GMM Estimation of Spatial Autoregressive Probit Models: An Analysis of the Implementation of the *District Planning System* in Japan

Tadao, HOSHINO*
Department of Social Engineering, Tokyo Institute of Technology

Abstract

This paper proposes a feasible generalized method of moments (GMM) estimator for spatial binary probit models containing both a spatial lag latent dependent variable and spatial autoregressive disturbances. Under empirically reasonable conditions, the estimator is consistent and asymptotically normal. We numerically assess the finite sample properties of our estimator using Monte Carlo experiments, and confirm the validity of the estimator. To demonstrate the availability of our proposed GMM estimator, we applied the technique to actual data for urban planning policies in Japan. In particular, this study focuses on the implementation of the District Planning System in Yokohama city, the capital city of Kanagawa prefecture. The District Planning System is a detailed land use management system introduced voluntarily by the local authorities according to the nature of each district, which reflects the demands of the residents for the local environment. Our results indicated the existence of positive spatial autocorrelations in the utilization of the system in terms of both the dependent variable and the omitted variables. This implies that the inhabitants' preferences for the local environment are spatially autocorrelated.

Keywords:
Generalized method of moments; Spatial autocorrelation; Probit models; District Planning System

JEL Classification: C31, C35, R52

*Correspondence: Tadao Hoshino; E-mail: hoshino.t.ai@m.titech.ac.jp; Address: 627, W9-89, Tokyo Institute of Technology, 2-12-1 Ookayama, Meguro, Tokyo, JAPAN.

1 Introduction

In statistical analyses of spatial data, the importance of spatial dependence among the observations of the consistency and efficiency of estimates has been emphasized. Since Anselin (1988), many theoretical and empirical studies on the econometric models that deal with spatial dependence have been undertaken. However, most of these studies consider linear models. The number of studies on spatial dependence in non-linear models, including discrete choice models, is significantly smaller than that of the linear cases. As Fleming (2002) states, this may be due to the added complexity that spatial dependence introduces into discrete choice models. With the presence of spatial dependence, i.e., spatial autocorrelation, in discrete choice models, the traditional maximum likelihood method is less practical because the likelihood function requires us to evaluate n-dimensional integration and the determinant of the $n \times n$ matrix, where n is the number of observations. To avoid the direct calculation of multi-dimensional integration, several techniques have been proposed (e.g., Beron and Vijverberg 2000; Bhat and Guo 2004; Bolduc et al. 1997; Klier and McMillen 2005; LeSage 2000; McMillen 1992; Pinkse and Slade 1998). These techniques are summarized justly in Fleming (2002). Among them, especially when the number of samples is large, the use of the generalized method of moments (GMM) is quite attractive in terms of its computational feasibility. Thus, the objective of this study is to develop a feasible GMM estimator of spatial autoregressive binary probit models.

To demonstrate the availability of our GMM estimator in empirical studies, we apply the technique to actual data from urban planning policies in Japan. More specifically, this study focuses on the implementation of the *District Planning System* by the districts of Yokohama city, the capital of Kanagawa prefecture. The District Planning System - *chiku keikaku* in Japanese - was introduced with the goal of micro-level land use management according to each district's nature by aggregating the demands of that area's residents regarding the local environment. Though the use of the system is not obligatory, many city districts across the country are utilizing it with the intention of improving the local environments. Our research hypothesis is that there are spatial autocorrelations in the implementation of the District Planning System in terms of both the latent dependent variable and the omitted variables. The former may be due to the presence of the interactions of people between neighbourhoods and the latter may be due to the presence of unobservable regional factors.

The importance of spatial dependence in the context of urban planning policies has been discussed often by urban sociologists and planners. An interesting and well-known example is "Broken Windows" by Wilson and Kelling (1982). The broken window theory states that when a neighborhood begins to decline visually, it can be an indication that no one cares about the area; then, more crime occurs in the area, and the neighbourhood will continue to deteriorate. Such multiplier effects, so-called neighbourhood effects, can be expected to function in a positive way as well. Accordingly, our research is considered as an examination of the positive side of the broken window theory; that is, we want to examine whether there are significant spatial endogenous effects in the inhabitants' demands on the improvement of local environments. The dependent variable used in this study is one when the city district contains at least one area that is somewhat preserved or restricted by the district planning and zero otherwise. The theoretical considerations of the identification and estimation of discrete choice models in the presence of neighbourhood effects are given by Brock and Durlauf (2001) on binary choice models and the paper by the same authors (2002) on multinomial choice models.

The remainder of this paper is organized as follows. In Section 2, we describe our GMM estimator. Further, the consistency and asymptotic normality of the estimator are established under certain conditions. In Section 3, we examine the properties of our estimator by a set of Monte Carlo experiments. Our empirical analysis is described in Section 4 containing an overview of the study area, an explanation of the data used, and the estimation results. Finally, in Section 5, we give our concluding remarks.

2 GMM Estimation of Spatial Autoregressive Probit Models

2.1 Consistency

In the binary probit models, we observe a dummy variable, y_i, defined by

$$\begin{aligned} y_i &= 1 \quad if \quad y_i^* \geq 0, \\ &= 0 \quad otherwise, \qquad (i = 1, ..., n) \end{aligned} \tag{1}$$

where y_i^* is a latent variable for an ith observation. To define our sampling space, we borrow the framework used in Conley (1999): Suppose that each observation i is located at one of a collection of points $\{s_i\}$ inside a sample region. Then assume

A.1 (a) The sample region $\Lambda_\tau \subset R^2$ is one of a sequence of compact convex regions $\{\Lambda_\tau\}$ which increases in area as $\tau \to \infty$; (b) Λ_τ grows uniformly in area in at least two non-opposing directions to increase the sample size as $\tau \to \infty$; and (c) the density of observation in a bounded area is bounded.

A.1 is similar in concept to the asymptotics in space that is referred to as *increasing domain* asymptotics.[1]

[1] A different way of characterizing asymptotics in space is called *infill* asymptotics. Infill asymptotics are appropriate when the spatial domain is bounded, and new observations are

Now, consider the spatial autoregressive model specified as

$$\begin{aligned} y^* &= \rho W_n y^* + X\beta + \varepsilon, \\ \varepsilon &= \lambda M_n \varepsilon + u, \end{aligned} \tag{2}$$

where ρ and λ are scalar autoregressive parameters indicating the degree of spatial dependence, W_n and M_n are $n \times n$ spatial weight matrices, $X = (x_1, ..., x_n)'$ is the $n \times k$ matrix of regressors, x_i being a $k \times 1$ vector, β is the $k \times 1$ vector of coefficients, $\varepsilon = (\varepsilon_1, ..., \varepsilon_n)'$ is the $n \times 1$ vector of omitted random variables, and $u = (u_1, ..., u_n)'$ is an $n \times 1$ vector of innovations. With regard to this model, consider the following conditions:

A.2 $\{u_i\}$ has an i.i.d. normal distribution with mean $E[u_i] = 0$, and variance $E[u_i^2] \equiv \sigma_u^2 < \infty$.

A.3 (a) The diagonal elements of W_n and M_n are zero; and (b) each of the off-diagonal elements of W_n and M_n is uniformly bounded in absolute value by one, and is defined by a non-random function of $w_{ij}(d_W(i,j))$ and $m_{ij}(d_M(i,j))$, respectively, such that $w_{ij}(\kappa), m_{ij}(\kappa) \to 0$ as $\kappa \uparrow \infty$, where $d.(i,j)$ is the distance between i and j measured by some metric that is positively correlated with the Euclidean distance between them.

A.4 (a) $(I_n - \rho W_n)$ and $(I_n - \lambda M_n)$ are nonsingular; and (b) $\rho \in (-1, 1)$, $\lambda \in (-1, 1)$.

A.5 (a) X has the full rank k; (b) the elements of X are uniformly bounded in absolute value; (c) the values of X_i are determined by a $k \times 1$ underlying random field $\mathscr{X}_s$ at location s_i that is spatially stationary and mixing; and (d) the sampling process is determined by a random process that is spatially stationary and mixing being independent of $\mathscr{X}_s$ (see the W_s *process* in Conley 1999 3.1.2).

The specification in A.3 (b) means that the processes of spatial dependence for y^* and ε are assumed to be spatially stationary and mixing. For the definition of mixing condition in this context, please see Appendix 2, and also refer to the definition in Bolthausen (1982) and Conley (1999).

Assuming A.3 and A.4, the following equations hold:

$$\begin{aligned} (I_n - \rho W_n)^{-1} &= I_n + \rho W_n + \rho^2 W_n W_n + ... \quad , \\ (I_n - \lambda M_n)^{-1} &= I_n + \lambda M_n + \lambda^2 M_n M_n + ... \quad . \end{aligned} \tag{3}$$

Since calculating the $n \times n$ inverse matrix is quite difficult when the number of samples is large, utilizing Eq.(3) is really important in practice to speed up the estimation. Note that in the following, one can always use Eq.(3) as substitutes of $(I_n - \rho W_n)^{-1}$ and $(I_n - \lambda M_n)^{-1}$, respectively.

added in between existing ones, generating a increasingly denser surface (Anselin 2003).

Now, y_i^* is re-written as

$$y^* = (I_n - \rho W_n)^{-1} X\beta + (I_n - \rho W_n)^{-1}(I_n - \lambda M_n)^{-1} u. \quad (4)$$

For simplicity of notation, let us define the following:

$$v \equiv (I_n - \rho W_n)^{-1}(I_n - \lambda M_n)^{-1} u,$$

$$X^\circ \equiv (I_n - \rho W_n)^{-1} X.$$

Then, the probability $P(y_i = 1)$ is also considered as the probability $P(v_i \geq -X_i^\circ \beta)$. Note that since the row sums of $(I_n - \rho W_n)^{-1}$ and $(I_n - \lambda M_n)^{-1}$ are uniformly bounded (Lemma 1),[2] and $E[u_i] = 0$ (A.2), applying the product rule, $E[v_i] = 0$. The variance-covariance matrix is given by

$$E[vv'] = (I_n - \rho W_n)^{-1}(I_n - \lambda M_n)^{-1}(I_n - \lambda M_n')^{-1}(I_n - \rho W_n')^{-1}\sigma_u^2. \quad (5)$$

Let us denote the diagonal element of Eq.(5) as σ_v^2 (formally, $\sigma_v^2(\rho, \lambda) = (\sigma_{v1}^2(\rho, \lambda), ..., \sigma_{vn}^2(\rho, \lambda))'$). We now have the expected value of v_i as follows:

$$E[v_i(\theta)|X_i, y_i = 1] = \sigma_{vi} \frac{\phi(X_i^\circ \beta / \sigma_{vi})}{\Phi(X_i^\circ \beta / \sigma_{vi})},$$

$$E[v_i(\theta)|X_i, y_i = 0] = -\sigma_{vi} \frac{\phi(X_i^\circ \beta / \sigma_{vi})}{1 - \Phi(X_i^\circ \beta / \sigma_{vi})}, \quad (6)$$

where θ is the $(k+2) \times 1$ vector of parameters $< \beta', \rho, \lambda >'$, $\phi(\cdot)$ is the normal density function, and $\Phi(\cdot)$ is the normal cumulative distribution function. In our model, v_i is replaced by its prediction $E[v_i(\theta)|X_i, y_i]$ - the generalized residuals (for details, please see, e.g., Cox and Snell 1968; Gourieroux et al. 1987). The GMM estimator for spatial probit models proposed by Pinkse and Slade (1998) constructed moment conditions based on the use of the generalized residuals.[3] Indeed, our estimator stated below is an extension of theirs.

In a regular probit model, the standard error is assumed to be homoskedastic or constant across observations. However, as Eq.(5) shows, the variances are heteroskedastic, being a function of ρ and λ. Although ignoring the heterosedasticity can achieve consistency in the context of linear regression models, this is not the case in the context of discrete choice models (Yatchew and Griliches 1985). Therefore, it is required to correct the heteroskedasticity to obtain unbiased and consistent estimates. Of course, σ_u^2 must be fixed in our model as well.

By using $q_i \equiv 2y_i - 1$, we re-write Eq.(6) as

$$\tilde{v}_i(\theta|X_i, y_i) \equiv q_i \sigma_{vi} \frac{\phi(X_i^\circ \beta / \sigma_{vi})}{\Phi(q_i X_i^\circ \beta / \sigma_{vi})}. \quad (7)$$

[2] All the lemmas are summarized in Appendix 1.

[3] Klier and McMillen (2005) also constructed the GMM estimator for spatial logit models based on the generalized residuals for logit models. Their estimator was a linearized logit version of Pinkse and Slade (1998)'s estimator.

We henceforth write $\tilde{v}_i(\theta|X_i, y_i)$ simply as $\tilde{v}_i(\theta)$. Now, suppose there is a moment function vector $g(Z, \theta)$ such that

$$\hat{g}_n(Z,\theta) \equiv n^{-1}\sum_{i=1}^{n} g(Z_i,\theta) = n^{-1}\sum_{i=1}^{n} Z_i'\tilde{v}_i(\theta), \tag{8}$$

where Z is a matrix of instruments. In the first step of the estimation, we take Z to be a fixed subset of linearly independent columns of $\{X$, W_nX, W_n^2X ,..., M_nX, $M_n^2X, ...\}$. As is well known, the matrix of optimal instruments must be sufficiently correlated with the Jacobian of $\tilde{v}(\theta_0)$. Accordingly, in the second step, we update the matrix of instruments to be optimal by utilizing the estimates from the first step of the estimation. Newey (1993)[4] shows that the optimal instruments in the case of conditional heteroskedasticity are given by

$$Z^* \equiv \Omega(\theta_0)^{-1}R(\theta_0)F, \tag{9}$$

where $\Omega(\theta_0) = E[\tilde{v}(\theta_0)\tilde{v}(\theta_0)']$, $R(\theta_0) = E[\partial\tilde{v}(\theta_0)/\partial\theta']$, and F is any nonsingular matrix representing the proportional transformaitons of the matrix of instruments. $R(\theta_0)$ is calculated by using the equations (A.1)-(A.3) in Appendix 1. In the GMM estimation of panel probit models, Inkmann (2000) shows that the efficiency gains from exploiting the optimal instruments and weight matrix are remarkably high.

Since $E[v_i]$ is zero, $E[\tilde{v}_i(\theta_0)]$ is also zero by the law of iterated expectations where θ_0 is a vector of true parameters. Hence, if the elements of Z are uniformly bounded (Lemma 3), $E[g(Z_i, \theta_0)] = 0$. Thus, a GMM estimator is one that solves the following:

$$\hat{\theta} = \underset{\theta\in\Theta}{argmin}\, \hat{Q}_n(Z,\theta), \tag{10}$$

where

$$\hat{Q}_n(Z,\theta) = \hat{g}_n(Z,\theta)'\hat{\Xi}\hat{g}_n(Z,\theta),$$

and where $\hat{\Xi}$ is a positive definite matrix. If $\hat{g}_n(Z,\theta) \xrightarrow{p} E[g(Z_i,\theta)]$, and $\hat{\Xi} \xrightarrow{p} \Xi$, then, $\hat{Q}_n(Z,\theta) \xrightarrow{p} Q_0(Z,\theta) = E[g(Z_i,\theta)]'\Xi E[g(Z_i,\theta)]$. To establish the consistency of this estimator, we consider the assumptions listed below:

A.6 Θ is compact.

A.7 $\hat{\Xi} \xrightarrow{p} \Xi$, where Ξ is a positive definite matrix.

A.8 $\Xi E[g(Z_i, \theta_0)] = 0$ uniquely at $\theta_0 \in \Theta$.

A.9 $g(Z_i, \theta)$ is measurable and differentiable on Z_i for all $\theta \in \Theta$.

[4]Reference quoted from Inkmann (2000).

PROPOSITION 1 [Consistency]: If A.1-9 hold, then $\hat{\theta} \xrightarrow{p} \theta_0$.

A key condition for the establishment of the consistency of the estimator is $\sup_{\theta\in\Theta} |\hat{Q}_n(Z,\theta) - Q_0(Z,\theta)| \xrightarrow{p} 0$. The proof given in Appendix 1.1 shows that the condition can be satisfied under the assumptions set out above. These assumptions can be relaxed in some ways at the expense of additional complexities.

2.2 Asymptotic normality

To establish the asymptotic normality for our estimator, we utilize Bolthausen (1982)'s central limit theorem for stationary mixing random fields. To use the theorem, we need several additional conditions.

N.1 $G'\Xi G$ and S are nonsingular, where $G \equiv E[\partial g(Z_i,\theta_0)/\partial\theta']$ and $S \equiv E[g(Z_i,\theta_0)g(Z_i,\theta_0)']$.

N.2 $\sum_{\omega=1}^{\infty} \omega\alpha_{k,l}(\omega) < \infty$, for $k + l \leq 4$.

N.3 $\alpha_{1,\infty}(\omega) = o(\omega^{-2})$.

The definition of the spatial mixing coefficient used in N.2 and N.3, namely, $\alpha_{k,l}(\omega)$, is found in Appendix 2.

PROPOSITION 2 [Asymptotic normality]: If the assumptions of Proposition 1 and N.1-3 hold, then, $\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N[0, (G'\Xi G)^{-1}G'\Xi S\Xi G(G'\Xi G)^{-1}]$.

For proof, please see Appendix 1.2. Hansen (1982) shows that the optimal choice of the weight matrix Ξ converges to S^{-1}. Hence, in this case, the asymptotic distribution of the estimator collapses to

$$\sqrt{n}(\hat{\theta} - \theta_0) \xrightarrow{d} N[0, (G'S^{-1}G)^{-1}]. \tag{11}$$

3 Monte Carlo Experiments

We conduct Monte Carlo experiments in order to assertain the properties of our estimator. The basis of the experiments is a simple version of Eq.(2) with two regressors including the intercept: $y^* = \rho W_n y^* + \alpha + x\beta + \varepsilon$, $\varepsilon = \lambda M_n \varepsilon + u$. u is drawn from $N[0,1]$ and x is drawn from $N[4,4]$. We set the true parameters as $\alpha_0 = -1.5$ and $\beta_0 = 0.5$. The spatial weight matrix for the latent dependent variable, namely W_n, for the experiments is created to be heterogenous as follows: We first draw two random numbers from the uniform distribution $U[0,1]$ for each observation. These numbers are used to specify the coordinates of each observation in the $[0,1] \times [0,1]$ plane. For $i \neq j$, we set $w_{ij} = 1/\#C$ if the Euclidean distance between the

ith point and the jth point is less than 0.1, where $\#C$ is the total number of points that meet the condition stated. Of course, the diagonal elements of the matrix is set to zero. The spatial weight matrix for the disturbances, namely M_n, is created as follows: For $i \neq 1$, n, $m_{ij} = 1/2$ if $j = i+1, i-1$, and $m_{ij} = 0$ otherwise; when $i = 1$, $m_{1,2} = 1/2$, $m_{1,n} = 1/2$, and $m_{1,j} = 0$ otherwise; and when $i = n$, $m_{n,n-1} = 1/2$, $m_{n,1} = 1/2$, and $m_{n,j} = 0$ otherwise. Thus, this matrix is defined in a circular world.

Based on the set up explained above, given a particular value of ρ_0 and λ_0, two sizes of samples are generated, namely, $n = 300$ and $n = 800$. We report the results of the cases when the magnitude of the autocorrelation for each process is 0.2 or 0.6. Then, based on the variations of the spatial autocorrelation parameters and the sample sizes, ten sets of experiments including two benchmark sets with $\rho = 0$ and $\lambda = 0$ for each sample size are carried out. The models to be tested are the standard non-spatial probit model (i.e., misspecified model with $\rho = 0$ and $\lambda = 0$) and the spatial probit model. We use the classical maximum likelihood method to estimate the standard probit models. Table 1-(1) summarizes the mean, standard deviation, and root mean square error (RMSE) statistics of each parameter in these models in each Monte Carlo experiment when $n = 300$; and Table 1-(2) reports these statistics when $n = 800$. Each experiment is based on 1000 replications.

When spatial autocorrelation does not exist, the both models show a good performance with almost no differences between their performances. On the other hand when spatial autocorrelation exists, regarding the non-spatial standard (i.e., misspecified) probit, we observe a trend that the accuracy of the predictions rapidly decreases especially for the intercept term as the true spatial autoregressive parameters ρ_0 and λ_0 increase; while in the case of the spatial probit, the increases of the true spatial autoregressive parameters do not influence much on the accuracy of the estimates of the spatial autoregressive parameters, the intercept, and the coefficient. In addition, although the estimates of spatial autocorrelation in the disturbances, λ, have smaller deviations and errors than that in the latent dependent variable, ρ, they are always underestimates (about 10%) of the true value λ_0, while the estimates of ρ are always very close to the true value ρ_0 on average. However, at present, we do not have a clear explanation for this phenomenon, this might be due to the property of finite samples because the underestimations are somewhat corrected by enlarging the sample size. For the both models, increasing the sample size decreases the RMSE statistics of parameters only except for ρ (rather very slightly worse performance than the small sample cases, but nevertheless very accurate on average). On the whole, the increase of the sample size would make the estimations more reliable because of the fact that the average value of the RMSE for the large population is measurably smaller than that for the small population.

Thus, from these reported results, we can conclude that ignoring the spatial effects is quite problematic especially when the magnitude of the spatial autocorrelations is not small; and, we may conclude that the validity of our GMM estimator is confirmed.

Table 1: (1) Results of the Monte Carlo Experiments: Sample size: $n = 300$, True parameters: $\alpha_0 = -1.5$, $\beta_0 = 0.5$

		Para-	Standard Probit			Spatial probit		
ρ_0	λ_0	meter	Mean	Std. dev.	RMSE	Mean	Std. dev.	RMSE
0	0	α	-1.5367	0.2238	0.2268	-1.5372	0.2276	0.2306
		β	0.5117	0.0602	0.0592	0.5132	0.0572	0.0587
		ρ				-0.0051	0.1069	0.1070
		λ				-0.0001	0.0004	0.0005
0.2	0.2	α	-1.3369	0.2089	0.2650	-1.5295	0.2100	0.2121
		β	0.4878	0.0579	0.0592	0.5126	0.0572	0.0585
		ρ				0.1971	0.1021	0.1022
		λ				0.1801	0.0574	0.0608
0.6	0.2	α	-0.8888	0.1688	0.6341	-1.5340	0.2226	0.2252
		β	0.4028	0.0413	0.1056	0.5135	0.0619	0.0634
		ρ				0.6001	0.1326	0.1326
		λ				0.1761	0.0661	0.0703
0.2	0.6	α	-1.1691	0.2022	0.3878	-1.5121	0.1546	0.1551
		β	0.4258	0.0488	0.0888	0.5077	0.0407	0.0414
		ρ				0.1986	0.1101	0.1101
		λ				0.5665	0.1094	0.1144
0.6	0.6	α	-0.8172	0.1825	0.7068	-1.5142	0.1724	0.1729
		β	0.3655	0.0385	0.1399	0.5058	0.0471	0.0475
		ρ				0.6003	0.1426	0.1426
		λ				0.5526	0.1125	0.1221
RMSE Col. Average					0.2673			0.1114

Table 1: (2) Results of the Monte Carlo Experiments: Sample size: $n = 800$, True parameters: $\alpha_0 = -1.5$, $\beta_0 = 0.5$

		Para-	Standard Probit			Spatial probit		
ρ_0	λ_0	meter	Mean	Std. dev.	RMSE	Mean	Std. dev.	RMSE
0	0	α	-1.5113	0.1321	0.1325	-1.5093	0.1411	0.1414
		β	0.5041	0.0346	0.0349	0.5050	0.0348	0.0352
		ρ				-0.0079	0.1109	0.1112
		λ				-2.92E-05	9.70E-05	0.0001
0.2	0.2	α	-1.3959	0.1311	0.1674	-1.5114	0.1329	0.1334
		β	0.4975	0.0339	0.0340	0.5057	0.0331	0.0336
		ρ				0.2016	0.1207	0.1208
		λ				0.1858	0.0448	0.0470
0.6	0.2	α	-1.1483	0.1240	0.3729	-1.5151	0.1256	0.1265
		β	0.4759	0.0321	0.0402	0.5065	0.0336	0.0343
		ρ				0.5993	0.1355	0.1355
		λ				0.1856	0.0422	0.0446
0.2	0.6	α	-1.2093	0.1244	0.3163	-1.4841	0.1023	0.1035
		β	0.4310	0.0293	0.0750	0.4985	0.0261	0.0262
		ρ				0.1932	0.1325	0.1326
		λ				0.5679	0.0865	0.0923
0.6	0.6	α	-1.0029	0.1327	0.5145	-1.4894	0.0956	0.0962
		β	0.4167	0.0285	0.0881	0.4990	0.0240	0.0240
		ρ				0.6001	0.1481	0.1481
		λ				0.5657	0.0852	0.0919
RMSE Col. Average					0.1776			0.0839

4 Empirical Study

4.1 The District Planning System

In Japan, the District Planning System - *chiku keikaku* in Japanese - has been stipulated in the Urban Planning Law from 1980, aiming for micro-level land use management according to each district's nature. The basic idea of this system was based on the detailed district planning scheme in Germany, *Bebauungsplan* (widely known as B-plan). The District Planning System regulates, for example, the development of green areas; the purpose and design of buildings; and the building-to-land ratio. Such district plans are introduced in relatively small areas. In drawing up a plan, the local authority must obtain the consent of the local residents living in that area. District plans can be proposed not only by local authorities but also by the inhabitants. Thus, the system is designed to reflect the aggregated demands of the inhabitants. Thus, if there are some spatial autocorrelations in characterizing the inhabitants' preferences for local environment, we can also observe the spatial dependence in the implementation of the District Planning System.

In order to preserve and create better regional environments and amenities, the system has been introduced mainly in residential districts, shopping and business districts near train stations, and bureaucratic districts. Since 1980, the total number of plans created has reached 4570 across the country (Chiku keikaku kenkyukai 2005).

4.2 Data

Situated 30 km south of the central Tokyo, Yokohama city is the capital of Kanagawa prefecture. The city's population is about 3.6 million (Population Census 2006), as it is a commercial hub of the Greater Tokyo area and one of the most populous bedroom suburbs where commuters to Tokyo reside. As of 2008, a total number of 90 district plans covering an area of 1573.7 ha had been proposed for the city.[5] For a single city, these figures are remarkaby high compared with other cities like Kyoto, one of the most historical and cultural cities in Japan.[6]

Yokohama city has 18 wards. In the Japanese addressing system, wards are divided into towns and towns may be subdivided into city districts, called *chome* in Japanese. In our empirical study, the binary dependent variable is defined as follows: one if a city district contains at least one area that utilizes the District Planning System and zero otherwise. It is often the case that a district plan intersects the boundaries of more than two distinct city districts. This presents minor complications throughout our study. Our definition of the dependent variable means that if a district plan crosses over city districts, we allot one to each of them despite the crossover. Based on this definition, Figure 1 shows the distribution of the dependent variable. The presence of a strong spatial correlation in the implementation of the District Planning System can be confirmed from the figure. Note that if we estimate the model without controlling the crossovers in the specification of the spatial weight matrices, we will obtain unreasonably high spatial autoregressive parameters. Our approach to deal with this issue is stated in the next subsection.

The variables used and their definitions are listed in Table A.1 in Appendix 3. Among 18 the wards, ward dummies that do not have a significant effect on the dependent variable are dropped. The total number of districts is 1758. We exclude non-residential areas including seas, non-residential islands, urban parks, and districts that have data with missing values from

[5] Yokohama city web page: http://www.city.yokohama.jp/me/toshi/

[6] The number of district plans in Kyoto city is 50 as of 2008 (Kyoto city web page: http://www.city.kyoto.lg.jp/tokei/).

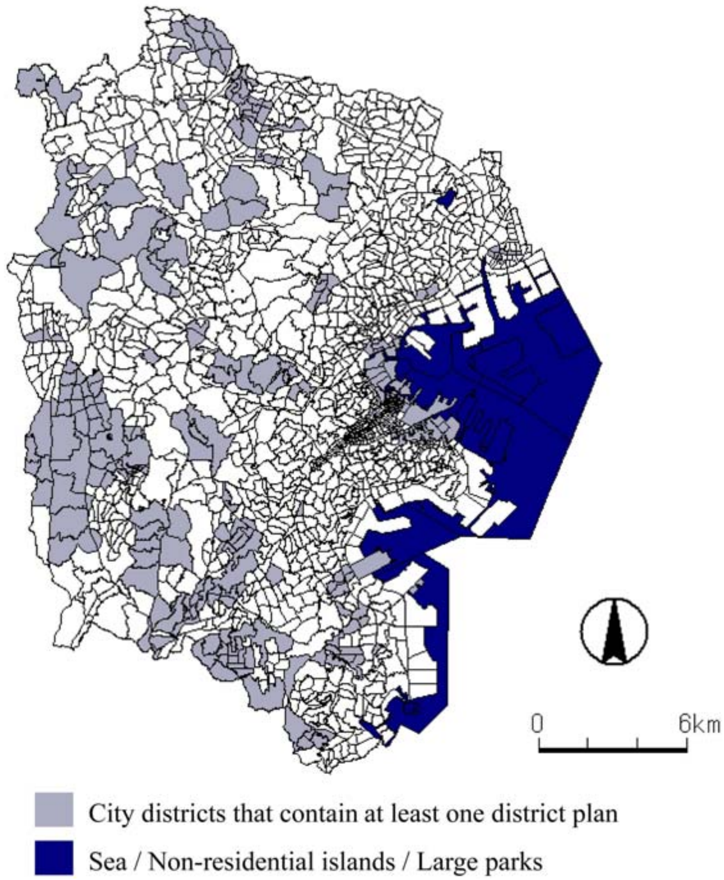

Figure 1: District plans in Yokohama city

the total. Consequently, the number of city districts used in this study is 1573. Table A.2 in Appendix 3 presents the descriptive statistics for the variables used. The variables are the most current data, which was made available in November 2008 from the Census and the website of Yokohama city's urban planning. It is important to note that we only use cross-sectional information. Although it takes some time for a district to be affected by the neighbouring districts' planning policies, and make a decision on the planning of the district itself, we ignore such time series properties on the variables. This is simply due to the limited availability of detailed data. Thus, the results presented in the subsection 4.4 should be interpreted as long term trends.

4.3 Spatial weight matrices

There are many possibilities in choosing the metrics to define the spatial weight matrices. As stated above, we need a special method to treat the crossovers of district plans lying on distinct city districts. In this study, we consider four types of spatial weight matrices. For notational simplicity, each specification is explained in terms of W_n.

1. **First order contiguity:** The first metric is the one that is most commonly used in studies on spatial analysis. With this metric, $w_{ij} = 1/\#C$ if the jth district is contiguous to the ith district, where $\#C$ is the total number of contiguous districts to i. It is easy to create this matrix by using packages for spatial data analysis such as Geoda (Anselin et al. 2006).

2. **First order contiguity $\times$ Euclidean distance:** This metric combines the Euclidean distance with the first specification; w_{ij} is the reciprocal of the Euclidean distance between the mid-point of the ith district and that of the jth district if j is contiguous to i and normalized such that $\sum_j^n w_{ij} = 1$.

3 4. **Modified version of 1 and 2:** There are district plans that are shared by several city districts, and necessarily, these districts are contiguous. Hence, if we do not control this, we will overestimate the degree of spatial dependence. This is particularly important for the estimation of ρ rather than λ. Our approach is simple: If the jth district shares the same district plan with the ith district, we set w_{ij} to zero. The third and fourth weight matrices are modified versions of the first and second weight matrices, respectively in this manner. We describe this weighting scheme by using the type of diagram presented below.

Figure 2: Example region

Figure 2 shows an example of the sample region. For this region, the modified version of the first order contiguity weight matrix is given by

$$W_n = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 0 & 1/3 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 1/3 & 1/3 & 1/3 & 0 \end{bmatrix}.$$

Let us simply call these four matrices as type 1, type 2, type 3, and type 4, respectively. When using the modified spatial weight matrix, namely type 3 or 4, the estimated magnitude of the spatial autoregressive parameters should be regarded as a lower bound.

4.4 Results

Table 2 reports the estimation results of the standard non-spatial probit model (i.e., misspecified model with $\rho = 0$ and $\lambda = 0$). The overall fitness of the model is very high despite the fact that the presence of spatial autocorrelations is not considered. LRI is 0.490, and more than 85 percent of the sample's behaviour is correctly predicted.

The results appear to show two different trends in the effects of the variables. Namely, the negative sign for Density and the positive sign for distance to Yokohama, and Residential Zone imply that the District Planning System tends to be utilized in the suburban residential areas. On the contrary, the positive sign for High-rise buildings, the number of large businesses, number of retailers, and Industrial Zone are evidences that the system is often introduced in the highly developed urban areas. This discrepancy may be the result of the heterogeneity in the aims of the district plans implemented by the heterogenous city districts.

All coefficient and parameter estimation results from the spatial probit models are reported in Table 3. In the estimation, we consider two versions of the models: model 1 with $W_n =$ type 3, and $M_n =$ type 1, and model 2 with $W_n =$ type 4, and $M_n =$ type 2. Regarding the overall estimation results, those concerning the relative importance of the variables do not differ much. The ratios of success predictions of the spatial models are very slightly higher than that of the standard probit case. In addition, using the combined metric of the first order contiguity and the Euclidean distance slightly improves the prediction. The values of the objective function to be minimized for the both models are almost zero, suggesting that the estimation was successfully conducted.

With regard to the spatial autoregressive parameters, both ρ and λ in the two models are significantly positively related to the dependent variable. The degree of the spatial lag effects, namely $W_n y^*$, is lying around 0.26 when

Table 2: Standard probit model

	Standard probit		
Variable	Coef.		t-value
Intercept	-1.572	***	-4.393
Density	-0.004	***	-4.919
# Owner	0.026	***	2.643
High-rise buil.	0.405	**	2.157
# Large bus.	0.016	***	3.991
# Retailers	0.077	***	4.467
# Wholesalers	0.024		0.627
Area of residence	-0.019		-0.558
Dist. to Yokohama	0.021		1.093
Ratio of Workers	0.221		1.062
Ratio of Managers	-4.470	*	-1.766
Ratio of Blue Collar	-3.189	***	-4.930
Ratio of Farmers	3.095		1.141
Residential Zone	0.019	***	2.924
Industrial Zone	0.024	***	3.547
Dummy Sakae	0.931	***	4.888
Dummy Kohoku	-1.436	***	-3.407
Dummy Izumi	0.525	**	2.375
Dummy Naka	0.935	***	6.349
Log-likelihood	-556.153		
LRI	0.490		
Correct preds.	1344 (85.4 %)		
Number of observations	1573		

*,**, and *** indicate significance at the 0.10, 0.05, and 0.01 levels.

the spatial weight matrix is the type 3, and 0.2 when it is the type 4; and the degree of spatial autocorrelations in the disturbances, namely $M_n\varepsilon$, is a littel under 0.8 for both the type 1 and 2 spatial weight matrices.

From these results, we may conclude that the results support our research hypothesis: the implementation of the District Planning System is spatially dependent; and therefore, the residents' preferences for the local environment should be inferred as spatially autocorrelated. This is likely because of the interactions between the people in contiguous neighbourhoods and unobservable regional common factors.

Table 3: Spatial probit models

	Spatial probit					
	W_n =type 3, M_n =type 1			W_n =type 4, M_n =type 2		
Variable	Coef.		t-value	Coef.		t-value
Intercept	-1.356	***	-4.893	-1.463	***	-5.309
Density	-0.004	***	-6.729	-0.004	***	-7.322
# Owner	0.033	***	4.117	0.033	***	3.935
High-rise buil.	0.415	***	2.638	0.440	***	2.799
# Large bus.	0.017	***	5.040	0.018	***	5.231
# Retailers	0.089	***	6.476	0.086	***	6.657
# Wholesalers	0.015		0.558	0.021		0.488
Area of residence	-0.028		-1.046	-0.031		-1.210
Dist. to Yokohama st.	0.016		1.464	0.018		1.524
Ratio of Workers	0.149		0.775	0.184		0.916
Ratio of Managers	-4.482	*	-1.894	-4.485	*	-1.901
Ratio of Blue Collar	-2.989	***	-5.906	-2.993	***	-6.048
Ratio of Farmers	3.104		1.176	3.106		1.190
Residential Zone	0.019	***	4.515	0.020	***	4.627
Industrial Zone	0.023	***	5.349	0.024	***	5.540
Dummy Sakae	0.908	***	7.787	0.978	***	8.511
Dummy Kohoku	-1.430	***	-3.865	-1.421	***	-3.864
Dummy Izumi	0.417	***	3.738	0.477	***	4.092
Dummy Naka	0.846	***	7.410	0.897	***	8.108
$W_n y^*$	0.259	***	3.351	0.200	***	2.984
$M_n \varepsilon$	0.769	***	5.516	0.786	***	8.789
$\hat{Q}_n(Z, \hat{\theta})$	2.683E-05			5.531E-05		
Correct preds.	1345 (85.5 %)			1346 (85.6 %)		
Number of observations	1573			1573		

*,**, and *** indicate significance at the 0.10, 0.05, and 0.01 levels.

5 Conclusions

The objective of this paper was to develop a feasible GMM estimator for spatial autoregressive probit models that involve both a saptial lag latent dependent variable and spatial autoregressive disturbances. The proposed estimator is computationally attractive because it allows us not to evaluate the complicated likelihood funcition containing n-dimensional integration and the determinant of the $n \times n$ matrix. Under empirically reasonable conditions, the estimator is consistent and asymptotically normal. In the Monte Carlo analysis, the validity of our estimator was confirmed.

We applied our proposed GMM estimator to examine the presence of spatial autocorrelations in the introduction of the District Planning System in Yokohama city. In the estimation, we formulated four types of spatial weight matrices in order to control the crossovers of particular district plans lying on distinct city districts. Our finding was that significant spatial effects exist in the implementation of the system, which implied that the inhabitants' preferences for the local environment are spatially autocorrelated. Although even the standard probit model achieved a good prediction performance, confirming the presence of such spatial endogenous effects is quite important in their implications in urban planning policies, as suggested in the broken window theory. Our results suggest that if the policy makers' objective is to create a better environment for a particular region, they do not have to construct a wide and detailed planning scenario that applies to the entire region in terms of the efficiency. Instead, for example, by designating one of the areas in that region as a "model case" and intensively improving its environment, the improvements would spread to the neighbourhoods to the area; further, the effects would spread to the neighbourhoods that are adjacent to those as well. At the end, the ultimate effects of the policy can be unexpectedly huge, even if the policy itself is small and limited.

As stated in the introduction, the number of theoretical and empirical studies on spatial discrete choice models is very small. Therefore, it is still too early to give a definitive appraisal of our estimator. Hence, comparing our approach with the other proposed estimators is a future task. Moreover, it may be necessary to create estimation techniques not only for spatial binary probit models but for a more general class of discrete choice models, such as semiparametric models.

Acknowledgements

I am particularly indebted to Hisatoshi Tanaka, Koichi Kuriyama, Dan Sasaki, and Tetsuya Shimane for valuable suggestions and discussion. I would like to thank seminar participants at Tokyo Institute of Technology and GRIPS. All remaining errors are solely my responsibility.

Appendix 1. Proofs of Propositions

Lemma 1 The row and column sums of the matrices $(I-\rho W_n)^{-1}$ and $(I-\lambda M_n)^{-1}$ are uniformly bounded in absolute value uniformly in ρ and λ, respectively.

Immediate from A.3, A.4, and Eq.(3). Q.E.D.

Lemma 2 Let A be a spatially stationary and mixing scalar sequence of $< A_1, ..., A_n >'$ such that the elements of A are uniformly bounded in absolute value, and U_i be the ith row of $(I_n - \rho W_n)^{-1}$ (or $(I_n - \lambda M_n)^{-1}$). Then, (1) $U_i A$ is uniformly bounded in absolute value uniformly in ρ (or λ); and (2) $n^{-1}\sum_{i=1}^{n} U_i A \xrightarrow{p} E[U_i A]$.

(1) By Lemma1, there exists a constant C independent of n such that $|U_iA| \leq |U_i||A| \leq \{\sup_{i\leq n}|U_i|\}|A| \leq C < \infty$.

(2) Given stationarity, the mixing condition implies the ergodicity (see Rosenblatt 1978).[7] Therefore, U_i is spatially ergodic by the definition of ρ, λ and W_n, M_n, and so is A. Hence, $\{U_iA\}$ is a spatially stationary and ergodic scalar sequence with $E[|U_iA|] < \infty$. Then, the result follows by the ergodic theorem. Q.E.D.

By applying these results and the assumptions A.2, A.5 and A.6, we obtain the following basic result: $\sup_{i\leq n}\sigma^2_{vi}(\rho,\lambda) < \infty$ for all $\rho, \lambda \in (-1, 1)$.

Lemma 3 $\sup_{\theta\in\Theta}\left\| n^{-1}\sum_{i=1}^{n}\frac{\partial g(Z_i,\theta)}{\partial\theta'}\right\| = O_p(1)$.

Given A.9, since $\partial g(Z_i,\theta)/\theta' = Z_i'\partial\tilde{v}_i(\theta)/\partial\theta'$, clearly Z_i used in the first step is uniformly bounded by definition, and Z_i^* involves $E[\partial\tilde{v}(\theta_0)/\partial\theta']$, we then just consider the boundedness of $\partial\tilde{v}_i(\theta)/\partial\theta'$. Now, for notational simplicity, let us define the followings: $(I_n-\rho W_n)^{-1}X\sigma_v^{-1} \equiv X^*$, and $(I_n - \rho W_n)^{-1}W_n(I_n-\rho W_n)^{-1}X \equiv X^{**}$; then we have the following results:

$$\frac{\partial}{\partial\beta_k}\begin{bmatrix}\tilde{v}_1(\theta)\\ \vdots\end{bmatrix} = -\begin{bmatrix}\tilde{v}_1(\theta)x^*_{k1}\left\{(X^*\beta)_1 + \dfrac{\tilde{v}_1(\theta)}{\sigma_{v1}}\right\}\\ \vdots\end{bmatrix}; \tag{A.1}$$

$$\frac{\partial}{\partial\rho}\begin{bmatrix}\tilde{v}_1(\theta)\\ \vdots\end{bmatrix} = \begin{bmatrix}\dfrac{\tilde{v}_1(\theta)}{\sigma_{v1}}\left\{\Delta_{\rho1} - [(X^{**}\beta)_1 - \Delta_{\rho1}(X^*\beta)_1]\left[(X^*\beta)_1 + \dfrac{\tilde{v}_1(\theta)}{\sigma_{v1}}\right]\right\}\\ \vdots\end{bmatrix};$$

[7] Reference quoted from Proposition 3.44 in White (2001).

(A.2)

and

$$\frac{\partial}{\partial\lambda}\left[\begin{array}{c}\tilde{v}_1(\theta)\\ \vdots\end{array}\right] = \left[\begin{array}{c}\Delta_{\lambda 1}\tilde{v}_1(\theta)\left\{\frac{1}{\sigma_{v1}} + \frac{(X^*\beta)_1}{\sigma_{v1}^2}[(X^*\beta)_1 + \tilde{v}_1(\theta)]\right\}\\ \vdots\end{array}\right], \quad (A.3)$$

where Δ_ρ is the derivative of the Eq.(5) with respect to ρ, and Δ_λ is similarly the derivative of the equation with respect to λ. Then, repeated application of Lemma 2 to the equations (A.1)-(A.3) yields $\sup_{\theta\in\Theta}||\partial\tilde{v}_i(\theta)/\partial\theta'|| = O_p(1)$. Hence, the result is verified. Q.E.D.

From this result, we obtain $\sup_{\theta\in\Theta}||g(Z_i,\theta)|| = O_p(1)$. Also, since $\hat{\Xi}$ is positive definite, $\sup_{\theta\in\Theta}||\partial\hat{Q}_n(Z,\theta)/\partial\theta'|| = O_p(1)$ holds as well.

Lemma 4 $\hat{g}_n(Z,\theta) \xrightarrow{p} E[g(Z_i,\theta)]$ for all $\theta \in \Theta$.

Given Lemma 2 and 3, the result holds. For detailed proof, please see the proposition 1 in Conley (1999). Further, by A.9, $E[g(Z_i,\theta)]$ is continuous for all $\theta \in \Theta$. Q.E.D.

Lemma 5 $\hat{Q}_n(Z,\theta) \xrightarrow{p} Q_0(Z,\theta)$ for all $\theta \in \Theta$.

Immediate from A.7 and Lemma 4. Hence, $Q_0(Z,\theta)$ is uniquely minimized at $\theta_0 \in \Theta$ and coninuous for all $\theta \in \Theta$. Q.E.D.

Lemma 6 $\hat{Q}_n(Z,\theta) - Q_0(Z,\theta)$ is stochastic equicontinuous.

To establish the stochastic equicontinuity for $\hat{Q}_n(Z,\theta)$, we consider the Lipschitz condition. The Lipschitz condition is that $|\hat{Q}_n(Z,\theta^*) - \hat{Q}_n(Z,\theta^{**})| \leq B(Z_i)h(e(\theta^*,\theta^{**}))$, $\forall\theta^*,\theta^{**} \in \Theta$ where $B : Z_i \to R^+$ and $h : R^+ \to R^+$ such that $h(\kappa) \downarrow 0$ as $\kappa \downarrow 0$, where e is a metric on Θ. By applying the mean value theorem and the Cauchy-Schwartz inequality,

$$|\hat{Q}_n(Z,\theta^*) - \hat{Q}_n(Z,\theta^{**})| \leq \sup_{\bar{\theta}\in\Theta}\left\|\frac{\partial\hat{Q}_n(Z,\bar{\theta})}{\partial\theta'}\right\| \|\theta^* - \theta^{**}\|, \quad (A.4)$$

where $\bar{\theta}$ is the mean value between θ^* and θ^{**}.

Setting $\sup_{\bar{\theta}\in\Theta}\left\|\partial\hat{Q}_n(Z,\bar{\theta})/\partial\theta'\right\| = B(Z_i) < \infty$ by Lemma 3, and $h(e(\theta^*,\theta)) = \|\theta^* - \theta^{**}\|$, $\hat{Q}_n(Z,\theta)$ is Lipschitz in $\theta \in \Theta$. The stochastic equicontinuity is immediate from the Lipschitz condition (see Lemma 1 in Andrews 1992). Since $Q_0(Z,\theta)$ is continuous by Lemma 5, $\hat{Q}_n(Z,\theta) - Q_0(Z,\theta)$ is stochastically equicontinuous. Q.E.D.

Lemma 7 $\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(Z_i,\theta_0) \xrightarrow{d} N[0,S]$.

According to Bolthausen (1982), if (1) $\sum_{\omega=1}^{\infty}\omega^{d-1}\tilde{\alpha}_{k,l}(\omega)<\infty$, for $k+l\leq 4$; (2) $\tilde{\alpha}_{1,\infty}(\omega)=o(\omega^{-d})$; (3) for some $\delta>0$, $E[\|g(Z_i,\theta_0)\|]^{2+\delta}<\infty$ and $\sum_{\omega=1}^{\infty}\omega^{d-1}\tilde{\alpha}_{1,1}(\omega)^{\delta/(2+\delta)}<\infty$; and (4) S is nonsingular, then the distribution of $n^{-\frac{1}{2}}\sum_{i=1}^{n} g(Z_i,\theta_0)$ converges to $N[0,S]$. Although the mixing coefficient in Bolthausen (1982), namely, $\tilde{\alpha}_{k,l}(\omega)$, is defined in terms of a metric of maximum coordinate wise distance, the same arguments can be applied to the Euclidean metric (Conley 1999). In our case, since the dimension d is 2, the assumptions N.2 and N.3 correspond to the Euclidean version of (1) and (2), respectively. Since $E[g(Z_i,\theta_0)]$ is uniformly bounded, (3) holds (see Bolthausen 1982 Lemma 1). (4) is assumed. Q.E.D.

1.1 Proof of Proposition 1

As, for example, Newey and McFadden (1994) proves in Theorem 2.1, if there is a function $Q_0(Z,\theta)$ such that (1) Θ is compact; (2) $Q_0(Z,\theta)$ is uniquely minimized at $\theta_0\in\Theta$; (3) $Q_0(Z,\theta)$ is continuous; and (4) $\sup_{\theta\in\Theta}|\hat{Q}_n(Z,\theta)-Q_0(Z,\theta)|\xrightarrow{p}0$, then, $\hat{\theta}\xrightarrow{p}\theta_0$. Conditions (1) is assumed. Conditions (2) and (3) are satisfied by Lemma 5. Condition (4) holds if (a) Θ is compact; (b) $\hat{Q}_n(Z,\theta)\xrightarrow{p}Q_0(Z,\theta)$ for all $\theta\in\Theta$; and (c) $\hat{Q}_n(Z,\theta)-Q_0(Z,\theta)$ is stochastic equicontinuous for all $\theta\in\Theta$ (Andrews 1992, Newey 1991). Condition (b) is satisfied by Lemma 5; and (c) is by Lemma 6. Q.E.D.

1.2 Proof of Proposition 2

The first order condition from minimizing Eq.(10) is

$$\hat{G}_n(Z,\hat{\theta})'\hat{\Xi}\hat{g}_n(Z,\hat{\theta})=0, \tag{A.5}$$

where $\hat{G}_n(Z,\theta)=\partial\hat{g}_n(Z,\theta)/\partial\theta'$. For Eq.(A.5), applying the mean value theorem to $\hat{g}_n(Z,\hat{\theta})$ yields

$$\hat{g}_n(Z,\hat{\theta})=\hat{g}_n(Z,\theta_0)+\hat{G}_n(Z,\bar{\theta})(\hat{\theta}-\theta_0), \tag{A.6}$$

where $\bar{\theta}$ is the mean value between $\hat{\theta}$ and θ_0. By solving Eq.(A.6) for $(\hat{\theta}-\theta_0)$, noting that $\hat{g}_n(Z,\theta)=n^{-1}\sum_{i=1}^{n} g(Z_i,\theta)$, we obtain the familiar expression,

$$\sqrt{n}(\hat{\theta}-\theta_0)=-[\hat{G}_n(Z,\hat{\theta})'\hat{\Xi}\hat{G}_n(Z,\bar{\theta})]^{-1}\hat{G}_n(Z,\hat{\theta})'\hat{\Xi}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(Z_i,\theta_0). \tag{A.7}$$

Since $\bar{\theta}$ is between $\hat{\theta}$ and θ_0, and $\hat{\theta}\xrightarrow{p}\theta_0$, both $\hat{G}_n(Z,\hat{\theta})$ and $\hat{G}_n(Z,\bar{\theta})$ converge to $G\equiv E[\partial g(Z_i,\theta_0)/\partial\theta']$ in probability. Hence, by the assumption A.7 and Lemma 4, we obtain

$$\sqrt{n}(\hat{\theta}-\theta_0)\xrightarrow{p}-(G'\Xi G)^{-1}G'\Xi\sqrt{n}E[g(Z_i,\theta_0)]=0. \tag{A.8}$$

Thus, the conclusion follows by Lemma 7. Q.E.D.

Appendix 2. Spatial Mixing Condition

Suppose $(\Omega, \mathscr{F}, P)$ is a probability space. Let $\mathscr{T}_{\Lambda_\tau} \subset \mathscr{F}$ be the σ-algebra generated by a random field $\mathscr{X}_{s_i}$, $s_i \in \Lambda_\tau$, and let $|\Lambda_\tau|$ be the number of $s_i \in \Lambda_\tau$. Let $\Gamma(\Lambda_1, \Lambda_2)$ denote the minimum Euclidean distance from an element of Λ_1 to an element Λ_2. The mixing coefficient is defined as follows,

$$\alpha_{k,l}(\eta) \equiv \sup_{A \in \mathscr{T}_{\Lambda_1}, B \in \mathscr{T}_{\Lambda_2}} |P(A \cap B) - P(A)P(B)|,$$

and

$$|\Lambda_1| \leq k, |\Lambda_2| \leq l, \Gamma(\Lambda_1, \Lambda_2) \geq \eta.$$

The strong mixing condition requires $\alpha_{k,l}(\eta)$ to converge to zero as $\eta \to \infty$.

Appendix 3. Tables

Table A.1: Variables and their definitions

Variable	Definition
District-level variables ($n = 1573$)	
Density	Population density (Pop/km^2)
# Owner	Number of house owners (100 persons)
High-rise buil.	Ratio of residential buildings that have more than six stories to those with five or less
# Large bus.	Number of businesses that have more than 100 employees
# Retailers	Number of retailers per hectare
# Wholesalers	Number of wholesalers per hectare
Area of residence	Square root of the average area of residence ($\sqrt{m^2}$)
Dist. to Yokohama	Distance to the Yokohama railway station from the city district's mid-point (km)
Ratio of Workers	Ratio of workers to the district's population
Ratio of Managers	Ratio of managers to the total number of workers
Ratio of Blue Collar	Ratio of blue collar workers to the total number of workers
Ratio of Farmers	Ratio of farmers to the total number of workers
*Ward-level variables** (# wards = 18)	
Residential Zone	Ratio of land used as Exclusively Residential Zone for Low-rise Buildings Class 1 or 2** (%)
Industrial Zone	Ratio of land used as Industrial Zone or Exclusively Industrial Zone** (%)
Dummy Sakae	Dummy variable: 1 when the city district is in Sakae ward; 0 otherwise
Dummy Kohoku	Dummy variable: 1 when the city district is in Kohoku ward; 0 otherwise
Dummy Izumi	Dummy variable: 1 when the city district is in Izumi ward; 0 otherwise
Dummy Naka	Dummy variable: 1 when the city district is in Naka ward; 0 otherwise

* "Ward-level" refers to that the number of variations for such variables that only pertain to the number of wards. ** Exclusively Residential Zone for Low-rise Buildings Class 1, Class 2, Industrial Zone, and Exclusively Industrial Zone are 4 of the 12 zones defined in the zoning system, Zones for Certain Uses (*yoto chiiki* in Japanese) that is stipulated in the Town Planning Law.

Table A.2: Descriptive statistics

Variable	Mean	Std. deviation	Min	Max
Dependent variable	0.179	0.383	0	1
District-level variables				
Density	132.796	100.164	1.092	1660.054
# Owner	5.199	4.860	0	42.360
High-rise buil.	0.210	0.266	0	1
# Large bus.	5.360	11.817	0	188
# Retailers	1.600	3.044	0	6.000
# Wholesalers	0.596	1.354	0	12.710
Area of residence	8.350	1.633	0	12.211
Dist. to Yokohama	7.066	4.235	0.109	18.603
Ratio of Workers	0.480	0.209	0	0.995
Ratio of Managers	0.029	0.021	0	0.333
Ratio of Blue Collar	0.201	0.082	0	0.824
Ratio of Farmers	0.005	0.016	0	0.333
Ward-level variables				
Residential Zone	37.862	16.718	2.778	70.000
Industrial Zone	10.426	10.746	0	15.000
Dummy Sakae	0.036	0.185	0	1
Dummy Kohoku	0.062	0.241	0	1
Dummy Izumi	0.032	0.175	0	1
Dummy Naka	0.140	0.348	0	1

References

1. Andrews, D.W.K., 1992. Generic Uniform Convergence. *Econometric Theory* 8, 241-257.

2. Anselin, L., 1988. Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht.

3. Anselin, L., 2003. Spatial Econometrics. In: Baltagi, B.H. (Eds.), A Companion to Theoretical Econometrics, Blackwell Companions to Contemporary Economics, Oxford.

4. Anselin, L., Syabri, I., Kho, Y., 2006. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* 38, 5-22.

5. Beron, K.J., Vijverberg, W.P., 2002. Probit in A Spatial Context: A Monte Carlo Approach. In: Anselin, L., Florax, R.J., Rey, S.J. (Eds.), Advances in Spatial Econometrics. Springer, Heidelberg.

6. Bhat, C.R., Guo, J., 2004. A Spatially Correlated Logit Model: Formulation and Application to Residential Choice Modelling. *Transportation Research Part B* 38, 147-168.

7. Bolduc, D., Fortin, B., Gordon, S., 1997. Multinomial Probit Estimation of Spatially Interdependent Choices: An Empirical Comparison of Two New Techniques. *International Regional Science Review* 20, 77-101.

8. Bolthausen, E., 1982. On the Central Limit Theorem for Stationary Mixing Random Fields. *The Annals of Probability* 4, 1047-1050.

9. Brock, W.A., Durlauf, S.N., 2001. Discrete Choice with Social Interactions. *The Review of Economic Studies* 68, 235-260.

10. Brock, W.A., Durlauf, S.N., 2002. A Multinomial-Choice Model of Neighbourhood Effects. *American Economic Review* 92, 298-303.

11. Chiku keikaku kenkyukai [Research group on the district planning system], 2005. Kaisetsu & Jirei, Chiku Keikaku no Tebiki [Theoretical and Practical Guidance on the District Planning System], Gyosei.

12. Conley, T.G., 1999. GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics* 92, 1-45.

13. Cox, D.R., Snell, E.J., 1968. A General Definition of Residuals. *Journal of Royal Statistical Society B* 30, 248-275.

14. Fleming, M., 2002. Techniques for Estimating Spatially Dependent Discrete Choice Models. In: Anselin, L., Florax, R.J., Rey, S.J. (Eds.), Advances in Spatial Econometrics. Springer, Heidelberg.

15. Gourieroux, C., Monfort, A., Renault, E., Trognon, A., 1987. Generalized Residuals. *Journal of Econometrics* 34, 5-32.

16. Hansen, L.P., 1982. Large Sample Properties of Generalized Method of Moments Estimators. *Econometrica* 50, 1029-1054.

17. Inkmann, J., 2000. Misspecified Heteroskedasticity in the Panel Probit Model: A Small Sample Comparison of GMM and SML Estimators. *Journal of Econometrics* 97, 227-259.

18. Klier, T., McMillen, D., 2005. Clustering of Auto Supplier Plants in the U.S.: GMM Spatial Logit for Large Samples. *Working Paper Series* WP-05-18, Federal Reserve Bank of Chicago.

19. LeSage, J.P., 2000. Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. *Geographical Analysis* 32, 19-35.

20. McMillen, D.P., 1992. Probit with Spatial Autocorrelation. *Journal of Regional Science* 3, 335-348.

21. Newey, W.K., 1991. Uniform Convergence in Probability and Stochastic Equicontinuity. *Econometrica* 59, 1161-1167.

22. Newey, W.K., 1993. Efficient Estimation of Models with Conditional Moment Restrictions. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), Handbook of Statistics Vol. 11, North-Holland.

23. Newey, W.K., McFadden, D., 1994. Large Sample Estimation and Hypothesis Testing. In: Engle, R.F., McFadden, D. (Eds.), Handbook of Econometrics Vol. 4, North-Holland.

24. Pinkse, J., Slade, M., 1998. Contracting in Space: An Application of Spatial Statistics to Discrete Choice Models. *Journal of Econometrics* 85, 125-154.

25. Rosenblatt, M., 1978. Dependence and Asymptotic Independence for Random Process. In: Rosenblatt, M. (Eds.), Studies in Probability Theory, Mathematical Association of America, Washington, D.C..

26. White, H., 2001. Asymptotic Theory for Econometricians: Revised Edition. Academic Press, Orlando, Florida.

27. Wilson, J.Q., and Kelling, G.L., 1982. Broken Windows: The Police and Neighborhood Safety. *The Atlantic Monthly* March, 29-38.

28. Yatchew, A., Griliches, Z., 1985. Specification Error in Probit Models. *Review of Economics and Statistics* 18, 134-139.