even if this distribution is not conjugate, sampling can be easily obtained from a standard inverse Wishart with degrees of freedom equal to $\nu = J + 1$ and then scaled back to a correlation matrix using the variance decomposition $(\tilde{\Sigma} = dRd)$. For the bivariate specification, we have $J = 2$ and henceforth we will refer to the correlation matrix R as the identified matrix Σ.

Parameter Expansion

In many situations, Gibbs sampling methods mix poorly and can be slow to converge. Parameter Expanded Data Augmentation or Marginal Augmentation is a technique introduced by Liu and Wu (1999) to improve convergence by adding an additional parameter.

Assume that a multivariate specification is represented by an $nJ \times 1$ vector of observed data Y that is incomplete and an $nJ \times 1$ vector $Y^\star$ represents its missing parts. Suppose we can identify an hidden parameter υ from the complete data model $p(Y, Y^\star|\theta)$ where θ represents the parameter of interest and $p(Y, Y^\star)$ is the marginal distribution of the complete data. Then we can generate new latent data W from the transformation of the latent data $Y^\star$ induced by the expansion parameter υ. This extension preserves the observed-data model:

$$\int p(Y, W|\theta, \upsilon)dW = p(Y|\theta).$$

The link between the newly generated data W and the missing data $Y^\star$ corresponds to a one-to-one differentiable mapping $Y^\star = t_\upsilon(W)$. In other words, the working parameter υ is identifiable given the observed data and augmented data but cannot be identified given the observed data alone. Talhouk et al. (2012) define the mapping function between $Y_i^\star$ and W_i as the diagonal matrix d of the unidentified variance matrix $\tilde{\Sigma}$ such that the ith element $d_i = \sqrt{\tilde{\Sigma}_{ii}}$. They conveniently pick $\upsilon = (\upsilon_1, \ldots, \upsilon_J)$ to be a function of d by taking $\upsilon_i = s^{ii}/(2d_i^2)$ where s^{ii} is the ith diagonal element of $\tilde{\Sigma}^{-1}$ and d_i is the ith diagonal element of $d = (d_1, \ldots, d_J)$. Talhouk et al. (2012) show that the transformation of the latent data can be defined as $Y^\star = t_\upsilon(W) = D^{-1}W$ where $D = I_n \otimes d$ and each element d_i^2 follows an inverted Gamma distribution:

$$d_i^2 \sim IG((J+1)/2, s^{ii}/2)$$

This function is conveniently chosen so that its combination with the prior of $\theta = (\beta, \Sigma)$, the transformed likelihood, and the Jacobian, will result in posterior distributions that are easy to sample from. Using the transformation $\tilde{\Sigma} = d\Sigma d$, Barnard et al. (2000) emphasize that if $\tilde{\Sigma}$ follows a standard inverse Wishart distribution then $\tilde{\Sigma}$ is equivalent to:

$$p(\tilde{\Sigma}) = p(d, \Sigma) \times |J : \tilde{\Sigma} \to d, \Sigma| = p(\Sigma)p(d|\Sigma)$$

where the prior distribution $p(\Sigma)$ has the attractive property of being marginally uniform for each element s_{ij} on the interval $[-1, 1]$. As previously stated, the prior distribution for $p(d|\Sigma)$ is a Gamma distribution whose shape and rate parameters are equal to $(J+1)/2$ and 1, respectively. Therefore, we obtain the desired prior distribution $p(d|\Sigma)p(\Sigma)$ by sampling $\tilde{\Sigma}$ from a standard inverse Wishart and using the transformation $\tilde{\Sigma} = d\Sigma d$.

For our bivariate specification defined in (23) below, we need to define a joint prior on d and $\theta = (\beta, \mu, \rho, \Sigma_\mu, \Sigma)$. We assume that β, ρ, μ, Σ_μ and Σ are independent a priori so that $p(\beta, \mu, \Sigma, \Sigma_\mu, \rho, d) = p(d|\Sigma)p(\Sigma)p(\beta)p(\mu)p(\rho)p(\Sigma_\mu)$.

As detailed in Talhouk et al. (2012), (β, μ, Σ) are sampled under the expanded data model. The transformation on the latent variables is obtained from $W = Y^\star D$ where $D = I_n \otimes d$ is the expansion parameter with $d_i > 0$. Therefore, following the model defined in (23), the expanded likelihood follows:

$$p(W|\beta, \mu, \Sigma, \rho, d) \sim N\left(B^{-1}(DX\beta + D\Delta\mu), B^{-1}(I_n \otimes d\Sigma d)B^{-1\prime}\right) \tag{19}$$

where $B^{-1} = (I_{2n} - \rho_1 \hat{W}_1 - \rho_2 \hat{W}_2)^{-1}$. It is important to note that ρ and Σ_μ can be directly sampled from the non-expanded identified specification. A summary of the algorithm we adopt is presented in Table 1.

Appendix D -Conditional Posterior Distributions

Based on (15), a formal statement of the spatial bivariate probit model with group or urban area effects can be written as:

$$\begin{aligned} BY^\star &= Xb + \Delta M + V && (20)\\ M &\sim N(M_0, I_{\bar{r}} \otimes \Sigma_\mu) && (21)\\ V &\sim N(0_{2n}, I_n \otimes \Sigma) && \end{aligned}$$

with X the $(2n \times 4K)$ matrix of explanatory variables and b the $(4K \times 1)$ vector of parameters of interest. Following Chamberlain (1982), correlated random effects $M = (m_1', \ldots, m_{\bar{r}}')'$ are defined for the small and large municipalities as

$$m_{j,r} = \alpha_j + \bar{\bar{x}}_{j,r}\kappa_j + \mu_{j,r} \tag{22}$$

Table 1: Full Parameter Expansion Sampling Scheme for Spatial Bivariate Probit Model

At iteration t

Data Augmentation Step

- Draw $Y^\star \sim TMVN(B^{-1}(x\beta + \Delta\mu), B^{-1}(I_n \otimes \Sigma)B^{-1\prime})$ distribution
- Draw $d_j^2 \sim IG(\frac{J+1}{2}, s^j/2)$, where s^j is the j^{th} diagonal element of Σ^{-1} for $j = (1, 2)$
- Set $W_i = dY_i^\star$ for $i = (1, \ldots, n)$

Posterior Distribution

- Draw $\tilde{\Sigma}|\rho, W$ from the inverse Wishart distribution (25)
- Draw $\tilde{\beta}|W, \tilde{\Sigma}, \rho, \Sigma_\mu$ from the multivariate Normal distribution (26)
- Draw $\tilde{\mu}|W, \tilde{\Sigma}, \rho, \Sigma_\mu$ from the multivariate Normal distribution (27)
- set $\Sigma = d^{-1}\tilde{\Sigma}d^{-1}$, $\beta = (\beta_1', \beta_2')' = (\tilde{\beta}_1'/d_1, \tilde{\beta}_2'/d_2)'$ and $\mu = (\mu_1', \mu_2')' = (\tilde{\mu}_1'/d_1, \tilde{\mu}_2'/d_2)'$
- Draw $\rho|W, \beta, \mu, \Sigma$ from a random walk-MH step and set $B = (I_{2n} - \rho_1\hat{W}_1 - \rho_2\hat{W}_2)$
- Draw $\Sigma_\mu|\mu$ from the inverted Gamma distribution (29)

Repeat until convergence

with $m_r = (m_{1,r}, m_{2,r})$ and $\bar{\tilde{x}}_{j,r} = \frac{1}{n_r}\sum_{i=1}^{n_r}(\tilde{x}_{ij,r})$ represents the municipality characteristics averaged over each urban area r. Note that for each urban area r, we set the unobserved effect m_r to be a vector of dimension (2×1) in order to differentiate unobserved preferences between small ($j = 1$) versus large ($j = 2$) municipalities. The error terms $\mu_r = (\mu_{1,r}, \mu_{2,r})$, are normally distributed with zero mean and conditional variance Σ_μ of dimension (2×2). The intercept $\alpha = (\alpha_1, \alpha_2)$ is of dimension (2×1) and κ_j represents the $(k_j \times 1)$ vector of parameters. Correlated random effects can be rewritten in matrix form as $M = (\iota_{\bar{r}} \otimes diag(\alpha_1, \alpha_2)) + \bar{X}\kappa + \mu$ where $\bar{X} = (diag(\bar{\tilde{x}}_{1,1}, \bar{\tilde{x}}_{2,1})', \ldots, diag(\bar{\tilde{x}}_{1,\bar{r}}, \bar{\tilde{x}}_{2,\bar{r}})')'$ is the $(\bar{r} \times K)$ matrix of characteristics averaged over each urban area r, $\kappa = (\kappa_1', \kappa_2')'$ is the $(K \times 1)$ vector of parameters of interest, and $\mu = (\mu_1', \ldots, \mu_{\bar{r}}')'$.

Block sampling can be achieved assuming that the group specific effect M_0 is associated with the individual specific effect X. We set the following concatenation $\hat{X} = \begin{bmatrix} X & \Delta(\iota_{\bar{r}} \otimes I_2) & \Delta\bar{X} \end{bmatrix}$ with the parameters $\beta = (b', \alpha', \kappa')'$ being of dimension $\bar{k} = 5K + 2$.The spatial bivariate probit model (20) can be written as:

$$\begin{aligned} BY^\star &= \hat{X}\beta + \Delta\mu + V \\ \mu &\sim N(0_{2\bar{r}}, I_{\bar{r}} \otimes \Sigma_\mu) \\ V &\sim N(0_{2n}, I_n \otimes \Sigma) \end{aligned} \tag{23}$$

The key step of the Bayesian approach to estimating discrete choice models relies on the estimation of the vector of latent utilities $Y^\star$. For our spatial bivariate probit model, we wish to sample from the following truncated multivariate normal distribution:

$$Y^\star|\beta, \Sigma, \rho_1, \rho_2, Y, \sim N(B^{-1}(\hat{X}\beta + \Delta\mu), B^{-1}(I_n \otimes \Sigma)B^{-1\prime}),$$

subject to linear inequality restrictions, where the truncation bounds correspond to the interval $[(B^{-1}(\hat{X}\beta + \Delta\mu))_{ij}, +\infty)$ if $Y_{ij} = 1$ and $(-\infty, (B^{-1}(\hat{X}\beta + \Delta\mu))_{ij}]$ if $Y_{ij} = 0$. Because of partial observability, $Y_{ij} = 0$ if either or both vote no. The truncated multivariate distribution is approximated using the sampling method proposed by Geweke (1991). This sampler produces an $m-$step Gibbs iteration procedure based on a series of univariate truncated Gaussian distributions. For efficiently sampling from these truncated normal distributions, we implement the simulation procedure proposed by Chopin (2011).

To sample (β, μ, Σ) we adopt the parameter expansion framework defined in (19) using the transformation on the latent variables $W = Y^{\star}D$ where $D = I_n \otimes d$ and $d = diag(d_1, d_2)$ represents the expansion parameter. As previously explained, the working parameter $d = diag(d_1, d_2)$ is generated from an inverted Gamma distribution $d_i^2 = IG((J+1)/2, s^i/2)$ where the notation s^i is used to refer to the ith diagonal element of the unidentified matrix $\tilde{\Sigma}^{-1}$.

It is important to note that prior distributions for Σ, β and μ have to be the same under the observed data model and the expanded model. This is a condition required for the parameter expansion with marginal augmentation.

As previously explained, using the prior distribution for the correlation matrix Σ defined in (18), Barnard et al. (2000) shows that $\tilde{\Sigma} = d\Sigma d \sim IW(J+1, I_J)$ where $IW(\nu, I_J)$ denotes the inverse Wishart distribution defined as:

$$p(\tilde{\Sigma}|\nu) = |\tilde{\Sigma}|^{-\frac{(\nu+J+1)}{2}} \exp\left\{-\frac{1}{2}tr(\tilde{\Sigma}^{-1})\right\}. \tag{24}$$

To reduce correlation between draws, we implement a block sampling scheme and jointly draw $(\tilde{\Sigma}, \tilde{\beta}, \tilde{\mu})$ such that

$$p(\tilde{\beta}, \tilde{\mu}, \tilde{\Sigma}|W) = p(\tilde{\Sigma}|W)p(\tilde{\beta}, \tilde{\mu}|W, \tilde{\Sigma})$$

then compute the identified parameters using the transformation $\Sigma = d^{-1}\tilde{\Sigma}d^{-1}$, $\beta = (\beta_1', \beta_2')' = (\tilde{\beta}_1'/d_1, \tilde{\beta}_2'/d_2)'$ and $\mu = (\mu_1', \mu_2')' = (\tilde{\mu}_1'/d_1, \tilde{\mu}_2'/d_2)'$. To be able to marginalize the conditional distribution of $\tilde{\mu}$ over the parameter $\tilde{\beta}$ we choose a non-informative uniform prior for β.[10]

Setting $\hat{\beta} = (\hat{X}'V^{-1}\hat{X})^{-1}\hat{X}'V^{-1}BW$ with $V = D\left[(I_n \otimes \Sigma + \Delta\left(I_{\bar{r}} \otimes \Sigma_\mu\right)\Delta'\right]D$ and $\hat{\mu} = L\Delta'V^{-1}(BW - \hat{X}\hat{\beta})$ with $L = \Lambda \otimes (d\Sigma_\mu d)$ we obtain the following posterior distributions:

$$\tilde{\Sigma}|W, \rho \sim IW(n+J+1; \sum_i^n U_{i,r}'U_{i,r} + I_J) \tag{25}$$

$$\tilde{\beta}|\tilde{\Sigma}, \Sigma_\mu, W, \rho \sim N(\hat{\beta}, (\hat{X}'V^{-1}\hat{X})^{-1}), \tag{26}$$

$$\tilde{\mu}|\tilde{\Sigma}, \Sigma_\mu, W, \rho \sim N\left(\hat{\mu}, \left\{L - L\Delta'\left[V^{-1} - V^{-1}\hat{X}'(\hat{X}'V^{-1}\hat{X})^{-1}\hat{X}'V^{-1}\right]\Delta L\right\}\right), \tag{27}$$

where $U_{i,r} = W_{i,r} - \rho_1\sum_j^n \hat{w}_{1ij}W_{j,r} - \rho_2\sum_j^n \hat{w}_{2ij}W_{j,r} - \hat{X}_{i,r}\hat{\beta} - \hat{\mu}_r$.

Unlike the previous estimation step for (Σ,μ,β), the parameters ρ and Σ_μ are directly identified. No transformation is required. As noted in LeSage and Pace (2009), the conditional posterior distribution for ρ_j, $(j = 1, 2)$ is not reducible to a standard distribution. A Metropolis-Hastings (MH) step is used to draw from the posterior distribution (28) that relies on a random walk proposal with normally distributed increments.

[10]Unlike the method proposed by McCulloch and Rossi (1994), Imai and van Dyk (2005) underline the flexibility of the parameter expanded scheme in allowing flat priors.

Since the posterior distribution for ρ_j, $(j = 1, 2)$ depends upon its Beta prior $p(\rho_j) = Beta(b_0, b_0)$ centered around 0, we obtain:

$$p(\rho_j|\rho_{(-j)}, \Sigma, \Sigma_\mu) \propto |B|\tilde{\Sigma}^{-n/2} \exp\left(-\frac{1}{2}U'(I_n \otimes \tilde{\Sigma}^{-1})U\right) p(\rho_j), \tag{28}$$

where $\rho_{(-j)}$ is the vector ρ_j except the jth element and $B = (I_{2n} - \rho_1 \hat{W}_1 - \rho_2 \hat{W}_2)$. The parameter setting for this prior distribution places little weight on the end points of the stationary interval for ρ_j (see LeSage and Pace (2009) for further discussion).

Finally using an inverted Wishart prior distribution for $\Sigma_\mu \sim IW(\bar{r}, I_J)$, we can simulate the variance Σ_μ by reordering the identified correlated random effects in an $(\bar{r} \times 2)$ matrix $\mu_0 = (\mu_1, \mu_2)$ that yields the following posterior distribution:

$$\Sigma_\mu \sim IW\left(\bar{r} + J + 1, I_J + \mu_0'\mu_0\right). \tag{29}$$

Appendix E - Legal and Geo-Political Environment

In Ohio and Texas, three or more school districts can combine to form a new school district (Ohio Revised Code 3311.26, Texas Codes 13.151). Princeton City School District in Cincinnati has seven municipality members, for example.

In Texas, a consolidation may be initiated by either a resolution of the board of trustees of the school district or by a petition of registered voters (Texas Codes 13.152), or by the county board of education with the consent of the boards of trustees of the school districts (Texas Revised Statutes 49.12, Article 2740b Section 2). Consolidation may only take place between contiguous school districts. Votes occur simultaneously in each affected school district, and a majority vote by people in each school district is required for passage. A consolidated school district may be dissolved by the same procedure (Texas Codes 13.157(a)), and a new school district may be carved out of an existing school district as long as it is big enough and populous enough (Texas Codes 13.101-13.105). A school district can add a neighbor if the neighbor votes to abolish itself, too, but consolidation in this manner still requires consent by the board of trustees of the receiving school district (Texas Codes 13.205(b)).

In addition to consolidation of entire school districts in Texas, territory from a school district can be detached and annexed to an adjacent school district. The procedures are similar to those for consolidation. A school district may be declared dormant by a county commissioner's court, and the school district will be annexed to an adjacent school district upon approval of the receiving board's trustees. A county commissioner may also force a school district rated academically unacceptable for two years to be shut down and annexed by a neighboring district or districts, as long as the receiving districts can still meet their educational and financial obligations to existing students.[11] Citizens may also petition to vote that their school district abolish itself. A majority vote carries the motion, but only if all territory of the abolished school district is annexed to contiguous school districts.

Between the 1983 and 2013 school years in Texas, there have been 35 consolidations and 27 annexations. Three new school districts formed, including a school district formed

[11] Receiving school districts get additional state aid.

by the detachment of territory from two different school districts. One school district was annexed to three existing districts. One school district was declared dormant, five districts were closed or abolished, one school district was forcibly consolidated with a neighbor by the county commissioner, and three annexations occurred by order of the county commissioner (Texas Education Agency, 2014).

Ohio consolidation law is similar to that of Texas, except that only school boards can initiate a vote, and no vote by the people is required, just approval by each pre-consolidation school board. The procedure for disbanding or splitting up a school district is similar to the procedure for consolidation (Brasington, 2003b). In Ohio since 1985, three consolidated school districts formed, one school district disbanded for lack of students, two school districts formed from detachment of territory from existing school districts, and three split into their pre-merger components.

Appendix F - Interpretation of Model Estimates

We extend the method of LeSage and Pace (2011) that produces useful summary measures of spatial spillover impacts that arise in probit models with a spatial lag in response to changes in the explanatory variables. For each municipality j of the pair i, the probability of consolidation is non-linear in its own as well as in neighboring characteristics. In fact, when allowing for spatial dependence, changes in the explanatory variables for municipality j in pair i will influence the willingness to consolidate of municipalities belonging to neighboring pairs.

In a simple bivariate probit without endogenous and contextual effects that same change will only affect the probability of consolidation of both municipalities of the pair i. By setting $\rho_1 = \rho_2 = 0$ in (10), we can evaluate for a pair i in urban area r the effect on the probability of consolidation arising from a change in the k^th explanatory variable $x^k_{i,r} = (x^k_{i1,r}, x^k_{i2,r})$, where $x^k_{ij,r}$ is the k^{th} explanatory variable of municipality j in pair i. Since $\tilde{x}_{ij,r}$ is an $2K-$dimensional vector defined in (9), the contextual effects can be removed by constraining $\beta_{j,2k} = 0$, $k = 1, \ldots, K$.[12] Assuming that the change can come from the small or the large municipality of each pair, the marginal effect is given as:

$$\begin{aligned}\frac{\delta Pr(Y_{i1r}=1, Y_{i2r}=1|X)}{\delta x^k_{i,r}} &= \phi(x_{i1,r}\beta_1 + m_{1,r})\Phi\left(\frac{x_{i2,r}\beta_2 + m_{2,r} - \varrho(x_{i1,r}\beta_1 + m_{1,r})}{\sqrt{1-\varrho^2}}\right)\beta_{1,k} + \\ &\quad \phi(x_{i2,r}\beta_2 + m_{2,r})\Phi\left(\frac{x_{i1,r}\beta_1 + m_{1,r} - \varrho(x_{i2,r}\beta_2 + m_{2,r})}{\sqrt{1-\varrho^2}}\right)\beta_{2,k},\end{aligned}$$

where ϱ is the covariance on the correlation matrix Σ, $m_{j,r}$ are the correlated random effects defined in (22), X is the $(2n\times 4K)$ vector of explanatory variables, β_1 is the $2K-$dimensional vector of parameter estimates measuring the effects of small and large municipality characteristics on the smaller municipality of each pair and β_2 is the $2K-$dimensional vector measuring the same characteristics on the larger municipality with $K = k_1 + k_2$.

In this case, the marginal effect of a change of the $2n-$dimensional variable x^k on the

[12]The first $K = k_1 + k_2$ elements represents the explanatory variables for the small and large municipalities and the second set of K elements represents the contextual effects.

probability of consolidation of all pairs is given by:

$$\frac{\delta Pr(Y_1 = 1, Y_2 = 1|X)}{\delta x^{k\prime}} = \sum_{j=1}^{2} g_j I_n \beta_{j,k}, \tag{30}$$

where $g_j = \phi(\eta_j)\Phi\left(\frac{\eta_{(-j)} - \varrho\eta_j}{\sqrt{1-\varrho^2}}\right)$, with $\eta_j = x_j\beta_j + m_j$ and $-j$ represents the other municipality (i.e. $\eta_{(-j)} = \eta_2$ if $j = 1$).

For our general model with contextual effects, we simply note that $(I_n\beta_{j,k})$ has to be replaced in (30) by $(I_n\beta_{j,k} + (\hat{W}_1 + \hat{W}_2)_j\beta_{j,2k})$, where $(\hat{W}_1 + \hat{W}_2)_j$ represents the $(n \times n)$ spatial weight matrix for $j = (1, 2)$.

LeSage et al. (2011) propose a method to analyse the direct and indirect effects in the case of a spatial autoregressive probit model. Since the small municipalities have a different spatial weight matrix than the larger municipalities, we rely on the work of Elhorst et al. (2012) who show that it is not possible to dissociate the direct and indirect effects between these two spatial weight matrices.

$$B^{-1}_{(\rho_1,\rho_2)} = (I_{2n} - \rho_1\hat{W}_1 - \rho_2\hat{W}_2)^{-1} = I_{2n} + \sum_{q=1}^{\infty}(\rho_1\hat{W}_1 + \rho_2\hat{W}_2)^q.$$

However we can still separate the direct effects based on the diagonal elements of $B^{-1}_{(\rho_1,\rho_2)}$ from the indirect effects calculated from the off-diagonal terms of that same matrix $B^{-1}_{(\rho_1,\rho_2)}$.

Unlike the simple probit model, we observe several non-linear transformations for the partial derivatives of the bivariate probit because of the presence of the cumulative normal distribution $\Phi(.)$ and its density function $\phi(.)$. These non-linearities are reinforced in the spatial case by the presence of the inverse of the spatial weight matrix. In fact, by analyzing the reduced form of (13), the expected value of the latent variables is equal to $\hat{Y}^\star = B^{-1}_{(\hat{\rho}_1,\hat{\rho}_2)}X\hat{\beta} + B^{-1}_{(\hat{\rho}_1,\hat{\rho}_2)}\hat{M}$.

Another contrast involves a change in location for observations having different explanatory variables. It seems impossible to disentangle the impact of a change in location from a change in characteristics. LeSage et al. (2011) argue this can be overcome by the fact that the same variables in neighboring locations might have similar values.

For computing marginal effects, we typically use the expression at the sample mean of the data but other pertinent values could be analyzed. In the univariate case, LeSage et al. (2011) propose different scalar summary measures that are based on the following $(n \times n)$ partial derivatives:

$$\begin{aligned} \frac{\delta Pr(Y = 1)}{\delta x^{k'}} &= D\left[\phi(\eta)\right] S_{(\rho)}\beta_k \\ &= \left\{D\left[\phi(\eta)\right] + \rho D\left[\phi(\eta)\right] W + \rho^2 D\left[\phi(\eta)\right] W^2 + \ldots\right\}\beta_k \end{aligned}$$

where $S_{(\rho)} = (I_n - \rho W)^{-1} = I_n + \rho W + \ldots + \rho^n W^n$, $\eta = S_{(\rho)}X\beta$ and $d(.)$ is an $n-$dimensional vector of a diagonal matrix $D(.)$.

The $(n \times 1)$ vector of cumulative total effects is equivalent to:

$$\begin{aligned}
\frac{\delta Pr(Y=1)}{\delta x^{k'}}\iota_n &= D\left[\phi(\eta)\right] S_{(\rho)}\iota_n\beta_k \\
&= D\left[\phi(\eta)\right]\iota_n(1-\rho)^{-1}\beta_k \\
&= d\left[\phi(\eta)\right](1-\rho)^{-1}\beta_k.
\end{aligned}$$

The average total effect can be summarized as $n^{-1}d\left[\phi(\eta)\right]'\iota_n(1-\rho)^{-1}\beta_k$.
And using the diagonal elements, it is easy to obtain the average direct effect:

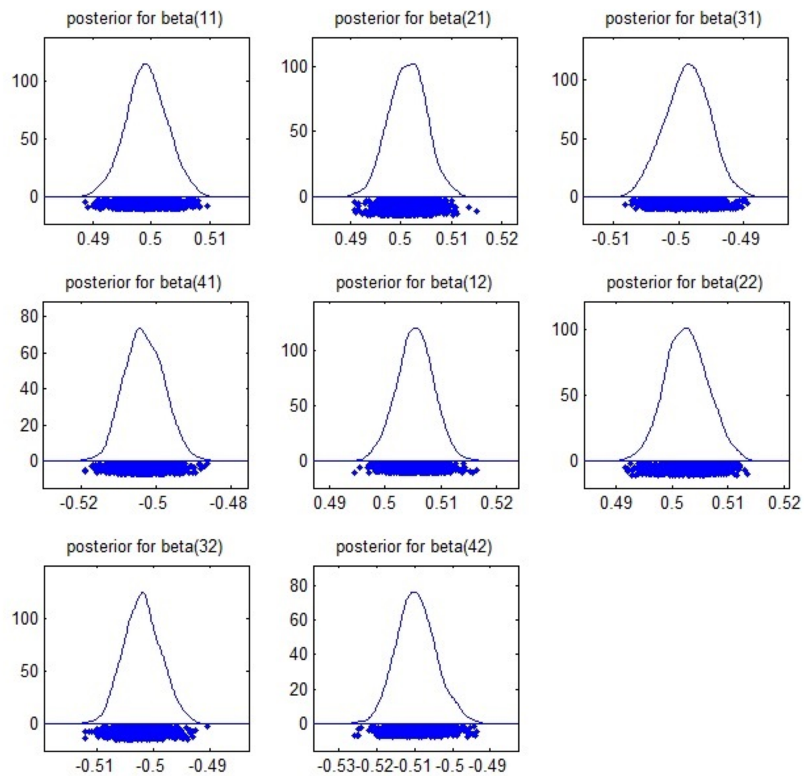
$$\begin{aligned}
n^{-1}tr\left\{\frac{\delta Pr(Y=1)}{\delta x^{k'}}\right\} &= \{tr(D\left[\phi(\eta)\right]) + \rho tr(D\left[\phi(\eta)\right])W + \\
&\quad \rho^2 tr(D\left[\phi(\eta)\right])W^2 + \ldots\}\,\beta_k/n.
\end{aligned}$$

We can extend these two summary measures to the bivariate case by replacing η with the expected value of the latent variables $\hat{Y}^{\star} = B^{-1}_{(\hat{\rho}_1,\hat{\rho}_2)}X\hat{\beta} + B^{-1}_{(\hat{\rho}_1,\hat{\rho}_2)}\hat{M}$ in (30) and obtain:

$$\begin{aligned}
(2n)^{-1}\iota_{2n}\frac{\delta Pr(Y_1=1,Y_2=1|X)}{\delta x_r^{k\prime}}\iota_{2n}' &= (2n)^{-1}\iota_{2n}D\left[g_1\beta_{1k}+g_2\beta_{2k}\right]S_{(\rho_1,\rho_2)}\iota_{2n} \\
&= (2n)^{-1}\iota_{2n}D\left[g_1\beta_{1k}+g_2\beta_{2k}\right]\iota_{2n}(1-\rho_1-\rho_2)^{-1} \\
&= (2n)^{-1}d\left[g_1\beta_{1k}+g_2\beta_{2k}\right]'\iota_{2n}(1-\rho_1-\rho_2)^{-1} \\
(2n)^{-1}tr\left\{\frac{\delta Pr(Y_1=1,Y_2=1|X)}{\delta x_r^{k\prime}}\right\} &= \{tr(D\left[g_1\beta_{1k}+g_2\beta_{2k}\right]) + (\rho_1+\rho_2)tr(D\left[g_1\beta_{1k}+g_2\beta_{2k}\right])W + \\
&\quad (\rho_1+\rho_2)^2 tr(D\left[g_1\beta_{1k}+g_2\beta_{2k}\right])W^2 + \ldots\}\,\beta_k/(2n).
\end{aligned}$$

NOT FOR PUBLICATION

Figure 2: posterior densities for β - First case



Tables

Figure 3: posterior draws for ρ - First case

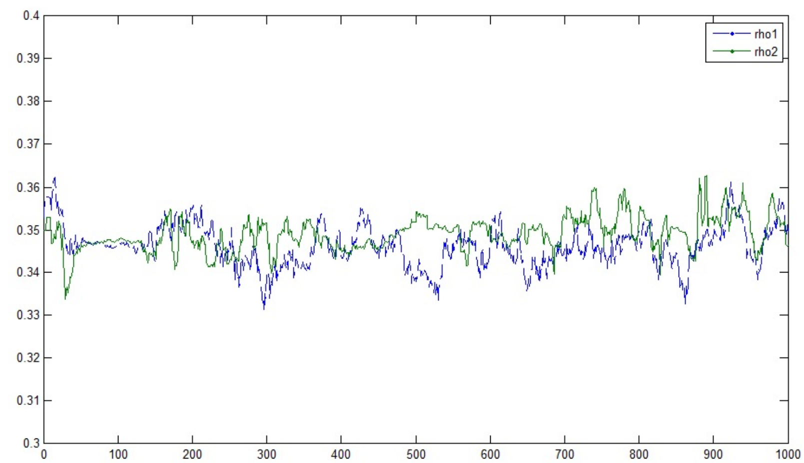


Figure 4: posterior draws for σ_{12} - First case

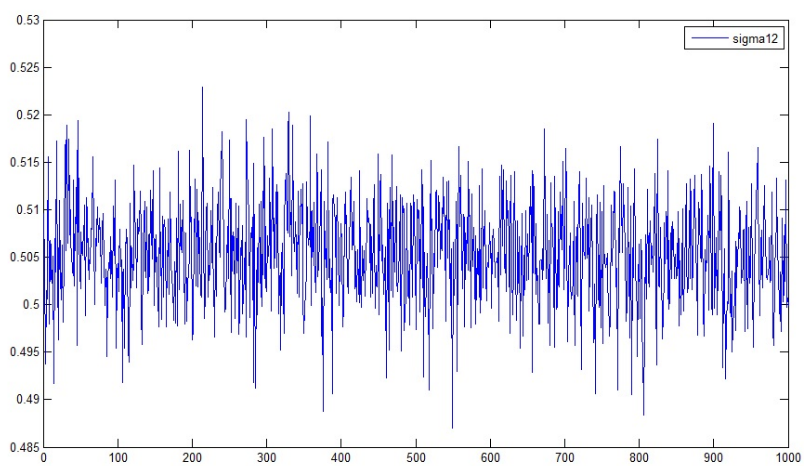


Table 2: Monte Carlo simulations

Parameter	True value	Mean	NSE	True value	Mean	NSE
	Small Urban Areas					
	Small municipalities ($j = 1$)			Large municipalities ($j = 2$)		
ρ_j	0.35	0.3590	0.0041	0.35	0.3524	0.0032
b_{1j}	0.5	0.4857	0.0090	0.5	0.4754	0.0091
b_{2j}	-0.5	-0.4736	0.0241	-0.5	-0.48456	0.0223
b_{3j}	0.5	0.4962	0.0078	0.5	0.4864	0.0075
b_{4j}	-0.5	-0.4762	0.0255	-0.5	-0.4895	0.0268
α_j	1	1.0105	0.0085	1	1.0124	0.0088
κ_j	2	2.9806	0.0127	2	2.0306	0.0125
σ_{12}	0.5	0.4720	0.0122			
	Large Urban Areas					
	Small municipalities ($j = 1$)			Large municipalities ($j = 2$)		
ρ_j	0.35	0.3519	0.0042	0.35	0.3521	0.0034
b_{1j}	0.5	0.4845	0.0084	0.5	0.4871	0.0075
b_{2j}	-0.5	-0.4815	0.0212	-0.5	-0.4876	0.0245
b_{3j}	0.5	0.4892	0.0084	0.5	0.4816	0.0075
b_{4j}	-0.5	-0.4782	0.0245	-0.5	-0.4812	0.0254
α_j	1	1.0120	0.0084	1	1.0125	0.0082
κ_j	2	0.9786	0.0115	2	1.97236	0.0119
σ_{12}	0.5	0.4880	0.0123			

Note: NSE: Numerical Standard Error (see Geweke, 1991)

Table 3: Variable Definition and Sources

Variable	Definition	Source[a]
pup	number of children enrolled in K-12 (public or private) in municipality (in thousands)	(1)
renterocc	percent of occupied housing units in municipality that are occupied by renters rather than owners	(1)
valpup_all	property value per pupil; for Ohio it's class 1 (agricultural and residential) and class 2 (industrial, commercial, mineral, and railroad) real market value of property in dollars divided by number of children enrolled in K-12 (public or private) in municipality; for Texas it's taxable property value in dollars divided by number of children enrolled in K-12 (public or private) in municipality (in hundreds of thousands)	(1),(2),(3)
pctbadenglish	percent of persons in municipality aged 5-17 years who speak English less than "very well"; p. B-29 of Technical Documentation: Census 2000 Summary File 3, U.S. Census Bureau, 2002	(1)
pcthisp	percent of municipality population that is Hispanic/Latino/Spanish, and may be of any race	(1)
loweduc	percentage of persons 25 years or older in municipality whose highest educational attainment is no more than a high school degree or equivalent	(1)
priv	percent of students in municipality enrolled in grades 1-12 who attend private schools	(1)
avinc	average income of households in municipality in tens of thousands of dollars, where household is householder and other individuals 15+ living in household	(1)
albundy	percent of households in municipality that are married with own child 0-17 years	(1)
leikrace	index created by Robert K. Leik (1966) measure of ethnic heterogeneity in municipality	(1)

[a]: (1) GeoLytics CensusCD 2000 Long Form Release 2.0, East Brunswick, NJ: 2002,
(2) "2000 Real Property Abstracts by Taxing District" from Ohio Department of Taxation,
(3) Texas Comptroller of Public Accounts, Carole Keeton Strayhorn, Property Tax Division "Annual property tax report tax year 2000 appendix D: City local self report data - 2000"

Table 4: Variable Mean and Standard Deviation

Variable	Ohio Small Mean	Ohio Small St.Dev.	Ohio Large Mean	Ohio Large St.Dev.	Texas Small Mean	Texas Small St.Dev.	Texas Large Mean	Texas Large St.Dev.
pup	1.586	1.839	11.130	21.947	2.786	6.600	23.154	48.002
renterocc	0.234	0.151	0.274	0.152	0.257	0.199	0.305	0.170
valpup_all	3.707	3.249	2.778	1.793	8.470	2.364	3.020	5.930
pctbadenglish	1.366	1.670	1.7093	1.4004	5.2521	6.6187	7.5830	6.6912
pcthisp	1.287	2.151	1.4937	2.0743	16.5981	17.3016	21.8775	17.6517
priv	16.688	14.089	15.8018	9.1368	12.2613	12.5846	8.9517	6.9336
leikrace	0.085	0.075	0.114	0.091	0.348	0.215	0.458	0.243
albundy	24.834	7.970	24.478	7.881	30.788	11.145	31.392	8.914
loweduc	46.028	18.186	45.374	15.680	39.115	19.400	40.936	17.085
avinc	6.470	2.509	6.132	2.156	7.161	3.020	6.537	2.386

Table 5: Estimation Results - Model 1

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Ohio - Smaller municipalities				Ohio - Larger municipalities			
pupS	-0.040	0.040	-0.104	0.025	-0.023	0.046	-0.095	0.058
pupL	-0.012	0.005	-0.021	-0.003	-0.012	0.006	-0.021	-0.003
valpup_allS	0.004	0.048	-0.072	0.084	-0.055	0.042	-0.125	0.016
valpup_allL	-0.073	0.054	-0.166	0.016	-0.136	0.058	-0.236	-0.041
albundyS	0.020	0.017	-0.009	0.048	-0.008	0.016	-0.034	0.018
albundyL	0.013	0.021	-0.021	0.046	0.003	0.022	-0.036	0.041
pcthispS	0.027	0.054	-0.057	0.119	-0.052	0.048	-0.129	0.022
pcthispL	-0.004	0.055	-0.094	0.091	-0.020	0.065	-0.118	0.091
renteroccS	0.006	0.008	-0.008	0.020	-0.004	0.008	-0.017	0.009
renteroccL	0.013	0.010	-0.005	0.029	0.004	0.013	-0.017	0.025
privS	-0.010	0.009	-0.024	0.006	0.010	0.009	-0.004	0.023
privL	0.005	0.011	-0.012	0.025	0.023	0.012	0.004	0.045
*W * pupS*	0.001	0.038	-0.065	0.068	-0.021	0.040	-0.086	0.050
*W * pupL*	0.006	0.005	-0.003	0.015	0.010	0.005	0.002	0.019
*W * valpup_allS*	0.023	0.033	-0.028	0.081	0.071	0.034	0.021	0.127
*W * valpup_allL*	0.081	0.040	0.013	0.148	0.030	0.042	-0.037	0.102
*W * albundyS*	-0.005	0.014	-0.029	0.017	0.012	0.014	-0.009	0.038
*W * albundyL*	-0.009	0.016	-0.036	0.016	0.003	0.015	-0.022	0.029
*W * pcthispS*	-0.026	0.047	-0.109	0.047	0.046	0.042	-0.023	0.110
*W * pcthispL*	0.066	0.043	-0.001	0.142	0.028	0.042	-0.051	0.097
*W * renteroccS*	0.001	0.007	-0.012	0.012	0.009	0.008	-0.002	0.023
*W * renteroccL*	-0.007	0.008	-0.021	0.005	-0.001	0.008	-0.013	0.013
*W * privS*	-0.005	0.007	-0.018	0.006	-0.010	0.007	-0.022	0.001
*W * privL*	-0.005	0.011	-0.022	0.012	-0.017	0.010	-0.034	0.001
	Texas - Smaller municipalities				Texas - Larger municipalities			
pupS	0.027	0.017	-0.002	0.052	0.008	0.013	-0.012	0.029
pupL	-0.006	0.003	-0.009	-0.002	0.015	0.001	0.012	0.018
valpup_allS	0.006	0.007	-0.006	0.018	0.002	0.006	-0.008	0.012
valpup_allL	0.034	0.028	-0.003	0.088	0.004	0.015	-0.019	0.030
albundyS	-0.033	0.015	-0.057	-0.008	-0.025	0.013	-0.048	-0.004
albundyL	-0.032	0.019	-0.063	0.002	-0.001	0.019	-0.034	0.032
pcthispS	-0.016	0.010	-0.034	0.002	0.003	0.009	-0.014	0.018
pcthispL	0.001	0.011	-0.016	0.019	0.011	0.014	-0.016	0.032
renteroccS	-0.013	0.008	-0.028	0.003	-0.006	0.007	-0.019	0.005
renteroccL	0.001	0.009	-0.013	0.016	0.001	0.011	-0.016	0.018
privS	0.005	0.014	-0.017	0.025	0.010	0.011	-0.008	0.029
privL	-0.015	0.015	-0.041	0.010	0.028	0.021	-0.006	0.063
*W * pupS*	-0.041	0.015	-0.066	-0.016	-0.031	0.013	-0.052	-0.010
*W * pupL*	-0.002	0.002	-0.006	0.002	-0.004	0.002	-0.008	0.001
*W * valpup_allS*	-0.002	0.005	-0.012	0.007	-0.003	0.006	-0.012	0.007
*W * valpup_allL*	-0.002	0.026	-0.047	0.043	-0.011	0.020	-0.043	0.023
*W * albundyS*	0.041	0.013	0.020	0.062	0.022	0.011	0.001	0.044
*W * albundyL*	0.025	0.009	0.009	0.040	0.005	0.014	-0.017	0.033
*W * pcthispS*	0.002	0.010	-0.014	0.019	-0.006	0.009	-0.021	0.009
*W * pcthispL*	0.018	0.008	0.006	0.033	0.011	0.009	-0.003	0.025
*W * renteroccS*	0.011	0.007	-0.001	0.022	0.010	0.007	-0.001	0.022
*W * renteroccL*	0.002	0.009	-0.012	0.017	-0.007	0.009	-0.020	0.008
*W * privS*	-0.007	0.012	-0.028	0.011	-0.011	0.011	-0.029	0.006
*W * privL*	0.015	0.014	-0.009	0.038	-0.018	0.014	-0.041	0.006

	Mean	s.d.	0.05	0.95
ρ_1	0.391	0.023	0.347	0.432
ρ_2	0.213	0.024	0.172	0.254
Σ_{12}	0.350	0.074	0.195	0.494

Table 6: Estimation Results - Direct effects - Model 1

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Ohio - Smaller municipalities				Ohio - Larger municipalities			
pupS	0.007	0.004	-0.001	0.016	-0.009	0.015	-0.038	0.021
pupL	-0.002	0.001	-0.003	-0.001	0.001	0.001	-0.002	0.001
valpup_allS	-0.009	0.002	-0.013	-0.006	0.003	0.005	-0.008	0.013
valpup_allL	0.003	0.003	-0.003	0.009	-0.013	0.004	-0.022	-0.005
albundyS	0.004	0.003	-0.002	0.009	0.004	0.001	0.002	0.007
albundyL	-0.001	0.002	-0.005	0.004	-0.005	0.002	-0.009	-0.001
pcthispS	0.003	0.005	-0.006	0.013	-0.003	0.001	-0.005	-0.001
pcthispL	-0.011	0.004	-0.018	-0.003	-0.004	0.008	-0.021	0.012
renteroccS	0.001	0.001	-0.001	0.003	-0.002	0.001	-0.004	-0.001
renteroccL	0.001	0.001	-0.001	0.003	-0.001	0.001	-0.003	0.001
privS	-0.003	0.001	-0.005	-0.001	0.001	0.001	-0.002	0.002
privL	0.002	0.001	-0.001	0.005	0.002	0.002	-0.001	0.005
	Texas - Smaller municipalities				Texas - Larger municipalities			
pupS	-0.002	0.002	-0.005	0.002	-0.010	0.006	-0.021	0.001
pupL	0.001	0.001	-0.001	0.001	-0.001	0.001	-0.003	0.001
valpup_allS	0.001	0.001	-0.001	0.003	0.002	0.001	0.001	0.003
valpup_allL	-0.003	0.003	-0.009	0.002	-0.001	0.002	-0.005	0.003
albundyS	-0.004	0.002	-0.008	-0.001	-0.003	0.003	-0.009	0.003
albundyL	0.004	0.002	0.001	0.008	-0.003	0.003	-0.009	0.004
pcthispS	-0.001	0.001	-0.003	0.001	-0.002	0.001	-0.003	-0.001
pcthispL	0.001	0.001	-0.002	0.002	-0.003	0.002	-0.007	0.001
renteroccS	0.002	0.001	0.001	0.003	-0.002	0.001	-0.003	-0.001
renteroccL	0.003	0.001	0.001	0.005	0.001	0.002	-0.003	0.003
privS	-0.002	0.001	-0.004	-0.001	0.005	0.001	0.002	0.007
privL	0.002	0.002	-0.002	0.007	0.009	0.004	0.001	0.017

Table 7: Estimation Results - Indirect effects - Model 1

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Ohio - Smaller municipalities				Ohio - Larger municipalities			
pupS	0.016	0.008	-0.001	0.033	0.017	0.023	-0.029	0.062
pupL	0.001	0.001	-0.001	0.003	0.007	0.001	0.005	0.009
valpup_allS	0.008	0.002	0.004	0.013	0.018	0.005	0.008	0.028
valpup_allL	0.016	0.006	0.004	0.029	-0.035	0.017	-0.070	-0.001
albundyS	0.004	0.003	-0.002	0.010	0.001	0.003	-0.006	0.007
albundyL	-0.004	0.004	-0.011	0.003	0.004	0.005	-0.005	0.014
pcthispS	0.013	0.010	-0.007	0.033	0.012	0.013	-0.013	0.038
pcthispL	-0.004	0.014	-0.032	0.023	0.016	0.011	-0.006	0.038
renteroccS	-0.003	0.001	-0.004	-0.002	0.003	0.002	-0.001	0.007
renteroccL	0.001	0.002	-0.003	0.004	-0.001	0.002	-0.006	0.003
privS	-0.002	0.001	-0.004	0.001	-0.001	0.002	-0.004	0.002
privL	-0.001	0.003	-0.007	0.005	0.003	0.003	-0.002	0.009
	Texas - Smaller municipalities				Texas - Larger municipalities			
pupS	-0.008	0.005	-0.017	0.001	-0.019	0.013	-0.046	0.007
pupL	0.001	0.001	-0.001	0.002	0.001	0.001	-0.001	0.002
valpup_allS	0.002	0.002	-0.002	0.005	0.003	0.001	0.001	0.005
valpup_allL	-0.017	0.010	-0.037	0.004	0.001	0.006	-0.012	0.013
albundyS	0.006	0.002	0.001	0.010	0.002	0.003	-0.004	0.008
albundyL	0.001	0.003	-0.005	0.007	-0.011	0.008	-0.027	0.004
pcthispS	-0.001	0.002	-0.005	0.002	-0.002	0.001	-0.005	0.001
pcthispL	0.001	0.002	-0.003	0.004	-0.003	0.003	-0.009	0.002
renteroccS	-0.001	0.001	-0.003	0.002	-0.001	0.001	-0.004	0.002
renteroccL	0.005	0.003	-0.001	0.011	0.001	0.003	-0.007	0.006
privS	-0.004	0.002	-0.009	0.001	-0.005	0.004	-0.013	0.002
privL	0.008	0.004	0.001	0.015	0.012	0.006	0.001	0.023

Table 8: Estimation Results - Total effects - Model 1

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Ohio - Smaller municipalities				Ohio - Larger municipalities			
pupS	0.024	0.013	-0.002	0.049	0.008	0.038	-0.068	0.083
pupL	0.001	0.002	-0.003	0.002	0.006	0.002	0.003	0.010
valpup_allS	-0.001	0.004	-0.009	0.007	0.021	0.010	0.001	0.041
valpup_allL	0.020	0.010	0.001	0.038	-0.049	0.022	-0.092	-0.006
albundyS	0.008	0.007	-0.004	0.019	0.005	0.005	-0.004	0.014
albundyL	-0.004	0.007	-0.017	0.007	-0.001	0.007	-0.014	0.012
pcthispS	0.017	0.015	-0.013	0.046	0.010	0.014	-0.018	0.037
pcthispL	-0.015	0.018	-0.050	0.019	0.012	0.019	-0.027	0.050
renteroccS	-0.002	0.003	-0.006	0.001	0.001	0.003	-0.005	0.007
renteroccL	0.002	0.003	-0.005	0.007	-0.002	0.004	-0.009	0.004
privS	-0.004	0.002	-0.009	0.001	-0.001	0.003	-0.007	0.004
privL	0.001	0.005	-0.008	0.010	0.005	0.004	-0.004	0.014
	Texas - Smaller municipalities				Texas - Larger municipalities			
pupS	-0.010	0.006	-0.023	0.003	-0.030	0.019	-0.067	0.007
pupL	0.001	0.001	-0.002	0.002	-0.001	0.001	-0.004	0.002
valpup_allS	0.003	0.002	-0.002	0.007	0.004	0.002	0.001	0.008
valpup_allL	-0.020	0.013	-0.046	0.005	0.001	0.008	-0.017	0.016
albundyS	0.002	0.004	-0.007	0.010	-0.001	0.006	-0.013	0.010
albundyL	0.004	0.006	-0.008	0.015	-0.014	0.011	-0.036	0.008
pcthispS	-0.002	0.003	-0.007	0.004	-0.004	0.002	-0.008	0.001
pcthispL	0.001	0.003	-0.006	0.006	-0.006	0.004	-0.015	0.002
renteroccS	0.001	0.002	-0.003	0.004	-0.003	0.002	-0.006	0.001
renteroccL	0.008	0.004	-0.001	0.016	0.001	0.005	-0.009	0.009
privS	-0.006	0.003	-0.013	0.001	-0.001	0.005	-0.011	0.009
privL	0.011	0.006	-0.001	0.022	0.021	0.010	0.001	0.040

Table 9: Estimation Results - Correlated Random Effects - Model 1

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Smaller municipalities				Larger municipalities			
Cleveland	0.267	0.747	-0.963	1.531	-0.226	0.785	-1.602	0.999
Cincinnati	0.536	0.840	-0.864	1.910	-0.355	0.901	-1.883	1.043
Colombus	0.368	0.814	-0.903	1.637	-0.402	0.879	-1.900	1.003
Dayton	0.422	0.765	-0.840	1.625	-0.172	0.807	-1.560	1.133
Toledo	-1.002	0.456	-1.914	-0.091	-1.215	0.688	-2.628	0.007
San Antonio	-1.785	0.802	-3.384	-0.176	-0.899	0.536	-1.967	0.169
Houston	-1.235	0.579	-2.398	-0.081	1.075	0.495	0.025	2.110
Dallas	-1.466	0.716	-2.894	-0.031	0.995	0.485	0.045	1.919

Table 10: Estimation Results - Model 2

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Ohio - Smaller municipalities				Ohio - Larger municipalities			
pupS	-0.037	0.022	-0.081	0.007	-0.035	0.021	-0.077	0.006
pupL	-0.014	0.002	-0.018	-0.011	0.116	0.029	0.058	0.174
avincS	-0.097	0.066	-0.035	0.229	-0.071	0.058	-0.187	0.045
avincL	0.025	0.065	-0.104	0.155	0.139	0.033	0.074	0.204
pctbadenglishS	-0.026	0.055	-0.136	0.083	-0.102	0.035	-0.173	-0.032
pctbadenglishL	-0.118	0.044	-0.205	-0.031	0.119	0.064	-0.009	0.246
loweducS	0.017	0.011	-0.004	0.038	-0.012	0.009	-0.031	0.006
loweducL	-0.011	0.011	-0.010	0.032	0.023	0.007	0.010	0.037
leikraceS	-0.312	1.296	-2.280	2.903	0.479	1.364	-2.249	3.207
leikraceL	-2.205	0.934	-4.072	-0.338	-1.789	1.385	-4.560	0.982
$W * pupS$	-0.024	0.009	-0.042	-0.006	0.019	0.036	-0.053	0.092
$W * pupL$	0.002	0.005	-0.008	0.011	-0.004	0.005	-0.014	0.005
$W * avincS$	0.038	0.056	-0.075	0.150	0.107	0.025	0.056	0.157
$W * avincL$	-0.075	0.053	-0.181	0.031	-0.036	0.052	-0.140	0.067
$W * pctbadenglishS$	0.052	0.044	-0.035	0.140	-0.088	0.034	-0.156	-0.020
$W * pctbadenglishL$	0.011	0.059	-0.106	0.128	0.002	0.054	-0.106	0.109
$W * loweducS$	0.008	0.009	-0.011	0.026	0.011	0.009	-0.006	0.029
$W * loweducL$	-0.020	0.008	-0.036	-0.004	-0.009	0.007	-0.023	0.006
$W * leikraceS$	-1.111	0.608	-2.327	0.105	1.270	1.035	-0.799	3.339
$W * leikraceL$	1.012	0.709	-0.406	2.430	0.982	1.067	-1.151	3.116
	Texas - Smaller municipalities				Texas - Larger municipalities			
pupS	0.044	0.026	-0.007	0.096	0.026	0.014	-0.002	0.053
pupL	0.001	0.003	-0.006	0.006	0.041	0.018	0.005	0.077
avincS	0.143	0.121	-0.098	0.385	0.029	0.055	-0.081	0.138
avincL	0.231	0.087	0.058	0.405	0.175	0.072	0.031	0.318
pctbadenglishS	0.007	0.027	-0.047	0.061	-0.024	0.025	-0.075	0.027
pctbadenglishL	0.018	0.029	-0.041	0.076	-0.027	0.028	-0.083	0.028
loweducS	0.012	0.011	-0.010	0.035	0.025	0.010	0.005	0.044
loweducL	-0.012	0.011	-0.011	0.034	0.031	0.013	0.005	0.058
leikraceS	0.548	0.656	-0.764	1.860	-0.606	0.560	-1.726	0.514
leikraceL	-0.439	0.657	-1.753	0.875	0.227	0.642	-1.056	1.510
$W * pupS$	-0.030	0.012	-0.054	-0.006	-0.018	0.014	-0.045	0.009
$W * pupL$	-0.001	0.003	-0.007	0.006	0.001	0.003	-0.005	0.007
$W * avincS$	-0.121	0.060	-0.241	-0.001	-0.032	0.054	-0.141	0.076
$W * avincL$	-0.036	0.057	-0.149	0.077	-0.014	0.062	-0.138	0.110
$W * pctbadenglishS$	-0.008	0.022	-0.051	0.036	0.016	0.020	-0.024	0.056
$W * pctbadenglishL$	-0.002	0.026	-0.055	0.050	-0.006	0.025	-0.056	0.042
$W * loweducS$	0.002	0.009	-0.016	0.020	-0.019	0.008	-0.036	-0.003
$W * loweducL$	-0.008	0.009	-0.026	0.010	-0.004	0.009	-0.022	0.014
$W * leikraceS$	-0.391	0.556	-1.504	0.722	0.804	0.472	-0.141	1.748
$W * leikraceL$	0.807	0.543	-0.279	1.894	-0.235	0.488	-1.212	0.741

	Mean	s.d.	0.05	0.95
ρ_1	0.321	0.024	0.285	0.361
ρ_2	0.201	0.028	0.256	0.151
Σ_{12}	0.327	0.074	0.175	0.464

Table 11: Estimation Results - Direct effects - Model 2

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Ohio - Smaller municipalities				Ohio - Larger municipalities			
pupS	0.015	0.012	-0.009	0.040	-0.003	0.011	-0.025	0.019
pupL	-0.003	0.001	-0.005	-0.001	0.003	0.001	0.001	0.004
avincS	0.052	0.027	-0.002	0.106	-0.008	0.014	-0.036	0.021
avincL	-0.019	0.017	-0.054	0.015	0.002	0.015	-0.027	0.032
pctbadenglishS	0.002	0.011	-0.020	0.025	-0.060	0.025	-0.110	-0.011
pctbadenglishL	-0.027	0.013	-0.052	-0.001	0.056	0.026	0.005	0.107
loweducS	0.013	0.004	0.006	0.020	0.001	0.002	-0.004	0.004
loweducL	-0.004	0.003	-0.011	0.003	0.001	0.003	-0.005	0.006
leikraceS	0.091	0.237	-0.383	0.566	0.803	0.496	-0.189	1.794
leikraceL	-0.342	0.203	-0.748	0.064	-0.648	0.330	-1.307	0.011
	Texas - Smaller municipalities				Texas - Larger municipalities			
pupS	0.009	0.002	0.004	0.013	-0.001	0.005	-0.011	0.008
pupL	0.001	0.001	-0.001	0.002	0.001	0.001	0.001	0.002
avincS	0.043	0.025	-0.007	0.093	-0.015	0.017	-0.050	0.019
avincL	0.047	0.020	0.006	0.087	0.076	0.040	-0.004	0.157
pctbadenglishS	-0.007	0.006	-0.019	0.005	-0.002	0.005	-0.011	0.007
pctbadenglishL	0.006	0.006	-0.007	0.018	-0.007	0.007	-0.022	0.008
loweducS	0.009	0.004	0.001	0.018	-0.002	0.004	-0.009	0.005
loweducL	-0.002	0.004	-0.011	0.006	0.013	0.006	0.002	0.025
leikraceS	0.011	0.117	-0.223	0.246	0.183	0.185	-0.187	0.553
leikraceL	0.268	0.218	-0.167	0.703	-0.112	0.169	-0.450	0.226

Table 12: Estimation Results - Indirect effects - Model 2

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Ohio - Smaller municipalities				Ohio - Larger municipalities			
pupS	0.014	0.012	-0.010	0.038	0.006	0.015	-0.024	0.035
pupL	0.001	0.001	-0.001	0.004	0.002	0.001	0.001	0.004
avincS	0.062	0.019	0.024	0.101	-0.008	0.016	-0.041	0.024
avincL	-0.016	0.017	-0.051	0.018	0.002	0.022	-0.041	0.045
pctbadenglishS	0.006	0.013	-0.021	0.033	-0.055	0.016	-0.087	-0.022
pctbadenglishL	-0.038	0.009	-0.056	-0.020	0.042	0.024	-0.006	0.091
loweducS	0.007	0.005	-0.003	0.016	0.001	0.003	-0.006	0.006
loweducL	-0.003	0.003	-0.010	0.004	0.001	0.004	-0.006	0.009
leikraceS	0.085	0.301	-0.518	0.688	0.767	0.454	-0.141	1.675
leikraceL	-0.644	0.220	-1.084	-0.204	-0.815	0.307	-1.430	-0.201
	Texas - Smaller municipalities				Texas - Larger municipalities			
pupS	0.009	0.007	-0.004	0.022	0.003	0.005	-0.008	0.013
pupL	0.001	0.001	-0.002	0.001	0.001	0.001	-0.001	0.002
avincS	0.052	0.030	-0.008	0.111	0.003	0.014	-0.026	0.031
avincL	0.075	0.030	0.015	0.135	0.061	0.028	0.006	0.117
pctbadenglishS	-0.002	0.007	-0.016	0.011	-0.001	0.007	-0.014	0.013
pctbadenglishL	0.010	0.009	-0.008	0.027	-0.010	0.010	-0.029	0.009
loweducS	0.006	0.003	0.001	0.013	0.003	0.003	-0.004	0.010
loweducL	0.002	0.004	-0.006	0.008	0.012	0.005	0.002	0.021
leikraceS	0.070	0.168	-0.266	0.405	0.052	0.197	-0.341	0.445
leikraceL	0.163	0.174	-0.184	0.511	0.024	0.173	-0.322	0.369

Table 13: Estimation Results - Total effects - Model 2

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Ohio - Smaller municipalities				Ohio - Larger municipalities			
pupS	0.029	0.024	-0.020	0.077	0.003	0.026	-0.048	0.054
pupL	-0.002	0.002	-0.006	0.002	0.004	0.002	0.001	0.008
avincS	0.115	0.046	0.022	0.207	-0.016	0.030	-0.077	0.044
avincL	-0.036	0.034	-0.105	0.033	0.004	0.036	-0.068	0.077
pctbadenglishS	0.009	0.025	-0.041	0.058	-0.115	0.041	-0.197	-0.033
pctbadenglishL	-0.065	0.022	-0.109	-0.021	0.098	0.050	-0.001	0.197
loweducS	0.019	0.008	0.002	0.036	0.001	0.005	-0.011	0.010
loweducL	-0.007	0.007	-0.021	0.006	0.001	0.007	-0.011	0.014
leikraceS	0.176	0.539	-0.901	1.254	1.570	0.950	-0.330	3.469
leikraceL	-0.986	0.423	-1.832	-0.140	-1.463	0.637	-2.736	-0.190
	Texas - Smaller municipalities				Texas - Larger municipalities			
pupS	0.018	0.009	0.001	0.035	0.005	0.008	-0.011	0.022
pupL	0.001	0.001	-0.003	0.003	0.001	0.002	-0.002	0.004
avincS	0.094	0.055	-0.016	0.204	-0.002	0.025	-0.051	0.047
avincL	0.121	0.050	0.021	0.222	0.133	0.060	0.013	0.253
pctbadenglishS	-0.010	0.013	-0.035	0.016	-0.003	0.009	-0.022	0.015
pctbadenglishL	0.015	0.015	-0.015	0.045	-0.015	0.016	-0.047	0.016
loweducS	0.016	0.007	0.001	0.030	0.003	0.004	-0.005	0.012
loweducL	-0.001	0.008	-0.016	0.014	0.024	0.009	0.005	0.042
leikraceS	0.081	0.285	-0.489	0.651	0.141	0.272	-0.402	0.684
leikraceL	0.432	0.391	-0.351	1.214	-0.052	0.313	-0.678	0.574

Table 14: Estimation Results - Correlated Random Effects - Model 2

Parameter	Mean	s.d.	0.05	0.95	Mean	s.d.	0.05	0.95
	Smaller municipalities				Larger municipalities			
Cleveland	0.652	0.534	-0.416	1.720	-0.677	0.480	-1.637	0.283
Cincinnati	0.485	0.563	-0.641	1.611	0.449	0.736	-1.022	1.920
Colombus	-1.390	0.610	-2.611	-0.170	0.467	0.989	-1.511	2.445
Dayton	-1.312	0.573	-2.458	-0.166	-0.244	0.722	-1.688	1.200
Toledo	-1.460	0.649	-2.758	-0.161	0.164	0.705	-1.246	1.574
San Antonio	-0.727	0.907	-2.541	1.086	0.293	0.696	-1.098	1.684
Houston	-1.894	0.766	-3.427	-0.361	-0.608	0.448	-1.504	0.287
Dallas	-1.870	0.613	-3.097	-0.644	-0.053	0.648	-1.350	1.243