Smoothed Spatial Maximum Score Estimation of Spatial Autoregressive Binary Choice Panel Models*

Jinghua Lei †

May 29, 2014

Abstract

This paper considers spatial autoregressive (SAR) binary choice models in the context of panel data with correlated random effects, where the latent dependent variables are spatially correlated and the individual effects are assumed to be stationary. Without imposing any parametric structure of the error terms, this paper proposes a smoothed spatial maximum score (SSpMS) estimator which consistently estimates the model parameters up to scale. The identification of parameters is obtained when the disturbances are time-stationary and the explanatory variables vary enough over time, along with an exogenous and time-invariant spatial weight matrix. Consistency and asymptotic distribution of the proposed estimator are also derived in this paper. Finally, a Monte Carlo study indicates that the SSpMS estimator performs quite well in finite samples.

JEL classification: C14 C21 C23 C25 R15

Keywords: Spatial Autoregressive Models, Binary Choice, Correlated Random Effects, Maximum Score Estimation.

*I am very grateful to my advisor Pavel Čížek for his enormously valuable advice, guidance and support. I also benefited from comments by participants at the 7-th World Conference of Spatial Econometrics Association (Washington DC, July 2013) and Econometrics Seminar at Tilburg University (Dec 2013). All errors are my responsibility.

†CentER, Department of Econometrics and Operation Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. Phone: +31-13-466-3399, E-mail: J.Lei@uvt.nl
1 Introduction

This paper considers a spatial autoregressive binary choice panel model of the form

\[ y_{it,n}^* = \lambda_0 \sum_{j=1}^{n} w_{ij,n} y_{jt,n}^* + x_{it,n} \beta_0 + \alpha_{i,n} + \epsilon_{it,n}, \quad i = 1, \ldots, n, \ t = 1, \ldots, T, \]  

(1)

where \( y_{it,n}^* \) is the latent dependent variable that links to the observed binary outcome \( y_{it,n} \) such that \( y_{it,n} = 1 \) if \( y_{it,n}^* > 0 \) and \( y_{it,n} = 0 \) otherwise. \( x_{it,n} \) are observed exogenous regressors for the individual \( i \) in the \( t \)-th period and \( W_n = (w_{ij,n}) \) is an \( n \times n \) time-invariant exogenous spatial weight matrix whose elements are all nonnegative. As usual, a proper normalization has \( w_{ii,n} = 0 \) for all \( i \). \( \lambda_0 \) is a parameter to capture the spatial effect, \( \alpha_{i,n} \) is the individual effect which is unobserved and allowed to be correlated with the regressors in an arbitrary way, and \( \epsilon_{it,n} \) is the idiosyncratic individual error term. This spatial model is an equilibrium model with endogenous interaction effects among the unobserved dependent variable.

The spatial binary choice model is useful when a researcher allows the outcome of an agent’s decision to be determined with that of neighboring agents. It has been increasingly used in spatial econometrics literature, especially for the spatial probit model. The most popular specification is the spatial lag probit model, which is the cross-sectional version of model (1). Many studies have considered this model from a methodological viewpoint: \cite{McMillen1992, LeSage2000, PaceLeSage2011} among others, there are also some empirical studies using this model, such as \cite{Beron2003} and \cite{MukherjeeSinger2008}. Moreover, \cite{KlierMcMillen2008} replace the probit by the logit specification. Most recently, \cite{QuLee2012} and \cite{Elhorst2013} conduct an important variant of the cross-sectional spatial lag probit model in the following form:

\[ y_{i,n}^* = \lambda_0 \sum_{j=1}^{n} w_{ij,n} y_{j,n} + x_i \beta_0 + \epsilon_{i,n}, \]

where the latent dependent variable \( y_{i,n}^* \) depends on observed choices represented by \( \sum_{j=1}^{n} w_{ij,n} y_{j,n} \) rather than unobserved ones.

Another specification is a linear regression model with spatially correlated errors:

\[ y_{i,n}^* = x_i \beta_0 + v_i, \quad v_i = \rho_0 \sum_{j=1}^{n} w_{ij,n} v_{j,n} + \epsilon_{i,n}, \]

where \( v_i \) reflects the spatially correlated errors with coefficient \( \rho_0 \), and \( \epsilon_{i,n} \) follows a normal distribution. In this model, the variance of

Note that \( x_{it,n} \) consists of time-varying covariates, as any time-invariant covariate would be absorbed into the individual effect \( \alpha_{i,n} \).
errors is usually normalized to one, as it cannot be separately identified with the parameter \( \beta_0 \). This spatial error probit model has been studied by Beron and Vijverberg (2004), Fleming (2004), Klier and McMillen (2008), Wang et al. (2013) among others. Moreover, Bolduc et al. (1997) consider the logit specification in their empirical application such that the probability of \( \Pr(y = 1) \) has an analytical solution.

The main assumption of existing estimation methods for the spatial binary choice models is that distribution of \( \epsilon_{i,n} \) conditional on \( \{x_{i,n}\}_{i=1}^{n} \) is known up to a finite set of parameters. For example, it is often assumed that \( \epsilon_{i,n} \) has either a normal or a logistic distribution. However, when the distribution of \( \epsilon_{i,n} \) is misspecified, estimation methods that require specifying the distribution of \( \epsilon_{i,n} \) yield inconsistent estimators. Furthermore, even if the model is correctly specified, likelihood based estimation methods may suffer from the multi-dimensional integration problem as the individual error terms are dependent on each other. Many attempts have been made to solve this problem, see Elhorst et al. (2013) for a careful review.

Moreover, estimation would become much more difficult if a context of panel data with fixed effects or correlated random effects is considered. Even if the distribution of the errors is correctly specified and there is no spatial dependence, consistently estimating parameters in binary choice panel models with fixed effect requires clever estimators, such as conditional logit estimation (Chamberlain, 1984) or maximum score estimator (Manski 1987). These methods could either generate a conditional likelihood function without fixed effects or eliminate the fixed effects by some relationship based on the expectation of dependent variables. However, to my best knowledge, whether these methods still work or not when there is spatial dependence is still unknown, and so far there have been no formal studies on the asymptotic properties of the estimators of model \( (1) \).

In this paper, I consider a correlated random effects spatial autoregressive (SAR) binary choice model \( (1) \), where the only assumption imposed on the errors is time stationarity rather than any parametric assumption. Based on this assumption and the exogeneity of a time-invariant spatial weight matrix, a similar condition to Lemma 1 in Manski (1987) is derived in this paper. Therefore, a spatial maximum score estimator is defined, analogous to that of Manski (1987), which can be smoothed by replacing the sign function with a
continuous function, as in Horowitz (1992). The proposed smoothed spatial maximum score (SSpMS) estimator is currently the only available consistent estimator for model (1). Although the SSpMS estimator cannot be extended to cross-sectional SAR binary choice models, it is applicable to correlated random effects SAR binary choice models with arbitrarily spatial correlation in the errors—when such spatial correlation is time-invariant and satisfies some "fading memory" property as described in section 3.2. Finally, the SSpMS estimator is a general estimator that includes Charlier et al. (1995)'s smoothed maximum score estimator without spatial correlation as a special case.

The rest of the paper is organized as follows. Section 2 provides the model specification and the suggested SSpMS estimator. Section 3 proves identification, consistency and asymptotic normality of the proposed estimator. Section 4 presents the results of a Monte Carlo investigation of finite-sample properties of the estimators and Section 5 concludes. All the proofs are provided in the appendices.

At this point, it is convenient to introduce some notation. To write model (1) in matrix notation, we denote

$$Y_{nt}^* = (y_{1t,n}, \ldots, y_{nt,n})^\top, Y_{nt} = (y_{1t,n}, \ldots, y_{nt,n})^\top, X_{nt} = (x_{1t,n}, \ldots, x_{nt,n})^\top, X = (X_{n1}, \ldots, X_{nT}), \alpha_n = (\alpha_{1,n}, \ldots, \alpha_{n,n})^\top, \text{ and } \epsilon_{nt} = (\epsilon_{1t,n}, \ldots, \epsilon_{nt,n})^\top.$$  

Denote $S_n(\lambda) = I_n - \lambda W_n$, and $S_n = I_n - \lambda_0 W_n$, where $I_n$ is the identity matrix.

2 Description of the Estimator

To motivate the estimator, we consider the simplest case when there are only two time periods ($t = 1, 2$). Suppose that the inverse of matrix $S_n$ exists, rearrange equation (1) and rewrite it in matrix notation. The equilibrium vector $Y_{nt}^*$ is then

$$Y_{nt}^* = (I_n - \lambda_0 W_n)^{-1}(X_{nt}\beta_0 + \alpha_n + \epsilon_{nt}) = S_n^{-1}X_{nt}\beta_0 + S_n^{-1}\alpha_n + S_n^{-1}\epsilon_{nt}. \quad (2)$$

Denote $\tilde{\epsilon}_{nt} = S_n^{-1}\epsilon_{nt}$, $\tilde{\epsilon}_{nt}$ is an $n \times 1$ vector of linear combinations of the error terms for all individuals. Let $e_{i,n}$ denote an $n \times 1$ vector with the $i$-th element equal to one and all other elements equal to zero, then $\tilde{\epsilon}_{it,n} = e_{i,n}^\top S_n^{-1}\epsilon_{nt}$ is the $i$-th element of $\tilde{\epsilon}_{nt}$. Under the conditional stationarity assumption that $\epsilon_{n1}$ and $\epsilon_{n2}$ are identically distributed conditional on $(\alpha_n, X)$, we know that $\tilde{\epsilon}_{n1}$ and $\tilde{\epsilon}_{n2}$ also have the same distribution. Therefore, we obtain the following relationship for each individual $i$, which is similar to Lemma 1 of
\[ E[y_{1,n} - y_{2,n} | \alpha_n, X] > 0 \quad \text{if and only if} \quad e_{i,n}^T S_n^{-1} X_{1,n} \beta_0 > e_{i,n}^T S_n^{-1} X_{2,n} \beta_0, \]

\[ E[y_{1,n} - y_{2,n} | \alpha_n, X] = 0 \quad \text{if and only if} \quad e_{i,n}^T S_n^{-1} X_{1,n} \beta_0 = e_{i,n}^T S_n^{-1} X_{2,n} \beta_0, \quad (3) \]

\[ E[y_{1,n} - y_{2,n} | \alpha_n, X] < 0 \quad \text{if and only if} \quad e_{i,n}^T S_n^{-1} X_{1,n} \beta_0 < e_{i,n}^T S_n^{-1} X_{2,n} \beta_0. \]

Similar to Manski (1987) Lemma 3, under Assumption I, \footnote{A similar result as Corollary of Manski (1987) could also be obtained immediately as}
nos comment

\[ G_{i,n}(\theta) = E[\Delta y_{i,n} \operatorname{sgn}\{e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta\}], \quad i = 1, \ldots, n, \]

where \( \theta = (\lambda, \beta^\top) \), \( \Delta y_{i,n} = (y_{1,n} - y_{2,n}) \), \( \Delta X_n = X_{1,n} - X_{2,n} \), and \( \operatorname{sgn}(x) = 1 \) if \( x \geq 0 \) and \(-1\) otherwise. Apparently, \( \theta_0 \) is also the unique maximizer of the average of \( G_i(\theta) \), that is

\[ \theta_0 = \arg\max_{\theta} \frac{1}{n} \sum_{i=1}^{n} G_{i,n}(\theta) = \arg\max_{\theta} \frac{1}{n} \sum_{i=1}^{n} E[\Delta y_{i,n} \operatorname{sgn}\{e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta\}]. \]

A consistent estimator of the parameter \( \theta_0 \) can be obtained by maximizing the following objective function:

\[ G_n^*(\theta) = \frac{1}{n} \sum_{i=1}^{n} \Delta y_{i,n} \operatorname{sgn}\left(e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta\right). \]

Observe that the behavior of \( G_n^*(\cdot) \) is unaffected by removing observations having \( y_{1,n} = y_{2,n} \), thus, the estimator maximizing \( G_n^*(\cdot) \) is a conditional maximum score estimator. However, it is difficult to derive its asymptotic distribution as the score function is a step function. Chamberlain (1986) has shown that there is no \( n^{-1/2} \)-consistent estimator of \( \beta_0 \) under Manski’s assumptions. Horowitz (1992) then modifies Manski’s maximum score estimator (Manski 1985), under somewhat stronger but still very weak assumptions, by smoothing the score function to be continuous and differentiable, and shows that the...
convergence rate of the smoothed maximum score estimator is at least as fast as $n^{-2/5}$ and, depending on how smooth the distribution of $\epsilon_{i,n}$ and $x_{it,n}\beta_0$ are, can be arbitrarily close to $n^{-1/2}$.[3] In the context of panel data models with fixed effects, Charlier et al. (1995) investigate the smoothed version of Manski (1987)'s estimator and indicate that maximizing $G^*_n(\theta)$ boils down to maximizing

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{2} \Delta y_{i,n} \left[ \text{sgn} \left( e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta \right) + 1 \right] = \frac{1}{n} \sum_{i=1}^{n} \Delta y_{i,n} \mathbb{I} \{ e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta \geq 0 \} \quad (6)$$

This objective function can then be smoothed by

$$G_n(\theta; \sigma_n) = \frac{1}{n} \sum_{i=1}^{n} \Delta y_{i,n} K \left( \frac{e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta}{\sigma_n} \right). \quad (7)$$

where $K(\cdot/\sigma_n)$ is some smooth function that converges to the indicator function as $n \to \infty$, and $\sigma_n$ is a sequence of strictly positive real numbers satisfying $\lim_{n \to \infty} \sigma_n = 0$. Note that $K(\cdot/\sigma_n)$ could be a cumulative distribution function such as the cumulative standard normal distribution function $\Phi(\cdot/\sigma_n)$, as in Horowitz (1992), and $\sigma_n$ can be viewed as the bandwidth.

**Remark 1.** Apparently, when there is no spatial effect ($\lambda_0 = 0$), then the SSpmS estimator that maximizes equation (7) degenerates to the standard smoothed maximum score estimator for panel models, as in Charlier et al. (1995).

Moreover, the identification and estimation strategy described above also works for models with arbitrarily spatially correlated errors, if the spatial correlation is time stationary and satisfies the "fading memory" property as stated in the next section. For example, the mixed spatial lag and spatial error binary choice models with correlated random effects:

$$Y^*_n = \lambda_0 W_{n,1} Y^*_n + X_n \beta_0 + \alpha_n + v_{nt}, \quad v_{nt} = \rho_0 W_{n,2} v_{nt} + \epsilon_{nt}. \quad (8)$$

Suppose that the inverse of matrices $S_{n1} = (I - \lambda_0 W_{n,1})$ and $(I - \rho_0 W_{n,2})$ exist and the spatial weight matrices $W_{n,1}$ and $W_{n,2}$ are time-invariant. Rearranging the above equation

[3] Recently, Jun et al. (2013) propose a classical Laplace estimator as alternative that provides a unified method of smoothing for a large class of $\sqrt{n}$ consistent estimators and which can have computational advantages, e.g. in the maximum score case.
and rewriting it in matrix notation, we have

\[ Y_{nt}^* = (I_n - \lambda_0 W_{n,1})^{-1} [X_{nt} \beta_0 + \alpha_n + (I_n - \rho_0 W_{n,2})^{-1} \epsilon_{nt}] \]

\[ = S_n^{-1}(X_{nt} \beta_0 + \alpha_n) + S_n^{-1}(I_n - \rho_0 W_{n,2})^{-1} \epsilon_{nt}. \]

As in equation (3), when there are only two time periods, \((t = 1, 2)\) and \(\epsilon_{n1}\) and \(\epsilon_{n2}\) are identically distributed conditional on \((\alpha_n, X)\), then \(\bar{\epsilon}_{n1}\) and \(\bar{\epsilon}_{n2}\) also have the same distribution, where \(\bar{\epsilon}_{nt} = e_i^\top S_n^{-1}(I_n - \rho_0 W_{n,2})^{-1} \epsilon_{nt}\). Therefore, we could also obtain the same relationship for each individual \(i\) as in equation (3). The identification and estimation strategy will be exactly the same as I discussed previously, however, the parameter \(\rho_0\) cannot be estimated in this case.

In addition, when \(\beta_0 = 0\), the model degenerates to a spatial binary choice models without covariates. In this case, the spatial effect \(\lambda_0\) is not identified without imposing additional assumptions on the error terms. Another point that we should notice is that the identification strategy described in this paper cannot be applied to the cross sectional spatial binary choice models directly.

Finally, when the time periods are more than two, but finite, the SSpMS estimator can be defined analogously to that of [Charlier et al. (1995)] by

\[ \hat{\theta}_{nT} = \arg \max_\theta \frac{1}{nT(T - 1)} \sum_{i=1}^n \sum_{s < t} c_{its}(y_{it,n} - y_{is,n})K\left( \frac{e_i^\top S_n^{-1}(\lambda)(X_{nt} - X_{ns}) \beta}{\sigma_n} \right), \]

where \(c_{its} = r_{it}r_{is}\), with \(r_{it} = 1\) if \(\{(y_{it,n}, x_{it,n})\}_{i=1, t=1}^{i=n, t=T}\) is observed, and zero otherwise. Therefore, \(c_{its} = 1\) if both \(\{(y_{it,n}, x_{it,n})\}\) and \(\{(y_{is,n}, x_{is,n})\}\) are observed and zero otherwise. The inclusion of \(c_{its}\) is to make the SSpMS estimator applicable to an unbalanced panel, which is common in applications.

**Remark 2.** Note that \(G_n^*(\theta)\) is equivalent to the absolute loss objective function

\[ \min_\theta \frac{1}{n} \sum_{i=1}^n \left| \Delta y_{i,n} - \text{sgn}\left(e_i^\top S_n^{-1}(\lambda)\Delta X_n \beta\right) \right| \cdot 1\{y_{i1,n} \neq y_{i2,n}\}, \]

and the squared loss objective function

\[ \min_\theta \frac{1}{n} \sum_{i=1}^n \left[ \Delta y_{i,n} - \text{sgn}\left(e_i^\top S_n^{-1}(\lambda)\Delta X_n \beta\right) \right]^2 \cdot 1\{y_{i1,n} \neq y_{i2,n}\}. \]
Motivated by Khan (2012), when a standard normal distribution $\Phi(\cdot)$ is applied for smoothing the objective function and observations with $\{y_{i1,n} = y_{i2,n}\}$ are included to improve efficiency, we can define a spatial nonlinear least square (SpNLLS) probit estimator as

$$\hat{\theta}_{SpNLLS} = \arg\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \left[ \Delta y_{i,n} - \Phi \left( \frac{e_{i,n}^\top S_n^{-1}(\lambda)\Delta X_n\beta}{\sigma_n} \right) \right]^2.$$

The main advantage of this procedure is that the standard NLLS objective function can be extended to the case with spatial correlation, and the standard software packages, such as Stata, can be easily adjusted to compute the SpNLLS probit estimator.

3 Identification and Asymptotic Properties

3.1 Identification

In this subsection, identification of parameters in model (1) is provided, and the definition of identification is similar as in Manski (1987). Consider $(\lambda, \beta^\top) \in \Lambda \times \mathbb{R}^q$, $(\lambda, \beta^\top)^T \neq (\lambda_0, \beta_0^\top)^T$, condition (3) (which holds for each individual $i$) distinguishes $(\lambda, \beta^\top)^T$ from $(\lambda_0, \beta_0^\top)^T$ if there exists a set of $\Delta X_n$ values having positive $F_{\Delta X_n}$ probability such that condition (3) does not hold for some $i$, when $(\lambda, \beta^\top)^T$ is substituted for $(\lambda_0, \beta_0^\top)^T$. Let

$$V_{i,(\lambda, \beta)} = \left\{ \Delta X_n \in \mathbb{R}^q : \text{sgn}(e_{i,n}^\top S_n^{-1}(\lambda)\Delta X_n\beta) \neq \text{sgn}(e_{i,n}^\top S_n^{-1}\Delta X_n\beta_0) \right\},$$

then $(\lambda_0, \beta_0^\top)^T$ is identified relative to $(\lambda, \beta^\top)^T$ if there is some $i = 1, \ldots, n$, such that

$$R_i(\lambda, \beta) = \int_{V_{i,(\lambda, \beta)}} dF_{\Delta X_n} > 0.$$

**Assumption 1.**

i). $F_{e_{i1}|a_n, X} = F_{e_{i2}|a_n, X}$ for all $i$ and $(a_n, X)$.

ii). The support of $F_{e_{i1}|a_n, X}$ is $\mathbb{R}$ for all $i$ and $(a_n, X)$.

**Assumption 2.**

i). The support of $F_{\Delta X_n}$ is not contained in any proper linear subspace of $\mathbb{R}^q$.

ii). There exists at least one $q' \in [1, 2, \ldots, q]$ such that $\beta_{0,q'} \neq 0$, and for almost every value of $\Delta \tilde{x}_{i,n} = (\Delta x_{i,1,n}, \Delta x_{i,2,n}, \ldots, \Delta x_{i,q'-1,n}, \Delta x_{i,q'+1,n}, \ldots, \Delta x_{i,q,n})$, the scalar random variable $\Delta x_{i,q',n}$ has everywhere positive Lebesgue density conditional on $\Delta \tilde{x}_{i,n}$ for all
Assumption 3. The matrix $S_n(\lambda) = I_n - \lambda W_n$ is nonsingular for all $\lambda \in \Lambda$.

Assumptions 1 and 2 have the same forms as Assumptions 1 and 2 in Manski (1987), except that we have different conditionings. As in assumption 1 i), $\epsilon_{it,n}$ is stationary not only conditional on its own characteristics, but also conditional on the characteristics of other individuals. Such conditioning also appears in Assumption 2, and is necessary because there is spatial correlation between individuals, which is obvious if we assume the process \{\(x_{it,n}, \alpha_{i,n}, \epsilon_{it,n}\}\} is strong mixing as in section 3.2. Assumption 1 ii) guarantees that the event $y_{i1,n} \neq y_{i2,n}$ occurs with positive probability for all $\alpha_n$. Assumption 2 i) is the familiar full-rank condition that prevents a global failure of identification, and part ii) is a substantive restriction, which implies that $\Delta X_n \beta$ has everywhere positive density for all $\beta$ such that $\beta_q' \neq 0$. Assumption 3 guarantees that the system (1) has an equilibrium and matrix $S_n(\lambda)$ is invertible.

Assumption 4. For $\lambda \neq \lambda_0$ and $\beta \neq 0$, it holds that

$$\liminf_{n \to \infty} \inf_{\lambda \in \Lambda} \frac{1}{n} \sum_{i=1}^{n} \mathbb{P} \left\{ \left[ e_{i,n}' S_n^{-1}(\lambda) - e_{i,n}' S_n^{-1} \right] \Delta X_n \beta \neq 0 \right\} > 0.$$

Although for any fixed sample size $n$, identification only requires that there is at least one $i$ satisfying condition (9), identification may fail for some extreme spatial weight matrices when $n$ tends to infinity. For instance, a trivial case is that identification of $\lambda_0$ fails if $W_n = 0$. Therefore, failure of identification of $\lambda_0$ may occur if the matrix $W_n$ becomes sufficiently sparse as $n \to \infty$. A sufficient technical condition to prevent identification failure is Assumption 4 which precludes the case that the fraction of individuals, with $e_{i,n}' S_n^{-1}(\lambda)$ and $e_{i,n}' S_n^{-1}$ have the same elements for some $\lambda \neq \lambda_0$, decays to zero as $n$ tends to infinity. Note that $\beta \neq 0$ is excluded in Assumption 4 otherwise, $\lambda_0$ is not identified as described in remark 1.

Identification of $\theta_0$ requires that there is a positive probability such that $e_{i,n}' S_n^{-1}(\lambda) \Delta X_n \beta$ has a different sign with $e_{i,n}' S_n^{-1} \Delta X_n \beta_0$. As Assumption 2 imposes no condition on the parameter vector $\theta_0$ except that $\beta_0,q' \neq 0$, it is possible for $e_{i,n}' S_n^{-1}(\lambda) \Delta X_n \beta$ to have bounded support for all $\theta$, given sharper bounds on $\theta_0$. Therefore, $\theta_0$ is identified, which is stated in
the following Lemma. Clearly, the scale of $\beta_0$ is not identified. To see this, we can simply set $\lambda = \lambda_0$, then the identification problem degenerates to that of Manski (1987). As in estimation, the usual way is to normalize $\beta_{0,q'} = 1$ such that the other elements of $\beta_0$ are point identified, we only consider the case $\beta_q \neq 0$.

**Lemma 1.** Under Assumptions 1-4, $(\lambda_0, \beta_0^\top)^\top$ is identified relative to $(\lambda, \beta^\top)^\top \in \Lambda \times \mathbb{R}^q$, where $\beta/||\beta|| \neq \beta_0/||\beta_0||$ and $\beta_q \neq 0$.

### 3.2 Consistency

In this subsection, consistency of estimator that maximizes the objective function (7) is established. The main difficulty to prove the consistency is that the objective function (7) is based on a dependent and heterogeneous process. Therefore, some appropriate "fading memory" property must be guaranteed to support laws of large numbers and uniform laws of large numbers, and the "fading memory" property in this paper is near epoch dependence, which is defined in Definition 1.

To proceed, we need to first define the space and metric (which are not restricted to physical space and distance) for the convenience of analyzing the spatial correlation structure. Following the setting in Jenish and Prucha (2009, 2012) on the development of statistical theory for spatial mixing and near epoch dependent (NED) processes, we list the following assumption.

**Assumption 5.** Individual units in the economy are located or living in a region $D_n \subset D \subset \mathbb{R}^d$, where the cardinality of $D_n$ satisfies $\lim_{n \to \infty} |D_n| = \infty$. The distance $d(l_i, l_j)$ between any two different individuals $i$'s location $l_i$ and $j$'s location $l_j$ is larger than or equal to a specific positive constant, without loss of generality, say, 1.

The assumption of a minimum distance ensures the growth of the sample size as the sample regions $D_n = \{l_1, \ldots, l_n\} \subset D$ expand, which means the asymptotic method in this paper is increasing domain asymptotic rather than infilled asymptotic.

The models considered in this paper are actually the Cliff and Ord (1981) type model, which is one of the common approaches to model cross-sectional dependence in the econometrics literature. In the Cliff-Ord type models, the spatial weights $w_{ij,n}$ depend on some

---

4This assumption can also be found in Xu and Lee (2014) Assumption 1.
measure of distance and decline as the distance increases. Under Assumption 3, model (1) is then
\[ y_{it,n} = 1 \{ e_i^\top S_n^{-1} X_{nt} \beta_0 + \alpha_i + \epsilon_{nt} \} > 0, \]
where \( e_i^\top S_n^{-1} \) could be denoted by a vector \( (a_{i1,n}, \ldots, a_{in,n}) \). Although the output process \( y_{it,n} \) only depends on a finite number of elements of the input process \( \eta_{it,n} = (x_{it,n}, \alpha_i, \epsilon_{it,n})^\top \) for fixed \( n \), the mixing property of \( \eta_{it,n} \) may not carry over to \( y_{it,n} \). The reason is that the number of elements composing the spatial lags grows unboundedly with the sample size so that the mixing property can break down in the limit. This is especially important when analyzing the asymptotic properties of Cliff-Ord type processes. Therefore, towards establishing that \( \{y_{it,n}, l_i \in D_n\} \) is NED on \( \{\eta_{it,n}, l_i \in D_n\} \), we maintain the following assumptions:

\[
\lim_{d \to \infty} \sup_n \sum_{1 \leq i \leq n, 1 \leq j \leq n, d(l_i, l_j) > d} |a_{ij,n}| = 0 \tag{10}
\]

and

\[
\sup_n \sup_{1 \leq i \leq n, t} \|\eta_{it,n}\|_p < \infty \quad \text{for some} \quad p \geq 1. \tag{11}
\]

Jenish and Prucha (2012) show that a sufficient condition for (10) is that for some \( \gamma > 0, \)

\[
\sup_n \sum_{j=1}^n |a_{ij,n}| d(l_i, l_j)^\gamma < \infty.
\]

A similar condition has been used recently by Kelejian and Prucha (2007), and should be satisfied in a wide range of applications. It is slightly stronger than the typical assumption in the Cliff-Ord literature which imposes that the row and column sums of the absolute elements of the matrix \( S_n^{-1} \) are uniformly bounded as in Assumption 7 ii).

Now I define the near epoch dependence of random variables \( y_{it,n} \) based on a process of random variables \( \eta_{it,n} \) as follows:

**Definition 1.** Random variables \( y_{it,n}, i = 1, \ldots, n, t = 1, 2 \) are called near epoch dependent (NED) on \( \eta_{it,n} \) if

\[
\sup_i \|y_{it,n} - E(y_{it,n}|z_{i,n}(m))\|_2 = d_t \nu(m) \to 0, \quad \text{as} \quad m \to \infty \tag{12}
\]

\(^5\)Working paper version, available on author’s website.
where \( d_t \) is a sequence of positive constants, \( \nu(m) \geq 0 \) with \( \lim_{m \to \infty} \nu(m) = 0 \), and \( \mathcal{D}_{i,n}(m) = \sigma(\eta_{it,n} : d(l_i, j) \leq m) \) is the \( \sigma \)-field generated by the random variables \( \eta_{it,n} \) located in the \( m \)-neighborhood of location \( l_i \).

The idea behind the near epoch dependence condition is that given the \( m \)-neighborhood of input variables \( \eta_{it,n}, y_{it,n} \) should be predictable up to arbitrary accuracy. That is, the approximation error declines "sufficiently fast" as the conditioning set of input variables expands. The base process \( \eta_{it,n} \) needs to satisfy a condition such as strong or uniform mixing or independence.

**Assumption 6.** \( \{\eta_{it,n}\}, \ i = 1, \ldots, n, \ t = 1, 2, \) is a strictly stationary strong mixing process with \( \alpha \)-mixing coefficient \( \alpha(m) \).

**Proposition 1.** Under Assumptions 1, 3-6 and conditions (10)-(11), the process \( \{y_{it,n}\}, \ \{\Delta y_{it,n}\}, \) and \( \{\text{sgn}(e_{i,n}^\top S_n^{-1}(\lambda) \Delta X_n / \beta)\} \) are uniformly NED on the process \( \{\eta_{it,n}\} \).

**Remark 3.** Proposition 1 shows that \( y_{it,n} \) is a sequence of \( 0/1 \) valued random variable that is near epoch dependent on \( \eta_{it,n} \). Then \( (y_{it,n}, \eta_{it,n}) \) is strong mixing by Theorem 2 of de Jong and Woutersen (2011), and the mixing property of \( (y_{it,n}, \eta_{it,n}) \) will be used in the proofs for consistency and asymptotic normality of the smoothed spatial maximum score estimator. Although \( \{y_{it,n}\} \) is strong mixing, it is not stationary as the inverse spatial weights \( e_{i,n}^\top S_n^{-1}(\lambda) \) are different for each individual \( i \) in general. One example for \( \{y_{it,n}\} \) to be stationary is where the spatial correlation only exists within groups of the same size, and equal weights are assigned for individuals in the same group.

**Assumption 7.**

1. \( |\beta_{0,q}| = 1 \) and \( \tilde{\beta}_0 = (\beta_{0,1}, \ldots, \beta_{0,q-1})^\top \) is contained in a compact subset \( B \) of \( \mathbb{R}_q^{-1} \);
2. The sequence \( \{W_n\} \) and \( \{S_n^{-1}\} \) are uniformly bounded in both row and column sums;\(^7\)

\(^6\)The definition of \( \alpha \)-mixing can be found in Definition 1 of Jenish and Prucha (2009). For \( U \subseteq D_n \) and \( V \subseteq D_n \), let \( \mathfrak{D}_n(U) = \sigma(\eta_{it,n} : i \in U) \), \( \mathfrak{D}_n(U, V) = \alpha(\mathfrak{D}_n(U), \mathfrak{D}_n(V)) \). Then the \( \alpha \)-mixing coefficient is defined as: \( \alpha_{\alpha, \mathfrak{D}_n}(r) = \sup(\alpha_{\mathfrak{D}_n}(U, V), |U| \leq k, |V| \leq l, d(U, V) \geq r) \), with \( k, l, r, n \in \mathbb{N} \).

\(^7\) The notions of uniform boundedness can be defined in terms of some matrix norms: the maximum column matrix norm \( \|\cdot\|_1 \) of a \( n \times n \) matrix \( A = (a_{ij}) \) is defined as \( \|A_n\|_1 = \max_{1 \leq i \leq n} \sum_{j=1}^n |a_{ij}| \), and the maximum row sum matrix norm \( \|\cdot\|_\infty \) is \( \|A_n\|_\infty = \max_{1 \leq j \leq n} \sum_{i=1}^n |a_{ij}| \) (see Horn and Johnson (1985), pp.294-295). The uniformly boundedness of \( \{A_n\} \) in (resp. row) sums is equivalent to the sequence \( \{|\|A_n\|_1\|_{\infty}\} \) (resp. \( \{|\|A_n\|_{\infty}\|_1\} \) ) being bounded.

Lemma A.2 of Lee (2004) shows that, for any weights matrix, \( \|\lambda_0 W_n\|_1 < 1 \) and \( \|\lambda_0 W_n\|_{\infty} < 1 \) for all \( n \), are sufficient conditions for \( S_n^{-1} \) to be uniformly bounded in both row and column sums.

Because a matrix norm \( \|\cdot\| \) has the submultiplicative property that \( \|A_n B_n\| \leq \|A_n\| \cdot \|B_n\| \), Assumption 7 guarantees that products of matrices in our analysis such as \( S_n^{-1} W_n S_n^{-1} \) and \( S_n^{-1} W_n S_n^{-1} W_n S_n^{-1} \), etc., will be uniformly bounded in row and column sums.
iii). \( \{S_n^{-1}(\lambda)\} \) are uniformly bounded in either row or column sums, uniformly in \( \lambda \) in a compact parameter space \( \Lambda \). The true parameter \( \lambda_0 \) is in the interior of \( \Lambda \).

The uniform boundedness condition of \( S_n^{-1} \) in Assumption 7 ii) implies that \( S_n^{-1}(\lambda) \) are uniformly bounded in both row and column sums uniformly in a neighborhood of \( \lambda_0 \) (Lee, 2004). Assumption 7 i) and iii) are needed to deal with the nonlinearity of \( K \left( e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta / \sigma_n \right) \) as a function of \( \lambda \) and \( \beta \). The parameter space \( \Lambda \times B \) is usually assumed to be a compact convex subset of \( \mathbb{R}^q \) for a nonlinear extremum estimation. This assumption is required for the uniform convergence of the sample average objective function in the proof of consistency (Amemiya, 1985). However, Wang and Lee (2013) mention that relaxation of this assumption would be an important issue of future research as it does not cover leading specification for the parameter space of \( \lambda \), which is often taken to be an open set, e.g., \((-1, 1)\).

Under Assumptions 1−7, the following theorem shows the consistency of the smoothed spatial maximum score estimator.

**Theorem 1.** Let Assumptions 1-7 hold. Let \( \theta_n \) be a solution to

\[
\max_{\theta} G_n(\theta; \sigma_n), \tag{13}
\]

where \( G_n(\theta; \sigma_n) = \frac{1}{n} \sum_{i=1}^{n} \Delta y_{i,n} K \left( \frac{e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta}{\sigma_n} \right) \) and \( \theta = (\lambda, \beta^\top)^\top \), then \( \theta_n \to^p \theta_0 \). If in addition the strong mixing coefficients satisfy \( \alpha(m) \leq C m^{-r} \) for positive constants \( C \) and \( r \), then \( \theta_n \to^{a.s.} \theta_0 \).

### 3.3 Asymptotic Normality

In this subsection, the asymptotic normality of the smoothed spatial maximum score estimator is established, and the approach is analogous to that of Horowitz (1992) and de Jong and Woutersen (2011) except that the asymptotic properties are built on a dependent and heterogeneous process while the process in Horowitz (1992) is i.i.d and the process in de Jong and Woutersen (2011) is dependent but stationary.

Let Assumptions 1-3 hold, and suppose \( K(\cdot) \) is twice differentiable everywhere. Then \( G_n(\theta; \sigma_n) \) is twice differentiable with respect to \( \bar{\theta} = (\lambda, \bar{\beta}^\top)^\top \), where \( \bar{\beta} = (\beta_1, \ldots, \beta_{q-1})^\top \).
Assumption 7 ensures that $\hat{\theta}_0$ is an interior point of $\hat{\Theta}$. Define $T_n(\theta; \sigma_n) = \partial G_n(\theta; \sigma_n)/\partial \tilde{\theta}$, and $Q_n(\theta; \sigma_n) = \partial^2 G_n(\theta; \sigma_n)/\partial \tilde{\theta} \partial \tilde{\theta}^\top$. Let $\theta_n \equiv (\tilde{\theta}_n^\top, \beta_{n,q})^\top$ denote a solution to problem 13, then with probability approaching 1 as $n \to \infty$, $\tilde{\theta}_n$ is an interior point of $\hat{\Theta}$, $\beta_{n,q} = \beta_{0,q} = \pm 1$ and $T_n(\theta_n; \sigma_n) = 0$. A Taylor series expansion of $T_n(\theta_n; \sigma_n)$ yields:

$$
T_n(\theta_n; \sigma_n) = T_n(\theta_0; \sigma_n) + Q_n(\theta_n^\ast; \sigma_n)(\tilde{\theta}_n - \tilde{\theta}_0) = 0,
$$

(14)

where $\theta_n^\ast$ is between $\theta_n$ and $\theta_0$. Similar to Horowitz (1992), if there is a real function $\rho(n)$ such that $\rho(n)T_n(\theta_0; \sigma_n)$ converges in distribution as $n \to \infty$, and suppose $Q_n(\theta_n^\ast; \sigma_n)$ converges in probability to a nonsingular and nonstochastic matrix $Q$. Then

$$
\rho(n)(\tilde{\theta}_n - \tilde{\theta}_0) = -Q^{-1}\rho(n)T_n(\theta_0; \sigma_n) + o_p(1).
$$

(15)

Thus, we know that $\tilde{\theta}_n - \tilde{\theta}_0$ converges to 0 at the rate of $\rho(n)^{-1}$, and $\rho(n)(\tilde{\theta}_n - \tilde{\theta}_0)$ is distributed asymptotically as $-Q^{-1}\rho(n)T_n(\theta_0; \sigma_n)$.

Let $z_{i,n} = e_{i,n}^\top S_n^{-1} \Delta X_n \beta_0 = e_{i,n}^\top S_n^{-1} \Delta \hat{X}_n \beta_0 + e_{i,n}^\top S_n^{-1} \Delta X_{n,q}$, then there is a one-to-one relation between $\left(\Delta \hat{X}_n, Z_n\right)$ and $\Delta X_n$ for any fixed $\theta_0$, where $Z_n = (z_{1,n}, \ldots, z_{n,n})^\top$. Denote $Z_{-i,n} = (z_{1,n}, \ldots, z_{i-1,n}, z_{i+1,n}, \ldots, z_{n,n})^\top$ and $\tilde{Z}_i = \{\Delta \hat{X}_n, Z_{-i}\}$. By Assumption 2, the distribution of $z_{i,n}$ conditional on $\tilde{Z}_i$ has everywhere positive density with respect to Lebesgue measure for almost every $\tilde{Z}_i$. Let $p_i(z_{i,n} | \tilde{Z}_i)$ denote this density. For each positive integer $j$, define $p_i^{(j)}(z_{i,n} | \tilde{Z}_i) = \partial^j p_i(z_{i,n} | \tilde{Z}_i)/\partial z_{i,n}^j$ whenever the derivative exists, and define $p_i^{(0)}(z_{i,n} | \tilde{Z}_i) = p_i(z_{i,n} | \tilde{Z}_i)$. Let $F_i$ denote the cumulative distribution function of $Z_i$, and let $F_i(\cdot | z_{i,n}, \tilde{Z}_i)$ denote the cumulative distribution of $\tilde{z}_{i,n} = e_{i,n}^\top S_n^{-1}(\epsilon_{n1} - \epsilon_{n2})$ conditional on $z_{i,n}$ and $\tilde{Z}_i$. For each positive integer $j$, define $F_i^{(j)}(-z_{i,n} | z_{i,n}, \tilde{Z}_i) = \partial^j F_i^{(j)}(-z_{i,n} | z_{i,n}, \tilde{Z}_i)/\partial z_{i,n}^j$ whenever the derivative exists. Define the scalar constants $\alpha_A$ and $\alpha_D$ by $\alpha_A = \int_{-\infty}^{\infty} v^h K'(v)dv$ and $\alpha_D = \int_{-\infty}^{\infty} [K'(v)]^2dv$ whenever these quantities exist. For each integer $h \geq 2$, define the $q \times 1$ vector $A$ and the $q \times q$ matrices $D$ and $Q$ by

$$
A = -2\alpha_A \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{k=1}^{h} \frac{1}{k!(h-k)!} E \left[ F_i^{(k)}(0 | \tilde{Z}_i) p_i^{(k)}(z_{i,n} | \tilde{Z}_i) \tilde{B}_{i,1} \right] \Pr(y_{i1,n} \neq y_{i2,n}),
$$

(16)

The Mean-Value Theorem is applied componentwise.
\[ D = \alpha_D \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} E \left[ p_i(0|\tilde{Z}_i)\tilde{B}_{1,i}\tilde{B}_{1,i}^\top \right] \Pr(y_{i1,n} \neq y_{i2,n}), \]

\[ Q = \lim_{n \to \infty} \frac{2}{n} \sum_{i=1}^{n} E \left[ \mathcal{F}(1)(0|0, Z_i)p_i(0|\tilde{Z}_i)\tilde{B}_{2,i}\tilde{B}_{2,i}^\top \right] \Pr(y_{i1,n} \neq y_{i2,n}), \]

where \( \tilde{B}_{1,i} = \left( e_{i,n}^\top S^{-1} W_n S^{-1} \Delta X_n \beta_0, e_{i,n}^\top S^{-1} W_n S^{-1} \Delta X_n \right) \) and

\[ \tilde{B}_{2,i} = \begin{pmatrix} \left( e_{i,n}^\top S^{-1} W_n S^{-1} \Delta X_n \beta_0 \right)^2 e_{i,n}^\top S^{-1} W_n S^{-1} \Delta X_n e_{i,n} S^{-1} \Delta \tilde{X}_n \\ \Delta \tilde{X}_n^\top [S^{-1}] e_{i,n} e_{i,n}^\top S^{-1} \Delta \tilde{X}_n \end{pmatrix}. \]

**Assumption 8.**

i) The \( \alpha \)-mixing coefficient satisfies \( \alpha(m) \leq C m^{-(2s-2)/(s-2)-\gamma} \) for positive constants \( C, \gamma, \) and \( s > 4. \)

ii) For some sequence \( m_n \geq 1 \) and \( s > 4, \) when \( n \to \infty, \)

\[ \sigma_n^{-3(p+q-1)} \sigma_n^{-2} n^{1/s} \alpha(m_n) + \sigma_n^{2(p+q-1)/\mu} n^{2/s} \alpha(m_n) + |\log(nm_n)| \left( n^{1-4/s} \sigma_n^{-4} m_n^{-2} \right)^{-1} \rightarrow 0. \]

**Assumption 9.** For all vectors \( \xi \) such that \( |\xi| = 1, \) \( E|\xi^\top \tilde{B}_{1,i}|^s < \infty \) for some \( s > 4 \) and all \( i. \)

These two assumptions are identical to Assumptions 6 and 7 of de Jong and Woutersen (2011). They strengthen the fading memory conditions of Assumption 6 in order to establish asymptotic normality.

The following assumptions are analogous to Assumptions 7-11 of Horowitz (1992):

**Assumption 10.**

i) \( K(\cdot) \) is twice differentiable everywhere, \( |K(\cdot)| \) and \( |K''(\cdot)| \) are uniformly bounded, and each of the following integrals over \( (-\infty, \infty) \) is finite: \( \int [K'(v)]^4 dv, \)
\( \int [K''(v)]^2 dv, \) and \( \int |v|^2 K''(v)|dv. \)

ii) For some integer \( h \geq 2 \) and each integer \( k \) (\( 1 \leq k \leq h \)), \( \int |v^k K'(v)|dv < \infty, \) and

\[ \int_{-\infty}^{\infty} v^k K'(v)dv = \begin{cases} 0 & \text{if } k < h, \\ d \text{ (nonzero) if } k = h. \end{cases} \]
iii) For any integer \( k \) between 0 and \( h \), any \( \gamma > 0 \), and any sequence \( \{\sigma_n\} \) converging to 0,

\[
\lim_{n \to \infty} \sigma_n^{k-h} \int_{|\sigma_n v| > \gamma} |v^k K'(v)| \, dv = 0, \quad \lim_{n \to \infty} \sigma_n^{-1} \int_{|\sigma_n v| > \gamma} |K''(v)| \, dv = 0.
\]

**Assumption 11.** For all \( i \) and each integer \( k \) such that \( 1 \leq k \leq h-1 \), all \( z_{i,n} \) in a neighborhood of 0, almost every \( \tilde{Z}_i \), and some \( M < \infty \), \( p_i^{(k)}(z_{i,n} | \tilde{Z}_i) \) exists and is a continuous function of \( z_{i,n} \) satisfying \( p_i^{(k)}(z_{i,n} | \tilde{Z}_i) < M \). In addition, \( |p_i(z_{i,n} | \tilde{Z}_i)| < M \) for all \( z_{i,n} \) and almost every \( \tilde{Z}_i \).

**Assumption 12.** For all \( i \) and each integer \( k \) such that \( 1 \leq k \leq h \), all \( z_{i,n} \) in a neighborhood of 0, almost every \( \tilde{Z}_i \), and some \( M < \infty \), \( F_i^{(k)}(-z_{i,n} | z_{i,n}, \tilde{Z}_i) \) exists and is a continuous function of \( z_{i,n} \) satisfying \( F_i^{(k)}(-z_{i,n} | z_{i,n}, \tilde{Z}_i) < M \).

**Assumption 13.** The true parameter \( \tilde{\theta}_0 \) is an interior point of \( \tilde{\Theta} \).

**Assumption 14.** The matrix \( Q \) is negative definite.

In addition to the above assumptions, we still need the following two assumptions that are similar to Assumptions 13 and 14 in de Jong and Woutersen (2011). The first assumption is needed to ensure proper behavior of covariance terms, and the second assumption on \( K''(\cdot) \) is needed to formally show a uniform law of large numbers for the second derivative of the objective function.

**Assumption 15.** The conditional joint density \( p(z_{i,n}, z_{j,n} | \tilde{Z}_i, \tilde{Z}_j) \) exists and is continuous at \( (z_{i,n}, z_{j,n}) = (0, 0) \) for all \( i \neq j \).

**Assumption 16.** \( K''(\cdot) \) satisfies, for some \( \mu \in (0, 1] \) and \( L \in [0, \infty) \) and all \( x, y \in \mathbb{R} \),

\[
|K''(x) - K''(y)| \leq L|x - y|^{\mu}.
\]

The main results concerning the asymptotic distribution of the smoothed spatial maximum score estimator are given by the following theorem.

**Theorem 2.** Let Assumptions [10] hold for some \( h \geq 2 \), then

(a) If \( n \sigma_n^{2h+1} \to \infty \) as \( n \to \infty \), \( \sigma_n^{-h}(\tilde{\theta}_n - \tilde{\theta}_0) \to^p -Q^{-1}A \).
(b) If $n\sigma_n^{2h+1}$ has a finite limit $\kappa$ as $n \to \infty$, then

$$\sqrt{n\sigma_n}(\hat{\theta}_n - \tilde{\theta}_0) \to^d N(-\kappa^{1/2}Q^{-1}A, \quad Q^{-1}DQ^{-1}).$$

In order to make the results of Theorem 2 useful in applications, the next theorem shows how $A$, $D$ and $Q$ could be consistently estimated from observations of $(Y_{nt}, X_{nt}, W_n)$.

**Theorem 3.** Let $\theta_n$ be a consistent smoothed spatial maximum score estimator based on $\sigma_n$ such that $\sigma_n \propto n^{-1/(2h+1)}$. For $\theta \in \{-1, 1\} \times \tilde{\Theta}$, define

$$t_{i,n}(\theta, \sigma) = \mathbb{1}\{y_{i1,n} \neq y_{i2,n}\} (2 \cdot \mathbb{1}\{y_{i1,n} = 1, y_{i2,n} = 0\} - 1) B^{(1)}_i(\theta, \sigma),$$

where $B^{(1)}_i(\theta, \sigma)$ is defined in Appendix A. Let $\sigma^*_n$ be such that $\sigma^*_n \propto n^{-\delta/(2h+1)}$, where $0 < \delta < 1$. Then: (a) $\hat{A}_n = (\sigma^*_n)^{-h}T_n(\theta_n, \sigma_n^*)$ converges in probability to $A$; (b) the matrix

$$\hat{D}_n = \frac{\sigma_n}{n} \sum_{i=1}^n t_{i,n}(\theta_n, \sigma_n) t_{i,n}(\theta_n, \sigma_n)^\top$$

converges in probability to $D$; (c) $Q_n(\theta_n, \sigma_n)$ converges in probability to $Q$.

Theorem 2 indicates that the asymptotic bias of $n^{h/(2h+1)}(\hat{\theta}_n - \tilde{\theta}_0)$ is $-\kappa^{h/(2h+1)}Q^{-1}A$ if $\sigma_n \propto n^{-1/(2h+1)}$, and this can be consistently estimated by $-\kappa^{h/(2h+1)}Q_n(\theta_n, \sigma_n)^{-1}\hat{A}_n$ by Theorem 3. Therefore, an asymptotically unbiased estimator of $\tilde{\theta}_0$, which is also called the bias-corrected smoothed spatial maximum score estimator, is

$$\hat{\theta}_{bc} = \hat{\theta}_n + (\kappa/n)^{h/(2h+1)}Q_n(\theta_n, \sigma_n)^{-1}\hat{A}_n. \quad (16)$$

Another important issue in applications is choosing the bandwidth $\sigma_n$, and no completely satisfactory solutions have been found for the well-known problem of bandwidth selection. Horowitz (1992) proposed that a possible choice of the bandwidth for smoothed maximum score estimator is $(\hat{\kappa}/n)^{1/(2h+1)}$, where $\hat{\kappa}$ is a consistent estimate of $\kappa^*$. $\kappa^*$ is the asymptotically optimal value of $\kappa$ that minimizes MSE of the smoothed maximum score estimator, as shown in part (c) of Theorem 2 in Horowitz (1992), $\kappa^* = |\text{trace}(Q^{-1}\Omega Q^{-1}D)|/(2h\Lambda^\top Q^{-1}\Omega Q^{-1}A)$ for any nonstochastic, positive semidefinite matrix such that $A^\top Q^{-1}\Omega Q^{-1}A \neq 0$. Therefore, the procedure of bandwidth selection is as
follows. Given $h$, first choose any $\sigma_n \propto n^{-1/(2h+1)}$ to compute the smoothed maximum score estimate $\hat{\theta}_n$, then use $\hat{\theta}_n$ and any $\sigma_n^* \propto n^{-\delta/(2h+1)}$, $0 < \delta < 1$ to compute $\hat{A}_n$, $\hat{D}_n$, and $Q_n(\theta_n, \sigma_n)$. After that, estimate $\kappa^*$ from the formula given above by replacing $A$, $D$, and $Q$ with $\hat{A}_n$, $\hat{D}_n$, and $Q_n(\theta_n, \sigma_n)$. Finally, the bandwidth is given by $(\kappa^*/n)^{1/(2h+1)}$.

In finite samples, $E\hat{A}_n \neq A$. The bias of $\hat{A}_n$ consists of two components: one component is due to the use of a nonzero bandwidth to estimate $A$, and the other is due to the use of an estimate of $\theta_0$ in the estimator of $A$. As suggested in Horowitz (1992), only the second component of the bias can be removed by a corrected estimator of $A$, which is given by

$$\hat{A}_n^* = \frac{\hat{A}_n}{1 - [\kappa^*-1n\sigma_n(\sigma_n^*)^{2h}]^{-1/2}}.$$

Note that, the use of $\hat{A}_n^*$ instead of $\hat{A}_n$ also improves the estimate of the asymptotically optimal bandwidth.

4 Monte Carlo Experiments

To investigate the finite sample properties of our estimator by a Monte Carlo study, the spatial binary choice SAR model is specified as

$$y_{it,n}^* = \lambda_0 \sum_{j=1}^{n} w_{ij,n} y_{jt,n}^* + x_{it,1,n} + x_{it,2,n} \beta_0 + \alpha_{i,n} + \epsilon_{it,n}$$

for $t = 1$ and $2$, where $x_{it,1,n}$ is drawn from the standard normal distribution $N(0, 1)$, and $x_{it,2,n}$ from the chi-square distribution with one degree of freedom, normalized to have zero mean and unit variance, independent of each other, $\alpha_{i,n} = \frac{1}{2}(x_{i1,2,n} + x_{i2,2,n}) + \gamma_{i,n}$ with $\gamma_{i,n}$ is from $N(0, 1)$ independent of other variables, and $\epsilon_{it,n}$ is drawn from $N(0, 1)$, independent of $(x_{it,1,n}, x_{it,2,n})$, the observed dependent variable $y_{it,n}$ is generated by $y_{it,n} = 1$ if $y_{it,n}^* > 0$ and $y_{it,n} = 0$ otherwise. When the sample size is $n = 49$, the spatial weights matrix $W_n$ corresponds to the weights matrix for the study of crimes across 49 districts in Columbus, Ohio in Anselin (1988). For large sample sizes of $n = 490$ and $n = 980$, the corresponding spatial weights matrices are block diagonal matrices with the preceding $49 \times 49$ matrix as their diagonal blocks, as in Lee (2007). These correspond to the pooling, respectively, of ten and twenty separate districts with similar neighboring structures in each district.
Given that the coefficients of $x$ could be estimated only up to scale, we set the coefficient of $x_{it,1,n}$ to one such that the coefficient of $x_{it,2,n}$ is point identified. In different cases of the Monte Carlo study, true parameters are $\beta_0 = 1$ and $\lambda_0 = 0.3, 0.7$, respectively. As the score function $G_n(\theta, \sigma_n)$ can have many local extrema, so it is necessary to use a global optimization method such as tunneling (Levy and Montalvo 1985) and generalized simulated annealing (Bohachevsky et al. 1986). However, results reported here are based on grid search, the number of grid is 200, and the parameters are searched in intervals centered at their true values with width 0.9.

Tables 1-3 report results for comparing the performance of estimators discussed in this paper: the maximum score (MS), smoothed maximum score (SMS), spatial maximum score (SpMS), smoothed spatial maximum score (SSpMS), and spatial nonlinear least square (SpNLLS) probit estimators. For SMS, SSpMS and SpNLLS estimators, the bandwidth for each sample is selected as follows. For SSpMS, a cumulative normal distribution function is used and and compared with two bandwidth selection procedures. For SSpMS-1, bandwidth $\sigma_n$ is selected according to the procedure suggested by Horowitz (1992) and discussed in section 3.3. For SSpMS-2, bandwidth is selected by using Silverman’s rule of thumb, $\sigma_n = 1.06 \cdot \hat{s} \cdot n^{-1/5}$, where $\hat{s}$ is the sample standard deviation of $y_{it,n}$. Finally, the bandwidth selection for SMS and SpNLLS estimators is also according to Silverman’s rule of thumb.

The number of repetitions is 1000 for each case in this Monte Carlo experiment. The regressors are randomly redrawn for each repetition. In each case, we report the mean bias, median bias, root mean square errors (RMSE), and mean absolute deviation (MAD) of the empirical distributions of the estimates.

Table 1 reports simulation results under homoskedasticity, where $\epsilon_{it,n}$ is drawn from $N(0, 1)$. In all the cases, the performance of the SpNLLS probit estimator is the worst, this is intuitive given that the SpNLLS probit estimator has a slower convergence rate and the bandwidth selection procedure may not be optimal. Although the spatial effects $\lambda$ can not be estimated, the MS and SMS estimators for the estimation of $\beta$ are better (for the case of small spatial correlation $\lambda_0 = 0.3$) than those of the SpMS and SSpMS estimators, respectively; nevertheless, the spatial estimators fight back for the case of large spatial correlation $\lambda_0 = 0.7$. The performances of SSpMS-1 and SSpMS-2 estimators are
almost the same, which suggests that Silverman’s rule of thumb is an effective bandwidth selection procedure for the SSpMS estimator. The SpMS estimator outperforms the SSpMS estimators, especially for a small sample size \((n = 49)\), where the ratios of RMSEs of SpMS estimator to those of the SSpMS estimators are roughly 50-60\% and 20-30\% for the estimates of \(\lambda\) and \(\beta\), respectively. However, this ratio increases as the \(n\) increases, which is consistent with the findings in Horowitz (1992) for the (smoothed) maximum score estimators without spatial effect. Finally, the RMSEs of SSpMS estimators for the estimation of \(\lambda\) is even around 30\% less than those of the SpMS estimator for the modest and large sample sizes.

The estimators are robust under heteroskedasticity and spatial errors. Table 2 reports simulation results under heteroskedasticity, where \(\epsilon_{it,n} = \left( 1 + 2z_{it,n}^2 + z_{it,n}^4 \right) u_{it,n}/4, z_{it,n} = x_{it,1,n} + x_{it,2,n}, u_{it,n}\) is logistic with median 0 and variance 1. Table 3 reports simulation results with spatial errors, where \(\epsilon_{nt} = \rho_0 W_n \epsilon_{nt} + v_{nt}, v_{it,n}\) are i.i.d. errors with distribution \(N(0, 1)\), \(\rho_0 = 0.5\), and weight matrix \(W_n\) is the same as in spatial lags. As we can see, the MS, SMS, SpMS and SSpMS estimators appear to perform better in the heteroskedasticity designs while the SpNLLS probit estimator stays the same, and all the estimators are robust under spatial errors.

In summary, the SpNLLS probit estimator performs the worst. Both the SpMS and the SSpMS estimators delivers a robust performance for various spatial autoregressive binary choice models. The SSpMS estimator can improve substantially and outperform the SpMS estimator with large sample sizes for the estimation of \(\lambda\).

5 Conclusion

In this paper, new estimation procedures for spatial autoregressive binary choice panel models were proposed. The estimators were based on a modification of the (smoothed) maximum score estimator to the correlated random effects binary choice models without spatial effect. Asymptotic properties of the SSpMS estimator were derived. A simulation study indicates these estimators perform quite well for various spatial models in finite samples.

The work here suggests areas for future research. Although both SpMS and SSpMS est-
Table 1: Simulation results with homoskedasticity

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>-0.0172</td>
<td>-0.0201</td>
<td>0.0601</td>
<td>0.1686</td>
<td>-0.0020</td>
<td>-0.0068</td>
<td>0.0628</td>
<td>0.2090</td>
</tr>
<tr>
<td>SMS</td>
<td>0.0564</td>
<td>0.2016</td>
<td>0.1097</td>
<td>0.2521</td>
<td>-0.0040</td>
<td>0.0158</td>
<td>0.0238</td>
<td>0.1187</td>
</tr>
</tbody>
</table>

$N = 490, \lambda_0 = 0.3$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>0.0065</td>
<td>0.0038</td>
<td>0.0150</td>
<td>0.0737</td>
<td>0.0008</td>
<td>0.0023</td>
<td>0.0484</td>
<td>0.1600</td>
</tr>
<tr>
<td>SMS</td>
<td>0.0190</td>
<td>0.0362</td>
<td>0.1509</td>
<td>0.3309</td>
<td>0.0190</td>
<td>0.0362</td>
<td>0.1509</td>
<td>0.3309</td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0176</td>
<td>0.0656</td>
<td>0.1226</td>
<td>0.2840</td>
<td>0.0176</td>
<td>0.0656</td>
<td>0.1226</td>
<td>0.2840</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0211</td>
<td>0.0701</td>
<td>0.1267</td>
<td>0.2842</td>
<td>0.0211</td>
<td>0.0701</td>
<td>0.1267</td>
<td>0.2842</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0067</td>
<td>-0.0023</td>
<td>0.0609</td>
<td>0.1913</td>
<td>0.0067</td>
<td>-0.0023</td>
<td>0.0609</td>
<td>0.1913</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>0.0404</td>
<td>0.0814</td>
<td>0.1671</td>
<td>0.3214</td>
<td>-0.0360</td>
<td>-0.0859</td>
<td>0.1631</td>
<td>0.3228</td>
</tr>
</tbody>
</table>

$N = 490, \lambda_0 = 0.7$

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>0.0158</td>
<td>0.0204</td>
<td>0.0855</td>
<td>0.2278</td>
<td>0.0158</td>
<td>0.0204</td>
<td>0.0855</td>
<td>0.2278</td>
</tr>
<tr>
<td>SMS</td>
<td>0.0249</td>
<td>0.0204</td>
<td>0.1149</td>
<td>0.2598</td>
<td>0.0249</td>
<td>0.0204</td>
<td>0.1149</td>
<td>0.2598</td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0178</td>
<td>-0.0027</td>
<td>0.0315</td>
<td>0.0937</td>
<td>-0.0178</td>
<td>-0.0027</td>
<td>0.0315</td>
<td>0.0937</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0064</td>
<td>0.0166</td>
<td>0.0184</td>
<td>0.0856</td>
<td>0.0064</td>
<td>0.0166</td>
<td>0.0184</td>
<td>0.0856</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>-0.0003</td>
<td>0.0102</td>
<td>0.0135</td>
<td>0.0904</td>
<td>-0.0003</td>
<td>0.0102</td>
<td>0.0135</td>
<td>0.0904</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>-0.0137</td>
<td>0.0006</td>
<td>0.0501</td>
<td>0.1618</td>
<td>-0.0137</td>
<td>0.0006</td>
<td>0.0501</td>
<td>0.1618</td>
</tr>
</tbody>
</table>

$N = 980, \lambda_0 = 0.3$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>-0.0172</td>
<td>-0.0068</td>
<td>0.0628</td>
<td>0.2090</td>
<td>-0.0172</td>
<td>-0.0068</td>
<td>0.0628</td>
<td>0.2090</td>
</tr>
<tr>
<td>SMS</td>
<td>0.0215</td>
<td>0.0000</td>
<td>0.0992</td>
<td>0.2380</td>
<td>0.0215</td>
<td>0.0000</td>
<td>0.0992</td>
<td>0.2380</td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0096</td>
<td>0.0006</td>
<td>0.0495</td>
<td>0.1641</td>
<td>-0.0096</td>
<td>0.0006</td>
<td>0.0495</td>
<td>0.1641</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>-0.0040</td>
<td>0.0068</td>
<td>0.0446</td>
<td>0.1650</td>
<td>-0.0040</td>
<td>0.0068</td>
<td>0.0446</td>
<td>0.1650</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>-0.0028</td>
<td>0.0068</td>
<td>0.0429</td>
<td>0.1633</td>
<td>-0.0028</td>
<td>0.0068</td>
<td>0.0429</td>
<td>0.1633</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>0.0404</td>
<td>0.0814</td>
<td>0.1671</td>
<td>0.3214</td>
<td>0.0404</td>
<td>0.0814</td>
<td>0.1671</td>
<td>0.3214</td>
</tr>
</tbody>
</table>

$N = 980, \lambda_0 = 0.7$

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>MS</td>
<td>0.0084</td>
<td>-0.0023</td>
<td>0.0714</td>
<td>0.2128</td>
<td>0.0084</td>
<td>-0.0023</td>
<td>0.0714</td>
<td>0.2128</td>
</tr>
<tr>
<td>SMS</td>
<td>0.0287</td>
<td>0.0158</td>
<td>0.0983</td>
<td>0.2235</td>
<td>0.0287</td>
<td>0.0158</td>
<td>0.0983</td>
<td>0.2235</td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0065</td>
<td>0.0038</td>
<td>0.0150</td>
<td>0.0737</td>
<td>-0.0065</td>
<td>0.0038</td>
<td>0.0150</td>
<td>0.0737</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0091</td>
<td>0.0166</td>
<td>0.0159</td>
<td>0.0652</td>
<td>0.0091</td>
<td>0.0166</td>
<td>0.0159</td>
<td>0.0652</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0030</td>
<td>0.0118</td>
<td>0.0108</td>
<td>0.0703</td>
<td>0.0030</td>
<td>0.0118</td>
<td>0.0108</td>
<td>0.0703</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>-0.0114</td>
<td>0.0038</td>
<td>0.0342</td>
<td>0.1230</td>
<td>-0.0114</td>
<td>0.0038</td>
<td>0.0342</td>
<td>0.1230</td>
</tr>
</tbody>
</table>
Table 2: Simulation results with heteroskedascity

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.0039</td>
<td>0.0023</td>
<td>0.0204</td>
<td>0.0678</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0707</td>
<td>0.1198</td>
<td>0.1946</td>
<td>0.3190</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>0.0043</td>
<td>-0.0023</td>
<td>0.0243</td>
<td>0.0978</td>
<td>-0.0014</td>
<td>0.0023</td>
<td>0.0156</td>
<td>0.0722</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0710</td>
<td>0.1153</td>
<td>0.1593</td>
<td>0.2523</td>
<td>0.1116</td>
<td>0.2013</td>
<td>0.2287</td>
<td>0.3090</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0685</td>
<td>0.1108</td>
<td>0.1574</td>
<td>0.2540</td>
<td>0.1047</td>
<td>0.1945</td>
<td>0.2242</td>
<td>0.3127</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>0.0574</td>
<td>0.0905</td>
<td>0.1759</td>
<td>0.3265</td>
<td>-0.0568</td>
<td>-0.1447</td>
<td>0.1828</td>
<td>0.3194</td>
</tr>
</tbody>
</table>

N = 49, \lambda_0 = 0.3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.0130</td>
<td>0.0113</td>
<td>0.0464</td>
<td>0.1434</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0278</td>
<td>0.0113</td>
<td>0.0652</td>
<td>0.1548</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0212</td>
<td>-0.0158</td>
<td>0.0338</td>
<td>0.0873</td>
<td>0.0150</td>
<td>0.0181</td>
<td>0.0368</td>
<td>0.1175</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0017</td>
<td>0.0090</td>
<td>0.0147</td>
<td>0.0891</td>
<td>0.0354</td>
<td>0.0226</td>
<td>0.0671</td>
<td>0.1390</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0013</td>
<td>0.0068</td>
<td>0.0146</td>
<td>0.0903</td>
<td>0.0347</td>
<td>0.0204</td>
<td>0.0669</td>
<td>0.1403</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>-0.0076</td>
<td>-0.0068</td>
<td>0.0618</td>
<td>0.1886</td>
<td>-0.0627</td>
<td>-0.0882</td>
<td>0.1365</td>
<td>0.2290</td>
</tr>
</tbody>
</table>

N = 490, \lambda_0 = 0.3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.0146</td>
<td>0.0113</td>
<td>0.0414</td>
<td>0.1316</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0215</td>
<td>0.0158</td>
<td>0.0461</td>
<td>0.1246</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0136</td>
<td>-0.0068</td>
<td>0.0218</td>
<td>0.0688</td>
<td>0.0212</td>
<td>0.0158</td>
<td>0.0388</td>
<td>0.1042</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>-0.0015</td>
<td>0.0023</td>
<td>0.0080</td>
<td>0.0616</td>
<td>0.0299</td>
<td>0.0226</td>
<td>0.0458</td>
<td>0.0982</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>-0.0019</td>
<td>0.0023</td>
<td>0.0084</td>
<td>0.0619</td>
<td>0.0298</td>
<td>0.0249</td>
<td>0.0457</td>
<td>0.0978</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>-0.0235</td>
<td>-0.0204</td>
<td>0.0597</td>
<td>0.1521</td>
<td>-0.0498</td>
<td>-0.0746</td>
<td>0.1080</td>
<td>0.1976</td>
</tr>
</tbody>
</table>

N = 980, \lambda_0 = 0.3

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>-0.0100</td>
<td>0.0023</td>
<td>0.0465</td>
<td>0.1352</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0085</td>
<td>-0.0113</td>
<td>0.1454</td>
<td>0.3396</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0163</td>
<td>-0.0332</td>
<td>0.0344</td>
<td>0.1085</td>
<td>-0.0106</td>
<td>0.0023</td>
<td>0.0420</td>
<td>0.1276</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0559</td>
<td>0.1035</td>
<td>0.0909</td>
<td>0.1525</td>
<td>0.0785</td>
<td>0.1357</td>
<td>0.2010</td>
<td>0.3124</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0525</td>
<td>0.1003</td>
<td>0.0879</td>
<td>0.1536</td>
<td>0.0682</td>
<td>0.1153</td>
<td>0.1880</td>
<td>0.3080</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>0.0453</td>
<td>0.2225</td>
<td>0.1175</td>
<td>0.2514</td>
<td>-0.0077</td>
<td>-0.0452</td>
<td>0.1269</td>
<td>0.3115</td>
</tr>
</tbody>
</table>

N = 49, \lambda_0 = 0.7

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.0061</td>
<td>-0.0068</td>
<td>0.0677</td>
<td>0.2104</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0234</td>
<td>0.0068</td>
<td>0.0976</td>
<td>0.2293</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0202</td>
<td>-0.0091</td>
<td>0.0284</td>
<td>0.0707</td>
<td>-0.0087</td>
<td>-0.0068</td>
<td>0.0481</td>
<td>0.1593</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0023</td>
<td>0.0070</td>
<td>0.0088</td>
<td>0.0622</td>
<td>0.0088</td>
<td>0.0023</td>
<td>0.0517</td>
<td>0.1658</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>-0.0030</td>
<td>0.0038</td>
<td>0.0103</td>
<td>0.0661</td>
<td>-0.0036</td>
<td>0.0023</td>
<td>0.0495</td>
<td>0.1720</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>0.0052</td>
<td>0.0166</td>
<td>0.0369</td>
<td>0.1483</td>
<td>-0.0524</td>
<td>-0.0927</td>
<td>0.1422</td>
<td>0.2583</td>
</tr>
</tbody>
</table>

N = 490, \lambda_0 = 0.7

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.0292</td>
<td>0.0249</td>
<td>0.0793</td>
<td>0.1858</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0295</td>
<td>0.0113</td>
<td>0.0791</td>
<td>0.1925</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0115</td>
<td>-0.0059</td>
<td>0.0159</td>
<td>0.0497</td>
<td>0.0072</td>
<td>0.0068</td>
<td>0.0346</td>
<td>0.1318</td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0012</td>
<td>0.0070</td>
<td>0.0046</td>
<td>0.0444</td>
<td>0.0180</td>
<td>0.0090</td>
<td>0.0431</td>
<td>0.1249</td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0007</td>
<td>0.0054</td>
<td>0.0044</td>
<td>0.0466</td>
<td>0.0119</td>
<td>0.0023</td>
<td>0.0395</td>
<td>0.1311</td>
</tr>
<tr>
<td>SpNLLS</td>
<td>-0.0168</td>
<td>-0.0059</td>
<td>0.0367</td>
<td>0.1133</td>
<td>-0.0372</td>
<td>-0.0543</td>
<td>0.1060</td>
<td>0.2217</td>
</tr>
</tbody>
</table>

N = 980, \lambda_0 = 0.7

22
Table 3: Simulation results with spatial errors

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\lambda = 0.3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>$\beta$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$N = 49, \lambda_0 = 0.3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.0054</td>
<td>0.0023</td>
<td>0.0525</td>
<td>0.1640</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0227</td>
<td>0.0339</td>
<td>0.1630</td>
<td>0.3451</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>0.0223</td>
<td>0.0158</td>
<td>0.0893</td>
<td>0.2221</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0326</td>
<td>0.0882</td>
<td>0.1497</td>
<td>0.3063</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0293</td>
<td>0.0814</td>
<td>0.1437</td>
<td>0.3010</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpNLLS</td>
<td>0.0484</td>
<td>0.1176</td>
<td>0.1748</td>
<td>0.3221</td>
<td>-0.0333</td>
<td>-0.0905</td>
<td>0.1669</td>
<td>0.3353</td>
<td></td>
</tr>
<tr>
<td>$N = 490, \lambda_0 = 0.3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>-0.0142</td>
<td>-0.0294</td>
<td>0.0814</td>
<td>0.2224</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0191</td>
<td>-0.0068</td>
<td>0.1040</td>
<td>0.2504</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0091</td>
<td>0.0113</td>
<td>0.0589</td>
<td>0.1862</td>
<td>-0.0108</td>
<td>-0.0294</td>
<td>0.0816</td>
<td>0.2293</td>
<td></td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0086</td>
<td>0.0294</td>
<td>0.0579</td>
<td>0.1813</td>
<td>0.0346</td>
<td>0.0271</td>
<td>0.1167</td>
<td>0.2467</td>
<td></td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0044</td>
<td>0.0249</td>
<td>0.0554</td>
<td>0.1857</td>
<td>0.0228</td>
<td>0.0113</td>
<td>0.1056</td>
<td>0.2473</td>
<td></td>
</tr>
<tr>
<td>SpNLLS</td>
<td>-0.0055</td>
<td>0.0023</td>
<td>0.0796</td>
<td>0.2319</td>
<td>-0.0473</td>
<td>-0.0837</td>
<td>0.1419</td>
<td>0.2669</td>
<td></td>
</tr>
<tr>
<td>$N = 980, \lambda_0 = 0.3$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.0076</td>
<td>-0.0023</td>
<td>0.0705</td>
<td>0.2134</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0318</td>
<td>0.0181</td>
<td>0.0970</td>
<td>0.2150</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0119</td>
<td>0.0023</td>
<td>0.0443</td>
<td>0.1468</td>
<td>0.0121</td>
<td>0.0023</td>
<td>0.0734</td>
<td>0.2104</td>
<td></td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>-0.0002</td>
<td>0.0023</td>
<td>0.0300</td>
<td>0.1386</td>
<td>0.0420</td>
<td>0.0339</td>
<td>0.1055</td>
<td>0.2099</td>
<td></td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>-0.0009</td>
<td>0.0113</td>
<td>0.0316</td>
<td>0.1411</td>
<td>0.0402</td>
<td>0.0384</td>
<td>0.1047</td>
<td>0.2115</td>
<td></td>
</tr>
<tr>
<td>SpNLLS</td>
<td>-0.0188</td>
<td>-0.0181</td>
<td>0.0723</td>
<td>0.1903</td>
<td>-0.0490</td>
<td>-0.0882</td>
<td>0.1330</td>
<td>0.2479</td>
<td></td>
</tr>
<tr>
<td>$N = 49, \lambda_0 = 0.7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>-0.0049</td>
<td>0.0023</td>
<td>0.0578</td>
<td>0.1825</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0139</td>
<td>0.0294</td>
<td>0.1604</td>
<td>0.3571</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>0.0031</td>
<td>0.0070</td>
<td>0.0415</td>
<td>0.1717</td>
<td>0.0037</td>
<td>0.0023</td>
<td>0.0678</td>
<td>0.2115</td>
<td></td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0352</td>
<td>0.1099</td>
<td>0.0878</td>
<td>0.2004</td>
<td>0.0662</td>
<td>0.1108</td>
<td>0.1996</td>
<td>0.3335</td>
<td></td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0323</td>
<td>0.1067</td>
<td>0.0850</td>
<td>0.2011</td>
<td>0.0580</td>
<td>0.0927</td>
<td>0.1901</td>
<td>0.3309</td>
<td></td>
</tr>
<tr>
<td>SpNLLS</td>
<td>0.0473</td>
<td>0.1983</td>
<td>0.1169</td>
<td>0.2428</td>
<td>0.0117</td>
<td>-0.0136</td>
<td>0.1358</td>
<td>0.3185</td>
<td></td>
</tr>
<tr>
<td>$N = 490, \lambda_0 = 0.7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>-0.0089</td>
<td>-0.0249</td>
<td>0.0835</td>
<td>0.2393</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0237</td>
<td>0.0204</td>
<td>0.1204</td>
<td>0.2725</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0111</td>
<td>0.0134</td>
<td>0.0341</td>
<td>0.1246</td>
<td>-0.0088</td>
<td>-0.0158</td>
<td>0.0894</td>
<td>0.2481</td>
<td></td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0167</td>
<td>0.0359</td>
<td>0.0378</td>
<td>0.1162</td>
<td>0.0091</td>
<td>0.0023</td>
<td>0.1060</td>
<td>0.2736</td>
<td></td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0122</td>
<td>0.0327</td>
<td>0.0341</td>
<td>0.1189</td>
<td>0.0033</td>
<td>-0.0090</td>
<td>0.1004</td>
<td>0.2744</td>
<td></td>
</tr>
<tr>
<td>SpNLLS</td>
<td>0.0157</td>
<td>0.0424</td>
<td>0.0593</td>
<td>0.1798</td>
<td>-0.0515</td>
<td>-0.1063</td>
<td>0.1600</td>
<td>0.2933</td>
<td></td>
</tr>
<tr>
<td>$N = 980, \lambda_0 = 0.7$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS</td>
<td>0.0032</td>
<td>0.0068</td>
<td>0.0754</td>
<td>0.2309</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SMS</td>
<td>0.0166</td>
<td>-0.0023</td>
<td>0.1021</td>
<td>0.2535</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpMS</td>
<td>-0.0109</td>
<td>0.0102</td>
<td>0.0257</td>
<td>0.0989</td>
<td>-0.0029</td>
<td>-0.0204</td>
<td>0.0696</td>
<td>0.2204</td>
<td></td>
</tr>
<tr>
<td>SSpMS-1</td>
<td>0.0071</td>
<td>0.0198</td>
<td>0.0208</td>
<td>0.0930</td>
<td>0.0243</td>
<td>0.0045</td>
<td>0.0986</td>
<td>0.2317</td>
<td></td>
</tr>
<tr>
<td>SSpMS-2</td>
<td>0.0016</td>
<td>0.0182</td>
<td>0.0162</td>
<td>0.0966</td>
<td>0.0163</td>
<td>-0.0023</td>
<td>0.0919</td>
<td>0.2351</td>
<td></td>
</tr>
<tr>
<td>SpNLLS</td>
<td>-0.0115</td>
<td>-0.0059</td>
<td>0.0422</td>
<td>0.1443</td>
<td>-0.0271</td>
<td>-0.0746</td>
<td>0.1162</td>
<td>0.2594</td>
<td></td>
</tr>
</tbody>
</table>
timators have desirable asymptotic properties and perform adequately well in finite samples, they may not be easy to implement in practice. The SpMS estimator has a discontinuous objective function, ruling out gradient-based optimization methods. The objective function of the SSpMS estimator can have several local maxima, and thus requires a global maximization algorithm that is not available in standard econometric software packages. The SpNLLS probit estimator may be an alternative in applications, although it has a slower rate of convergence, non-Gaussian limiting distribution (Blevins and Khan, 2013), and relatively worse finite sample performance. Therefore, using bias correction procedures for the SpNLLS probit estimator or deriving other competing estimators may be the direction for future research.

Furthermore, it would be useful to explore a more effective bandwidth selection procedure than that suggested in (Horowitz, 1992), as the SSpMS, especially for the bias-corrected SSpMS, estimators are quite sensitive to the choice of bandwidth in finite samples.

Finally, Horowitz (2002) shows that the differences between the true and nominal levels of tests based on smoothed maximum score estimates can be very large in finite samples when first order asymptotics are used to obtain critical values, and the bootstrap provides asymptotic refinements. Thus, it is natural to ask whether this property carries on for the SSpMS estimator or not.
Appendix A: Notations

As $\frac{\partial S^{-1}}{\partial x} = -S^{-1} \frac{\partial S}{\partial x} S^{-1}$, we know that $\frac{\partial e_{i,n}^{T} S^{-1}(\lambda)}{\partial x} = e_{i,n}^{T} S^{-1}(\lambda) W_{n} S^{-1}(\lambda)$.

Denote $B_{i} = K \left( \frac{e_{i,n}^{T} S^{-1}(\lambda) \Delta X_{n}}{\sigma_{n}} \right) = K \left( \frac{z_{i,n}}{\sigma_{n}} \right)$ where $z_{i,n} = e_{i,n}^{T} S^{-1}(\lambda) \Delta X_{n}$, then

$$\frac{\partial B_{i}}{\partial \lambda} = K' \left( \frac{z_{i,n}}{\sigma_{n}} \right) \frac{e_{i,n}^{T} S^{-1}(\lambda) W_{n} S^{-1}(\lambda) \Delta X_{n}}{\sigma_{n}};$$

$$\frac{\partial B_{i}}{\partial \beta^{T}} = K' \left( \frac{z_{i,n}}{\sigma_{n}} \right) \frac{e_{i,n}^{T} S^{-1}(\lambda) \Delta \tilde{X}_{n}}{\sigma_{n}};$$

$$B_{i}^{(1)}(\theta, \sigma_{n}) = (\partial B_{i}/\partial \lambda, \partial B_{i}/\partial \beta^{T})^{T};$$

$$\frac{\partial^{2} B_{i}}{\partial \lambda^{2}} = K'' \left( \frac{z_{i,n}}{\sigma_{n}} \right) \left[ \frac{e_{i,n}^{T} S^{-1}(\lambda) W_{n} S^{-1}(\lambda) \Delta X_{n}}{\sigma_{n}} \right]^{2} + 2K' \left( \frac{z_{i,n}}{\sigma_{n}} \right) \frac{e_{i,n}^{T} S^{-1}(\lambda) W_{n} S^{-1}(\lambda) W_{n} S^{-1}(\lambda) \Delta X_{n}}{\sigma_{n}};$$

$$\frac{\partial^{2} B_{i}}{\partial \beta^{T} \partial \lambda} = K'' \left( \frac{z_{i,n}}{\sigma_{n}} \right) \frac{e_{i,n}^{T} S^{-1}(\lambda) W_{n} S^{-1}(\lambda) \Delta X_{n}}{\sigma_{n}} \frac{e_{i,n}^{T} S^{-1}(\lambda) \Delta \tilde{X}_{n}}{\sigma_{n}} + K' \left( \frac{z_{i,n}}{\sigma_{n}} \right) \frac{e_{i,n}^{T} S^{-1}(\lambda) W_{n} S^{-1}(\lambda) \Delta \tilde{X}_{n}}{\sigma_{n}};$$

$$\frac{\partial^{2} B_{i}}{\partial \beta^{T} \partial \beta} = K'' \left( \frac{z_{i,n}}{\sigma_{n}} \right) \Delta \tilde{X}_{n}^{T} S^{-1}(\lambda) e_{i,n} e_{i,n}^{T} S^{-1}(\lambda) \Delta \tilde{X}_{n};$$

$$B_{i}^{(2)}(\theta, \sigma_{n}) = \begin{pmatrix} \frac{\partial B_{i}}{\partial \lambda^{2}} & \frac{\partial B_{i}}{\partial \beta^{T} \partial \lambda} \\ \frac{\partial B_{i}}{\partial \beta^{T}} & \frac{\partial B_{i}}{\partial \beta^{T} \partial \beta} \end{pmatrix};$$

Recall that

$$G_{n}(\theta; \sigma_{n}) = \frac{1}{n} \sum_{i=1}^{n} \Delta y_{i,n} K \left( \frac{e_{i,n}^{T} S^{-1}(\lambda) \Delta X_{n}}{\sigma_{n}} \right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1} \{ y_{i,n} \neq y_{i,2,n} \} \left( 1 - 2 \cdot \mathbb{1} \{ y_{i,1,n} = 0, y_{i,2,n} = 1 \} \right) B_{i},$$

then we have

$$T_{n}(\theta, \sigma_{n}) = \frac{\partial G_{n}(\theta, \sigma_{n})}{\partial \theta} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1} \{ y_{i,1,n} \neq y_{i,2,n} \} \left( 1 - 2 \cdot \mathbb{1} \{ y_{i,1,n} = 0, y_{i,2,n} = 1 \} \right) B_{i}^{(1)}(\theta, \sigma_{n});$$

$$Q_{n}(\theta, \sigma_{n}) = \frac{\partial^{2} G_{n}(\theta, \sigma_{n})}{\partial \theta \partial \theta^{T}} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1} \{ y_{i,1,n} \neq y_{i,2,n} \} \left( 1 - 2 \cdot \mathbb{1} \{ y_{i,1,n} = 0, y_{i,2,n} = 1 \} \right) B_{i}^{(2)}(\theta, \sigma_{n});$$

25
Appendix B: Proofs

Proof of Lemma 1. Without loss of generality, let \( q' = q \) and consider the case in which \( \beta_{0,q} > 0 \) (the case \( \beta_{0,q} < 0 \) is symmetric). For any \( (\lambda, \beta) \in \Lambda \times \mathbb{R}^q \), let \( \tilde{\beta} = (\beta_1, \ldots, \beta_{q-1}) \) and \( \tilde{\beta}_0 = (\beta_{0,1}, \ldots, \beta_{0,q-1}) \). Denote \( e_{i,n}^T S_n^{-1}(\lambda) = (a_{i1,n}(\lambda), a_{i2,n}(\lambda), \ldots, a_{in,n}(\lambda)) \) and \( e_{i,n}^T S_n^{-1} = (b_{i1,n}(\lambda_0), b_{i2,n}(\lambda_0), \ldots, b_{in,n}(\lambda_0)) \), as the inverse matrices \( S_n^{-1}(\lambda) \) and \( S_n^{-1} \) exist, so there exists at least one \( i \) such that vectors \( e_{i,n}^T S_n^{-1}(\lambda) \) and \( e_{i,n}^T S_n^{-1} \) do not have identical elements. For any fixed sample size \( n \), identification only requires that there is at least one \( i \) such that \( R_i(\lambda, \beta) > 0 \), for all \( (\lambda, \beta) \in \Lambda \times \mathbb{R}^q \) with \( \beta/||\beta|| \neq \beta_0/||\beta_0|| \) and \( \beta_q \neq 0 \). Assumption 4 which guarantees the fraction of individuals satisfying condition (9) does not decay to zero, prevents the failure of identification when \( n \) goes to infinity. Therefore, to show the identification of \( \theta_0 \) or \( R_i(\lambda, \beta) > 0 \), it is sufficient to show that, either

\[
\Pr(e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta < 0 < e_{i,n}^T S_n^{-1} \Delta X_n \beta) \text{ or } \Pr(e_{i,n}^T S_n^{-1} \Delta X_n \beta_0 < 0 < e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta)
\]

or both.

Given the existence of \( S_n^{-1}(\lambda) \) and \( S_n^{-1} \), again there exists at least one element \( a_{ij,n}(\lambda) \neq 0 \) and one element \( b_{ij',n}(\lambda_0) \neq 0 \). Apparently, there are four possible index sets: \( J, J', K, K' \), where \( a_{ij,n}(\lambda) \neq 0, b_{ij,n}(\lambda_0) = 0 \) for all \( j \in J \); \( a_{ij',n}(\lambda) = 0, b_{ij',n}(\lambda_0) \neq 0 \) for all \( j' \in J' \); \( a_{ik,n}(\lambda) \neq 0, b_{ik,n}(\lambda_0) \neq 0 \) for all \( k \in K \); \( a_{ik,n}(\lambda) = 0, b_{ik,n}(\lambda_0) = 0 \) for all \( k' \in K' \).

It is easy to see that

\[
\Pr(e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta < 0 < e_{i,n}^T S_n^{-1} \Delta X_n \beta) = \Pr\left(\sum_{j \in J} a_{ij,n}(\lambda) \Delta x_{j,n} \beta + \sum_{k \in K} a_{ik,n}(\lambda) \Delta x_{k,n} \beta < 0 < \sum_{j' \in J'} b_{ij',n}(\lambda_0) \Delta x_{j',n} \beta_0 + \sum_{k' \in K'} b_{ik',n}(\lambda_0) \Delta x_{k',n} \beta_0\right)
\]

\[
\geq \Pr(A_{i,j} \cap B_{i,j'} \cap C_{i,k})
\]

(A.1)

where

\[
A_{i,j} = \{a_{ij,n}(\lambda) \Delta x_{j,n} \beta < 0\} = \{a_{ij,n}(\lambda) \Delta \tilde{x}_{j,n} \tilde{\beta} + a_{ij,n}(\lambda) \Delta x_{j,q,n} \beta_q < 0\},
\]

\[
B_{i,j'} = \{b_{ij',n}(\lambda_0) \Delta x_{j',n}\beta_0 > 0\} = \{b_{ij',n}(\lambda_0) \Delta \tilde{x}_{j',n} \tilde{\beta}_0 + b_{ij',n}(\lambda_0) \Delta x_{j',q,n} \beta_{0,q} > 0\},
\]

\[
C_{i,k} = \{a_{ik,n}(\lambda) \Delta \tilde{x}_{k,n} \tilde{\beta} + a_{ik,n}(\lambda) \Delta x_{k,q,n} \beta_q < 0 < b_{ik,n}(\lambda_0) \Delta \tilde{x}_{k,n} \tilde{\beta}_0 + b_{ik,n}(\lambda_0) \Delta x_{k,q,n} \beta_{0,q}\}
\]

for all \( j, j' \) and \( k \). For \( \Pr(e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta_0 < 0 < e_{i,n}^T S_n^{-1}(\lambda) \Delta X_n \beta) \), we have

\[
A'_{i,j} = \{a_{ij,n}(\lambda) \Delta x_{j,n} \beta > 0\} = \{a_{ij,n}(\lambda) \Delta \tilde{x}_{j,n} \tilde{\beta} + a_{ij,n}(\lambda) \Delta x_{j,q,n} \beta_q > 0\},
\]
\[ B'_{i,j'} = \{ b_{ij',n}(\lambda_0)\Delta x_{j',n}\beta_0 < 0 \} = \{ b_{ij',n}(\lambda_0)\Delta x_{j',n}\beta_0 + b_{ij',n}(\lambda_0)\Delta x_{j',n}\beta_{0,q} < 0 \}, \]

\[ C'_{i,k} = \{ b_{ik,n}(\lambda_0)\Delta x_{k,n}\beta_0 + b_{ik,n}(\lambda_0)\Delta x_{k,n}\beta_{0,q} < 0 \} < a_{ik,n}(\lambda)\Delta x_{k,n}\beta + a_{ik,n}(\lambda_0)\Delta x_{k,n}\beta_q \}

Under Assumption 2 and using the same argument as in the proof of Lemma 2 in Manski (1985), we know that the conditional probabilities of \( A_{i,j}, A'_{i,j}, B_{i,j'}, \) and \( B'_{i,j'} \) are always positive, given that \( \beta_q \neq 0. \) For the positive conditional probability of \( C_{i,k} \) and/or \( C'_{i,k} \), there are four cases to consider, as we need to consider the different signs of \( a_{ik,n}(\lambda) \) and \( b_{ik,n}(\lambda_0) \):

(i) Case \( a_{ik,n}(\lambda)\beta_q < 0 \) and \( b_{ik,n}(\lambda_0) > 0 \):

\[ C_{i,k} = \left[ \Delta x_{k,q,n} > \max \left( -\Delta x_{k,n}\tilde{\beta}/\beta_q, -\Delta x_{k,n}\tilde{\beta}_0/\beta_{0,q} \right) \right] ; \]

\[ C'_{i,k} = \left[ \Delta x_{k,q,n} < \min \left( -\Delta x_{k,n}\tilde{\beta}/\beta_q, -\Delta x_{k,n}\tilde{\beta}_0/\beta_{0,q} \right) \right] ; \]

(ii) Case \( a_{ik,n}(\lambda)\beta_q < 0 \) and \( b_{ik,n}(\lambda_0) < 0 \):

\[ C_{i,k} = \left[ -\Delta x_{k,n}\tilde{\beta}/\beta_q < \Delta x_{k,q,n} < -\Delta x_{k,n}\tilde{\beta}_0/\beta_{0,q} \right] ; \]

\[ C'_{i,k} = \left[ -\Delta x_{k,n}\tilde{\beta}_0/\beta_{0,q} < \Delta x_{k,q,n} < -\Delta x_{k,n}\tilde{\beta}/\beta_q \right] ; \]

(iii) Case \( a_{ik,n}(\lambda)\beta_q > 0 \) and \( b_{ik,n}(\lambda_0) > 0 \):

\[ C_{i,k} = \left[ -\Delta x_{k,n}\tilde{\beta}/\beta_q < \Delta x_{k,q,n} < -\Delta x_{k,n}\tilde{\beta}_0/\beta_{0,q} \right] ; \]

\[ C'_{i,k} = \left[ -\Delta x_{k,n}\tilde{\beta}_0/\beta_{0,q} < \Delta x_{k,q,n} < -\Delta x_{k,n}\tilde{\beta}/\beta_q \right] ; \]

(iv) Case \( a_{ik,n}(\lambda)\beta_q > 0 \) and \( b_{ik,n}(\lambda_0) < 0 \):

\[ C_{i,k} = \left[ \Delta x_{k,q,n} < \min \left( -\Delta x_{k,n}\tilde{\beta}/\beta_q, -\Delta x_{k,n}\tilde{\beta}_0/\beta_{0,q} \right) \right] ; \]

\[ C'_{i,k} = \left[ \Delta x_{k,q,n} > \max \left( -\Delta x_{k,n}\tilde{\beta}/\beta_q, -\Delta x_{k,n}\tilde{\beta}_0/\beta_{0,q} \right) \right] . \]

Under Assumption 2 and using the same argument as in the proof of Lemma 2 in Manski (1985), we know that the conditional probabilities of \( C_{i,k} \) and \( C'_{i,k} \) in cases (i) and (iv) are always positive. In cases (ii) and (iii), either the probability of \( C_{i,k} \) or the probability of \( C'_{i,k} \) is positive. Without loss of generality, we assume the probability of \( C_{i,k} \) is positive.
Recall that in equation (A.1),
\[
\Pr\left(e_{i,n}^\top S_n^{-1}(\lambda)\Delta X_n^{\beta} < 0 < e_{i,n}^\top S_n^{-1}\Delta X_n^{\beta_0}\right)
\geq \Pr\left(A_{i,j} \in J \cap B_{i,j'} \in J' \cap C_{i,k} \in K\right)
= \Pr\left(A_{i,j} \in J | B_{i,j'} \in J', C_{i,k} \in K\right) \Pr\left(B_{i,j'} \in J' | C_{i,k} \in K\right) \Pr\left(C_{i,k} \in K\right)
\]
(A.2)
as we just argued that each conditional probability in equation (A.2) is positive under Assumption 2, so we have
\[
\Pr\left(e_{i,n}^\top S_n^{-1}(\lambda)\Delta X_n^{\beta} < 0 < e_{i,n}^\top S_n^{-1}\Delta X_n^{\beta_0}\right) > 0, \quad R(\lambda, \beta) > 0
\]
Therefore, \((\lambda_0, \beta_0)\) is identified relative to \((\lambda, \beta)\) except those \(\beta\) that are scalar multiples of \(\beta_0\).

For the proof of Theorem 1, we need Proposition 1 and the following Lemmas.

**Proof of Proposition 1.** To prove the NED of \(\{y_{it,n}\}\), we first show the NED of
\[
y_{it,n}^* = e_{i,n}^\top S_n^{-1}(\lambda) (X_{nt}^{\beta} + \alpha_{n} + \epsilon_{nt}) = \sum_{j=1}^{n} a_{ij,n}(\lambda) (x_{jt,n}^{\beta} + \alpha_{j,n} + \epsilon_{jt,n}).
\]
By Assumption 6 and Theorem 14.1 of Davidson (1994), the process \(v_{it,n} = x_{it,n}^{\beta} + \alpha_{i,n} + \epsilon_{it,n}\) is strong mixing with \(\alpha\)-mixing coefficient \(\alpha(m)\). Then, by the Minkowski inequality,
\[
\left\|y_{it,n}^* - E(y_{it,n}^*|3_{i,n}(m))\right\|_2 = \left\|\sum_{j,d(l_i,l_j)>m} a_{ij,n}(\lambda) (v_{jt,n} - E(v_{jt,n}|3_{i,n}(m)))\right\|_2 \leq d_t \nu(m),
\]
(A.3)
where \(\nu(m) = \sup_{j,d(l_i,l_j)>m} |a_{ij,n}(\lambda)|\), and \(d_t = 2 \sup_j ||v_{jt,n}||_2\). Therefore, \(\{y_{it,n}^*\}\) is NED because \(\nu(m) \to 0\) as \(m \to \infty\) by equation (10).

For any \(\epsilon > 0\), let \(\delta_\epsilon(0)\) denote the \(\epsilon\)-neighborhood of 0, then we have the following
inequality for the indicator function:

\[
|\mathbbm{1}\{x_1 > 0\} - \mathbbm{1}\{x_2 > 0\}| \\
\leq \frac{|x_1 - x_2|}{\epsilon} \mathbbm{1}\{x_1 \not\in \delta_\epsilon(0)\} \text{ or/and } \mathbbm{1}\{x_2 \not\in \delta_\epsilon(0)\} + \mathbbm{1}\{x_1 \in \delta_\epsilon(0), x_2 \in \delta_\epsilon(0)\}.
\]

(A.4)

Denote \( B = \{y_{it,n}^* \in \delta_\epsilon(0), \mathbb{E}(y_{it,n}^*|I_{i,n}(m)) \in \delta_\epsilon(0)\} \), then we have

\[
\left| \mathbbm{1}\{y_{it,n}^* > 0\} - \mathbb{E}(\mathbbm{1}\{y_{it,n}^* > 0\}|I_{i,n}(m)) \right|_2 \\
\leq \left( \mathbb{E} \left( \mathbbm{1}\{y_{it,n}^* > 0\} - \mathbb{E}(y_{it,n}^*|I_{i,n}(m)) > 0 \right)^2 \right)^{1/2} \\
\leq \left( \frac{1}{\epsilon^2} \int_{B^c} |y_{it,n}^* - \mathbb{E}(y_{it,n}^*|I_{i,n}(m))|^2 dP + \int_B dP \right)^{1/2} \\
\leq \frac{1}{\epsilon} \left( \int_{B^c} |y_{it,n}^* - \mathbb{E}(y_{it,n}^*|I_{i,n}(m))|^2 dP \right)^{1/2} + \left( \int_B dP \right)^{1/2} \\
\leq \frac{1}{\epsilon} \left| y_{it,n} - \mathbb{E}(y_{it,n}^*|I_{i,n}(m)) \right|_2 + \left( \int_B dP \right)^{1/2} \\
\leq \frac{1}{\epsilon} \mathbb{d} \nu(m) + \left( \int_B dP \right)^{1/2},
\]

where the first inequality is followed by Theorem 10.12 of Davidson (1994), the third line is by definition, the fourth line is by equation (A.4), the last line is followed by equation (A.3). As these two terms converge to 0 when \( \epsilon \) converges to 0 at a slower rate than \( \nu(m) \), so the process \( \{y_{it,n}\} \) is near epoch dependent.

The NED of process \( \{\Delta y_{it,n}\} \) follows from Davidson (1994) Theorem 17.8, which is also applicable under spatial dependence, and the NED of \( \{\text{sgn}(e_{i,n}^\top S_n^{-1}(\lambda)\Delta X_n \beta)\} \) could be shown similarly as \( \{y_{it,n}\} \).

\[ \blacksquare \]

**Lemma 2.** Under Assumptions \[14\] define

\[ G(\theta) = \frac{1}{n} \sum_{i=1}^n G_{i,n}(\theta) = \frac{1}{n} \sum_{i=1}^n \mathbb{E} \left[ \Delta y_{i,n} \text{sgn}(\epsilon_{i,n}^\top S_n^{-1}(\lambda)\Delta X_n \beta) \right], \]

then \( G(\theta_0) > G(\theta) \) for all \( \theta = (\lambda, \beta) \in \Lambda \times \mathbb{R}^q \) and \( n = 1, \ldots, \infty \), where \( \beta/||\beta|| \neq \beta_0/||\beta_0|| \) and \( \beta_q \neq 0 \).

\[ ^{10} \{E(y_{it,n}^*|I_{i,n}(m)) > 0\} \text{ is a } \mathcal{I}_{i,n}(m) \text{ measurable approximation to } \mathbbm{1}\{y_{it,n}^* > 0\}. \]
Proof of Lemma 2. As in Manski (1987) Lemma 3, for all \( \theta \in \Theta \),

\[
G(\theta_0) - G(\theta) = \frac{1}{n} \sum_{i=1}^{n} E \left[ \Delta y_{i,n} \left( \text{sgn} \{ e_{i,n}^T S_n^{-1} \Delta X_n \beta_0 \} - \text{sgn} \{ e_{i,n}^T S_n^{-1} (\lambda) \Delta X_n \beta \} \right) \right].
\]

\[
= 2 \frac{1}{n} \sum_{i=1}^{n} \int_{V(\lambda, \beta)} \text{sgn} \{ e_{i,n}^T S_n^{-1} \Delta X_n \beta_0 \} E(\Delta y_{i,n} | \Delta X_n) dF_{\Delta X_n}
\]

Conditions (3) imply that for all \( \Delta X_n \), \( \text{sgn} \{ e_{i,n}^T S_n^{-1} \Delta X_n \beta_0 \} E(\Delta y_{i,n} | \Delta X_n) = |E(\Delta y_{i,n} | \Delta X_n)| \).

Therefore,

\[
G(\theta_0) - G(\theta) = 2 \frac{1}{n} \sum_{i=1}^{n} \int_{V(\lambda, \beta)} |E(\Delta y_{i,n} | \Delta X_n)| dF_{\Delta X_n} \geq 0.
\]

Under Assumptions 1-4, \( E(\Delta y_{i,n} | \Delta X_n) \neq 0 \) for almost all \( \Delta X_n \). It now follows from Lemma 1 that \( G(\theta_0) > G(\theta) \) whenever \( \beta/||\beta|| \neq \beta_0/||\beta_0|| \) and \( \beta_q \neq 0 \). Note that it also holds for \( n \to \infty \), as Lemma 1 shows that identification holds in the limit.

\[\blacksquare\]

Lemma 3. For all \( c \in \mathbb{R} \) if \( (\Delta y_{i,n}, \Delta x_{i,n}) \) is strong mixing, then

\[
\sup_{\theta \in \Theta} \left| \frac{1}{n} \sum_{i=1}^{n} \left( \Delta y_{i,n} 1 \{ e_{i,n}^T S_n^{-1} (\lambda) \Delta X_n \beta \leq c \} - E \Delta y_{i,n} 1 \{ e_{i,n}^T S_n^{-1} (\lambda) \Delta X_n \beta \leq c \} \right) \right| \rightarrow^p 0.
\]

In addition, if \( \alpha(m) \leq Cm^{-r} \) for positive constants \( C \) and \( r \), then convergence is almost surely.

Proof of Lemma 3. The proof is similar to the proof of Lemma 4 in de Jong and Woutersen (2011), except that \( \{\Delta y_{i,n}\} \) is a heterogenous rather than stationary strong mixing process. We also apply the generic uniform law of large numbers of the Theorem of Andrews (1987).

It requires compactness of the parameter space \( \Theta \), which is assumed by Assumptions 3 and 7 the summands \( q_i(w_i, \theta), q_i^*(w_i, \theta) = \sup \{ q_i(w_i, \theta') : \theta' \in \Theta, d(\theta, \theta') < \rho \} \) and \( q_i(w_i, \theta) = \inf \{ q_i(w_i, \theta') : \theta' \in \Theta, d(\theta, \theta') < \rho \} \) are well-defined and satisfy a (respectively weak or strong) law of large numbers; and for all \( \theta \in \Theta \),

\[
\lim_{\rho \to 0} \sup_i \left| \frac{1}{n} \sum_{i=1}^{n} (E q_i^*(w_i, \theta) - E q_i(w_i, \theta)) \right| = 0.
\]
Here we show the last result, denote \( q_i^1(w_i, \theta) = e_i^\top S_n^{-1}(\lambda) \Delta X_n \beta \) and 

\[ K = \sup_i \sup_{\theta', d(\theta, \theta') < \rho} \frac{\partial q_i^1(w_i, \theta)}{\partial \theta} |_{\theta = \theta'} \], we have

\[
\lim_{\rho \to 0} \sup_n \left| \frac{1}{n} \sum_{i=1}^n (E q_i^* (w_i, \theta) - E q_{*1} (w_i, \theta)) \right|
\]

\[
= \lim_{\rho \to 0} \sup_n \left| \frac{1}{n} \sum_{i=1}^n \left( E \Delta y_{i,n} \mathbb{1} \{ q_i^1(w_i, \theta') \leq c \} - E \Delta y_{i,n} \mathbb{1} \{ \inf_{\theta': d(\theta, \theta') < \rho} q_i^1(w_i, \theta') \leq c \} \right) \right|
\]

\[
\leq \lim_{K \to \infty} \limsup_{\rho \to 0} \sup_n \left| \frac{1}{n} \sum_{i=1}^n \left[ E \Delta y_{i,n} \mathbb{1} \{ q_i^1(w_i, \theta) \leq c + \rho K \} - \mathbb{1} \{ q_i^1(w_i, \theta') \leq c - \rho K \} \right] \mathbb{1} \{ \left| w_i \right| \leq K \} \right|
\]

\[
+ \lim_{K \to \infty} \limsup_{\rho \to 0} \sup_n \left| \frac{1}{n} \sum_{i=1}^n \left[ E \Delta y_{i,n} \mathbb{1} \{ q_i^1(w_i, \theta) \leq c + \rho K \} - \mathbb{1} \{ q_i^1(w_i, \theta') \leq c - \rho K \} \right] \mathbb{1} \{ \left| w_i \right| > K \} \right|
\]

\[
\leq \lim_{K \to \infty} \limsup_{\rho \to 0} \sup_n \left| \frac{1}{n} \sum_{i=1}^n \left( \mathbb{P} \{ q_i^1(w_i, \theta) \leq c + \rho K \} - \mathbb{P} \{ q_i^1(w_i, \theta') \leq c - \rho K \} \right) \right| = 0,
\]

because \( \Delta x_{i,q,n} \) has a continuous distribution. Moreover, note that \( q_i(w_i, \theta) \), \( q_i^*(w_i, \theta) \) and 

\( q_{*1}(w_i, \theta) \) are all well-defined strong mixing random variables and satisfy a strong law of 

large numbers of Theorem 4 of [De Jong (1995)] if \( \alpha(m) + \nu(m) \leq C m^{-r} \) for some positive 

constants \( C \) and \( r \).  

**Lemma 4.** Under Assumptions 3, \( |G_n^*(\theta) - G(\theta)| \rightarrow_p 0 \) uniformly over \( \theta \in \Theta \), where 

\( G_n^*(\theta) = \frac{1}{n} \sum_{i=1}^n \Delta y_{i,n} \text{sgn} \left( e_i^\top S_n^{-1}(\lambda) \Delta X_n \beta \right) \). In addition, if \( \alpha(m) \leq C m^{-r} \) for positive 

constants \( C \) and \( r \), then convergence is almost surely.

**Proof of Lemma 4.** By equation 6, we have

\[
G_n^*(\theta) = \frac{1}{n} \sum_{i=1}^n \Delta y_{i,n} \text{sgn} \left( e_i^\top S_n^{-1}(\lambda) \Delta X_n \beta \right)
\]

\[
= \frac{2}{n} \sum_{i=1}^n \Delta y_{i,n} \mathbb{1} \{ e_i^\top S_n^{-1}(\lambda) \Delta X_n \beta \geq 0 \} - \frac{1}{n} \sum_{i=1}^n \Delta y_{i,n},
\]

both terms satisfy a weak or strong uniform law of large numbers by Lemma 3.

**Lemma 5.** \( \{ G(\theta) \}_{n=1}^\infty \) is continuous at all \( \theta = (\lambda, \beta^\top)^\top \) such that \( \beta_q \neq 0 \).

**Proof of Lemma 5.** \( G(\theta) \) is continuous on \( \theta \in \Theta \), uniformly over \( n \geq 1 \), by the Theorem of 

Lemma 6. Under Assumptions 1-7, \( |G_n(\theta; \sigma_n) - G^*_n(\theta)| \to^p 0 \) uniformly over \( \theta \in \Theta \). In addition, if \( \alpha(m) \leq C m^{-r} \) for positive constants \( C \) and \( r \), then convergence is almost surely.

Proof of Lemma 6. As in [Charlier et al. (1995)], here we actually adjusted the definition of \( G^*_n(\theta) \) such that \( G^*_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \Delta y_{i,n} \{ e_{i,n}^T S_n^{-1}(\lambda) \Delta X_{n} \beta \geq 0 \} \) as in equation (6).

\[
\begin{align*}
|G_n(\theta; \sigma_n) - G^*_n(\theta)| &= \frac{1}{n} \left| \sum_{i=1}^{n} \Delta y_{i,n} \left[ 1 \{ e_{i,n}^T S_n^{-1}(\lambda) \Delta X_{n} \beta \geq 0 \} - K \left( \frac{e_{i,n}^T S_n^{-1}(\lambda) \Delta X_{n} \beta}{\sigma_n} \right) \right] \right| \\
&\leq \frac{1}{n} \left| \sum_{i=1}^{n} \left[ 1 \{ e_{i,n}^T S_n^{-1}(\lambda) \Delta X_{n} \beta \geq 0 \} - K \left( \frac{e_{i,n}^T S_n^{-1}(\lambda) \Delta X_{n} \beta}{\sigma_n} \right) \right] \right|
\end{align*}
\]

Under the uniform weak or strong law of large numbers for \( \frac{1}{n} \sum_{i=1}^{n} 1 \{ |e_{i,n}^T S_n^{-1}(\lambda) \Delta X_{n} \beta| < c \} \) converging to \( \frac{1}{n} \sum_{i=1}^{n} \Pr\{ |e_{i,n}^T S_n^{-1}(\lambda) \Delta X_{n} \beta| < c \} \) (implied by Lemma 3), similar to Horowitz (1992 Lemma 4), we can easily show that \( |G_n(\theta; \sigma_n) - G^*_n(\theta)| \to 0 \) almost surely uniformly over \( \theta \in \Theta \) as \( n \to \infty \). 

Proof of Theorem 1. For weak or strong consistency of \( (\theta_n \to \theta_0) \), it is sufficient to verify the following conditions: (i) \( \{G(\theta)\}_{n=1}^{\infty} \) has a unique maximum at \( \theta_0 \); (ii) The parameter space \( \Theta \) is compact; (iii) \( \{G(\theta)\}_{n=1}^{\infty} \) is continuous; (iv) \( |G_n(\theta) - G(\theta)| \to^p 0 \) uniformly over \( \theta \in \Theta \), and strong consistency can be obtained if this is replaced by \( \sup_{\theta \in \Theta} |G_n(\theta) - G(\theta)| \to^{a.s.} 0 \).

Condition (i) is satisfied by Lemmas 1 and 2, condition (ii) is provided by Assumptions 3 and 7, condition (iii) is proved by Lemma 5, and condition (iv) is obtained by Lemmas 4 and 6.

For the proofs of Theorems 2 and 3, we need the following Lemmas:

Lemma 7. Let Assumptions 1, 12, and 15 hold. Then

\[
(a) \ \lim_{n \to \infty} E \left[ \sigma_n^{-b} T_n(\theta_0; \sigma_n) \right] = A; \quad (b) \ \lim_{n \to \infty} \text{Var} \left[ (n \sigma_n)^{1/2} T_n(\theta_0; \sigma_n) \right] = D.
\]
Proof of Lemma 7. As we know that

\[ T_n(\theta_0, \sigma_n) = \frac{1}{n} \sum_{i=1}^{n} \Delta y_{i,n} B_i^{(1)}(\theta_0, \sigma_n) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{y_{i,n} \neq y_{i,2,n}\} (1 - 2 \cdot \mathbb{1}\{y_{i1,n} = 0, y_{i2,n} = 1\}) B_i^{(1)}(\theta_0, \sigma_n), \]

then

\[ E_n(T) = E[\sigma_n^{-1} T_n(\theta_0, \sigma_n)] \]

\[ = \sigma_n^{-1} \frac{1}{n} \sum_{i=1}^{n} \Pr\{y_{i1,n} \neq y_{i2,n}\} \int \left[ 1 - 2F_i(-z_{i,n}|z_{i,n}, \tilde{Z}_i) \right] B_i^{(1)}(\theta_0, \sigma_n)p_i(z_{i,n}|\tilde{Z}_i)dz_{i,n}dP(\tilde{Z}_i). \]

By Assumption 1, condition 3 and the Corollary of Manski (1987), we can easily derive that Med \((y_{i1,n} - y_{i2,n})|\Delta X_n, y_{i1,n} \neq y_{i2,n}\) = sgn\{z_{i,n}\}, so Med \((\Delta\tilde{\epsilon}_{i,n})|\Delta X_n, y_{i1,n} \neq y_{i2,n}\) = 0 and \(F_i(0|0, \tilde{Z}_i) = 0.5\) for almost every \(\tilde{Z}_i\) and \(i = 1, \ldots, n\).

The proof of part (a) is analogous to that of Lemma 5 in Horowitz (1992), the only adjustment is that we need the boundedness of matrices \(S_n^{-1}\) and \(S_n^{-1}W_n S_n^{-1}\) to guarantee the boundedness of \(\tilde{B}_i\) for applying Lebesgue’s dominated convergence theorem. This is immediately from Assumption 7 and footnote 7.

To prove part (b), let first denote \(t_n(\theta_0, \sigma_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}\{y_{i1,n} \neq y_{i2,n}\} B_i^{(1)}(\theta_0, \sigma_n), \)

then

\[ V_n(T) = \text{Var} \left[ (n\sigma_n)^{1/2} T_n(\theta_0, \sigma_n) \right] \]

\[ = n\sigma_n E \left[ t_n(\theta_0, \sigma_n)t_n(\theta_0, \sigma_n)^\top \right] + o(1) \]

\[ = \frac{\sigma_n}{n} \sum_{i=1}^{n} E \left[ B_i^{(1)}(\theta_0, \sigma_n)B_i^{(1)}(\theta_0, \sigma_n)^\top \right] \Pr\{y_{i1,n} \neq y_{i2,n}\} \]

\[ + \frac{\sigma_n}{n} \sum_{i=1}^{n} \sum_{j \neq i} E \left[ B_i^{(1)}(\theta_0, \sigma_n)B_j^{(1)}(\theta_0, \sigma_n)^\top \right] \Pr\{y_{i1,n} \neq y_{i2,n}\} \Pr\{y_{j1,n} \neq y_{j2,n}\} + o(1) \]

\[ = D_{n1} + D_{n2} + o(1). \]
Similar to Lemma 5 of Horowitz (1992),

\[ D_{n1} = \frac{1}{n\sigma_n} \sum_{i=1}^{n} \mathbb{E} \left[ K' \left( \frac{z_{i,n}}{\sigma_n} \right) \right] \left( \tilde{B}_{1,i} \tilde{B}_{1,i}^\top \right) \Pr\{y_{i1,n} \neq y_{i2,n}\} \]

\[ = \frac{1}{n\sigma_n} \sum_{i=1}^{n} \Pr\{y_{i1,n} \neq y_{i2,n}\} \int \left[ K' \left( \frac{z_{i,n}}{\sigma_n} \right) \right] \left( \tilde{B}_{1,i} \tilde{B}_{1,i}^\top \right) p_i(z_{i,n}) d\tilde{z}_{i,n} dP_i(\tilde{Z}_i) \rightarrow D \]

by Lebesgue’s dominated convergence theorem and Assumptions 8-11. Lemma 7 of de Jong and Woutersen (2011) shows that \( D_{n2} \) is asymptotically negligible. This finishes the proof of part (b).

Lemma 8. Let Assumptions 1-12 and 15 hold. (a) If \( n\sigma_n^{2h+1} \rightarrow \infty \) as \( n \rightarrow \infty \), \( \sigma_n^{-h} T_n(\theta_0; \sigma_n) \) converges in probability to \( A \). (b) If \( n\sigma_n^{2h+1} \rightarrow \infty \) has a finite limit \( \kappa \) as \( n \rightarrow \infty \), \( (n\sigma_n)^{1/2} T_n(\theta_0; \sigma_n) \) converges in distribution to \( \text{MVN}(\kappa^{1/2} A, D) \).

Analogously to Horowitz (1992) and de Jong and Woutersen (2011), define

\[ g_i(\zeta) = 1 \{y_{i1,n} \neq y_{i2,n}\} (2 \cdot 1 \{y_{i1,n} = 1, y_{i2,n} = 0\} - 1) \tilde{B}_{1,i} K' \left( \frac{z_{i,n}}{\sigma_n} + \zeta \tilde{B}_{1,i} \right) . \]

Lemma 9. If \((y_{it,n}, x_{it,n})\) is strong mixing with strong mixing sequence \( \alpha(m) \), and there exists a sequence \( m_n \geq 1 \) such that

\[ \sigma_n^{-3(p+q-1)} \sigma_n^{-2} n^{1/s} \alpha(m_n) + (\log(nm_n)) \left( n^{1-2/s} \sigma_n^{-4} m_n^{-2} \right)^{-1} \rightarrow 0. \]

then

\[ \sup_{\zeta} \left| \frac{1}{n\sigma_n^2} \sum_{i=1}^{n} [g_i(\zeta) - \mathbb{E}g_i(\zeta)] \right| \rightarrow^p 0 \]

Proofs of Lemmas 8 and 9. The proofs are identical to the proofs of Lemma 8 and 11 of de Jong and Woutersen (2011) except that we have a different score function.

Lemma 10. Let Assumptions 1-16 hold, and define \( \phi_n = (\tilde{\theta}_n - \tilde{\theta}_0)/\sigma_n \), where \( \theta_n \) is a smoothed spatial maximum score estimator. Then \( \text{plim}_{n \rightarrow \infty} \phi_n = 0 \).

Proof of Lemma 10. This follows from Lemma 8 and the reasoning of Lemma 8 in Horowitz (1992).
Lemma 11. Let Assumptions 1-16 hold. Let \( \{ \theta'_{n} \} = \{ \tilde{\theta}'_{n}, \beta'_{n,q} \} \) be any sequence in \( \Theta \) such that \( (\theta'_{n} - \theta_{0})/\sigma_{n} \to 0 \) as \( n \to \infty \). Then \( \text{plim}_{n \to \infty} Q_{n}(\theta'; \sigma_{n}) = Q \).

Proof of Lemma 11. We can separately show that the elements of \( Q_{n}(\theta; \sigma_{n}) \) follow a uniformly law of larger numbers. The proof is then analogous to the proof of Lemma 13 in de Jong and Woutersen (2011), except that we have different objective functions. ■

Proof of Theorem 2. The proof is identical to that of Theorem 2 in Horowitz (1992), where we need Lemmas 10 and 11 instead of Lemmas 8 and 9 in Horowitz (1992). ■

Proof of Theorem 3. The proof is identical to that of Theorem 7 in de Jong and Woutersen (2011), where we need Lemmas 10 and 11 instead of Lemmas 12 and 13 in de Jong and Woutersen (2011). ■

References


Blevins, J. R., Khan, S., 2013. Local NLLS estimation of semi-parametric binary choice models. The Econometrics Journal 16 (2), 135–160. 5


Pace, R., LeSage, J., 2011. Fast simulated maximum likelihood estimation of the spatial probit model capable of handling large samples. Available at SSRN 1966039.


