Partially Linear Spatial Probit Models

Mohamed-Salem AHMED
University of Lille, LEM-CNRS 9221
Lille, France
mohamed-salem.ahmed@univ-lille.fr

Sophie DABO
INRIA-MODAL
University of Lille LEM-CNRS 9221
Lille, France
sophie.dabo@univ-lille.fr

Abstract
A partially linear probit model for spatially dependent data is considered. A triangular array setting is used to cover various patterns of spatial data. Conditional spatial heteroscedasticity and non-identically distributed observations and a linear process for disturbances are assumed, allowing various spatial dependencies. The estimation procedure is a combination of a weighted likelihood and a generalized method of moments. The procedure first fixes the parametric components of the model and then estimates the non-parametric part using weighted likelihood; the obtained estimate is then used to construct a GMM parametric component estimate. The consistency and asymptotic distribution of the estimators are established under sufficient conditions. Some simulation experiments are provided to investigate the finite sample performance of the estimators.

keyword: Binary choice model, GMM, non-parametric statistics, spatial econometrics, spatial statistics.

Introduction
Agriculture, economics, environmental sciences, urban systems, and epidemiology activities often utilize spatially dependent data. Therefore, modelling such activities requires one to find a type of correlation between some random variables in one location with other variables in neighbouring locations; see for instance [Pinkse & Slade (1998)]. This is a significant feature of spatial data analysis. Spatial/Econometrics statistics provides tools to perform such modelling. Many studies on spatial effects in statistics and econometrics using many diverse models have been published; see

1
Cressie (2015), Anselin (2010), Anselin (2013) and Arbia (2006) for a review.

Two main methods of incorporating a spatially dependent structure (see for instance Cressie 2015) can essentially be distinguished as between geostatistics and lattice data. In the domain of geostatistics, the spatial location is valued in a continuous set of $\mathbb{R}^N$, $N \geq 2$. However, for many activities, the spatial index or location does not vary continuously and may be of the lattice type, the baseline of this current work. In image analysis, remote sensing from satellites, agriculture etc., data are often received as a regular lattice and identified as the centroids of square pixels, whereas a mapping often forms an irregular lattice. Basically, statistical models for lattice data are linked to nearest neighbours to express the fact that data are nearby.

Two popular spatial dependence models have received substantial attention for lattice data, the spatial autoregressive (SAR) dependent variable model and the spatial autoregressive error model (SAE, where the model error is an SAR), which extend the regression in a time series to spatial data.

From a theoretical point of view, various linear spatial regression SAR and SAE models as well as their identification and estimation methods, e.g., two-stage least squares (2SLS), three-stage least squares (3SLS), maximum likelihood (ML) or quasi-maximum likelihood (QML) and the generalized method of moments (GMM), have been developed and summarized by many authors such as Anselin (2013), Kelejian & Prucha (1998), Kelejian & Prucha (1999), Conley (1999), Cressie (2015), Case (1993), Lee (2004), Lee (2007), Lin & Lee (2010), Zheng & Zhu (2012), Malikov & Sun (2017), Garthoff & Otto (2017), Yang & Lee (2017). Introducing nonlinearity into the field of spatial linear lattice models has attracted less attention; see for instance Robinson (2011), who generalized kernel regression estimation to spatial lattice data. Su (2012) proposed a semi-parametric GMM estimation for some semi-parametric SAR models. Extending these models and methods to discrete choice spatial models has seen less attention; only a few papers were concerned with this topic in recent years. This may be, as noted by Fleming (2004) (see also Smirnov (2010) and Bille (2014)), due to the "added complexity that spatial dependence introduces into discrete choice models". Estimating the model parameters with a full ML approach in spatially discrete choice models often requires solving a very computationally demanding problem of $n$-dimensional integration, where $n$ is the sample size.

For linear models, many discrete choice models are fully linear and utilize a continuous latent variable; see for instance Smirnov (2010), Wang et al. (2013) and Martinetti & Geniaux (2017), who proposed pseudo-ML methods, and Pinkse & Slade (1998), who studied a method based on the GMM approach. Also, others methodologies of estimation are emerged like, EM algorithm (McMillen, 1992) and Gibbs sampling approach (LeSage, 2000).

When the relationship between the discrete choice variable and some explanatory variables is not linear, a semi-parametric model may represent an alternative to fully parametric models. This type of model is known in the literature as partially linear choice spatial models and is the baseline of this current work. When the data are independent, these choice models can be viewed as special cases of the famous generalized additive models (Hastie & Tibshirani, 1990) and have received substantial attention in the literature, and various estimation methods have been explored (see for instance Hunsberger 1994, Severini & Staniswalis 1994, Carroll et al. 1997).

To the best of our knowledge, semi-parametric spatial choice models have not yet been investigated
from a theoretical point of view. To fill this gap, this work addresses an SAE spatial probit model for when the spatial dependence structure is integrated in a disturbance term of the studied model. We propose a semi-parametric estimation method combining the GMM approach and the weighted likelihood method. The method consists of first fixing the parametric components of the model and non-parametrically estimating the non-linear component by weighted likelihood (Staniswalis, 1989). The obtained estimator depending on the values at which the parametric components are fixed is used to construct a GMM estimator (Pinkse & Slade, 1998) of these components.

The remainder of this paper is organized as follows. In Section 1, we introduce the studied spatial model and the estimation procedure. Section 2 is devoted to hypotheses and asymptotic results, while Section 3 reports a discussion and computation of the estimates. Section 4 gives some numerical results based on simulated data to illustrate the performance of the proposed estimators. The last section presents the proofs of the main results.

1 Model

We consider that at \( n \) spatial locations \( \{s_1, s_2, \ldots, s_n\} \) satisfying \( \|s_i - s_j\| > \rho \) with \( \rho > 0 \), observations of a random vector \((Y, X, Z)\) are available. Assume that these observations are considered as triangular arrays (Robinson, 2011) and follow the partially linear model of a latent dependent variable \( Y^* \):

\[
Y^*_i = X^T \beta_0 + g_0(Z_{in}) + U_{in}, \quad 1 \leq i \leq n, \quad n = 1, 2, \ldots \tag{1}
\]

with

\[
Y_{in} = I(Y^*_i \geq 0), \quad 1 \leq i \leq n, \quad n = 1, 2, \ldots \tag{2}
\]

where \( I(\cdot) \) is the indicator function; \( X \) and \( Z \) are explanatory random variables taking values in the two compact subsets \( \mathcal{X} \subset \mathbb{R}^p (p \geq 1) \) and \( \mathcal{Z} \subset \mathbb{R}^d (d \geq 1) \), respectively; the parameter \( \beta_0 \) is an unknown \( p \times 1 \) vector that belongs to a compact subset \( \Theta_\beta \subset \mathbb{R}^p \); and \( g_0(\cdot) \) is an unknown smooth function valued in the space of functions \( \mathcal{G} = \{ g \in C^2(\mathcal{Z}) : \|g\| = \sup_{z \in \mathcal{Z}} |g(z)| < C \} \), with \( C^2(\mathcal{Z}) \) the space of twice differentiable functions from \( \mathcal{Z} \) to \( \mathbb{R} \) and \( C \) a positive constant. In model (1), \( \beta_0 \) and \( g_0(\cdot) \) are constant over \( i \) (and \( n \)). Assume that the disturbance term \( U_{in} \) in (2) is modelled by the following spatial autoregressive process (SAR):

\[
U_{in} = \lambda_0 \sum_{j=1}^{n} W_{ijn} U_{jn} + \varepsilon_{in}, \quad 1 \leq i \leq n, \quad n = 1, 2, \ldots \tag{3}
\]

where \( \lambda_0 \) is the autoregressive parameter, valued in the compact subset \( \Theta_\lambda \subset \mathbb{R} \), \( W_{ijn}, j = 1, \ldots, n \) are the elements in the \( i \)-th row of a non-stochastic \( n \times n \) spatial weight matrix \( W_n \), which contains the information on the spatial relationship between observations. This spatial weight matrix is usually constructed as a function of the distances (with respect to some metric) between locations; see Pinkse & Slade (1998) for additional details. The \( n \times n \) matrix \( (I_n - \lambda_0 W_n) \) is assumed to be non-singular for all \( n \), where \( I_n \) denotes the \( n \times n \) identity matrix and \( \{\varepsilon_{in}, 1 \leq i \leq n\} \) are assumed to
be independent random Gaussian variables; $E(\varepsilon_{in}) = 0$ and $E(\varepsilon_{in}^2) = 1$ for $i = 1, \ldots, n$, $n = 1, 2, \ldots$

Note that one can rewrite (3) as

$$U_n = (I_n - \lambda_0 W_n)^{-1} \varepsilon_n, \quad n = 1, 2, \ldots$$

where $U_n = (U_{n1}, \ldots, U_{nn})^T$ and $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^T$. Therefore, the variance-covariance matrix of $U_n$ is

$$V_n(\lambda_0) \equiv \text{Var}(U_n) = (I_n - \lambda_0 W_n)^{-1} \left\{(I_n - \lambda_0 W_n)^T\right\}^{-1}, \quad n = 1, 2, \ldots$$

This matrix allows one to describe the cross-sectional spatial dependencies between the $n$ observations. Furthermore, the fact that the diagonal elements of $V_n(\lambda_0)$ depend on $\lambda_0$ and particularly on $i$ and $n$ allows some spatial heteroscedasticity. These spatial dependences and heteroscedasticity depend on the neighbourhood structure established by the spatial weight matrix $W_n$.

Before proceeding further, let us give some particular cases of the model.

If one consider i.i.d observations, that is, $V_n(\lambda_0) = \sigma^2 I_n$, with $\sigma$ depending on $\lambda_0$, the obtained model may be viewed as a special case of classical generalized partially linear models (e.g. Severini & Staniswalis, 1994) or the classical generalized additive model (Hastie & Tibshirani, 1990). Several approaches for estimating this particular model have been developed; among these methods, we cite that of Severini & Staniswalis (1994) based on the concept of the generalized profile likelihood (e.g. Severini & Wong, 1992). This approach consists of first fixing the parametric parameter $\beta$ and non-parametrically estimating $g_0(\cdot)$ using the weighted likelihood method. This last estimate is then used to construct a profile likelihood to estimate $\beta_0$.

When $g_0 \equiv 0$ (or is an affine function), that is, without a non-parametric component, several approaches have been developed to estimate the parameters $\beta_0$ and $\lambda_0$. The basic difficulty encountered is that the likelihood function of this model involves an $n$-dimensional normal integral; thus, when $n$ is high, the computation or asymptotic properties of the estimates may present difficulties (e.g. Poirier & Ruud, 1988). Various approaches have been proposed to addressed this difficulty; among these approaches, we cite the following:

- Feasible Maximum Likelihood approach: this approach consists of replacing the true likelihood function by a pseudo-likelihood function constructed via marginal likelihood functions. Smirnov (2010) proposed a pseudo-likelihood function obtained by replacing $V_n(\lambda_0)$ by some diagonal matrix with the diagonal elements of $V_n(\lambda_0)$. Alternatively, Wang et al. (2013) proposed to divide the observations by pairwise groups, where the latter are assumed to be independent with a bivariate normal distribution in each group, and estimate $\beta_0$ and $\lambda_0$ by maximizing the likelihood of these groups. Recently Martinetti & Geniaux (2017) proposed a pseudo-likelihood function defined as an approximation of the likelihood function where the latter is inspired by some univariate conditioning procedure.

- Generalized Method of Moments (GMM) approach used by Pinkse & Slade (1998). These authors used the generalized residuals defined by $\tilde{U}_{in}(\beta, \lambda) = E(U_{in}|Y_{in}, \beta, \lambda), \quad 1 \leq i \leq n, \quad n = 1, 2, \ldots$ with some instrumental variables to construct moment equations to define the GMM estimators of $\beta_0$ and $\lambda_0$.
In what follows, using the \( n \) observations \( (X_{in}, Y_{in}, Z_{in}), i = 1, \ldots, n \), we propose parametric estimators of \( \beta_0, \lambda_0 \) and a non-parametric estimator of the smooth function \( g_0(\cdot) \).

To this end, we assume that, for all \( n = 1, 2, \ldots, \{\varepsilon_i \in 1 \leq i \leq n\} \) is independent of \( \{X_{in}, 1 \leq i \leq n\} \) and \( \{Z_{in}, 1 \leq i \leq n\} \) is independent of \( \{Z_{in}, 1 \leq i \leq n\} \).

We give asymptotic results according to increasing domain asymptotic. This consists of a sampling structure whereby new observations are added at the edges (boundary points) to compare to the infill asymptotic, which consists of a sampling structure whereby new observations are added in between existing observations. A typical example of an increasing domain is lattice data. An infill asymptotic is appropriate when the spatial locations are in a bounded domain.

### 1.1 Estimation Procedure

We propose an estimation procedure based on a combination of a weighted likelihood method and a generalized method of moments. We first fix the parametric components \( \beta \) and the non-parametric component using a weighted likelihood. The obtained estimate is then used to construct generalized residuals, where the latter are combined with the instrumental variables to propose GMM parametric estimates. This approach will be described as follow.

By equation (2), we have

\[ \mathbb{E}_0(Y_{in}|X_{in}, Z_{in}) = \Phi \left( (v_{in}(\lambda_0))^{-1} (X_{in}^T \beta_0 + g_0(Z_{in})) \right), \quad 1 \leq i \leq n, \quad n = 1, 2, \ldots \]  

where \( \mathbb{E}_0 \) denotes the expectation under the true parameters (i.e., \( \beta_0, \lambda_0 \) and \( g_0(\cdot) \)), \( \Phi(\cdot) \) is the cumulative distribution function of a standard normal distribution, and \( (v_{in}(\lambda_0))^2 = V_{in}(\lambda_0), 1 \leq i \leq n, n = 1, 2, \ldots \) are the diagonal elements of \( V_n(\lambda_0) \).

For each \( \beta \in \Theta_\beta, \lambda \in \Theta_\lambda, z \in \mathcal{Z} \) and \( \eta \in \mathbb{R} \), we define the conditional expectation on \( Z_{in} \) of the log-likelihood of \( Y_{in} \) given \((X_{in}, Z_{in})\) for \( 1 \leq i \leq n, n = 1, 2, \ldots \) as

\[ H(\eta; \beta, \lambda, z) = \mathbb{E}_0 \left( \mathcal{L}(\Phi \left( (v_{in}(\lambda))^{-1} (\eta + X_{in}^T \beta) \right); Y_{in}) \bigg| Z_{in} = z \right), \]

with \( \mathcal{L}(u; v) = \log (u^v(1-u)^{1-v}) \). Note that \( H(\eta; \beta, \lambda, z) \) is assumed to be constant over \( i \) (and \( n \)). For each fixed \( \beta \in \Theta_\beta, \lambda \in \Theta_\lambda \) and \( z \in \mathcal{Z} \), \( g_{\beta, \lambda}(z) \) denotes the solution in \( \eta \) of

\[ \frac{\partial}{\partial \eta} H(\eta; \beta, \lambda, z) = 0. \]

Then, we have \( g_{\beta_0, \lambda_0}(z) = g_0(z) \) for all \( z \in \mathcal{Z} \).

Now, using \( g_{\beta, \lambda}(\cdot) \), we construct the GMM estimates of \( \beta_0 \) and \( \lambda_0 \) as in Pinkse & Slade (1998). For that, we define the generalized residuals, replacing \( g_0(Z_{in}) \) in (1) by \( g_{\beta, \lambda}(Z_{in}) \):

\[ \tilde{U}_{in}(\beta, \lambda, g_{\beta, \lambda}) = \mathbb{E}(U_{in}|Y_{in}, \beta, \lambda) \]

\[ = \frac{\phi \left( G_{in}(\beta, \lambda, g_{\beta, \lambda}) \right) \left( Y_{in} - \Phi \left( G_{in}(\beta, \lambda, g_{\beta, \lambda}) \right) \right)}{\Phi \left( G_{in}(\beta, \lambda, g_{\beta, \lambda}) \right) \left( 1 - \Phi \left( G_{in}(\beta, \lambda, g_{\beta, \lambda}) \right) \right)}, \]

where \( \phi(\cdot) \) is the density of the standard normal distribution and

\[ G_{in}(\beta, \lambda, g_{\beta, \lambda}) = (v_{in}(\lambda))^{-1} (X_{in}^T \beta + g_{\beta, \lambda}(Z_{in})). \]
For simplicity of notation, we write $\theta = (\beta^T, \lambda)^T \in \Theta \equiv \Theta_\beta \times \Theta_\lambda$ when possible. Note that in (9), the generalized residual $\tilde{U}_n(\cdot, \cdot)$ is calculated by conditioning only on $Y_{in}$ and not on the entire sample $\{Y_{in}, i = 1, 2, \ldots, n, n = 1, \ldots\}$ or a subset of it. This of course will influence the efficiency of the estimators of $\theta$ obtained by these generalized residuals, but it allows one to avoid a complex computation; see [Poirier & Ruud, 1988] for additional details. To address this loss of efficiency, let us follow Pinkse & Slade (1998)’s procedure, which consists of employing some instrumental variables to create some moment conditions, and use a random matrix to define a criterion function. Both the instrumental variables and the random matrix permit one to consider more information about the spatial dependences and heteroscedasticity characterizing the dataset.

Let us now detail the estimation procedure. Let

$$ S_n(\theta, g_\theta) = n^{-1} \xi_n^T \tilde{U}_n(\theta, g_\theta), \tag{10} $$

where $\tilde{U}_n(\theta, g_\theta)$ is an $n \times 1$ vector, composed of $\tilde{U}_i(\theta, g_\theta)$, $1 \leq i \leq n$ and $\xi_n$ is an $n \times q$ matrix of instrumental variables, whose $i$th row is given by the $1 \times q$ random vector $\xi_{in}$. The latter may depend on $g_\theta(\cdot)$ and $\theta$. We assume that $\xi_{in}$ is $\sigma(X_{in}, Z_{in})$, measurable for each $i = 1, \ldots, n$, $n = 1, 2, \ldots$. We suppress the possible dependence of the instrumental variables on the parameters for notational simplicity. The GMM approach consists of minimizing the following sample criterion function:

$$ Q_n(\theta, g_\theta) = S_n^T(\theta, g_\theta) M_n S_n(\theta, g_\theta), \tag{11} $$

where $M_n$ is some positive-definite $q \times q$ weight matrix that may depend on the sample information. The choice of the instrumental variables and weight matrix characterizes the difference between GMM estimator and all pseudo-maximum likelihood estimators. For instance, if one takes

$$ \xi_{in}(\theta, g_\theta) = \frac{\partial G_{in}(\theta, \eta_i)}{\partial \theta} + \frac{\partial G_{in}(\theta, \eta_i)}{\partial \eta} \frac{\partial g_\theta(Z_{in})}{\partial \theta}, \tag{12} $$

with $\eta_i = g_\theta(Z_{in})$, $G_{in}(\theta, \eta_i) = (v_{in}(\lambda))^{-1} (X_{in}^T \beta + \eta_i)$, and $M_n = I_q$ with $q = p + 1$, then the GMM estimator of $\theta$ is equal to a pseudo-maximum profile likelihood estimator of $\theta$, accounting only for the spatial heteroscedasticity.

Now, let

$$ S(\theta, g_\theta) = \lim_{n \to \infty} \mathbb{E}_0 (S_n(\theta, g_\theta)), \tag{13} $$

and

$$ Q(\theta, g_\theta) = S^T(\theta, g_\theta) M S(\theta, g_\theta), $$

where $M$, the limit of the sequence $M_n$, is a nonrandom positive-definite matrix. The functions $S_n(\cdot, \cdot)$ and $Q_n(\cdot, \cdot)$ are viewed as empirical counterparts of $S(\cdot, \cdot)$ and $Q(\cdot, \cdot)$, respectively. Clearly, $g_\theta(\cdot)$ is not available in practice. However, we need to estimate it, particularly by an asymptotically efficient estimate. By (8) and for fixed $\theta^T = (\beta^T, \lambda) \in \Theta$, an estimator of $g_\theta(z)$, for $z \in Z$, can be given by $\hat{g}_\theta(z)$, which denotes the solution in $\eta$ of

$$ \sum_{i=1}^n \frac{\partial}{\partial \eta} \mathcal{L}_i (\Phi(G_{in}(\theta, \eta)) ; Y_{in}) K \left( \frac{z - Z_{in}}{b_n} \right) = 0 \tag{14} $$
where $K(\cdot)$ is a kernel from $\mathbb{R}^d$ to $\mathbb{R}_+$ and $b_n$ is a bandwidth depending on $n$.

Now, replacing $g_\theta(\cdot)$ in (11) by the estimator $\hat{g}_\theta(\cdot)$ permits one to obtain the GMM estimator $\hat{\theta}$ of $\theta$ as

$$
\hat{\theta} = \arg\min_{\theta \in \Theta} Q_n(\theta, \hat{g}_\theta).
$$

A classical inconvenience of the estimator $\hat{g}_\theta(z)$ proposed in (14) is that the bias of $\hat{g}_\theta(z)$ is high for $z$ near the boundary of $Z$. Of course, this bias will affect the estimator of $\theta$ given in (15) when some of the observations $Z_{in}$ are near the boundary of $Z$. A local linear method, or more generally the local polynomial method (Fan & Gijbels, 1996), can be used to reduce this bias. Another alternative is to use trimming (Severini & Staniswalis, 1994), in which the function $S_n(\theta, g_\theta)$ is computed using only observations associated with $Z_{in}$ that are away from the boundary. The advantage of this approach is that the theoretical results can be presented in a clear form, but it is less tractable from a practical point of view, in particular, for small sample sizes.

2 Large sample properties

We now turn to the asymptotic properties of the estimators derived in the previous section: $\hat{\theta}^T = (\hat{\beta}^T, \lambda)$ and $\hat{g}_\theta(\cdot)$. Let us use the following notation: $\frac{d}{d\theta} S(\theta, g_\theta)$ means that we differentiate $S(\cdot, \cdot)$ with respect to $\theta$, and $\frac{\partial}{\partial g} S(\theta, g_\theta)$ is the partial derivative of $S(\cdot, \cdot)$ w.r.t the first variable. The partial derivative of $S_n(\theta, g)$ w.r.t $g$, for any function $v \in \mathcal{G}$, is

$$
\frac{\partial S_n}{\partial g}(\theta, g)(v) = n^{-1} \sum_{i=1}^n \xi_{in} \frac{\partial U_{in}}{\partial \eta}(\theta, \eta_i)v(Z_{in}).
$$

Without ambiguity, $\|a\|$ denotes $\sup_t|a(t)|$ when $a$ is a function, $(\sum a_t^2)^{1/2}$ when $a$ is a vector, and $(\sum \sum a_{ij}^2)^{1/2}$ when $a$ is a matrix.

Let the following matrices be needed in the asymptotic variance-covariance matrix of $\hat{\theta}$:

$$
B_1(\theta_0) = \lim_{n \to \infty} \mathbb{E}_0 \left(nS_n(\theta_0, g_0)S_n^T(\theta_0, g_0)\right), \quad B_2(\theta_0) = \left\{ \frac{d}{d\theta} S^T(\theta, g_\theta) \bigg|_{\theta = \theta_0} \right\} M \left\{ \frac{d}{d\theta} S(\theta, g_\theta) \bigg|_{\theta = \theta_0} \right\},
$$

with

$$
\frac{d}{d\theta} S(\theta, g_\theta) = \frac{\partial S}{\partial \theta}(\theta, g_\theta) + \frac{\partial S}{\partial g}(\theta, g_\theta) \frac{\partial}{\partial g} g_\theta,
$$

and

$$
\Omega(\theta_0) = \left\{ B_2(\theta_0) \right\}^{-1} \left\{ \frac{d}{d\theta} S^T(\theta, g_\theta) \bigg|_{\theta = \theta_0} \right\} M B_1(\theta_0) M \left\{ \frac{d}{d\theta} S(\theta, g_\theta) \bigg|_{\theta = \theta_0} \right\} \left\{ B_2(\theta_0) \right\}^{-1}.
$$

The following assumptions are required to establish the asymptotic results.

**Assumption A1. (Smoothing condition).** For each fixed $\theta \in \Theta$ and $z \in \mathcal{Z}$, let $g_\theta(z)$ denote
the unique solution with respect to $\eta$ of
\[
\frac{\partial}{\partial \eta} H(\eta; \theta, z) = 0.
\]

For any $\varepsilon > 0$ and $g \in \mathcal{G}$, there exists $\gamma > 0$ such that
\[
\sup_{\theta \in \Theta, z \in \mathcal{Z}} \left| \frac{\partial}{\partial \eta} H(g(z); \theta, z) \right| \leq \gamma \quad \Longrightarrow \quad \sup_{\theta \in \Theta, z \in \mathcal{Z}} |g(z) - g_\theta(z)| \leq \varepsilon.
\]

**Assumption A2. (Local dependence).** The density $f_{in}(\cdot)$ of $Z_{in}$ exists, is continuous on $\mathcal{Z}$ uniformly on $i$ and $n$ and satisfies
\[
\lim_{n \to \infty} \inf_{z \in \mathcal{Z}} \frac{1}{n} \sum_{i=1}^{n} f_{in}(z) > 0.
\]

The joint probability density $f_{ijn}(\cdot)$ of $(Z_{in}, Z_{jn})$ exists and is bounded on $\mathcal{Z} \times \mathcal{Z}$ uniformly on $i \neq j$ and $n$.

**Assumption A3. (Spatial dependence).** Let $h_{in}^{\theta, \eta}(\cdot, \cdot)$ denote the conditional log likelihood function of $Y_{in}$ given $(X_{in}, Z_{in})$, where $\eta_i = g(Z_{in})$. Let $T_{in}$ be the vector $(Y_{in}, X_{in}, Z_{in})$, $i = 1, \ldots, n$, $n = 1, 2, \ldots, \hat{p} = p + 1$, and assume that for all $i, l = 1, \ldots, n$,
\[
|\text{Cov}_0(\psi(T_{in}), \psi(T_{ln}))| \leq \text{Var}_0(\psi(T_{in})) \text{Var}_0(\psi(T_{ln}))^{1/2} \alpha_{ijn},
\]
with
\[
\psi(T_{in}) = K \left( \frac{z - Z_{in}}{b_n} \right) \quad \text{or} \quad \psi(T_{in}) = K \left( \frac{z - Z_{in}}{b_n} \right) \frac{\partial^{j_1 + \cdots + j_\hat{p} + r}}{\partial \theta_{1}^{j_1} \cdots \partial \theta_{\hat{p}}^{j_\hat{p}} \partial \eta^{r}} h_{in}^{\theta, \eta}(Y_{in} \mid X_{in}, Z_{in} = z),
\]
for all $z \in \mathcal{Z}$, $\theta \in \Theta, \eta = g(z)$ with $g \in \mathcal{G}$, and for all nonnegative integers $j_1, \ldots, j_\hat{p} = 0, 1, 2$ and $r = 0, \ldots, 4$, such that $j_1 + \cdots + j_\hat{p} + r \leq 6$.

We assume that
\[
\left| \text{Cov}_0 \left( \xi_{itn} \tilde{U}_{in}(\theta, g_\theta), \xi_{jsn} \tilde{U}_{jn}(\theta, g_\theta) \right) \right| \leq \left\{ \text{Var}_0 \left( \xi_{itn} \tilde{U}_{in}(\theta, g_\theta) \right) \text{Var}_0 \left( \xi_{jsn} \tilde{U}_{jn}(\theta, g_\theta) \right) \right\}^{1/2} \alpha_{ijn},
\]
for all $\theta \in \Theta, i, j = 1, \ldots, n$, $n = 1, 2, \ldots$ and for any $s, t = 1, \ldots, q$, and
\[
\left| \text{Cov}_0 \left( \xi_{in}^{(2)}(\theta_0, \eta_0^i), \xi_{jn}^{(2)}(\theta_0, \eta_0^j) \right) \right| \leq \left\{ \text{Var}_0 \left( \xi_{in}^{(2)}(\theta_0, \eta_0^i) \right) \text{Var}_0 \left( \xi_{jn}^{(2)}(\theta_0, \eta_0^j) \right) \right\}^{1/2} \alpha_{ijn},
\]
with
\[
\xi_{in}^{(2)}(\theta_0, \eta_0^i) := w^T \xi_{i} \Lambda \left( G_{in}(\theta_0, \eta_0^i) \right) \phi \left( G_{in}(\theta_0, \eta_0^i) \right) \frac{\partial G_{in}}{\partial \theta}(\theta_0, \eta_0^i),
\]
where $\eta_0^i = g_0(Z_i)$ for each $w \in \mathbb{R}^q$ such that $\|w\| = 1$. 

8
In addition, assume that there is a decreasing (to 0) positive function \( \varphi(\cdot) \) such that \( \alpha_{ijn} = O(\varphi(\|s_i - s_j\|)) \), \( r^2 \varphi(rr^*)/\varphi(r^*) = \omega(1) \), as \( r \to 0 \), for all fixed \( r^* > 0 \), where \( s_i \) and \( s_j \) are spatial coordinates associated with observations \( i \) and \( j \), respectively.

**Assumption A4.** The kernel \( K \) satisfies \( \int K(u)du = 1 \). It is Lipschitzian, i.e., there is a positive constant \( C \) such that

\[
|K(u) - K(v)| \leq C\|u - v\| \quad \text{for all } u, v \in \mathbb{R}^d.
\]

**Assumption A5.** The bandwidth \( b_n \) satisfies \( b_n \to 0 \) and \( nb_n^{d+1} \to \infty \) as \( n \to \infty \).

**Assumption A6.** The instrumental variables satisfy \( \sup_{i,n} \|\xi_{in}\| = O_p(1) \), where \( \xi_{in} \) is the i-th column of the \( n \times q \) matrix of instrumental variables \( \xi_n \).

**Assumption A7.** \( \theta^T = (\beta^T, \lambda) \) takes values in a compact and convex set \( \Theta = \Theta_\beta \times \Theta_\lambda \subset \mathbb{R}^p \times \mathbb{R} \), and \( \theta_0^T = (\beta_0^T, \lambda_0) \) is in the interior of \( \Theta \).

**Assumption A8.** \( S(\cdot, \cdot) \) is continuous on both arguments \( \theta \) and \( g \), and \( Q(\cdot, g) \) attains a unique minimum over \( \Theta \) at \( \theta_0 \).

**Assumption A9.** The square root of the diagonal elements of \( V_n(\lambda) \) are twice continuous differentiable functions with respect to \( \lambda \) and \( \sup_{\lambda \in \Theta_\lambda} \left| v_{in}^{-1}(\lambda) + \frac{d}{d\lambda} v_{in}(\lambda) + \frac{d^2}{d\lambda^2} v_{in}(\lambda) \right| < \infty \) uniformly on \( i \) and \( n \).

**Assumption A10.** \( B_1(\theta_0) \) and \( B_2(\theta_0) \) are positive-definite matrices, and \( M_n - M = o_p(1) \).

**Remark 1** Assumption A1 ensures the smoothness of \( H(\cdot, \cdot, \cdot) \) around its extrema point \( g_0(\cdot) \); see Severini & Staniswalis (1997). Assumption A2 is a decay of the local independence condition of the covariates \( Z_{in} \), meaning that these variables are not identically distributed; a similar condition can be found in Robinson (2011). Condition (18) generalizes the classical assumption \( \inf_z f(z) > 0 \) used in the case of estimating the density function \( f(\cdot) \) with identically distributed or stationary random variables. This assumption has been used in Robinson (2011) (Assumption A7(z), p. 8). Assumption A3 describes the spatial dependence structure. The processes that we use are not assumed stationary; this allows for greater generalizability and the dependence structure to change with the sample size \( n \) (see Pinkse & Slade (1998) for more discussion). Conditions (19) and (20) and (21) are not restrictive. When the regressors and instrumental variables are deterministic, conditions (19) and (20) are equivalent to \( |\text{Cov}(Y_{in}, Y_{in})| \leq \alpha_{ijn} \). The condition on \( \varphi(\cdot) \) is satisfied when the latter tends to zero at a polynomial rate, i.e., \( \varphi(t) = O(t^{-\tau}) \), for all \( \tau > 2 \), as in the case of mixing random variables.

Assumption A6 requires that the instruments and explanatory variables be bounded uniformly on \( i \) and \( n \). In addition, when the instruments depend on \( \theta \) and \( g(\cdot) \), they are also uniformly bounded with respect to these parameters. The compactness condition in Assumption A7 is standard, and the convexity is somewhat unusual; however, it is reasonable in most applications. Condition A8 is necessary to ensure the identification of the true parameters \( \theta_0 \). Assumption A9 requires the standard deviations of the errors to be uniformly bounded away from zero with bounded derivatives. This has been considered by Pinkse & Slade (1998). Assumption A10 is classic (Pinkse & Slade...
and required in the proof of Theorem 2.2. Those authors noted that in their model (without a non-parametric component), when the autoregressive parameter $λ_0 = 0$, $B_2(θ_0)$ is not invertible, regardless of the choice of $M_n$. This is also the case in our context because for each $g_θ(z)$ solution of $\{8\}$, $θ ∈ Θ$ and $z ∈ Z$, we have

$$\frac{∂g_θ}{∂β}(z) = -\frac{E(Γ_{jn}(θ, g_θ(z))X_{jn}|Z_{jn} = z)}{E(Γ_{jn}(θ, g_θ(z)))Z_{jn} = z},$$

and

$$\frac{∂g_θ}{∂λ}(z) = \frac{v′_{jn}(λ)}{v_{jn}(λ)} E\left(Γ_{jn}(θ, g_θ(z))\left(X^T_{jn}β + g_θ(z)\right)|Z_{jn} = z\right)\left(g_θ(z) - β^T\frac{∂g_θ}{∂β}(z)\right),$$

where $v′_{jn}(λ) = \frac{d}{dλ}v_{jn}(λ) = v_{jn}(λ)\left[W_nS^{-1}_n(λ)V_n(λ)\right]_{jj}$,

$$Γ_{jn}(·) = Λ′(G_{jn}(·))\left[Y_{jn} - Φ(G_{jn}(·))\right] - Λ(G_{jn}(·))Φ(G_{jn}(·))$$

and $Λ(·) = φ(·)/(1 - Φ(·))Φ(·)$. However

$$\left.\frac{∂g_θ}{∂λ}(z)\right|_{λ=0} = 0 \text{ because } v′_{jn}(0) = 0,$$

then $B_2(θ_0)$ will be singular when $λ_0 = 0$.

With these assumptions in place, we are able to give some asymptotic results. The weak consistencies of the proposed estimators are given in the following two results. The first theorem and corollary below establish the consistency of our estimators, whereas the second theorem addresses the question of convergence to a normal distribution of the parametric component when it is properly standardized.

**Theorem 2.1** Under Assumptions A1-A10, we have

$$\hat{θ} - θ_0 = o_p(1).$$

**Corollary 2.1** If the assumptions of Theorem 2.1 are satisfied, then we have

$$\|\hat{g}_θ - g_θ\| = o_p(1).$$

**Proof of Corollary 2.1** Note that

$$\|\hat{g}_θ - g_θ\| \leq \|\hat{g}_θ - g_θ\| + \|g_θ - g_θ\| \leq \sup_θ\|\hat{g}_θ - g_θ\| + \sup_θ\left\|\frac{∂g_θ}{∂θ}\right\| \|\hat{θ} - θ_0\| = o_p(1),$$

since, by the assumptions of Theorem 2.1, $\sup_θ\|\hat{g}_θ - g_θ\| = o_p(1)$ and $\sup_θ\left\|\frac{∂g_θ}{∂θ}\right\| < ∞$. The following gives an asymptotic normality result of $\hat{θ}$. 10
Theorem 2.2 Under assumptions A1-A10, we have

\[ \sqrt{n} \left( \hat{\theta} - \theta_0 \right) \to \mathcal{N} (0, \Omega(\theta_0)) \]

Remark 2 In practice, the previous asymptotic normality result can be used to construct asymptotic confidence intervals and build hypothesis tests when a consistent estimate of the asymptotic covariance matrix \( \Omega(\theta_0) \) is available. To estimate this matrix, let us follow the idea of Pinkse and Slade (1998) and define the estimator

\[ \Omega_n(\hat{\theta}) = \left\{ B_{2n}(\hat{\theta}) \right\}^{-1} \left\{ \frac{d}{d\theta} S_n^T (\theta, \hat{\eta}) \right\}_{\theta = \hat{\theta}} M_n B_{1n}(\hat{\theta}) M_n \left\{ \frac{d}{d\theta} S_n (\theta, \hat{\eta}) \right\}_{\theta = \hat{\theta}} \left\{ B_{2n}(\hat{\theta}) \right\}^{-1}, \]

with

\[ B_{1n}(\theta) = n S_n(\theta, \hat{\eta}) S_n^T(\theta, \hat{\eta}) \quad \text{and} \quad B_{2n}(\theta) = \left\{ \frac{d}{d\theta} S_n^T(\theta, \hat{\eta}) \right\} M_n \left\{ \frac{d}{d\theta} S_n (\theta, \hat{\eta}) \right\}. \]

The consistency of \( \Omega_n(\hat{\theta}) \) will be based on that of \( B_{1n}(\hat{\theta}) \) and \( B_{2n}(\hat{\theta}) \), the estimators of \( B_1(\theta_0) \) and \( B_2(\theta_0) \), respectively. Note that the consistency of \( B_{2n}(\theta) \) is relatively easy to establish. On the other hand, that of \( B_{1n}(\hat{\theta}) \) asks for additional assumptions and an adaptation of the proof of Theorem 3 of Pinkse and Slade (1998 p.134) to our case; this is of interest to future research.

3 Computation of the estimates

The aim of this section is to outline in detail how the regression parameters \( \beta \), the spatial autocorrelation parameter \( \lambda \) and the non-linear function \( g_\theta \) can be estimated. We begin with the computation of \( \hat{g}_\theta(z) \), which will play a crucial role in what follows.

3.1 Computation of the estimate of the non-parametric component

An iterative method is needed to compute the \( \hat{g}_\theta(z) \) solution of \( (14) \) for each fixed \( \theta \in \Theta \) and \( z \in \mathbb{Z} \). For fixed \( \theta^T = (\beta, \lambda) \in \Theta \) and \( z \in \mathbb{Z} \), let \( \eta_\theta = g_\theta(z) \) and \( \psi(\eta; \theta, z) \) denote the left-hand side of \( (14) \), which can be rewritten as

\[ \psi(\eta; \theta, z) = \sum_{i=1}^{n} \left[ v_{\eta_\theta}(\lambda) \right]^{-1} \Lambda (G_{\eta_\theta}(\theta, \eta)) \left[ Y_{\eta_\theta} - \Phi(G_{\eta_\theta}(\theta, \eta)) \right] K \left( \frac{z - Z_{\eta_\theta}}{b_{\eta_\theta}} \right). \]  

Consider the Fisher information:

\[ \Psi(\eta_\theta; \theta, z) = E_0 \left( \frac{\partial}{\partial \theta} \psi(\eta; \theta, z) \right) \bigg|_{\eta = \eta_\theta} \{(X_{\eta_\theta}, Z_{\eta_\theta}), 1 \leq i \leq n, n = 1, \ldots\} \]

\[ = - \sum_{i=1}^{n} \left[ v_{\eta_\theta}(\lambda) \right]^{-2} \Lambda (G_{\eta_\theta}(\theta, \eta_\theta)) \phi(G_{\eta_\theta}(\theta, \eta_\theta)) K \left( \frac{z - Z_{\eta_\theta}}{b_{\eta_\theta}} \right) + \]

\[ + \sum_{i=1}^{n} \left[ v_{\eta_\theta}(\lambda) \right]^{-2} \Lambda^* (G_{\eta_\theta}(\theta, \eta_\theta)) \left[ \Phi(G_{\eta_\theta}(\theta_0, \eta_\theta)) - \Phi(G_{\eta_\theta}(\theta, \eta_\theta)) \right] K \left( \frac{z - Z_{\eta_\theta}}{b_{\eta_\theta}} \right). \]
Note that the second term in the RHS of (23) is negligible when \( \theta \) is near the true parameter \( \theta_0 \). Because \( \psi(\eta; \theta, z) = 0 \) for \( \eta = \hat{g}_\theta(z) \), an initial estimate \( \tilde{\eta} \) can be updated to \( \eta^\dagger \) using Fisher’s scoring method:

\[
\eta^\dagger = \tilde{\eta} - \frac{\psi(\tilde{\eta}; \theta, z)}{\Psi(\tilde{\eta}; \theta, z)},
\]

The iteration procedure (24) requests some starting value \( \tilde{\eta} = \tilde{\eta}_0 \) to ensure convergence of the algorithm. To this end, let us adapt the approach of Severini & Staniswalis (1994), which consists of supposing that for fixed \( \theta \in \Theta \), there exists a \( \tilde{\eta}_0 \) satisfying \( G_{in}(\theta, \tilde{\eta}_0) = \Phi^{-1}(Y_{in}) \) for \( i = 1, \ldots, n \). Knowing that \( G_{in}(\theta, \tilde{\eta}_0) = (v_{in}(\lambda))^{-1}(X_{in}^T \beta + \tilde{\eta}_0) \), we have \( \tilde{\eta}_0 = v_{in}(\lambda)\Phi^{-1}(Y_{in}) - X_{in}^T \beta \). Then, (24) can be updated using the following initial value:

\[
\eta^\dagger_0 = \tilde{\eta}_0 - \frac{\psi(\tilde{\eta}_0; \theta, z)}{\Psi(\tilde{\eta}_0; \theta, z)} = \frac{\sum_{i=1}^n [v_{in}(\lambda)]^{-1} \Lambda(C_{in}) \phi(C_{in}) [C_{in}^{-1} - [v_{in}(\lambda)]^{-1} X_{in}^T \beta] K \left( \frac{z - Z_{in}}{b_n} \right)}{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Lambda(C_{in}) \phi(C_{in}) K \left( \frac{z - Z_{in}}{b_n} \right)},
\]

where \( C_{in} = \Phi^{-1}(Y_{in}), i = 1, \ldots, n \), is computed using a slight adjustment because \( Y_{in} \in \{0, 1\} \).

With this initial value, the algorithm iterates until convergence.

**Selection of the bandwidth**

A critical step (in non- or semi-parametric models) is the choice of the bandwidth parameter \( b_n \), which is usually selected by applying some cross-validation approach. The latter was adapted by Su (2012) in the case of a spatial semi-parametric model. Because cross-validation may be very time consuming, which is true in the case of our model, we adopt the following approach used in Severini & Staniswalis (1994) to achieve greater flexibility:

1. Consider the linear regression of \( C_{in} \) on \( X_{in}, i = 1, \ldots, n \), without an intercept term, and let \( R_{1n}, \ldots, R_{nn} \) denote the corresponding residuals.

2. Since we expect \( \mathbb{E}(R_{in}|Z_{in} = z) \) to have similar smoothness properties as \( g_0(.) \), the optimal bandwidth \( b_n \) is that of the non-parametric regression of the \( \{R_{in}\}_{i=1, \ldots, n} \) on \( \{Z_{in}\}_{i=1, \ldots, n} \), chosen by applying any non-parametric regression bandwidth selection method. For that, we use the cross-validation method in the np R Package.

**3.2 Computation of \( \hat{\theta} \)**

The parametric component \( \beta \) and the spatial autoregressive parameter \( \lambda \) are computed as mentioned above by a GMM approach based on some instrumental variables \( \xi_n \) and the weight matrix \( M_n \). The choices of these instrumental variables and weight matrix \( M_n \) are as follows. Because \( \psi(\hat{g}_\theta(z); \theta, z) = 0 \), if we differentiate the latter with respect to \( \beta \) and \( \lambda \), we have

\[
\frac{\partial}{\partial \beta} \hat{g}_\theta(z) = -\frac{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) X_{in} K \left( \frac{z - Z_{in}}{b_n} \right)}{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) K \left( \frac{z - Z_{in}}{b_n} \right)},
\]

\[
\frac{\partial}{\partial \lambda} \hat{g}_\theta(z) = -\frac{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) X_{in} K \left( \frac{z - Z_{in}}{b_n} \right)}{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) K \left( \frac{z - Z_{in}}{b_n} \right)},
\]
and

\[
\frac{\partial}{\partial \lambda} \hat{g}_\theta(z) = \frac{\sum_{i=1}^n [v_{in}(\lambda)]^{-1} v_{in}'(\lambda) \Delta_{in}(\theta, z) [X_{in}^T \beta + \hat{g}_\theta(z)] K \left( \frac{z - Z_{in}}{b_n} \right)}{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) K \left( \frac{z - Z_{in}}{b_n} \right)} + \frac{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} v_{in}'(\lambda) \Lambda (G_{in}(\theta, \hat{g}_\theta(z))) [Y_{in} - \Phi (G_{in}(\theta, \hat{g}_\theta(z)))] K \left( \frac{z - Z_{in}}{b_n} \right)}{\sum_{i=1}^n [v_{in}(\lambda)]^{-2} \Delta_{in}(\theta, z) K \left( \frac{z - Z_{in}}{b_n} \right)},
\]

with

\[
\Delta_{in}(\theta, z) = \Lambda' (G_{in}(\theta, \hat{g}_\theta(z))) [Y_{in} - \Phi (G_{in}(\theta, \hat{g}_\theta(z)))] - \Lambda (G_{ni}(\theta, \hat{g}_\theta(z))) \phi (G_{in}(\theta, \hat{g}_\theta(z))).
\]

Then, the previous result is used to define the following instrumental variables:

\[
\xi_{in}(\theta, \hat{g}_\theta) = \frac{\partial G_{in}(\theta, \hat{\eta}_i)}{\partial \theta} + \frac{\partial G_{in}(\theta, \hat{\eta}_i)}{\partial \eta} \frac{\partial}{\partial \theta} \hat{g}_\theta(Z_{in}),
\]

with \( \hat{\eta}_i = \hat{g}_\theta(Z_{in}) \).

For the weight matrix, we use (as in Pinkse & Slade (1998)) \( M_n = I_q \) with \( q = p + 1 \). Then, the obtained GMM estimator of \( \theta \) with this choice of \( M_n \) is equal to the pseudo-profile maximum likelihood estimator of \( \theta \), accounting only for the spatial heteroscedasticity.

The final step is to plug in the GMM estimator \( \hat{\theta} \) to obtain \( \hat{g}_\theta \).

### 4 Simulation study

In this section, we study the performance of the proposed model based on some numerical results, which highlight the importance of considering the spatial dependence and the partial linearity. We simulated some semi-parametric models and estimated them using our proposed method, i.e., the method that does not account for the spatial dependence (using the same estimation procedure above based on the partially linear probit model (PLPM)), and using a fully linear SAE probit (LSAEP) method. The latter method can account for the spatial dependence but ignores the partial linearity. The ProbitSpatial R package [Martinetti & Geniaux, 2016] is used to provide estimates for the LSAEP model. We generate observations from the following spatial latent partial linear model:

\[
Y_{in}^* = \beta_1 X_{in}^{(1)} + \beta_2 X_{in}^{(2)} + g(Z_{in}) + U_{in}; \quad Y_{in} = \mathbb{I}(Y_{in}^* > 0), \quad i = 1, \ldots, n \quad (25)
\]

\[
U_n = (I_n - \lambda W_n)^{-1} \varepsilon_n \quad (26)
\]

where \( U_n \sim \mathcal{N}(0, I_n) \) and \( W_n \) is the spatial weight matrix associated with \( n \) locations chosen randomly in a 60 \( \times \) 60 regular grid based on the 6 nearest neighbours of each unit. To observe the effect of partial linearity when we compare our estimation procedure to that based on LSAEP models, we will consider the following two cases:
Case 1: The explanatory variables \( X^{(1)} \) and \( X^{(2)} \) are generated as pseudo \( B(0.7) \) and \( U[-2,2] \), respectively, and the other explanatory variable \( Z \) is equal to the sum of 48 independent random variables, each uniformly distributed over \([-0.25,0.25]\). Here, we use the non-linear function \( g(t) = t + 2 \cos(0.5 \pi t) \).

Case 2: The explanatory variables \( X^{(1)}, X^{(2)} \) and \( Z \) are generated as pseudo \( \mathcal{N}(0,1) \), and we consider the linear function \( g(t) = 1 + 0.5 t \).

We take \( \beta_1 = -1, \beta_2 = 1 \) and different values of the spatial parameter \( \lambda \), that is, \( \lambda \in \{0.2, 0.5, 0.8\} \). The bandwidth \( b_n \) is selected using Severini & Staniswalis (1994)’s approach detailed previously with \( C_{ni} = \Phi^{-1}(0.9 Y_{ni} + 0.1(1 - Y_{ni})), i = 1, \ldots, n \). A Gaussian kernel will be considered: \( K(t) = (2\pi^{-1/2}) \exp(-t^2/2) \). As mentioned above, the instrumental variables are the trivial choice, and the weight matrix \( M_n = I_3 \) is the identity matrix.

The two studied cases are replicated 200 times for a sample size \( n = 200 \), and the results are presented in Tables 1 and 2. In each table, the columns titles Mean, Median and SD give the average, median and standard deviation, respectively, over these 200 replications associated with each estimation method.

First, when we compare the estimators based on our approach (PLSPM) with those based on the LSAEP model, we notice that the latter yields more biased estimators of the coefficients \( \beta_1 \) and \( \beta_2 \), in particular in Case 1. It makes sense that ignoring the partial linearity (see also Figure 1) weakens the quality of the estimation of the coefficients \( \beta_1 \) and \( \beta_2 \). In Case 2, these two approaches yield similar results in term of consistency, but our approach seems to be less efficient.

Second, note that for the two cases (Table 1 and Table 2), the LSAEP and PLPM estimates are similar in the case of low spatial dependence (\( \lambda = 0.2 \)). However, this is not the case for the large spatial dependence (\( \lambda = 0.8 \)) framework, where in this case the estimation procedure based on PLPM models yields inconsistent estimates of the parameters \( \beta_1 \) and \( \beta_2 \) and the smooth function \( g(\cdot) \) (see the right panel in Figure 1). It makes sense that considering the spatial dependence does not allow one to find consistent estimates of the coefficients \( \beta_1 \) and \( \beta_2 \) and the smooth function \( g(\cdot) \).

Note that this approach is less efficient; this can be realized when observing the differences between the mean and median (or the high values of the standard deviation) associated with our estimators in Tables 1 and 2. However, this is eventually due to the use of the GMM approach with the trivial choice of the weight matrix \( M_n = I_n \). In addition, when estimating the spatial parameter \( \lambda \), our procedure yields biased estimators; this may be related to the considered choice of IVs. Better choices of the weight matrix and instrumental variables have to be investigated in future research.

Discussion

In this manuscript, we have proposed a spatial semi-parametric probit model for identifying risk factors at onset and with spatial heterogeneity. The parameters involved in the models are estimated using weighted likelihood and generalized method of moment methods. A technique based on dependent random arrays facilitates the estimation and derivation of asymptotic properties, which otherwise would have been difficult to perform due to the complexity introduced by the
Table 1: Case 1 with \( n = 200 \) and 200 replications.

| \( \lambda \) | Methods | \( \beta_1 = -1 \) | \( \beta_2 = 1 \) |
|-------------|---------|-----------------|-----------------|
|             | Mean    | Median | SD  | Mean    | Median | SD  |
| 0.20        | PLSPM  | -1.08  | -1.00 | 0.53 | 1.07   | 0.99  | 0.33 |
|             | LSAEP  | -0.67  | -0.69 | 0.25 | 0.67   | 0.66  | 0.11 |
|             | PLPM   | -0.98  | -0.99 | 0.32 | 0.98   | 0.96  | 0.15 |
| 0.50        | PLSPM  | -1.13  | -0.96 | 0.67 | 1.08   | 0.98  | 0.40 |
|             | LSAEP  | -0.65  | -0.64 | 0.24 | 0.66   | 0.65  | 0.12 |
|             | PLPM   | -0.90  | -0.88 | 0.30 | 0.90   | 0.89  | 0.15 |
| 0.80        | PLSPM  | -1.12  | -0.86 | 0.86 | 1.08   | 0.89  | 0.55 |
|             | LSAEP  | -0.57  | -0.56 | 0.25 | 0.61   | 0.60  | 0.12 |
|             | PLPM   | -0.65  | -0.66 | 0.25 | 0.65   | 0.63  | 0.13 |

Figure 1: Case 1 with \( n = 200 \) and 200 replications.

Table 2: Case 2 with \( n = 200 \) and 200 replications.

| \( \lambda \) | Methods | \( \beta_1 = -1 \) | \( \beta_2 = 1 \) |
|-------------|---------|-----------------|-----------------|
|             | Mean    | Median | SD  | Mean    | Median | SD  |
| 0.20        | PLSPM  | -1.12  | -1.05 | 0.32 | 1.13   | 1.06  | 0.30 |
|             | LSAEP  | -1.08  | -1.06 | 0.19 | 1.09   | 1.07  | 0.20 |
|             | PLPM   | -1.00  | -0.99 | 0.20 | 0.99   | 0.98  | 0.14 |
| 0.50        | PLSPM  | -1.08  | -1.03 | 0.37 | 1.06   | 0.99  | 0.31 |
|             | LSAEP  | -1.06  | -1.06 | 0.21 | 1.05   | 1.01  | 0.19 |
|             | PLPM   | -0.95  | -0.94 | 0.21 | 0.93   | 0.91  | 0.18 |
| 0.80        | PLSPM  | -1.02  | -0.91 | 0.44 | 1.01   | 0.86  | 0.43 |
|             | LSAEP  | -0.88  | -0.87 | 0.19 | 0.87   | 0.86  | 0.20 |
|             | PLPM   | -0.66  | -0.65 | 0.15 | 0.66   | 0.65  | 0.16 |
spatial dependence to the model and high-dimensional integration required by a full maximum likelihood approach. Moreover, the technique yields consistent estimates through proper choices of the bandwidth, weight matrix, and instrumental variables. The proposed models provide a general framework and tools for researchers and practitioners when addressing binary semi-parametric choice models in the presence of spatial correlation. Although they provide significant contributions to the body of knowledge, to the best of our knowledge, additional work needs to be done. As indicated, the weights are used to improve the efficiency and convergence. It would be interesting to develop criteria for the choices of optimal weights toward achieving better performance. For instance, the performance may be improved by choosing, for instance, a weight matrix \( M_n \) as a consistent estimator \( \hat{B}_1(\theta_0) \) of the matrix \( B_1(\theta_0) \). Another empirical choice could be the idea of continuously updating the GMM estimator (one-step GMM) used in Pinkse et al. (2006):

\[
M_n(\theta) = n^{-1} \sum_{i,j=1}^{n} \delta_{ij} \xi_{ni} \xi_{nj} \tilde{U}_{in}(\theta, \hat{g}_\theta) \tilde{U}_{jn}(\theta, \hat{g}_\theta)
\]

with the weights

\[
\delta_{ij} = \frac{\sum_{r=1}^{n} \tau_{ri} \tau_{rj}}{\left[ \sum_{r=1}^{n} \tau_{ri}^2 \sum_{r=1}^{n} \tau_{rj}^2 \right]^{1/2}} \quad \text{for } i, j = 1, \ldots, n,
\]

where \( \tau_{ij} \) is a number depending on \( W_{ni,j} \). The nearer \( i \) is to \( j \), the larger \( \tau_{ij} \) is. Another topic of future research is in allowing some spatial dependency in the covariates (SAR models) and the response (endogenous models) for greater generality. These topics will be of interest in future research.

5 Appendix

**Proposition 5.1** Under Assumptions A1-A6, for \( \theta \in \Theta \) and \( z \in Z \), the functions \( g_\theta(z) \) and \( \hat{g}_\theta(z) \), solutions of (8) and (14), respectively, satisfy

1. for all \( i, j = 0, 1, 2, i + j \leq 2 \),

\[
\frac{\partial^{i+j}}{\partial \theta^i \partial \theta^j} g_\theta(z) \quad \text{and} \quad \frac{\partial^{i+j}}{\partial \theta^i \partial \theta^j} \hat{g}_\theta(z)
\]

exist and are finite for all \( 1 \leq l, r \leq p + 1 \).

2. \( \sup_{\theta \in \Theta} \left\| \hat{g}_\theta - g_\theta \right\| \), \( \sup_{\theta \in \Theta} \max_{j=1,\ldots,p+1} \left\| \frac{\partial}{\partial \theta_j} (\hat{g}_\theta - g_\theta) \right\| \) and \( \sup_{\theta \in \Theta} \max_{1 \leq i, j \leq p+1} \left\| \frac{\partial^2}{\partial \theta_i \partial \theta_j} (\hat{g}_\theta - g_\theta) \right\| \),

are all order \( o_p(1) \) as \( n \to \infty \).

Without loss of generality, the proof of this proposition is ensured by Lemma 5.2 in the univariate case i.e., \( \Theta, Z \subset \mathbb{R} \).

The following lemma is useful in the proof of Lemma 5.2 It is an extension of Lemma 8 in Severini & Wong (1992) to spatially dependent data.
Lemma 5.1 Let $\zeta_\theta(Y_i)$ denote a scalar function of $Y_{in}$, $i = 1, \ldots, n$, $n = 1, 2, \ldots$, depending on a scalar parameter $\theta \in \Theta$, and for $j = 0, 1, 2$, let

$$
\zeta^{(j)}_\theta(Y_{in}) = \frac{\partial^j}{\partial \theta^j} \zeta_\theta(Y_{in}), \quad i = 1, \ldots, n, \ n = 1, 2, \ldots
$$

Let $f_i(\cdot)$ denote the density of $Z_{in}$ (given in Assumption A2), and let $\bar{f}(z) = \frac{1}{n} \sum_{i=1}^{n} f(z)$. Assume that

$H.1$ $\sup_{\theta} \sup_{1 \leq i \leq n} \left| \zeta^{(j)}_\theta(Y_{in}) \right| < \infty$ for $j = 0, \ldots, 3$.

$H.2$ For all $\theta \in \Theta$, $j = 0, 1, 2$, and $1 \leq i, l \leq n$:

$$
\left| \text{Cov} \left( \zeta^{(j)}_\theta(Y_{in}), K_{in}(z) \right) \right| \leq \left\{ \text{Var} \left( \zeta^{(j)}_\theta(Y_{in}) \right) \right\}^{1/2} \varphi (\|s_i - s_l\|),
$$

(27)

$$
\left| \text{Cov} \left( \zeta^{(j)}_\theta(Y_{in})K_{in}(z), \zeta^{(j)}_\theta(Y_{ln})K_{ln}(z) \right) \right| \leq \left\{ \text{Var} \left( \zeta^{(j)}_\theta(Y_{in}) \right) \text{Var} \left( \zeta^{(j)}_\theta(Y_{ln}) \right) \right\}^{1/2} \varphi (\|s_i - s_l\|),
$$

(28)

with $K_{in}(z) = K ((z - Z_{in})/b)$.

Let $m_\theta(z) = \mathbb{E} (\zeta_\theta(Y_{in})|Z_{in} = z)$ for $z \in Z$, and assume that $\frac{\partial^j}{\partial \theta^j} m_\theta(\cdot)$ is continuous on $Z$, $j = 0, 1, 2$.

For each fixed $\theta \in \Theta$ and $z \in Z$, let the kernel estimator $\hat{m}_\theta(z)$ of $m_\theta(z)$ be defined by

$$
\hat{m}_\theta(z) = \sum_{i=1}^{n} \zeta_\theta(Y_{in})K_{in}(z).
$$

If Assumptions A2, A4, and A5 are satisfied, then

$$
\sup_{\theta \in \Theta} \sup_{z \in Z} \left| \frac{\partial^j}{\partial \theta^j} \hat{m}_\theta(z) - \frac{\partial^j}{\partial \theta^j} m_\theta(z) \right| = o_p(1),
$$

for $j = 0, 1, 2$.

Lemma 5.1 generalizes Lemma 8 in Severini & Wong (1992) to spatially dependent data.

Proof of Lemma 5.1

We give the proof in the case where $j = 0$, corresponding to the study of the uniform consistency of the kernel estimator of the regression function of $\zeta_\theta(Y_{in})$ on $Z_{in}$. The other cases are similar to this case and thus are omitted.

Let

$$
\hat{v}_\theta(z) = \frac{1}{nb^d} \sum_{i=1}^{n} \zeta_\theta(Y_{in})K_{in}(z); \quad \hat{f}(z) = \frac{1}{nb^d} \sum_{i=1}^{n} K_{in}(z),
$$

17
\[
v_\theta(z) = m_\theta(z)\bar{f}(z).
\]

We have to show that
\[
\sup_{\theta} \sup_{z} |\hat{v}_\theta(z) - v_\theta(z)| = o_p(1) \quad (29)
\]
and
\[
\sup_{z} |\bar{f}(z) - \bar{f}(z)| = o_p(1) \quad (30)
\]
We give the proof of (29), and that of (30) is similar.

**Asymptotic behavior of \( |\hat{v}_\theta(z) - v_\theta(z)| \)**

Let us first consider the bias \( |E(\hat{v}_\theta(z)) - v_\theta(z)| \). We have
\[
E(\hat{v}_\theta(z)) = (nb^d)^{-1} \sum_{i=1}^{n} \int K\left( \frac{z - u}{b} \right) m_\theta(u) f_i(u) du
\]
\[
= b^{-d} \int v_\theta(u) K\left( \frac{z - u}{b} \right) du;
\]
\[
= \int v_\theta(z - bu) K(u) du
\]
thus,
\[
E(\hat{v}_\theta(z)) - v_\theta(z) = \int (v_\theta(z - bu) - v_\theta(z)) K(u) du = o(1)
\]
by Assumption A4, the continuity of \( f_i(\cdot) \) (see A2) and \( m_\theta(\cdot) \), and the compactness of \( Z \). Clearly, the bias term does not depend on \( \theta \) or \( z \).

Let us now treat \( |\hat{v}_\theta(z) - E(\hat{v}_\theta(z))| \). Consider the sum of variances
\[
S_n = (nb^d)^{-2} \sum_{i=1}^{n} \text{Var} (\zeta_\theta(Y_{in}) K_{in}(z)).
\]

We have
\[
\text{Var} (\zeta_\theta(Y_{in}) K_{in}(z)) \leq E \left( \zeta_\theta^2(Y_{in}) K_{in}^2(z) \right)
\]
\[
\leq CE \left( K_{in}^2(z) \right) = Cb^d \sum_{i=1}^{n} \int K^2(u) f_i(z - ub) du
\]
\[
= Cb^d \sup_{u} |K(u)|^2 \int f_i(z - ub) du = Cb^d \sup_{u} |K(u)|^2,
\]
\[
18
\]
because $\zeta_\theta(Y_{in})$ is bounded uniformly on $i$ and $\theta$ by assumption H.1, $\int f_i(z - ub)du \leq C$ (see assumption A2) and $\sup_n |K(u)|^2 < \infty$ (see Assumption A4 and the compactness of $Z$). Then, we have

$$S_n = O \left( (nb^d)^{-1} \right). \quad (32)$$

Now, consider the covariance term

$$R_n = (nb^d)^{-2} \sum_{i=1}^{n} \sum_{j=1 \atop j \neq i}^{n} \text{Cov} \left( \zeta_\theta(Y_{in})K_{in}(z), \zeta_\theta(Y_{jn})K_{jn}(z) \right).$$

Let us partition the spatial locations of the observations using

$$D_n = \{ 1 \leq i, j \leq n : \rho < \|s_i - s_j\| \leq c_n \}$$

with $c_n$ being the sequence of integers going to $\infty$, and let $\tilde{D}_n$ denote the complement of $D_n$ in the set of locations $\{s_i, i = 1, ..., n\}$. On the one hand, let

$$R_n^{(1)} = (nb^d)^{-2} \sum_{i,j \in D_n} |\text{Cov} \left( \zeta_\theta(Y_{in})K_{in}(z), \zeta_\theta(Y_{jn})K_{jn}(z) \right)| = (nb^d)^{-2} \sum_{i,j \in D_n} |A - B|,$$

with

$$|A| = |E(\zeta_\theta(Y_{in})K_{in}(z)\zeta_\theta(Y_{jn})K_{jn}(z))|,$$

$$\leq C \int K \left( \frac{z - u}{b} \right)K \left( \frac{z - v}{b} \right)f_{i,j}(u,v)dudv,$$

$$\leq C b^{2d} \left| \int K(u)K(v)f_{i,j}(z - bu, z - bv)dudv \right|,$$

$$\leq C b^{2d} \left( \sup_u |K(u)| \right)^2 \left| \int f_{i,j}(z - bu, z - bv)dudv \right| = C b^{2d},$$

by Assumption H.1, $\sup_u |K(u)| < \infty$ (Assumption A4 and the compactness of $Z$), with $f_{i,j}$ being the joint density (Assumption A2 and the compactness of $Z$). Note that the second term $B$ is

$$B = E(\zeta_\theta(Y_{in})K_{in}(z)) E(\zeta_\theta(Y_{jn})K_{jn}(z))$$

Using similar arguments as above, we have $|B| \leq C b^{2d}$ by Assumptions A2 and A4, the compactness of $Z$ and the continuity of $m_\theta(\cdot)$. Thus, we have

$$R_n^{(1)} \leq C n^{-2} \sum_{i,j \in D_n} C \leq C n^2 - \rho^2 \frac{c_n^2}{n} = O \left( \frac{c_n^2}{n} \right). \quad (33)$$

On the other hand, let
By Assumption H.2 combined with (31), we have for all \( \theta \in \Theta \) and \( i, j = 1, \ldots, n \),

\[
|\text{Cov} (\zeta_\theta(Y_{in})K_{in}(z), \zeta_\theta(Y_{jn})K_{jn}(z))| \leq C b^d \varphi(\|s_i - s_j\|).
\]

Then, we have

\[
R_n^{(2)} \leq C (nb^d)^{-2} \sum_{i,j \in D_n} |\zeta_\theta(Y_{in})K_{in}(z), \zeta_\theta(Y_{jn})K_{jn}(z)|.
\]

By Assumption H.2 combined with (31), we have for all \( \theta \in \Theta \) and \( i, j = 1, \ldots, n \),

\[
R_n^{(2)} \leq C (nb^d)^{-1} \sum_{i>c/n} i \varphi(i\rho).
\]

Thus, we derive the following result:

\[
R_n = R_n^{(1)} + R_n^{(2)} = O \left( n^{-1} \left\{ c_2^2 + b^{-d} \sum_{i>c/n} i \varphi(i\rho) \right\} \right).
\]

The following steps of the proof are inspired by the proof of Lemma 8 in Severini & Wong (1992) (p. 1800–1801). Let

\[
\tilde{v}_\theta(z) = \frac{1}{n} b^{-d} \sum_{i=1}^{n} \left\{ \zeta_\theta(Y_{in})K_{in}(z) - \mathbb{E} (\zeta_\theta(Y_{in})K_{in}(z)) \right\}.
\]

For some \( \epsilon > 0 \), Markov’s inequality yields

\[
P (|\tilde{v}_\theta(z)| > \epsilon) \leq \frac{R_n + S_n}{\epsilon^2}.
\]

Now, let \( \theta_1 \) and \( \theta_2 \) be two elements in \( \Theta \); because \( \mathbb{E} \left( \sup_{1 \leq i \leq n, n} |\zeta_\theta^{(1)}(Y_{in})| \right) < \infty \) (by H.1), there exists a random triangular array (see Severini & Wong, 1992 p.1801) \( \{W_{in}^{(1)}, 1 \leq i \leq n, n = 1, 2 \ldots \} \) not depending on \( \theta_1 \) and \( \theta_2 \) such that \( \sup_{1 \leq i \leq n, n} \mathbb{E} \left( |W_{in}^{(1)}| \right) < \infty \) and

\[
\sup_z |\tilde{v}_{\theta_1}(z) - \tilde{v}_{\theta_2}(z)| \leq \sup_z |K(z)| \frac{|\theta_2 - \theta_1|}{b^d} \frac{1}{n} \sum_{i=1}^{n} W_{in}^{(1)}.
\]

Similarly, for all \( z^{(1)} \) and \( z^{(2)} \) in \( \mathcal{Z} \), there exists a random triangular array \( \{W_{in}^{(2)}, 1 \leq i \leq n, n = 1, 2 \ldots \} \) not depending on \( z^{(1)} \) and \( z^{(2)} \) such that \( \sup_{1 \leq i \leq n, n} \mathbb{E} \left( |W_{in}^{(2)}| \right) < \infty \) and

\[
\sup_{\theta} |\tilde{v}_\theta(z^{(2)}) - \tilde{v}_\theta(z^{(1)})| \leq C \frac{|z^{(2)} - z^{(1)}|}{b^{d+1}} \frac{1}{n} \sum_{i=1}^{n} W_{in}^{(2)}.
\]
because $K(\cdot)$ is Lipschitzian (see Assumption H.2).
Hence, there exists a random triangular array $\{W_in, \, 1 \leq i \leq n, \, n = 1, 2, \ldots\}$ such that $\sup_{1 \leq i \leq n, \, n} \mathbb{E}(|W_in|) < \infty$ and
\[
\sup_{\|z^{(2)} - z^{(1)}\| < \delta_1} \sup_{|\theta_2 - \theta_1| < \delta_2} \left| \tilde{v}_{\theta_2}(z^{(2)}) - \tilde{v}_{\theta_1}(z^{(1)}) \right| \leq C \left( b^{-d} \delta_2 + b^{-(d+1)} \delta_1 \right) \frac{1}{n} \sum_{i=1}^{n} W_in,
\]
for some $\delta_1 > 0, \, \delta_2 > 0$ and large $n$.

Because $Z$ is compact, one can define a real number $\delta_1 > 0$, an integer $l_n$ such that $l_n \delta_1 < C$ with $l_n = \lfloor \gamma_n b^{-(d+1)} \rfloor$ and
\[
Z \subset \bigcup_{j=1}^{l_n} B(z^{(j)}, \delta_1),
\]
where $B(z, \delta)$ is the closed ball in $\mathbb{R}^d$ with center $z$ and radius $\delta > 0$.
In addition, because $\Theta$ is compact, one can cover it by $r_n = \lfloor \gamma_n b^{-d} \rfloor$ finite intervals of centers $\theta_i$ with the same half length $\delta_2 = O(1/r_n)$.
With these coverings, we have
\[
\mathbb{P} \left( \sup_{\theta, z} |\tilde{v}_\theta(z)| > \epsilon \right) \leq \mathbb{P} \left( \max_{j \leq r_n} \max_{k \leq l_n} |\tilde{v}_{\theta_j}(z^{(k)})| > \epsilon/2 \right) + \mathbb{P} \left( \sup_{\|z^{(2)} - z^{(1)}\| < \delta_1} \sup_{|\theta_2 - \theta_1| < \delta_2} \left| \tilde{v}_{\theta_2}(z^{(2)}) - \tilde{v}_{\theta_1}(z^{(1)}) \right| > \epsilon/2 \right)
\]
\[
\leq r_n l_n \mathbb{P} (|\tilde{v}_\theta(z)| > \epsilon/2) + Cb^{-d} (\delta_2 + \delta_1 b^{-1}) = C r_n l_n (S_n + R_n) + Cb^{-d} (\delta_2 + \delta_1 b^{-1})
\]
\[
=: I^{(1)} + I^{(2)} + I^{(3)},
\]
where
\[
I^{(1)} = O \left( \frac{\gamma_n^2}{nb^{2d+1}} \left( c_n^2 + b^{-d} \sum_{i \geq c_n / \rho} \varphi(ip) \right) \right), \quad I^{(2)} = O \left( \gamma_n^{-1} \right), \quad I^{(3)} = O \left( \frac{\gamma_n^2}{nb^{2d+1}} \right).
\]
If we take $c_n = o(b^{-d/2})$ and $\gamma_n^2 = o(nb^{2d+1})$, then $I^{(1)}, I^{(2)}$ and $I^{(3)}$ are all of order $o(1)$ by Assumption A5 and by the fact that $\varphi(t) \to 0$ as $t \to \infty$ by Assumption A3. This yields the proof. \qed

**Lemma 5.2** For each $\theta \in \Theta$ and $z \in Z$, let
\[
H(\eta; \theta, z) = \mathbb{E}_0 (h^{\theta, \eta}_{in}(Y_{in} | X_{in}, Z_{in}) | Z_{in} = z), \, 1 \leq i \leq n, \, n = 1, 2, \ldots
\]
where $\eta = g(z), \, g \in G$ and $h^{\theta, \eta}_{in}(\cdot | \cdot, \cdot)$ is defined in Assumption A3.
**Condition I.** For fixed but arbitrary $\theta_1 \in \Theta$ and $\eta_1 \in \Pi$ with $\Pi = g_0(Z)$, let

$$\vartheta(\theta, \eta) = \int h_{\theta, \eta}^{\theta_1, \eta_1}(y|x, z) \exp(h_{\theta, \eta}^{\theta_1, \eta_1}(y|x, z)) dy, \quad \theta \in \Theta, \eta \in \Pi, (x, z) \in \mathcal{X} \times \mathcal{Z}$$

where $\{\exp(h_{\theta, \eta}^{\theta_1, \eta_1}(y|x, z)), \theta \in \Theta, \eta \in \Pi\}$ denotes the family of conditional density functions (indexed by the parameters $\theta$ and $\eta$) of $Y_{in}$ given $(X_{in}, Z_{in}) = (x, z) \in \mathcal{X} \times \mathcal{Z}$. For each $\theta \neq \theta_1$, assume that $\vartheta(\theta, \eta) < \vartheta(\theta_1, \eta_1)$.

**Condition S.** Let $\tilde{p} = p + 1$, and for all nonnegative integers $j_1, \ldots, j_{\tilde{p}} = 0, 1, 2$ and $r = 0, \ldots, 4$, such that $j_1 + \cdots + j_{\tilde{p}} + r \leq 6$, assume that the derivative

$$\frac{\partial^{j_1 + \cdots + j_{\tilde{p}} + r} h_{\theta, \eta}^{\theta_1, \eta_1}(y|x, z)}{\partial \theta_1^{j_1} \cdots \partial \theta_{\tilde{p}}^{j_{\tilde{p}}} \partial \eta^r}(y|x, z),$$

exists for almost all $y$ and that

$$E_0 \left( \sup_{i, n} \sup_{\theta \in \Theta} \left| \sup_{g \in G} \left| \frac{\partial^{j_1 + \cdots + j_{\tilde{p}} + r} h_{\theta, \eta}^{\theta_1, \eta_1}(Y_{in}|X_{in}, Z_{in})}{\partial \theta_1^{j_1} \cdots \partial \theta_{\tilde{p}}^{j_{\tilde{p}}} \partial \eta^r} \right| \right|^2 \right) < \infty,$$

with $\eta_i = g(Z_{in})$.

Assume that

$$\sup_{z} \sup_{\theta} \sup_{\eta} \left| \frac{\partial^j}{\partial \eta^k} H^{(k)}(\eta; \theta, z) \right| < \infty, \quad (37)$$

for $j = 0, 1, 2$ and $k = 2, 3, 4$ such that $j + k \leq 4$, with

$$H^{(k)}(\eta; \theta, z) = \frac{\partial^k}{\partial \eta^k} H(\eta; \theta, z).$$

Let

$$\hat{H}(\eta; \theta, z) = \sum_{i=1}^{n} \frac{h_{\theta, \eta}^{\theta_1, \eta_1}(Y_{in}|X_{in}, z) K_{in}(z)}{\sum_{i=1}^{n} K_{in}(z)};$$

then, $\hat{g}_\theta(z)$ is a solution of $\hat{H}^{(1)}(\eta; \theta, z) = 0$ with respect to $\eta$ for each fixed $\theta \in \Theta$ and $z \in \mathcal{Z}$.

If we assume that Assumptions A1-A6 are satisfied, then we have, for all $j = 0, 1, 2$,

$$\sup_{\theta} \sup_{z} \left| \frac{\partial^j}{\partial \theta^j} (\hat{g}_\theta(z) - g_\theta(z)) \right| = o_p(1). \quad (38)$$

The assumptions used in the previous lemma are satisfied under the conditions used in the main results. **Condition I** is needed to ensure the identifiability of the arbitrary parameter $\theta_1$ (it plays the role of the true parameter $\theta_0$). This condition is verified when $\theta_1 = \theta_0$ by the identifiability of our model [1]. **Condition S** allows integrals to be interchanged with differentiation; this will be combined with the implicit function theorem (see Saaty & Bram [2012]) to ensure the differentiability...
of \( \hat{g}_\theta(z) \) with respect to \( \theta \).

Knowing that \( \Phi(\cdot) \) is a smooth function on \( \mathbb{R} \) and \( h_{in}^{\theta,\eta}(\cdot, \cdot) \) is

\[
h_{in}^{\theta,\eta}(Y_{in}|X_{in}, Z_{in}) = Y_{in} \log \left( \frac{\Phi(G_{in}(\theta, \eta))}{1 - \Phi(G_{in}(\theta, \eta))} \right) - \log (1 - \Phi(G_{in}(\theta, \eta))) ,
\]

**Condition S** and Assumption \((37)\) are satisfied under the continuity condition of \( \Phi(\cdot) \) and \( \phi(\cdot) \), Assumption A9 and the compactness of \( \mathcal{X} \) and \( \mathcal{Z} \).

**Proof of Lemma 5.2**

The proof of this lemma is similar to that of Lemma 5 in [Severini & Wong (1992)](#). Let us follow similar lines as in the proof of Lemma 5.1 above, replacing \( \zeta_j(\cdot) \theta \) by \( \zeta_{j,k}(\cdot) \theta, \eta \) in

\[
\zeta_{j,k}(Y_{in}, X_{in}) = \frac{\partial^j}{\partial \theta^j} \frac{\partial^k}{\partial \eta^k} h_{in}^{\theta,\eta}(Y_{in}|X_{in}, z).
\]

and Assumptions **H.1** and **H.2** in Lemma 5.1 by the following:

**H.1’** \[ \sup_{\theta, \eta} \sup_{i,n} \| \zeta_{j,k}(Y_{in}, X_{in}) \| < \infty, \text{ for } j = 0, \ldots, 3, \ k = 0, \ldots, 5 \]

**H.2’** For all \( k = 0, \ldots, 4, \ j = 0, 1, 2 \) and \( \theta \in \Theta, \ z \in \mathcal{Z}, \) \((27)\) is satisfied and \((28)\) holds with \( \zeta_j(\cdot) (Y_{in}) \) replaced by \( \zeta_{j,k}(\cdot) (Y_{in}, X_{in}). \)

Under the conditions used in the lemma, it is clear that **H.1’** is verified, and **H.2’** is also satisfied by Assumption A3 (in particular, conditions \((19)\)).

Using the results of Lemma 5.1 we have the following for all \( j = 0, 1, 2 \):

\[
\sup_{\theta, \eta, z} \left| \frac{\partial^j}{\partial \theta^j} \left( \hat{H}_n^{(1)}(\eta; \theta, z) - H^{(1)}(\eta; \theta, z) \right) \right| \leq o_p(1), \tag{39}
\]

\[
\sup_{\theta, \eta, z} \left| \frac{\partial^j}{\partial \theta^j} \left( \hat{H}_n^{(2)}(\eta; \theta, z) - H^{(2)}(\eta; \theta, z) \right) \right| \leq o_p(1), \tag{40}
\]

\[
\sup_{\theta, \eta, z} \left| \frac{\partial^j}{\partial \theta^j} \left( \hat{H}_n^{(3)}(\eta; \theta, z) - H^{(3)}(\eta; \theta, z) \right) \right| \leq o_p(1), \tag{41}
\]

\[
\sup_{\theta, \eta, z} \left| \frac{\partial^j}{\partial \theta^j} \left( \hat{H}_n^{(4)}(\eta; \theta, z) - H^{(4)}(\eta; \theta, z) \right) \right| \leq o_p(1). \tag{42}
\]

Under Assumption A1, for any \( \epsilon > 0 \), there exists \( \gamma > 0 \) such that

\[
P \left( \sup_{\theta, z} |\hat{g}_\theta(z) - g_\theta(z)| > \epsilon \right) \leq P \left( \sup_{\theta, z} |H^{(1)}(\theta, \hat{g}_\theta(z); z)| > \gamma \right) \leq P \left( \sup_{\theta, z} |\hat{H}^{(1)}(\eta; \theta, z) - H^{(1)}(\eta; \theta, z)| > \gamma \right).
\]

\[
23
\]
Hence,  
\[
\sup_{\theta, z} |\hat{g}_\theta(z) - g_\theta(z)| = o_p(1) \tag{43}
\]

The remainder of the proof is very similar to that of Lemma 5 in Severini & Wong (1992) (p. 1798–1799); for the sake of completeness, we present the details. We have by **Condition I**
\[
\inf_{\theta} \inf_{z} -H^{(2)}(g_\theta(z); \theta, z) > 0.
\]

In addition, by **Condition S**, for every \( \delta > 0 \), there exists \( \epsilon > 0 \) such that
\[
\sup \sup_{\theta} \sup_{z} \sup_{\eta_1, \eta_2: |\eta_1 - \eta_2| \leq \epsilon} |H^{(2)}(\eta_2; \theta, z) - H^{(2)}(\eta_1; \theta, z)| < \delta.
\]
Hence, there exists \( \epsilon > 0 \) such that
\[
\inf_{\theta} \inf_{z} \inf_{|\eta - g_\theta(z)| \leq \epsilon} |H^{(2)}(\eta; \theta, z)| > 0. \tag{44}
\]
Because \( g_\theta(z) \) and \( \hat{g}_\theta(z) \) satisfy
\[
H^{(1)}(g_\theta(z); \theta, z) = 0 \quad \text{and} \quad \hat{H}^{(1)}(\hat{g}_\theta(z); \theta, z) = 0,
\]
respectively, for each \( \theta \) and \( z \), it follows that
\[
0 = \hat{H}^{(1)}(\hat{g}_\theta(z); \theta, z) - H^{(1)}(g_\theta(z); \theta, z) = H^{(1)}(\hat{g}_\theta(z); \theta, z) + H^{(1)}(g_\theta(z); \theta, z) - H^{(1)}(g_\theta(z); \theta, z)
\]
\[
= r_n(\theta, z) + d_n(\theta, z) (\hat{g}_\theta(z) - g_\theta(z)), \tag{45}
\]
for each \( \theta, z \), where
\[
r_n(\theta, z) = H^{(1)}(\hat{g}_\theta(z); \theta, z) - H^{(1)}(\hat{g}_\theta(z); \theta, z)
\]
and
\[
d_n(\theta, z) = \int_{0}^{1} H^{(2)}(tg_\theta(z) + (1-t)\hat{g}_\theta(z); \theta, z) dt.
\]

Note that by (44) and \( \sup_{\theta} \|\hat{g}_\theta - g_\theta\| = o_p(1) \), we have
\[
\lim \inf \inf_{\theta} \inf_{z} |\hat{H}^{(2)}(\hat{g}_\theta(z); \theta, z)| > 0 \quad \text{and} \quad \lim \inf \inf_{\theta} \inf_{z} |d_n(\theta, z)| > 0 \quad \text{as} \quad n \to \infty. \tag{46}
\]
Because
\[
\hat{H}^{(1)}(\hat{g}_\theta(z); \theta, z) = 0,
\]
for all \( \theta, z \), we have
\[
\hat{H}^{(2)}(\hat{g}_\theta(z); \theta, z) \frac{\partial \hat{g}_\theta(z)}{\partial \theta} + \frac{\partial \hat{H}^{(1)}(\hat{g}_\theta(z); \theta, z)}{\partial \theta} = 0.
\]
Then, we can deduce from (46), (39), and (40) that
\[
\sup_{\theta} \sup_{z} \left| \frac{\partial \hat{g}_\theta(z)}{\partial \theta} \right| = O_p(1).
\]
Similarly, we have
\[ \sup_{\theta} \sup_{z} \left| \frac{\partial^j \tilde{g}_{\theta}}{\partial \theta^j} (z) \right| = O_p(1), \quad j = 0, 1, 2. \tag{47} \]

Then, (47) and (39)–(42) yield
\[ \sup_{\theta} \sup_{z} \left| \frac{\partial^j r_n(\theta, z)}{\partial \theta^j} \right| = o_p(1), \quad \text{and} \quad \sup_{\theta} \sup_{z} \left| \frac{\partial^j d_n(\theta, z)}{\partial \theta^j} \right| = O_p(1), \quad j = 0, 1, 2. \tag{48} \]

Now, differentiating (45) with respect to \( \theta \) yields
\[ \frac{\partial r_n}{\partial \theta}(\theta, z) + (\tilde{g}_{\theta}(z) - g_{\theta}(z)) \frac{\partial d_n}{\partial \theta}(\theta, z) + d_n(\theta, z) \left( \frac{\partial \tilde{g}_{\theta}}{\partial \theta}(z) - \frac{\partial g_{\theta}}{\partial \theta}(z) \right) = 0. \tag{49} \]

Then, by (39)–(48),
\[ \sup_{\theta} \sup_{z} \left| \frac{\partial \tilde{g}_{\theta}}{\partial \theta}(z) - \frac{\partial g_{\theta}}{\partial \theta}(z) \right| = o_p(1). \]

One can similarly obtain
\[ \sup_{\theta} \sup_{z} \left| \frac{\partial^2 \tilde{g}_{\theta}}{\partial \theta^2}(z) - \frac{\partial^2 g_{\theta}}{\partial \theta^2}(z) \right| = o_p(1). \]

This completes the proof. \( \square \)

**Proof of Theorem 2.1**

By Lemmas 5.3 and 5.4, \( Q_n \) converges to \( Q \) in probability uniformly, i.e.,
\[ \sup_{\theta \in \Theta} |Q_n(\theta, g_{\theta}) - Q(\theta, g_{\theta})| = o_p(1). \tag{50} \]

This result allows one to obtain
\[ |Q(\tilde{\theta}, g_{\tilde{\theta}}) - Q(\theta_0, g_{\theta_0})| = o_p(1). \tag{51} \]

Indeed, using \( |\sup a - \sup b| \leq \sup |a - b| \), we have
\[
|Q(\tilde{\theta}, g_{\tilde{\theta}}) - Q(\theta_0, g_{\theta_0})| \leq \sup_{\theta} |Q_n(\theta, g_{\theta}) - Q(\theta_0, g_{\theta})| + |\sup_{\theta} Q_n(\tilde{\theta}, g_{\tilde{\theta}}) - \sup_{\theta} Q_n(\theta, g_{\theta})| \\
\leq 2 \sup_{\theta} |Q_n(\theta, g_{\theta}) - Q(\theta, g_{\theta})| + 2 \sup_{\theta} |Q_n(\theta, g_{\theta}) - Q(\theta, g_{\theta})| \\
= o_p(1),
\]

25
by Lemma 5.5 and \( \sup_{\theta} Q(\theta, g_{\theta}) = Q(\theta_0, g_0) \) (see Assumption A8).

By Assumption A8, we have for a given \( \theta \in \Theta \) that there exists \( \varepsilon > 0 \) and an open neighbourhood \( N_{\theta} \) such that
\[
\inf_{\theta_1 \in N_{\theta}} |Q(\theta_1, g_{\theta_1}) - Q(\theta_0, g_0)| > \varepsilon. \tag{52}
\]
This and \( Q_n(\theta, g_\theta) - Q(\theta, g_\theta) = o_p(1) \), as \( n \to \infty \).

Let \( N_0 \) be an open neighbourhood of \( \theta_0 \), and consider the compact set \( \Theta_0 = \Theta \setminus N_0 \). Let \( \{N_{\theta} : \theta \in \Theta, \theta \neq \theta_0\} \) denote the open covering of \( \Theta_0 \) by the procedure given above (each neighbourhood \( N_{\theta} \) satisfies \( 52 \)). By the compactness of \( \Theta_0 \), let \( \{N_{\theta_1}, \ldots, N_{\theta_r}\} \) be a finite sub-covering; then,
\[
\mathbb{P}_0 \left( \hat{\theta} \notin N_0 \right) = \mathbb{P}_0 \left( \hat{\theta} \in \Theta_0 \right) \leq \sum_{j=1}^{r} \mathbb{P}_0 \left( \hat{\theta} \in N_{\theta_j} \right) \to 0, \text{ as } n \to \infty, \tag{53}
\]
by \( 53 \). Therefore, we can conclude that
\[
\hat{\theta} - \theta_0 = o_p(1), \quad \text{as} \quad n \to \infty.
\]
This yields the proof of Theorem 2.1. □

**Lemmas 5.3-5.5**

We use the following notation:
\[
\eta_i = g(Z_{in}); \quad \bar{U}_{in} = U_{in}(\theta, \eta_i); \quad \Phi_{in} = \Phi(G_{in}(\theta, g_{\theta})); \quad \Lambda_{in} = \Lambda(G_{in}(\theta, g_{\theta})),
\]
for all \( \theta \in \Theta, 1 \leq i \leq n, n = 1, 2, \ldots, \) with \( \Lambda(\cdot) = \phi(\cdot)/\Phi(\cdot)(1 - \Phi(\cdot)) \).

The partial derivatives of \( S_n(\theta, g) \) with respect to \( g \) of order \( s = 1, 2, \ldots, \) for any functions \( v_1, \ldots, v_s \) in \( G \), are given by
\[
\frac{\partial^s S_n(\theta, g)}{\partial g^s}(v_1, \ldots, v_s) = n^{-1} \sum_{i=1}^{n} \xi_{in} \frac{\partial^s \bar{U}_{in}(\theta, \eta_i)}{\partial \eta^s}(v_1(Z_{in}) \cdots v_s(Z_{in})).
\]

**Lemma 5.3** Under Assumptions A3, A6 and A9, we have for all \( \theta \in \Theta \),
\[
S_n(\theta, g_\theta) - S'(\theta, g_\theta) = o_p(1). \tag{54}
\]

In addition, we have
\[
Q_n(\theta, g_\theta) - Q(\theta, g_\theta) = o_p(1), \tag{55}
\]
if \( M_n - M = o_p(1) \).

Note that if Assumption A10 is satisfied, then \( M_n - M = o_p(1) \).
Proof of Lemma 5.3

Let us start with the proof of (54). We remark that

\[ S_n(\theta, g_\theta) = n^{-1} \xi^T \tilde{U}_n(\theta, g_\theta) = n^{-1} \sum_{i=1}^{n} \xi_i \tilde{U}_{in}(\theta, g_\theta), \]

where \( \xi_i \) is the \( q \times 1 \) vector representing the \( i \)th row in the matrix of instrumental variables. By definition (see (13)), we have \( E_0(S_n(\theta, g_\theta)) - S(\theta, g_\theta) = o(1) \). Then, it suffices to show that

\[ S_n(\theta, g_\theta) - E_0(S_n(\theta, g_\theta)) = o_p(1). \] (56)

Indeed (omitting the \((\theta, g_\theta)\)-arguments to simplify the notation), we have

\[
E_0\left( \|S_n - E_0(S_n)\|^2 \right) = n^{-2} \sum_{i,j=1}^{n} E_0\left( \left( \xi_{in} \tilde{U}_{in} - E_0(\xi_{in} \tilde{U}_{in}) \right)^T \left( \xi_{jn} \tilde{U}_{jn} - E_0(\xi_{jn} \tilde{U}_{jn}) \right) \right) \\
\leq n^{-2} \sum_{i,j=1}^{n} \alpha_{ijn} \sum_{t=1}^{q} \left\{ \text{Var}_0\left( \xi_{itn} \tilde{U}_{in} \right) \text{Var}_0\left( \xi_{jtn} \tilde{U}_{jn} \right) \right\}^{1/2} \\
\leq C n^{-2} \sum_{i,j=1}^{n} \alpha_{ijn} = O\left( n^{-1} \sum_{s=1}^{\sqrt{n}} s \varphi(s) \right) = o(1),
\]

because \( \text{Var}_0(\xi_{itn} \tilde{U}_{in}) \) is bounded uniformly on \( \theta, i, \) and \( t = 1, \ldots, q \) (by Assumption A6) and because \( \varphi(s) \to as s \to +\infty \) (by assumption A3). This completes the proof of (56) and thus that of (54).

The proof of (55) is made straightforward by combining (54) with Assumption A10. \( \square \)

Lemma 5.4 Under Assumptions A6-A9, we have \( S_n(\cdot, \cdot) - S(\cdot, \cdot) \) is stochastically equicontinuous on \( \Theta \).

In addition, if \( M_n - M = o_p(1) \), then we have \( Q_n(\cdot, \cdot) - Q(\cdot, \cdot) \) is also stochastically equicontinuous on \( \Theta \).

Proof of Lemma 5.4

Stochastic equicontinuity in \( \Theta \) can be obtained by proving that \( S_n(\theta, g_\theta) \) satisfies a stochastic Lipschitz-type condition on \( \theta \) (see Mátýás, 1999, p. 17).

Let us show that \( S_n(\cdot, \cdot) \) is stochastically equicontinuous on \( \theta \) because \( S(\cdot, \cdot) \) is continuous by Assumption A8. It suffices to show that (Andrews, 1992) for each \( \theta_1, \theta_2 \in \Theta \):

\[
\|S_n(\theta_1, g_{\theta_1}) - S_n(\theta_2, g_{\theta_2})\| = O_p(\|\theta_1 - \theta_2\|).
\]

(57)
Indeed, for θ₁, θ₂ ∈ Θ,
\[
\|S_n(θ_1, g_{θ_1}) - S_n(θ_2, g_{θ_2})\| \leq n^{-1} \sup_{i, n} \| ξ_{in}\| \sum_{i=1}^{n} \left\| \tilde{U}_{in}(θ_1, g_{θ_1}) - \tilde{U}_{in}(θ_2, g_{θ_2}) \right\|
\]
\[
\leq n^{-1} \sup_{i, n} \| ξ_{in}\| \sum_{i=1}^{n} \left\{ \sup_{θ, η} \left\| \frac{∂\tilde{U}_{in}}{∂θ}(θ, η) \right\| \|θ_1 - θ_2\| + \sup_{θ, η} \left\| \frac{∂\tilde{U}_{in}}{∂η}(θ, η) \right\| \|g_{θ_1} - g_{θ_2}\| \right\}
\]
\[
\leq n^{-1} \sup_{i, n} \| ξ_{in}\| \sum_{i=1}^{n} \left\{ \sup_{θ, η} \left\| \frac{∂\tilde{U}_{in}}{∂θ}(θ, η) \right\| + \sup_{θ} \left\| \frac{∂g_θ}{∂θ} \right\| \sup_{θ, η} \left\| \frac{∂\tilde{U}_{in}}{∂η}(θ, η) \right\| \right\} \|θ_1 - θ_2\|.
\]

By Assumption A6 and Proposition 5.1, we have that sup_i, n \| ξ_{in}\| is bounded and sup_θ \left\| \frac{∂g_θ}{∂θ} \right\| is finite, respectively. Then, we have to show that
\[
\sum_{i=1}^{n} \sup_{θ, η} \left\| \frac{∂\tilde{U}_{in}}{∂θ}(θ, η) \right\| + \sup_{θ, η} \left\| \frac{∂\tilde{U}_{in}}{∂η}(θ, η) \right\| = O_p(1);
\] (58)

This is equivalent to
\[
\sup_{θ, η} \left\| \frac{∂\tilde{U}_{in}}{∂θ}(θ, η) \right\| = O_p(1), \quad 1 \leq i \leq n, n = 1, 2, \ldots
\] (59)

and
\[
\sup_{θ, η} \left\| \frac{∂\tilde{U}_{in}}{∂η}(θ, η) \right\| = O_p(1), \quad 1 \leq i \leq n, n = 1, 2, \ldots
\] (60)

Let us prove (59) in the following. The proof of (60) follows the same lines and is thus omitted.

Proof of (59):
Recall that
\[
Λ(t) = \frac{φ(t)}{Φ(t)(1 - Φ(t))}.
\]

By definition, we have
\[
\tilde{U}_{in}(θ, η) = Λ(G_{in}(θ, η)) (Y_{in} - Φ(G_{in}(θ, η))) ,
\]
with G_{in}(θ, η) = a_{in}(θ)b_{in}(θ, η), where a_{in}(·) and b_{in}(·) are defined by
\[
a_{in}(θ) := (v_{in}(λ))^{-1} \quad \text{and} \quad b_{in}(θ, η) := X_{in}^T β + η, \quad 1 \leq i \leq n, n = 1, 2, \ldots
\] (61)
with \( \theta^T = (\beta^T, \lambda) \). We have

\[
\frac{\partial \tilde{U}_{in}}{\partial \theta} (\theta, \eta) = \left\{ \Lambda' (G_{in}(\theta, \eta))(Y_{in} - \Phi(G_{in}(\theta, \eta))) - \Lambda(G_{in}(\theta, \eta))\phi(G_{in}(\theta, \eta)) \right\} \frac{\partial G_{in}}{\partial \theta}(\theta, \eta)
\]

(62)

where \( \Lambda' (\cdot) \) denotes the derivative of \( \Lambda(\cdot) \).

Let us first establish that

\[
\sup_{t \in \mathcal{M}, y \in \{0, 1\}} \| \Lambda'(t)(y - \Phi(t)) - \phi(t)\Lambda(t) \| < \infty,
\]

(63)

which is equivalent to showing that \( \Lambda'(t) \) and \( \phi(t)\Lambda(t) \) are bounded uniformly in \( t \in \mathcal{M} \) (the definition of \( \mathcal{M} \) is given in A.1). Because \( \phi'(t) = -t\phi(t) \), we can rewrite \( \Lambda'(t) \) as

\[
\Lambda'(t) = \frac{1}{\Phi(t)} \left\{ \frac{\phi(t)}{1 - \Phi(t)} \left( \frac{\phi(t)}{1 - \Phi(t)} - t \right) \right\} - \frac{\phi^2(t)}{\Phi^2(t)(1 - \Phi(t))}.
\]

(64)

Notice that \( \Lambda(\cdot) \) and \( \Lambda'(\cdot) \) may be unbounded only at \( \pm \infty \), and because \( \mathcal{M} \) is a compact subset of \( \mathbb{R} \), these functions are bounded on \( \mathbb{R} \). This establishes (63).

We remark that

\[
\left\| \frac{\partial G_{in}(\theta, \eta)}{\partial \theta} \right\| \leq \left\| \frac{\partial a_{in}(\theta)}{\partial \theta} \right\| \| b_{in}(\theta, \eta) \| + \left\| \frac{\partial b_{in}(\theta, \eta)}{\partial \theta} \right\| |a_{in}(\theta)|.
\]

(65)

Then, \( \left\| \frac{\partial G_{in}(\theta, \eta)}{\partial \theta} \right\| \) is bounded uniformly in \( i, n, \theta, \eta \) by Assumptions A6 and A9 and the compactness of \( \Theta \) (see assumption A7). This completes the proof of (59); hence, (57) is proved.

\[\square\]

**Lemma 5.5** Under the assumptions of Proposition 5.1 and Assumptions A6 and A9, we have

\[
\sup_{\theta \in \Theta} \| S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta) \| = o_p(1).
\]

(66)

If in addition \( M_n - M = o_p(1) \), then we have

\[
\sup_{\theta \in \Theta} \| Q_n(\theta, \hat{g}_\theta) - Q_n(\theta, g_\theta) \| = o_p(1).
\]

(67)

**Proof of Lemma 5.5**

Let us prove (66). For each \( \theta \in \Theta \)

\[
\| S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta) \| = n^{-1} \left\| \sum_{i=1}^{n} \xi_i \left( \tilde{U}_{in}(\theta, \hat{g}_\theta) - \tilde{U}_{in}(\theta, g_\theta) \right) \right\|
\]

\[
\leq n^{-1} \sum_{i=1}^{n} \sup_{i, n} \| \xi_i \| \| \tilde{U}_{in}(\theta, \hat{g}_\theta) - \tilde{U}_{in}(\theta, g_\theta) \|
\]

\[
\leq n^{-1} \sum_{i=1}^{n} \sup_{i, n} \| \xi_i \| \sup_{\theta, \eta} \left\| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right\| \| \hat{g}_\theta - g_\theta \|
\]

\[
= o_p(1),
\]
because $\sup_{i,n} \|\xi_{in}\| = O_p(1)$ (by Assumption A6), $\sup_{\theta} \|\hat{g}_\theta - g_\theta\| = o_p(1)$ (see Proposition 5.1) and $\sup_{\theta,\eta} \left| \frac{\partial Q_{in}}{\partial \eta} (\theta, \eta) \right| = O_p(1)$ uniformly on $i$ and $n$ (see the proof of Lemma 5.4).

The proof of (67) is made trivial by combining (66) with Assumption A10. □

**Proof of Theorem 2.2**

Recall that $\frac{d}{d\theta} Q_n(\theta, g_\theta)$ denotes differentiation with respect to $\theta$, while $\frac{\partial}{\partial \theta} Q_n(\theta, g_\theta)$ denotes the partial derivative with respect to $\theta$.

Using a Taylor’s series expansion and the fact that $\frac{d}{d\theta} Q_n(\theta, \hat{g}_\theta) \bigg|_{\theta = \hat{\theta}} = 0$, we have

$$\hat{\theta} - \theta_0 = - \left\{ \frac{d^2}{d\theta d\theta^T} Q_n(\theta, \hat{g}_\theta) \right\}^{-1} \left\{ \frac{d}{d\theta} Q_n(\theta, \hat{g}_\theta) \right\} \bigg|_{\theta = \theta_0}, \quad (68)$$

for some $\theta^*$ between $\theta_0$ and $\hat{\theta}$.

First, we would like to replace $\hat{g}_\theta(.)$ in (68) with $g_\theta(.)$. For this, let us show that $\frac{d}{d\theta} Q_n(\theta, \hat{g}_\theta)$ (resp. $\frac{d^2}{d\theta d\theta^T} Q_n(\theta, \hat{g}_\theta)$) and $\frac{d}{d\theta} Q_n(\theta, g_\theta)$ (resp. $\frac{d^2}{d\theta d\theta^T} Q_n(\theta, g_\theta)$) have the same behavior as a function of $\theta$ in a neighbourhood of $\theta_0$. In other words,

$$\sup_{\theta} \left\| \frac{d^2}{d\theta d\theta^T} Q_n(\theta, \hat{g}_\theta) - \frac{d^2}{d\theta d\theta^T} Q_n(\theta, g_\theta) \right\| = o_p(1) \quad (69)$$

and

$$\frac{d}{d\theta} Q_n(\theta, \hat{g}_\theta) \bigg|_{\theta = \theta_0} - \frac{d}{d\theta} Q_n(\theta, g_\theta) \bigg|_{\theta = \theta_0} = o_p(1). \quad (70)$$

We remark that (69) is equivalent to

$$\sup_{\theta} \left\| \frac{d}{d\theta} S_n(\theta, \hat{g}_\theta) - \frac{d}{d\theta} S_n(\theta, g_\theta) \right\| = o_p(1) \quad (71)$$

and

$$\sup_{\theta} \left\| \frac{d^2}{d\theta d\theta^T} S_n(\theta, \hat{g}_\theta) - \frac{d^2}{d\theta d\theta^T} S_n(\theta, g_\theta) \right\| = o_p(1) \quad (72)$$

by (11) (because $M_n - M = o_p(1)$ thanks to Assumption A10) and

$$\sup_{\theta} \| S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta) \| = o_p(1)$$
(see Lemma 5.5). Then, (71) and (72) follow immediately from Lemma 5.8.

To prove (70), we have the following Taylor expansion

\[
\frac{d}{d\theta} (Q_n(\theta, \hat{g}_\theta) - Q_n(\theta, g_\theta)) = \frac{d}{d\theta} \left( \frac{\partial Q_n(\theta, g_\theta) (\hat{g}_\theta - g_\theta) + \tilde{r}_n(\theta)}{\partial g} \right),
\]

where

\[
\tilde{r}_n(\theta) = \int_0^1 \frac{\partial^2 Q_n}{\partial g^2} (\theta, g_\theta + t(\hat{g}_\theta - g_\theta))(\hat{g}_\theta - g_\theta)^2 dt.
\]

We have

\[
\frac{d}{d\theta} \tilde{r}_n(\theta)\bigg|_{\theta = \theta_0} = o_p(1),
\]

using similar arguments as for the terms \( \frac{d^j}{d\theta^j} r_n^{(1)}(\theta) \) for \( j = 0, 1 \) and \( \frac{d^2}{d\theta^2} r_n^{(1)}(\theta) \) in Lemma 5.8 below (see (90)). Therefore, we obtain

\[
\frac{d}{d\theta} Q_n(\theta, \hat{g}_\theta)\bigg|_{\theta = \theta_0} = \frac{d}{d\theta} Q_n(\theta, g_\theta)\bigg|_{\theta = \theta_0} (\hat{g}_\theta - g_\theta) + \frac{d}{d\theta} r_n(\theta)\bigg|_{\theta = \theta_0},
\]

by Lemma 5.7 where \( g'_\theta(\cdot) = \frac{g_\theta(\cdot)}{\partial \theta^T} \bigg|_{\theta = \theta_0} \).

Consequently, we obtain

\[
\hat{\theta} - \theta_0 = -\left\{ \frac{d^2}{d\theta^2} Q_n(\theta, g_\theta)\bigg|_{\theta = \theta^*} \right\}^{-1} \left\{ \frac{d}{d\theta} Q_n(\theta, g_\theta)\bigg|_{\theta = \theta_0} \right\} + o_p(1) \tag{73}
\]

where \( \theta^* \) is between \( \hat{\theta} \) and \( \theta_0 \).

Let us show that for each \( \theta^* \) lying between \( \theta_0 \) and \( \hat{\theta} \),

\[
\frac{d^2}{d\theta^2} Q_n(\theta, g_\theta)\bigg|_{\theta = \theta^*} = 2 B_2(\theta_0) + o_p(1),
\]

to replace the Hessian matrix in the right-hand side of (73) by its limit \( B_2(\theta_0) \).

Let us consider the first- and second-order differentials of \( Q_n(\theta, g_\theta) \) with respect to \( \theta \):

\[
\frac{d}{d\theta} Q_n(\theta, g_\theta) = 2 S_n^T(\theta, g_\theta) M_n \left\{ \frac{\partial S_n}{\partial \theta} (\theta, g_\theta) + \frac{\partial S_n}{\partial g} (\theta, g_\theta) g_\theta' \right\} \tag{74}
\]

with \( g'_\theta \) being a \( 1 \times \tilde{p} \) (\( \tilde{p} = p + 1 \)) matrix given by \( \frac{\partial g_\theta}{\partial \theta^T} \) and

\[
\frac{d^2}{d\theta^2} Q_n(\theta, g_\theta) = 2 \left\{ \frac{\partial S_n}{\partial \theta} (\theta, g_\theta) + \frac{\partial S_n}{\partial g} (\theta, g_\theta) g_\theta' \right\}^T M_n \left\{ \frac{\partial S_n}{\partial \theta} (\theta, g_\theta) + \frac{\partial S_n}{\partial g} (\theta, g_\theta) g_\theta' \right\} + 2 S_n^T(\theta, g_\theta) M_n \frac{d}{d\theta} \left\{ \frac{\partial S_n}{\partial \theta} (\theta, g_\theta) + \frac{\partial S_n}{\partial g} (\theta, g_\theta) g_\theta' \right\} \tag{75}
\]
with
\[
\frac{d}{d\theta^T} \frac{\partial S_n}{\partial g}(\theta, g_\theta) = \frac{\partial^2 S_n}{\partial \theta \partial g^T}(\theta, g_\theta) + \frac{\partial^2 S_n}{\partial \theta \partial g}(\theta, g_\theta) g_\theta,
\]
\[
\frac{d}{d\theta^T} \frac{\partial S_n}{\partial g}(\theta, g_\theta) = \frac{\partial^2 S_n}{\partial \theta \partial g^T}(\theta, g_\theta) + \frac{\partial^2 S_n}{\partial g^2}(\theta, g_\theta) \frac{\partial g}{\partial \theta}.
\]

Note that
\[
S_n(\theta^*, g_{\theta^*}) = S_n(\theta^*, g_{\theta^*}) - S_n(\theta_0, g_0) + S_n(\theta_0, g_0) - S(\theta_0, g_0) = o_p(1),
\]
because \(S(\theta_0, g_0) = 0\) and by Lemmas 5.3-5.4
\[
S_n(\theta_0, g_0) - S(\theta_0, g_0) = o_p(1),
\]
and because \(\theta^*\) lies between \(\hat{\theta}\) and \(\theta_0\), by Lemma 5.4
\[
S_n(\theta^*, g_{\theta^*}) - S_n(\theta_0, g_0) = o_p(1).
\]

Using similar arguments as in the proof of (59) in Lemma 5.4 using Assumption A9 to ensure the boundedness when differentiating twice with respect to \(\theta\), we have
\[
\left\| \frac{d}{d\theta^T} \frac{\partial S_n}{\partial g}(\theta, g_\theta) \right\| = O_p(1) \quad \text{and} \quad \left\| \frac{d}{d\theta^T} \frac{\partial S_n}{\partial g}(\theta, g_\theta) g_\theta \right\| = O_p(1).
\]

Then, we can ignore the second term in the right-hand side of (75) at \(\theta = \theta^*\). Hence, by Lemma 5.6 and \(\theta^* - \theta_0 = o_p(1)\) (thanks to Theorem 2.1), we have
\[
\frac{\partial S_n}{\partial \theta}(\theta_0, g_0) - \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) = o_p(1)
\]
and
\[
\frac{\partial S_n}{\partial g}(\theta^*, g_{\theta^*}) g_\theta' - \frac{\partial S_n}{\partial g}(\theta_0, g_0) g_\theta' = o_p(1),
\]
with
\[
g_{\theta^*} = \frac{g_\theta}{\theta = \theta^*}.
\]

In addition, if \(M_n - M = o_p(1)\), we deduce that
\[
\frac{d^2}{d\theta d\theta^T} Q_n(\theta, g_\theta) \bigg|_{\theta = \theta^*} = 2 \left\{ \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) + \frac{\partial S_n}{\partial g}(\theta_0, g_0) g_\theta' \right\}^T M \left\{ \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) + \frac{\partial S_n}{\partial g}(\theta_0, g_0) g_\theta' \right\} + o_p(1) = 2 B_2(\theta_0) + o_p(1).
\]

We remark that
\[
\frac{d}{d\theta} Q_n(\theta, g_\theta) \bigg|_{\theta = \theta_0} = 2 S_n^T(\theta_0, g_0) M_n \left\{ \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) + \frac{\partial S_n}{\partial g}(\theta_0, g_0) g_\theta' \right\}.
\]

Then, by (80) (see the proof of Lemma 5.6), we have
\[
\frac{\partial S_n}{\partial \theta}(\theta_0, g_0) - \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) = o_p(1) \quad \text{and} \quad \frac{\partial S_n}{\partial g}(\theta_0, g_0) g_\theta' - \frac{\partial S_n}{\partial g}(\theta_0, g_0) g_\theta' = o_p(1).
\]
Consequently, we obtain

\[
\frac{d}{d\theta} Q_n(\theta, g_\theta) \bigg|_{\theta = \theta_0} = 2S_n^T(\theta_0, g_0)M \left\{ \frac{\partial S}{\partial \theta}(\theta_0, g_0) + \frac{\partial S}{\partial g}(\theta_0, g_0)g_0' \right\} + o_p(1).
\]

Then, we have

\[
\hat{\theta} - \theta_0 = -\{B_2(\theta_0)\}^{-1} \left\{ \frac{\partial S}{\partial \theta}(\theta_0, g_0) + \frac{\partial S}{\partial g}(\theta_0, g_0)g_0' \right\}^T M S_n(\theta_0, g_0) + o_p(1).
\]

To end the proof, it remains to be shown that

\[
\sqrt{n}B_1(\theta_0)^{-1/2}S_n(\theta_0, g_0) \xrightarrow{\text{dist.}} N(0, \mathbb{I}_q).
\]

Consider, for all \( w \in \mathbb{R}^q \) such that \( \|w\| = 1 \),

\[
A_n = w^T \{E_0( nS_n(\theta_0, g_0)S_n^T(\theta_0, g_0)) \}^{-1/2} \sqrt{n}S_n(\theta_0, g_0)
\]

with

\[
B_{in} = w^T \{E_0( nS_n(\theta_0, g_0)S_n^T(\theta_0, g_0)) \}^{-1/2} \xi_{in} \tilde{U}_{in}(\theta_0, g_0).
\]

By the Cramer-Wold device, it suffices to show that \( A_n \) converges asymptotically to a standard normal distribution, for all \( w \in \mathbb{R}^q \), such that \( \|w\| = 1 \).

To prove this, we will use the central theorem limit (CTL) proposed by Pinkse et al. (2007). These authors used an idea of Bernstein (1927) based on partitioning the observations into \( J \) groups \( G_{jn}, 1 \leq J < \infty \), which are divided up into mutually exclusive subgroups \( G_{jtn}, 1 \leq t \leq m_{jn}, j = 1, \ldots, J \). Each observation belongs to one subgroup, and its membership can vary with the sample size \( n \), as can the number of subgroups \( m_{jn} \) in group \( j \). We assume that the partition is constructed such that

\[
m_{jn}/m_{1n} = o(1) \quad j = 2, \ldots, J
\]

and

\[
\text{Card}(G_{irn}) = O(\text{Card}(G_{jtn})), \quad \forall i, j = 1, \ldots, J, r = 1, \ldots, m_{in}, t = 1, \ldots, m_{jn}.
\]

Partial sums over elements in groups and subgroups are denoted by \( A_{nj} \) and \( A_{jtn} \), respectively. Thus, we have

\[
A_n = \sum_{j=1}^J A_{jn} = \sum_{j=1}^J \sum_{t=1}^{m_{jn}} A_{jtn}, \quad A_{jtn} = n^{-1/2} \sum_{i \in G_{jtn}} B_{in}.
\]

Let us recall in the following the assumptions under which the CTL of Pinkse et al. (2007) holds.

**Assumption A.** For any \( j = 1, \ldots, J \), let \( G^*, G^{**} \subset G_{jn} \) be any sets for which

\[
\forall t = 1, \ldots, m_{jn} : G^* \cap G_{jtn} \neq \emptyset \quad \Rightarrow \quad G^{**} \cap G_{jtn} = \emptyset.
\]
Then, for any function $f$ in $\mathcal{F} = \{ f : \forall t \in \mathbb{R} f(t) = t \text{ or } \exists \nu \in \mathbb{R} : \forall t \in \mathbb{R} f(t) = e^{\nu t} \}$, where $\nu$ is the imaginary number $i$.

$$\left| \text{Cov} \left( f \left( \sum_{i \in \mathcal{G}^*} B_{in} \right), f \left( \sum_{i \in \mathcal{G}^{**}} B_{in} \right) \right) \right| \leq \left\{ \begin{array}{l} \text{Var} \left( f \left( \sum_{i \in \mathcal{G}^*} B_{in} \right) \right) \text{Var} \left( f \left( \sum_{i \in \mathcal{G}^{**}} B_{in} \right) \right) \end{array} \right\}^{1/2} \alpha_{jn},$$

for some mixing numbers $\alpha_{jn}$ with

$$\lim_{n \to \infty} \sum_{j=1}^{J} m_{jn}^2 \alpha_{jn} = 0.$$

**Assumption B.**

$$\lim_{n \to \infty} \max_{1 \leq m_{jn}} \frac{\sigma_{jn}}{\gamma_{jn}} = 0, \quad j = 1, \ldots, J, \quad \lim_{n \to \infty} \frac{\gamma_{jn}}{\gamma_{1n}} = 0, \quad j = 2, \ldots, J,$$

where

$$\sigma_{jn}^2 = \mathbb{E}_0(A_{jn}^2), \quad \text{and} \quad \gamma_{jn}^2 = \sum_{t=1}^{m_{jn}} \sigma_{jn}^2.$$

**Assumption C.** For some $\tau > 1$

$$\mathbb{E}_0 \left( |A_{jn}|^{2\tau} \right) = o \left( \sigma_{jn}^2 \gamma_{jn}^{2\tau-2} \right), \quad j = 1, \ldots, J, \quad t = 1, \ldots, m_{jn}.$$

If assumptions $A - C$ hold, then by Theorem 1 in [Pinkse et al. (2007)], we have $A_n \to \mathcal{N}(0,1)$.

Thus, to complete the proof, we have to check these assumptions in our context.

**Assumption A:** This holds under (20) (Assumption A3).

Let us choose for instance $J = 2$ groups, each with $m_{1n}, m_{2n}$ subgroups such that $m_{2n} = o(m_{1n})$.

Each subgroup is viewed as an area of size $O(\sqrt{c_n} \times \sqrt{c_n})$ such that $(m_{1n} + m_{2n})c_n = O(n)$. Because $\varphi(\cdot)$ is a decreasing function (Assumption A3), $\alpha_{jn} = O(\varphi(\sqrt{c_n}))$ for $j = 1, 2$. The sequence $c_n$ must be such that $c_n = O(n^{-\nu+1/2})$ for some $0 < \nu < 1/2$ and $n^{\nu+1/2} \varphi(\sqrt{c_n}) \to 0$ as $n \to \infty$.

If for instance $\varphi(t) = O(t^{-\nu})$, then $n^{\nu+1/2} \varphi(\sqrt{c_n}) = O(n^{\nu-1/4+(1+\nu)/2})$; this tends to 0 for each $\nu > 2(1+\nu)/(1-4\nu)$.

**Assumption B:** By assumption A10, $B_1(\theta_0)$ is positive definite and by definition is the limit of $\mathbb{E}_0 \left( n S_n(\theta_0, g_0) S_n^T(\theta_0, g_0) \right)$. Then, for sufficiently large $n$, the last matrix is positive definite, and its inverse is $O(1)$. Therefore, $B_{in}$ is bounded uniformly on $i$ and $n$ because $\xi_{in}$ is bounded uniformly on $i$ and $n$ by Assumption A6, as is $\tilde{U}_m(\theta_0, g_0)$. Then, for all $j = 1, \ldots, J$ and $t = 1, \ldots, m_{jn}$,

$$\sigma_{jn} = \left\{ n^{-1} \mathbb{E}_0 \left( \sum_{i \in \mathcal{G}_{jn}} B_{in} \right) \right\}^{1/2} = O \left( n^{-1/2} \text{Card}(\mathcal{G}_{jn}) \right).$$
\[ \gamma_{jn} = O \left( \frac{m_{jn}}{\sqrt{n}} \max_{t \leq m_{jn}} \text{Card}(G_{jtn}) \right). \]

Therefore,
\[ \frac{\sigma_{jtn}}{\gamma_{jn}} = O(1/m_{jn}) \to 0 \text{ as } n \to \infty, \]

for all \( j = 1, \ldots, J \) and \( t = 1, \ldots, m_{jn} \).

Now, consider the second limit in Assumption B. We have for all \( j = 2, \ldots, J \)
\[ \gamma_{jn} = O \left( \frac{m_{jn}}{m_{1n}} \max_{t \leq m_{jn}} \text{Card}(G_{jtn}) \right) = O \left( \frac{m_{jn}}{m_{1n}} \right) \to 0 \text{ as } n \to \infty, \]

because \( m_{jn}/m_{1n} = o(1) \) for all \( j = 2, \ldots, J \) as \( n \to \infty \).

**Assumption C** : By an easy calculation, we can show that
\[ \frac{\mathbb{E}_\theta (|A_{jtn}|^{2r})}{\sigma_{jtn}^{2r-2}} = O(m_{jn}^{2-2r}) \to 0 \text{ as } n \to \infty. \]

**Lemma 5.6** Under the assumptions of Theorem 2.2 and for any \( \tilde{\theta} \) such that \( \tilde{\theta} - \theta_0 = o_p(1) \), we have
\[ \frac{\partial S_n}{\partial \theta}(\tilde{\theta}, g_{\tilde{\theta}}) - \frac{\partial S}{\partial \theta}(\theta_0, g_0) = o_p(1) \quad (77) \]

and
\[ \frac{\partial S_n}{\partial g}(\tilde{\theta}, g_{\tilde{\theta}}) g_{\tilde{\theta}}' - \frac{\partial S}{\partial g}(\theta_0, g_0) g_0' = o_p(1), \quad (78) \]

with \( g_{\tilde{\theta}}'(\cdot) = \frac{g_{\theta}}{\partial \theta} |_{\theta = \tilde{\theta}} \).

**Proof of Lemma 5.6**

To prove (77), we need to show that for all \( w \in \mathbb{R}^q \) with \( \|w\| = 1 \),
\[ w^T \left\{ \frac{\partial S_n}{\partial \theta}(\tilde{\theta}, g_{\tilde{\theta}}) - \frac{\partial S}{\partial \theta}(\theta_0, g_0) \right\} = o_p(1), \]

which is equivalent to
\[ w^T \left\{ \frac{\partial S_n}{\partial \theta}(\tilde{\theta}, g_{\tilde{\theta}}) - \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) \right\} = o_p(1) \quad (79) \]

and
\[ w^T \left\{ \frac{\partial S_n}{\partial g}(\theta_0, g_0) g_0' - \frac{\partial S}{\partial g}(\theta_0, g_0) g_0' \right\} = o_p(1). \quad (80) \]

The proof of (79) is similar to that of (57), using the fact that
\[ \sup_{\theta, \eta} \left\| \frac{\partial^2 \tilde{U}_i}{\partial \theta \partial \eta^2}(\theta, \eta) \right\| \quad \text{and} \quad \sup_{\theta, \eta} \left\| \frac{\partial^2 \tilde{U}_i}{\partial \theta^2}(\theta, \eta) \right\| \]
are bounded uniformly on $i$ and $n$, and $\bar{\theta} - \theta_0 = o_p(1)$.

Now, let us prove (80). By the definition of $S(\cdot, \cdot)$ (see 13)

$$\lim_{n \to \infty} E_0 \left( \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) \right) = \frac{\partial S}{\partial \theta}(\theta_0, g_0).$$

Thus, it suffices to prove that

$$w^T \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) - w^T E_0 \left( \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) \right) = o_p(1).$$

Let

$$w^T \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) = n^{-1} w^T \xi_n \frac{\partial U_n}{\partial \theta}(\theta_0, \eta^0_i), = \Delta_{n1} - \Delta_{n2},$$

where

$$\Delta_{n1} = n^{-1} \sum_{i=1}^{n} \xi_n^{(1)}(\theta_0, \eta^0_i) \left( Y_{in} - \Phi \left( G_{in}(\theta_0, \eta^0_i) \right) \right) \quad \text{and} \quad \Delta_{n2} = n^{-1} \sum_{i=1}^{n} \xi_n^{(2)}(\theta_0, \eta^0_i),$$

with

$$\xi_n^{(1)}(\theta_0, \eta^0_i) := w^T \xi_i \Lambda' \left( G_{in}(\theta_0, \eta^0_i) \right) \frac{\partial G}{\partial \theta}(\theta_0, \eta^0_i),$$

$$\xi_n^{(2)}(\theta_0, \eta^0_i) := w^T \xi_i \Lambda \left( G_{in}(\theta_0, \eta^0_i) \right) \phi \left( G_{in}(\theta_0, \eta^0_i) \right) \frac{\partial G}{\partial \theta}(\theta_0, \eta^0_i),$$

and $\eta^0_i = g_0(Z_{in}).$

The proof of (81) is then reduced to proving

$$E_0 \left( \| \Delta_{n1} \|^2 \right) = o(1) \quad \text{and} \quad E_0 \left( \| \Delta_{n2} - E_0(\Delta_{n2}) \|^2 \right) = o(1).$$

This last part is trivial because $\xi_n^{(1)}$ and $\xi_n^{(2)}$ are bounded uniformly on $i$ and $n$ (see Assumption A6 and the compactness of $\Theta$, $X$, and $Z$) and by use of the mixing condition (20) and (21) in Assumption A3. This completes the proof of (77).

To prove (78), we remark that

$$\frac{\partial S_n}{\partial \theta}(\bar{\theta}, g_0) g_0' - \frac{\partial S}{\partial \theta}(\theta_0, g_0) g_0' = \left\{ \frac{\partial S_n}{\partial \theta}(\bar{\theta}, g_0) - \frac{\partial S}{\partial \theta}(\theta_0, g_0) \right\} g_0' + \frac{\partial S}{\partial \theta}(\theta_0, g_0) \left( g_0' - g_0' \right).$$

Consider the second term on the right-hand side in (84), where we remark that because $\left\| \frac{\partial S}{\partial \theta}(\theta_0, g_0) \right\|$ and $\sup_{\theta} \sup_{z} \left\| \frac{\partial g_0(z)}{\partial \theta \partial \theta'} \right\|$ are finite and $\bar{\theta} - \theta_0 = o_p(1),$

$$\frac{\partial S}{\partial \theta}(\theta_0, g_0) \left( g_0' - g_0' \right) = (\bar{\theta} - \theta_0) \left( \left\| \frac{\partial S}{\partial \theta}(\theta_0, g_0) \right\| \sup_{\theta} \sup_{z} \left\| \frac{\partial g_0(z)}{\partial \theta \partial \theta'} \right\| \right) = o_p(1).$$
For the first term on the right-hand side in (84), because $\hat{g}_0 = O_p(1)$ by Proposition 5.1, using similar arguments as when proving (77) permits one to obtain

$$\frac{\partial S_n}{\partial g}(\hat{\theta}, g_0) - \frac{\partial S_n}{\partial g}(\theta_0, g_0) = o_p(1).$$

This yields the proof of (78). □

Lemma 5.7 Under the assumptions of Theorem 2.2, we have

$$(i) \quad \frac{d}{d\theta} \left. \frac{\partial Q_n}{\partial g}(\theta, g_0) \right|_{\theta = \theta_0} (\hat{g}_0 - g_0) = O_p(1)$$

$$(ii) \quad \left. \frac{\partial Q_n}{\partial g}(\theta, g_0) \right|_{\theta = \theta_0} (\hat{g}_0' - g_0') = o_p(1),$$

where $\hat{g}_0'(.) = \left. \frac{\partial \hat{g}_0}{\partial \theta}(.) \right|_{\theta = \theta_0}$ and $g_0'(.) = \left. \frac{\partial g_0}{\partial \theta}(.) \right|_{\theta = \theta_0}$.

Proof of Lemma 5.7

To prove (i), and we note that

$$\frac{d}{d\theta} \left. \frac{\partial Q_n}{\partial g}(\theta, g_0) \right|_{\theta = \theta_0} = \frac{d}{d\theta} \left\{ S_n^T(\theta, g_0)M_n \frac{\partial S_n}{\partial g}(\theta, g_0) \right\}$$

$$= \frac{d}{d\theta} \left( S_n^T(\theta, g_0)M_n \frac{\partial S_n}{\partial g}(\theta, g_0) \right) + 2S_n^T(\theta, g_0)M_n \frac{d}{d\theta} \frac{\partial S_n}{\partial g}(\theta, g_0).$$

One can easily see that

$$\frac{d}{d\theta} S_n(\theta, g_0) = \frac{\partial S_n}{\partial \theta}(\theta, g_0) + \frac{\partial S_n}{\partial g}(\theta, g_0) \hat{g}_0'$$

and

$$\frac{d}{d\theta} \frac{\partial S_n}{\partial g}(\theta, g_0) = \frac{\partial^2 S_n}{\partial \theta \partial g}(\theta, g_0) + \frac{\partial^2 S_n}{\partial g^2}(\theta, g_0) g_0'.$$

Therefore, we have

$$\frac{d}{d\theta} \left. \frac{\partial Q_n}{\partial g}(\theta, g_0) \right|_{\theta = \theta_0} =$$

$$2S_n^T(\theta_0, g_0)M_n \left\{ \frac{\partial^2 S_n}{\partial \theta \partial g}(\theta_0, g_0) + \frac{\partial^2 S_n}{\partial g^2}(\theta_0, g_0) \hat{g}_0' \right\} (\hat{g}_0 - g_0)$$

$$+ 2 \frac{\partial S_n}{\partial g}(\theta_0, g_0)M_n \left\{ \frac{\partial S_n}{\partial \theta}(\theta_0, g_0) + \frac{\partial S_n}{\partial g}(\theta_0, g_0) g_0' \right\} (\hat{g}_0 - g_0).$$
By Lemma 5.3 and $S(\theta_0, g_0) = 0$, we obtain

$$S_n(\theta_0, g_0) = S_n(\theta_0, g_0) - S(\theta_0, g_0) = o_p(1).$$

(85)

In addition, we have

$$\left\| \frac{\partial^2 S_n}{\partial \theta \partial g}(\theta_0, g_0)(\hat{g}_0 - g_0) \right\| = n^{-1} \left\| \sum \xi_{in} \frac{\partial^2 \hat{U}_{in}}{\partial \theta \partial \eta}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in})) \right\|
\leq n^{-1} \sum \sup_i \|\xi_{in}\| \sup_{\eta} \left\| \frac{\partial^2 \hat{U}_{in}}{\partial \theta \partial \eta}(\theta_0, \eta) \right\| \|\hat{g}_0 - g_0\|
= o_p(1),$$

(86)

because $\xi_i$ is bounded uniformly on $i, n$ and $\theta$ (Assumption A6), $\|\hat{g}_0 - g_0\| = o_p(1)$ by Proposition 5.1, and

$$\sup_{i,n} \sup_{\eta} \left\| \frac{\partial^2 \hat{U}_{in}}{\partial \theta \partial \eta}(\theta_0, \eta) \right\| < \infty.$$

Using similar arguments as in the proof of (86), we obtain

$$\left\| \frac{\partial^2 S_n}{\partial g^2}(\theta_0, g_0)(\hat{g}_0 - g_0)\right\|' = n^{-1} \left\| \sum \xi_{in} \frac{\partial^2 \hat{U}_{in}}{\partial \eta^2}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))\right\|' = o_p(1),$$

(87)

$$\left\| \frac{\partial S_n}{\partial g}(\theta_0, g_0)(\hat{g}_0 - g_0)\right\|' = n^{-1} \left\| \sum \xi_{in} \frac{\partial \hat{U}_{in}}{\partial \eta}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))\right\|' = o_p(1),$$

(88)

and

$$\left\| \frac{\partial S_n}{\partial \theta}(\theta_0, g_0)(\hat{g}_0 - g_0)\right\| = n^{-1} \left\| \sum \xi_{in} \frac{\partial \hat{U}_{in}}{\partial \theta}(\theta_0, \eta_i)(\hat{g}_0(Z_{in}) - g_0(Z_{in}))\right\| = o_p(1).$$

(89)

Combining (85)-(89) with Assumption A10 permits one to have

$$\frac{d}{d\theta} \frac{\partial Q_n}{\partial g}(\theta, g_0) \bigg|_{\theta = \theta_0} (\hat{g}_0 - g_0) = o_p(1).$$

This yields the proof of (i).

The proof of (ii) follows along similar lines as (i) and hence is omitted. \hfill \Box

Lemma 5.8 Under the assumptions of Theorem 2.2, we have

$$S_n(\hat{\theta}, \hat{g}_0) - S_n(\theta, g_0) = r_n^{(1)}(\theta),$$

where

$$\sup_{\theta} \left\| \frac{\partial}{\partial \theta} r_n^{(1)}(\theta) \right\| = o_p(1), \quad \text{and} \quad \sup_{\theta} \left\| \frac{\partial^2}{\partial \theta \partial \theta} r_n^{(1)}(\theta) \right\| = o_p(1)$$

38
Proof of Lemma 5.8

By applying Taylor’s theorem to $\tilde{U}_i(\theta, \cdot)$ for each $\theta \in \Theta$, we obtain

$$S_n(\theta, \hat{g}_\theta) - S_n(\theta, g_\theta) = n^{-1} \sum_{i=1}^n \xi_{in} \left( \tilde{U}_{in}(\theta, \hat{g}_\theta) - \tilde{U}_{in}(\theta, g_\theta) \right)$$

$$= n^{-1} \sum_{i=1}^n \xi_{in} (\hat{g}_\theta(Z_{in}) - g_\theta(Z_{in}))$$

$$\times \int_0^1 \frac{\partial \tilde{U}_{in}}{\partial \eta} (\theta, g_\theta(Z_{in}) + t (\hat{g}_\theta(Z_{in}) - g_\theta(Z_{in}))) dt$$

$$:= r_{in}^{(1)}(\theta).$$

Because the instrumental variables are bounded uniformly on $i$, $n$, and $\theta$ (Assumption A6), $\sup_{\theta \in \Theta} \| \hat{g}_\theta - g_\theta \|$, $\sup_{\theta \in \Theta} \max_{j=1, \ldots, p+1} \left\| \frac{\partial}{\partial \theta} (\hat{g}_\theta - g_\theta) \right\|$ and $\sup_{\theta \in \Theta} \max_{1 \leq i, j \leq p+1} \left\| \frac{\partial^2}{\partial \theta_i \partial \theta_j} (\hat{g}_\theta - g_\theta) \right\|$ are all of order $o_p(1)$ by Proposition 5.1, it suffices to show that

$$\sup_{\theta, \eta} \sup_{i} \left\| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right\| = O_p(1) \quad (90)$$

$$\sup_{\theta, \eta} \sup_{i} \left\| \frac{\partial \tilde{U}_{in}}{\partial \eta}(\theta, \eta) \right\| = O_p(1) \quad \text{and} \quad \sup_{\theta, \eta} \sup_{i} \left\| \frac{\partial^2 \tilde{U}_{in}}{\partial \theta_i \partial \theta_j \partial \eta}(\theta, \eta) \right\| = O_p(1). \quad (91)$$

Equation (90) is already proved in the proof of Lemma 5.4 (see (60)). The proof of (91) can be established in a similar manner and is thus omitted. □

References

Andrews, D. W. (1992). Generic uniform convergence. *Econometric theory*, 8, 241–257.

Anselin, L. (2010). Thirty years of spatial econometrics. *Papers in regional science*, 89, 3–25.

Anselin, L. (2013). *Spatial econometrics: methods and models* volume 4. Springer Science & Business Media.

Arbia, G. (2006). *Spatial econometrics: statistical foundations and applications to regional convergence*. Springer Science & Business Media.

Bernstein, S. (1927). Sur l’extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Mathematische Annalen*, 97, 1–59.

Billé, A. G. (2014). Computational issues in the estimation of the spatial probit model: A comparison of various estimators. *The Review of Regional Studies*, 43, 131–154.
Carroll, R. J., Fan, J., Gijbels, I., & Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of the American Statistical Association, 92*, 477–489.

Case, A. (1993). Spatial patterns in household demand. *Econometrica, 52*, 285–307.

Conley, T. G. (1999). Gmm estimation with cross sectional dependence. *Journal of econometrics, 92*, 1–45.

Cressie, N. (2015). *Statistics for spatial data*. John Wiley & Sons.

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications: monographs on statistics and applied probability 66* volume 66. CRC Press.

Fleming, M. M. (2004). Techniques for estimating spatially dependent discrete choice models. In *Advances in spatial econometrics* (pp. 145–168). Springer.

Garthoff, R., & Otto, P. (2017). Control charts for multivariate spatial autoregressive models. *AStA Adv. Stat. Anal., 101*, 67–94. URL: http://dx.doi.org/10.1007/s10182-016-0276-x. doi:10.1007/s10182-016-0276-x.

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models* volume 43. CRC Press.

Hunsberger, S. (1994). Semiparametric regression in likelihood-based models. *Journal of the American Statistical Association, 89*, 1354–1365.

Kelejian, H. H., & Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics, 17*, 99–121.

Kelejian, H. H., & Prucha, I. R. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *Internat. Econom. Rev., 40*, 509–533. URL: http://dx.doi.org/10.1111/1468-2354.00027. doi:10.1111/1468-2354.00027.

Lee, L.-F. (2004). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica, 72*, 1899–1925. URL: http://dx.doi.org/10.1111/j.1468-0262.2004.00558.x. doi:10.1111/j.1468-0262.2004.00558.x.

Lee, L.-f. (2007). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *J. Econometrics, 137*, 489–514. URL: http://dx.doi.org/10.1016/j.jeconom.2005.10.004. doi:10.1016/j.jeconom.2005.10.004.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models. *Geographical Analysis, 32*, 19–35.

Lin, X., & Lee, L.-f. (2010). GMM estimation of spatial autoregressive models with unknown heteroskedasticity. *J. Econometrics, 157*, 34–52. URL: http://dx.doi.org/10.1016/j.jeconom.2009.10.035. doi:10.1016/j.jeconom.2009.10.035.
Malikov, E., & Sun, Y. (2017). Semiparametric estimation and testing of smooth coefficient spatial autoregressive models. *J. Econometrics, 199*, 12–34. URL: [http://dx.doi.org/10.1016/j.jeconom.2017.02.005](http://dx.doi.org/10.1016/j.jeconom.2017.02.005)
doi: [10.1016/j.jeconom.2017.02.005](10.1016/j.jeconom.2017.02.005).

Martinetti, D., & Geniaux, G. (2016). Probitspatial: Probit with spatial dependence, sar and sem models. CRAN, . URL: [https://CRAN.R-project.org/package=ProbitSpatial](https://CRAN.R-project.org/package=ProbitSpatial).

Martinetti, D., & Geniaux, G. (2017). Approximate likelihood estimation of spatial probit models. *Regional Science and Urban Economics, 64*, 30–45.

Mátyás, L. (1999). *Generalized method of moments estimation* volume 5. Cambridge University Press.

McMillen, D. P. (1992). Probit with spatial autocorrelation. *Journal of Regional Science,* 32, 335–348.

Pinkse, J., Shen, L., & Slade, M. (2007). A central limit theorem for endogenous locations and complex spatial interactions. *Journal of Econometrics, 140*, 215–225.

Pinkse, J., Slade, M., & Shen, L. (2006). Dynamic spatial discrete choice using one-step gmm: an application to mine operating decisions. *Spatial Economic Analysis, 1*, 53–99.

Pinkse, J., & Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models. *Journal of Econometrics, 85*, 125–154.

Poirier, D. J., & Ruud, P. A. (1988). Probit with dependent observations. *The Review of Economic Studies, 55*, 593–614.

Robinson, P. M. (2011). Asymptotic theory for nonparametric regression with spatial data. *Journal of Econometrics, 165*, 5–19.

Saaty, T. L., & Bram, J. (2012). *Nonlinear mathematics*. Courier Corporation.

Severini, T. A., & Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric models. *Journal of the American statistical Association, 89*, 501–511.

Severini, T. A., & Wong, W. H. (1992). Profile likelihood and conditionally parametric models. *The Annals of statistics*, (pp. 1768–1802).

Smirnov, O. A. (2010). Modeling spatial discrete choice. *Regional science and urban economics, 40*, 292–298.

Staniswalis, J. G. (1989). The kernel estimate of a regression function in likelihood-based models. *Journal of the American Statistical Association, 84*, 276–283.

Su, L. (2012). Semiparametric gmm estimation of spatial autoregressive models. *Journal of Econometrics, 167*, 543–560.
Wang, H., Iglesias, E. M., & Wooldridge, J. M. (2013). Partial maximum likelihood estimation of spatial probit models. *Journal of Econometrics, 172*, 77–89.

Yang, K., & Lee, L.-f. (2017). Identification and QML estimation of multivariate and simultaneous equations spatial autoregressive models. *J. Econometrics, 196*, 196–214. URL: [http://dx.doi.org/10.1016/j.jeconom.2016.04.019](http://dx.doi.org/10.1016/j.jeconom.2016.04.019) doi:[10.1016/j.jeconom.2016.04.019](10.1016/j.jeconom.2016.04.019).

Zheng, Y., & Zhu, J. (2012). On the asymptotics of maximum likelihood estimation for spatial linear models on a lattice. *Sankhya A, 74*, 29–56. URL: [http://dx.doi.org/10.1007/s13171-012-0009-5](http://dx.doi.org/10.1007/s13171-012-0009-5) doi:[10.1007/s13171-012-0009-5](10.1007/s13171-012-0009-5).