ECONSTOR
Make Your Publications Visible.

A Service of
ZBW
Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Arduini, Tiziano

Working Paper

Distribution Free Estimation of Spatial Autoregressive Binary Choice Panel Data Models

Quaderni - Working Paper DSE, No. 1052

Provided in Cooperation with:
University of Bologna, Department of Economics

Suggested Citation: Arduini, Tiziano (2016) : Distribution Free Estimation of Spatial Autoregressive Binary Choice Panel Data Models, Quaderni - Working Paper DSE, No. 1052, Alma Mater Studiorum - Università di Bologna, Dipartimento di Scienze Economiche (DSE), Bologna, https://doi.org/10.6092/unibo/amsacta/4501

This Version is available at:
https://hdl.handle.net/10419/159890

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

https://creativecommons.org/licenses/by-nc/3.0/

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

ISSN 2282-6483

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Distribution Free Estimation of Spatial Autoregressive Binary Choice Panel Data Models

Tiziano Arduini

Quaderni - Working Paper DSE N°1052

Distribution Free Estimation of Spatial Autoregressive Binary Choice Panel Data Models

Tiziano Arduini*

Abstract

This paper proposes a semiparametric estimator for spatial autoregressive (SAR) binary choice models in the context of panel data with fixed effects. The estimation procedure is based on the observational equivalence between distribution free models with a conditional median restriction and parametric models (such as Logit/Probit) exhibiting (multiplicative) heteroskedasticity and autocorrelation. Without imposing any parametric structure on the error terms, we consider the semiparametric nonlinear least squares (NLLS) estimator for this model and analyze its asymptotic properties under spatial near-epoch dependence. The main advantage of our method over the existing estimators is that it consistently estimates choice probabilities. The finite-dimensional estimator is shown to be consistent and root-n asymptotically normal under some reasonable conditions. Finally, a Monte Carlo study indicates that the estimator performs quite well in finite samples.

Key Words: Spatial Autoregressive Model, Binary Choice, Fixed Effects, Non-linear least squares, Semiparametric Estimation.

JEL Codes: C14, C21, C23, C25, R15,

*Department of Economics, University of Bologna. E-mail: tiziano.arduini@unibo.it.

1 Introduction

The spatial econometric literature has increasingly focused on spatial models with limited dependent variables. As Pinkse and Slade (2010) pointed out, the discrete-choice spatial model is a new direction for the future of spatial econometrics. For example, the spatial probit model has been widely studied by McMillen (1992), Pinkse and Slade (1998), LeSage (2000), Pinkse et al. (2006), Beron and Vijverberg (2000), Pace and LeSage (2011), and Wang et al. (2013), among others. These papers propose estimation methods such as an expectation-maximization algorithm, the general method of moments, Bayesian simulation approaches, and partial maximum likelihood estimation. The main assumption of the model is that the distribution of errors is known, and often assumed to be normal or logistic. However, when the distribution of errors is misspecified, this kind of parametric methods yield inconsistent estimates.[1] Furthermore, even if the model is correctly specified, likelihood-based methods may suffer from the multidimensional integration problem as individual disturbances depend on each other. Finally, estimation would become much more difficult with longitudinal data and in the presence of unobserved heterogeneity. In such a case, one needs a clever estimator, such as the conditional logit estimation (Chamberlain, 1984). For these reasons, it may be useful to consider a semiparametric approach.

This paper develops a semiparametric or "distribution free" estimator of binary response spatial autoregressive models for panel data. The model is allowed to have both a spatial lag and fixed or random individual effects in the data generating process. The parametric part estimates the regression coefficients using a probit-logit criterion function in an NLLS minimization procedure. The nonparametric part estimates the scale function of the (heteroskedastic) error terms through a kernel regression as in Jenish (2014). We use a two-stage procedure: in the first step we estimate the scale function and in the second step, the regression coefficients. The main aim of this paper is to show the consistency and asymptotic normality of the proposed semiparametric NLLS estimator.

Semiparametric estimators of spatial binary choice models are still scarce in the literature. Lei (2013) extends the smoothed maximum score estimator (SMS) proposed by Horowitz (1992) to SAR models in the panel data context. He solved the main issues related to these models, i.e., the inconsistency of the estimators in the presence of either heteroskedastic conditional error terms or a misspecification of the error distribution, or both. However, the Lei (2013) approach does not estimate the choice probabilities. Indeed, choice probabilities and marginal effects are of interest in most practical applications as well as for policy purposes.

For the standard binary choice model, Khan (2013) proposes a semiparametric estimator which consistently estimates the choice probabilities.

Khan (2013) shows that the binary response model $y_i = I[x_i'\beta - \epsilon_i]$ (where $I(\cdot)$ is an in-

[1] According to Pinkse and Slade (2010), spatial dependence can lead to heteroskedasticity.

dicator function) with a null conditional-median restriction for identification is "observationally" equivalent to a multiplicative heteroskedastic probit (or logit) model, up to an unknown infinite-parameter scale function.[2] Thanks to this equivalence, he constructs a sieve semiparametric estimator to consistently estimate both the regression coefficients and the choice probabilities.

The present paper shows such an observational equivalence for a larger class of models (SAR models) with heteroskedasticity and autocorrelation in the disturbances. Consequently, it proposes a two-stage semiparametric estimator for a fixed effects SAR binary choice model which also allows the estimation of the choice probabilities.

Andrews (1994) provides a general framework for showing the consistency and asymptotic normality of semiparametric estimators that minimize a criterion function that depends on a preliminary infinite-dimensional nuisance parameter estimator. Jenish (2014) extends this general framework to spatial models with endogeneous regressors and spatial lags of the dependent variable. Our asymptotic theory relies on this framework. However, verification of the high-level assumptions of Jenish (2014) is nontrivial. The finite-dimensional estimator is shown to be consistent and root-n asymptotically normal under some regularity conditions. The spatial weight matrix is assumed to be exogenous and time invariant.

The rest of this paper is organized as follows. Section 2 proves the equivalence between a spatial model with a conditional median restriction and an autocorrelated-heteroschedastic spatial probit model in terms of conditional probabilities. Section 3 defines our proposed estimator. Section 4 proves the consistency and asymptotic normality of the estimator. Section 5 investigates its finite sample properties and Section 6 concludes. Appendix A contains some basic results on stochastic equicontinuity and a functional central limit theorem. Appendix B includes proofs of our main results.[3]

2 Spatial Autoregressive Binary Choice Models

2.1 A class of Spatial Dual Models

As in Lei (2013), the SAR binary choice model is

$$y_{it}^* = \lambda_0 \sum_{j=1}^{n} w_{ij} y_{it}^* + \beta x_{it} + \alpha_i + \epsilon_{it},\ i = 1, \ldots, n\ t = 1, \ldots, T, \tag{1}$$

where y_{it}^* is a latent random variable which has the observation rule $y_{it} = 1$ if $y_{it}^* > 0$ and $y_{it} = 0$ otherwise. The x_{it} are the regressors at time t for each individual i, and w_{ij} is the element of

[2] That is, $P(y_i = 1|x_i = x)$ is the same in both models.

[3] The first theorem on the observational equivalence between the models is left in the main text to make the definition of the proposed estimator easier.

the spatial matrix W_n that is fixed across time. The parameter of interest is λ_0, which captures the spatial dependence, α_i is the unobserved individual heterogeneity constant across time[4], and ϵ_{it} is a disturbance with conditional distribution F_{ϵ_n} given (α, x) with $x = (x_{n1}, \ldots, x_{nT})$.

We consider the case when there are only two periods.[5] Suppose that the inverse of matrix $S_n(\lambda_0) = S_n = (I_n - \lambda_0 W_n)$ exists. We can rewrite Equation (1) as

$$y^*_{nt} = (I_n - \lambda_0 W)^{-1}(\beta_0 x_{nt} + \alpha_n + \epsilon_{nt}) = S_n^{-1}(\beta_0 x_{nt} + \alpha_n + \epsilon_{nt}), \;\; t = 1, 2.$$

Taking the time differences of this model and omitting the subscript n to make the notation simpler, we have

$$\Delta y^*_i = S_i^{-1}(\beta_0 \Delta x + \Delta\epsilon), \tag{2}$$

where S_i is the first row of the matrix S. Lei (2013) finds conditions similar to Lemma 1 of Manski (1987) that allow the parameters of model (1) to be identified up to scale. See Section 4.1 (*Identification*) for further details.

Khan (2013) proves the observational equivalence between a distribution-free model under a conditional median restriction (e.g., Manski, 1975) and a (multiplicative) heteroskedastic parametric (e.g., probit-logit) model when the observations are i.i.d. In what follows, Khan's duality is shown to hold even in an SAR binary choice model where the observation are not i.i.d. To prove the result, the following assumptions are needed.

- Model 1: SAR binary choice model in Equation (2):

  1. $x_i \in \mathcal{R}^k$ is assumed to have a density with respect to Lebesgue measure, which is positive on the set $\chi \subseteq \mathcal{R}^k$.

  2. $F_0(c, x)$ is the CDF $Pr(S_i^{-1}\Delta\epsilon \leq c|x)$. The density of $S_i^{-1}\Delta\epsilon$ is symmetric around zero.

- Model 2: SAR Autocorrelated-Heteroschedastic probit-logit Model:

  a. 1.

  b. $\epsilon_i = \sigma_0(S_i, x) \cdot \eta_{i1} - \sigma_0(S_i, x) \cdot \eta_{i2}$ where $\sigma(\cdot)$ is continuous and positive on χ and η_{it} is independent of x, and with any known (logistic, normal) distribution.

Proposition 1. *Under Assumptions 1 and 2, a and b, Models 1 and 2 are observationally equivalent.*

[4]Note that we don't impose any restriction on the distribution or dependence between α_i and x_{it}

[5]The extension to multiple periods extremely complicates the notational burden, but the theoretical results remain basically unchanged. See Charlier et al. (1995) for a multiple time period SMS estimator.

Proof. Note that the assumption in Model 2 implies that the assumptions in Model 1 are satisfied.

Assuming that the assumptions of Model 1 are satisfied, we show that there exists a scale function $\sigma_0(\cdot,\cdot)$ which satisfies Assumption b such that the two models are observationally equivalent, that is, $Pr(\Delta y_i = 1|\Delta x)$ is the same in both models.

Let $F_0(x) = F_0(S_i^{-1}\beta_0\Delta x, x)$ be the probability function of Model 1. Let us define the scale function as

$$\sigma_0(S_i, x) = S_i^{-1}(\beta_0\Delta x)/\Phi^{-1}(F_0(x))I[S_i^{-1}(\beta_0\Delta x_n) \neq 0], \tag{3}$$

where $\Phi(\cdot)$ denotes the CDF of $S_i^{-1}\Delta\eta$.

The scale function is positive for all x, so that $S_i^{-1}(\beta_0\Delta x) \neq 0$.

Thus, for Model 2 we have

$$\begin{aligned} Pr(\Delta y_i = 1|x) &= Pr(S_i^{-1}\epsilon \leq S_i^{-1}\beta_0\Delta x) \\ &= Pr(S_i^{-1}\sigma_0(S_i, x)\Delta\eta \leq S_i^{-1}\beta_0\Delta x) \\ = \Phi((S_i^{-1}\beta_0\Delta x/\sigma_0(S_i, x)) &= \Phi((\Phi^{-1}(F_0(x)))) = F_0(x), \end{aligned} \tag{4}$$

since the event $\{S_i^{-1}(\beta_0\Delta x_n) = 0\}$ holds with probability 0 under assumption 1. □

Comments

The equivalence result exploits the fact that the normal density has median zero and is positive everywhere. Also the positivity of the scale function is required to establish this class of "dual models." The first model can be estimated using the estimator proposed by Lei (2013). Following Manski (1987), Lei (2013) shows that under some regularity conditions, $\theta_0 = (\lambda_0, \beta_0')'$ is the unique maximizer of the following function

$$\theta_0 = \underset{\theta}{\text{argmax}} \frac{1}{n}\sum_{i=1}^{n} E\left\{\Delta y_i \, sign\left[S_i^{-1}(\lambda)\beta\Delta x\right]\right\}. \tag{5}$$

Thus, Lei (2013)'s estimator is based on the maximization of the sample analog (and a smoothed version) of the function (5). Following the author, we call it a smoothed spatial maximum score estimator (SSMS).

This is a standard approach in the sense that models with conditional median restrictions are often estimated by minimizing least absolute deviation (LAD) objective functions. Lei's method does not permit estimating probabilities, which are often needed by a policy maker. The observational equivalence of Model 1 with Model 2 allows us to propose a new estimator

that allows the estimation of predicted probabilities. There are other semiparametric estimators which allow estimating choice probabilities but they have the disadvantage of assuming the independence of the errors from the covariates, precluding any form of heteroskedasticity. See Klein and Spady (1993), among others. In the next section, we suggest a different estimator and we describe its advantages over the aforementioned existing estimators.

3 A Spatial Semiparamentric Nonlinear Least Square Estimator

In the previous section, we have shown that the SAR binary response model with conditional median restriction is observationally equivalent to an SAR autocorrelated-heteroskedastic probit-logit model up to an unknown infinite-parameter scale function. This equivalence, as Khan (2013) pointed out, suggests that an econometrician could estimate a probit model that is "distribution free." Thus, we use the result of Proposition 1 to construct a semiparametric estimator using a probit criterion function. In particular, we define the criterion function as

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \left\{ \Delta y_i - \Phi \left[S_i^{-1}(\lambda)\beta\Delta x \, \exp(l(S_i, x)) \right] \right\}, \tag{6}$$

where $l(S_i, x) = log\left(\frac{1}{\sigma(S_i,x)}\right)$. Recall the scale function defined in Equation (3). Given this definition, the spatial semiparametric NLLS (hereafter, SSNLLS) is unfeasible, as the values of the conditional probabilities $F_0(x)$ are unknown. Therefore, we propose a two-stage approach to estimate Equation (6).

The first step consists in the estimation of the unknown scale function.

Following Chen and Khan (2003), we use a nonparametric estimation of $F_0(x) = Pr(S_i^{-1}\Delta\epsilon \leq S_i^{-1}\beta\Delta x) = E(\Delta y_i | S_i^{-1}\Delta x)$. Observe that $z_\infty = S_i^{-1}\Delta x = (I - \lambda W_n)^{-1}\Delta x = \sum_{k=1}^{\infty}(\lambda W_n)^k \Delta x$, provided $||\lambda W_n||_\infty < 1$, where $||\cdot||_\infty$ is the row-sum matrix norm. Following standard practice in spatial econometrics, we can replace z_∞ by a linear combination. The conditioning matrix becomes $z_\infty = [(I, W_n, W_n^2, \ldots)\Delta x]$. We can select a submatrix $z_i = (\Delta x_i, W_i\Delta x, W_i^2\Delta x)$, where W_i and W_i^2 are the $i-$th rows of the matrix W_n and W_n^2 respectively. In particular, we propose to estimate the nonparametric regression

$$\min_{F \in \mathcal{S}} \frac{1}{n} \sum_{i=1}^{n} \left(\Delta y_i - F(z_i) \right)^2. \tag{7}$$

Using the definition of the scale function (3) and taking the logarithm, we have

$$\hat{l}(S_i, x) = log\left(\frac{1}{\hat{\sigma}(S_i, x)}\right) = log\left(\frac{\Phi^{-1}(\hat{F}(z_i))}{S_i^{-1}(\lambda_2)(\beta_2'\Delta x_n)}\right).$$

Observe that λ_2, β_2 are nuisance parameters estimated in the second stage. In particular, to estimate Equation (7), we use the Nadaraya–Watson kernel estimator, as in Jenish (2014).

$$\hat{f}(z) = \frac{1}{n}\sum_{i=1}^{n}\hat{K}\left(\frac{z - z_i}{\hat{h}_n}\right)\Big/\hat{h}_n \tag{8}$$

$$\hat{F}(z) = \left[\frac{1}{n}\sum_{i=1}^{n}\Delta y_i\hat{K}\left(\frac{z - z_i}{\hat{h}}\right)\Big/\hat{h}_n\right]\Big/\hat{f}(z),$$

where the kernel $K(\cdot)$ is a non-random real function on $\mathcal{R}^k$, and the bandwith parameter $\hat{h}_n$ is a positive constant.

As the second stage, we propose the semiparametric NLLS estimator

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}}\ \frac{1}{n}\sum_{i=1}^{n}\xi(z_i)\left\{\Delta y_i - \Phi\left[S_i^{-1}(\lambda)\beta'\Delta x_i\ \hat{g}_n(S_i, x)\right]\right\}^2, \tag{9}$$

where $\hat{g}_n(S_i, x) = exp\left[\hat{l}(S_i, x)\right]$ to constrain the scale function to be positive and $\xi(z_i) = 1(z_i \in \chi^*)$, where $\chi^* \subseteq \chi$. Φ is the normal CDF. The estimated vector $\hat{\theta}_n$ contains both the estimates of β and of the parameter of interest λ. Observe that the estimator $\hat{\theta}_n$ is defined using a trimming function $\xi(z_i)$ since the regression model (2) is only assumed to hold on a compact subset of χ, χ^*. In the next sections, we drop the trimming function to make the notation easier. Trimming is useful in this context mainly for two reasons. First, it can eliminate observations from the computation of $\hat{\theta}_n$ for which the nuisance parameter estimator is estimated without precision. Second, it makes it easier to derive the asymptotic properties of $\hat{\theta}_n$, since one can obtain uniform consistency of $\hat{F}(z)$ for $F(z)$ over a compact subset of χ under suitable conditions, but not over unbounded sets in general.

4 Asymptotic properties

4.1 Consistency

Identitfication

We list a set of assumptions that we need in order to show consistency.

Assumption 1. *i) For all i and (x, α), $F_{\epsilon_1|x,\alpha} = F_{\epsilon_2|x,\alpha}$ where $F_{\epsilon_t|x,\alpha}$ denotes the distribution of ϵ conditional on (x, α). ii) The support of $F_{\epsilon_1|x,\alpha}$ is $\mathcal{R}$ for all i and (x, α).*

Observe that under the conditional stationarity Assumption 1 i), also $S_i^{-1}\epsilon_1$ and $S_i^{-1}\epsilon_2$ have the same distribution.

Assumption 2. *i) The support of $F_{\Delta x}$ is not contained in any proper linear subspace of $\mathcal{R}^q$. ii) There exists at least one $q' \in [1, 2, \ldots, q]$ such that $\beta_{0,q'} \neq 0$, and for almost every value of $\Delta\tilde{x}_i = (\Delta x_{i,1}, \ldots, \Delta x_{i,q'-1}, \Delta x_{i,q'+1}, \ldots, \Delta x_{i,q})'$, the scalar random variable $\Delta x_{i,q'}$ has everywhere positive Lebesgue density conditional on $\Delta\tilde{x}_i$ for all $i = 1, 2, \ldots, n$ and conditional on $\Delta x_{j,q'}$ for all $j \neq i$.*

Assumptions 1 and 2 are very similar to those in Manski (1987), but the conditioning sets are different. In this framework, one needs to condition on other members to take into account the spatial correlation between individuals. For a further discussion of these assumptions, see Lei (2013). Observe that in the SAR model (2), $median(\Delta y_i | x_i, y_1 \neq y_2) = sgn(S_i^{-1}(\beta'\Delta x))$ due to the stationarity of the conditional error distributions. This means that the identification result of the binary response model under the conditional median restriction in Manski (1987) remains almost the same if one adds a spatial lag to the model. Lemma 1 in Lei (2013) proves that under these conditions, $\theta_0 = (\lambda_0, \beta_0')'$ is identified relative to $\theta = (\lambda, \beta')' \in \Lambda \times \mathcal{R}^q$, where $\beta/||\beta|| \neq \beta_0/||\beta_0||$.

Let us introduce some notation regarding the space and metric that are convenient when we are dealing with a spatial correlation structure. Following Jenish and Prucha (2009, 2012), we consider a spatial process located on an unevenly spaced lattice that satisfies the following assumption.

Assumption 3. *The lattice $D \subseteq \mathcal{Z}^d$, $d \geq 1$ is infinitely countable. All elements in D are located at a distance of at least $d_0 > 0$ from each other. For example, denoting by l_i the location of the corresponding unit i, we have that for all l_i, $l_j \in D : d(l_i, l_j) \geq d_0$. We assume $d_0 = 1$ without loss of generality.*

This assumption ensures the growth of the sample size as the sample regions $D_n = l_1, \ldots, l_n \subset D$ expand. This means that we are using increasing domain asymptotics.

Assumption 4. *$\zeta \equiv \lambda_m \sup_n ||W_n||_\infty < 1$ and $\Lambda = [-\lambda_m, \lambda_m]$ is the compat parameter space of λ on the real line.*

Assumption 4 is related to the stability of the linear SAR model, the model coherency for a nonlinear simultaneous system of equations, and the weakness of the spatial correlation between the variables when their spatial distance is large. See Xu and Lee (2015) for an exhaustive discussion.

Moment and NED properties of the key variables

In order to study the asymptotic properties of the SNLLS, some moment and NED properties are needed. Some structure for the spatial weights is needed in order to establish that our dependent variable is NED on some input process.

Assumption 5. *In addition to the diagonal element of W being normalized to zero, the weights w_{ij} satisfy at least one of the following two conditions:*

(a) If $d(i,j) > d_c$, we have that $w_{ij} = 0$. Without loss of generality we set $d_c > 1$. Only individuals whose distances are less than or equal to some specific constant may affect each other directly.

(b) (i) Let us define N as a nonnegative integer that does not depend on n. For every n, the number of columns of W_n with $|\lambda_0| \sum_{i=1}^{n} |w_{ij,n}| > \zeta$ is less than or equal to N. (ii) There exists an $\alpha > d$ and a constant C_0 such that $|w_{ij,n}| \leq C_0/d(i,j)^{\alpha}$.

The distance plays a crucial role in Assumption 5. In (a), agents might be linked to other agents over a wide area, but once the geographic distance between two agents exceeds a threshold, the two units do not spatially interact. In (b), the units might be linked in general but the spatial weights decline sufficiently fast at a certain rate as the distance increases.

Following Jenish and Prucha (2012), we first review the definition and some properties of NED random fields.

Definition 1. *For any random variable x, $||x||_p = [E|x|^p]^{1/p}$ denotes its L_p-norm. Let $\{x_{i,n},\, i \in D_n,\, n \geq 1\}$ and $\{\eta_{i,n},\, i \in D_n,\, n \geq 1\}$ be two random fields, where the D_n satisfy Assumption 1. Then, $x_{i,n}$ is called near epoch dependent on $\eta_{i,n}$ if*

$$\sup_{i,n} ||x_{i,n} - E(x_{i,n}|\Im_{i,n}(s))||_p = d_{i,n} v(s) \to 0, \;\; as\; s \to \infty$$

where d_i (the NED scaling factors) is a sequence of positive constants (scale factors) and $v(s)$ (the NED coefficients) can be taken to be non-increasing without loss of generality. $\Im_{i,n}(s) = \sigma(\eta_{j,n} : d(l_i, l_j) \leq s)$ is the σ field generated by the random variables $\eta_{j,n}$ located in the m-neighborhood of location i. $x_{i,n}$ is L_q- NED of size $-\iota$ if $v(s) = O(s^{-\iota-\delta})$ for some $\delta > 0$. To give an intuition of the definition, the random field $x_{i,n}$ should be predictable enough, once we condition on the m-neighborhood of the input process $\eta_{i,n}$. The forecast error declines quickly as the conditioning set of input variables expands. The NED property is preserved under summation, product, and Lipschitz transformations. The NED random field is uniform if and only if $\sup_n \sup_i d_{i,n} < \infty$, and it is called geometric if and only if $v(s) = O(\rho^s)$ for some $0 < \rho < 1$.

As in Xu and Lee (2015), we need the following moment conditions on the input process in order to prove the NED properties of some variables.

Assumption 6. $\sup_{1\leq k\leq K,i,n} E|(x_{ik,n})|^2 < \infty$ *and* $\sup_{i,n} E|(\epsilon_{i,n})|^2 < \infty$.

In what follows, we summarize the properties of the moments and the NED properties of the latent and observed dependent variables.

Proposition 2. *(a) Under Assumption 4, if* $\sup_{1\leq k\leq K,i,n} E|(x_{ik,n})|^p < \infty$ *and* $\sup_{i,n} E|(\epsilon_{i,n})|^p < \infty$*, for some* $p \geq 1$*, then* $\{y^*_{i,n}\}_{i=1}^n$ *and* $\{y_{i,n}\}_{i=1}^n$ *are uniformly* L_p *bounded.*

(b) Under Assumptions 3–5(a) and 6, $\{y^*_{i,n}\}_{i=1}^n$ *and* $\{y_{i,n}\}_{i=1}^n$ *are geometrically* L_2- *NED on* $\eta : ||y_{it} - E(y_{it}|\Im_{i,n}(m))||_2 \leq C\zeta^{s/d_c}$ *for some* $C > 0$ *that does not depend on* i *and* n*.*

(b) Under Assumptions 3–5(b) and 6, $\{y^*_{i,n}\}_{i=1}^n$ *and* $\{y_{i,n}\}_{i=1}^n$ *are* L_2-*NED on* $\eta : ||y_{it} - E(y_{it}|\Im_{i,n}(m))||_2 \leq C/s^{\alpha-d}$ *for some* $C > 0$ *that does not depend on* i *and* n*.*

To prove consistency we also need the objective function to be an NED random field. This result is summarized in the following proposition.

Proposition 3. *Under Assumptions 1–6, the processes* $\{\Delta y_{it}\}$, $\left\{\Phi\left(S_i^{-1}(\lambda)\beta'\Delta x_i \exp(l(S_i,x))\right)\right\}$, *and* $\left\{\left(\frac{\partial\Phi_i(\theta,\sigma)}{\theta}\right)\right\}$, *are* L_2-*NED on the process* $\{\eta_{it}\}$ *of the same size as* $\{y^*_{i,n}\}_{i=1}^n$ *and* $\{y_{i,n}\}_{i=1}^n$.

First-step estimation and consistency of SSNLLS estimator

In what follows, we establish the consistency of the nonparametric first-step estimator. Here, we introduce some notation which will be used in imposing smoothness and compactness conditions. For any $k \times 1$ vector $v = (v_1, v_2, \ldots, v_k)'$, let $|v| = \sum_{i=1}^k v_i$. Let $h(\cdot)$ denote any function on χ. We denote the $|v|-$th derivative of $h(\cdot)$ by

$$D^v h(x) = \frac{\partial^{|v|}}{\partial x_1^{v_1}, \ldots, \partial x_k^{v_k}} h(x).$$

In order to prove consistency, we consider q-times continuously differentiable functions with finite Sobolev norm

$$||h||_{q,r,\mathcal{U}} = \sum_{|v|\leq q} \left(\int_{\mathcal{U}} |D^v h(x)|^r dx\right)^{1/r}. \tag{10}$$

Let us define $\mathcal{T}^{q,r}(\mathcal{U})$, $1 \leq r < \infty$, as the Sobolev space endowed with the above norm, and let C^ω denote the space of $\omega-$times continuously differentiable functions on $\mathcal{U}$.

The consistency of a finite-dimensional parameter estimator relies heavily on the uniform consistency of the nonparametric estimator $\hat{F}(\cdot)$.

The uniform consistency of $\hat{F}$ is based on the following conditions.

Assumption 7. *For some large finite B and $q > 0$, $F \in \mathcal{S}$ where*

$$\mathcal{S} = \{F : ||F(\cdot)||_{q,2,\chi^*} \leq B\}$$

Assumption 8. $\eta = \{x_{it}, \epsilon_{it}\}_{i=1}^n$. $\{x_{it}\}$, $i = 1, ..., n$, $t = 1, 2$, *is a strictly stationary strongly mixing process with α-mixing coefficient $\alpha(k, l, r) \leq (k+l)^c \hat{\alpha}(r)$, $c > 0$, $\hat{\alpha}$ s.t.$\sum_{r=1}^{\infty} r^{d(c+1)-1}\hat{\alpha} < \infty$. ϵ_{it} are i.i.d. random variables. x_{it} and ϵ_{it} are independent.*

Assumption 9. *(a) The density of z, $f_i(z)$, is in C^ω, with $\omega \geq 1 + q$.*

(b) $\sup_n \sup_{z \in \mathcal{R}^k} |\frac{1}{n}\sum_{i=1}^n D^v f_i(z)| < \infty \; \forall v$ *with* $|v| \leq \omega$.

Assumption 10. *(b) $F(z)f_i(z) \in C^\omega$, with $\omega \geq 1 + q$.*

(c) $\sup_n \sup_{x \in \mathcal{R}^k} |\frac{1}{n}\sum_{i=1}^n D^v[F(z)f_i(z)]| < \infty \; \forall v$ *with* $|v| \leq \omega$.

Assumption 11. *(a) $\int K(z)dz = 1$, $\int z^v K(z)dz = 0 \; \forall \; 1 \leq |v| \leq \omega - 1$, $\int |z^v K(z)dz| < \infty \; \forall \; |v| = \omega$. For each $|v| \leq q$, $D^v K(z) \to 0$ as $||z|| \to \infty$.*

(b) $D^v K(z)$ is absolutely integrable and has a Fourier transform

$$\Psi_v(r) = \int exp(ir'z) D^v K(z) dz$$

that satisfies $\int (1 + ||r||) \sup_{b \geq 1} |\Psi_v(br)| dr < \infty \; \forall v \leq q$, where i is the imaginary unit.

Assumption 7 characterizes the set of nonparametric functions. We need to restrict this set in order to obtain the stochastic equicontinuity of the moment functions in F. Assumption 8 characterizes the dependency structure of the input process and serves to provide the NED properties of the objective function and the applicability of the weak LLN. Assumption 9 assumes that the regressors are continuous. This assumption can be relaxed to allow for mixed discrete continuous regressors. Assumption 10 serves to quantify the rates of convergence of the first-step estimator. Assumption 11 (a), requires bias-reducing kernels in order to obtain a proper rate of convergence. Assumption 11 (b) is a technical condition used to prove the following proposition. Recall that k in the following proposition is the dimension of the x and $|v|$ is the degree of smoothness of $F(\cdot)$.

Proposition 4. *Under Assumptions 7 and 12 together with the Assumptions 3 and 8, as $n \to \infty$ we have that $sup_\chi ||D^v \hat{\sigma} - D^v \sigma_0|| = O_p(n^{-\iota/(2\iota+d)} h_n^{-k-|v|-d/(2\iota+d)}) + O_p(h_n^{\omega-|v|})$, for $|v| \leq q$, provided that the R.H.S of the equality is $o_p(1)$.*

Proposition 4 establishes the consistency of the first-step nonparametric estimator in Equation (7). See Theorem 1 in Jenish (2014) for a detailed proof. Thus, the convergence is slower,

the stronger is the dependence (ι) and the higher the dimension of the lattice (d). As in Jenish (2014), to ensure uniform convergence of the first-step estimator, we use the following assumption.

Assumption 12. *The data-dependent bandwidth parameter $h_n = cn^{-\gamma}$, with $c > 0$ and $0 < \gamma < \iota/[(2\iota + d)(k + q) + d]$ where ι is the rate of decay of the NED coefficient of the process $\{\Delta y_i\}_{i=1}^n$.*

Assumption 12 is not restrictive for most application. For example, in Proposition 2 (b), ι is equal to $\alpha - d$. If we set the dimension of the lattice $d = 2$, $\alpha = 3$, $k = 1$, and $q = 3$, we have that γ must be between 0 and roughly 0.06. We need other assumptions in order to derive the consistency of the SSNLLS estimator in Equation (9).

Assumption 13. $Pr(\Delta y_i = 1|x) = \Phi\left(S_i^{-1}(\lambda)\beta'\Delta x_i \exp(l(S_i, x))\right)$.

Assumption 14. *There exists a function $G(\theta, \sigma)$ such that*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\Delta(S_i, x)' \left\{[\Delta y_i - \Phi_i(\theta, \sigma)]\right\},$$

where $\Delta(S_i, x) = E\left(\frac{\partial \Phi_i(\theta,\sigma)}{\partial \theta}\right)$ and $\Phi_i(\theta, \sigma) = \left\{\left[\Phi\left(\frac{S_i^{-1}(\lambda)\beta'\Delta x_i}{\sigma_n(S_i,x)}\right)\right]\right\}$, uniformly over $(\theta, \sigma) \in \Theta \times \mathcal{S}$

Theorem 1. *Under Assumptions 1–14, let*

$$\hat{\theta}_n = \underset{\theta}{\operatorname{argmin}}\, G_n(\theta, g_n), \tag{11}$$

where $G_n(\theta, g_n) = \frac{1}{n}\sum_{i=1}^n \xi(z_i)\left\{\Delta y_i - \Phi\left[S_i^{-1}(\lambda)\beta'\Delta x_i\ \hat{g}_n(S_i, x)\right]\right\}^2$ and $\theta = (\lambda, \beta')'$. Then we have $\hat{\theta} - \theta_0 = o_p(1)$.

The main motivation for the SSNLLS estimator proposed here is to consistently estimate the choice probability function. Observe that, under Assumptions 1–14, the choice probabilities $\Phi_i(\hat{\theta}, \hat{\sigma})$ are also consistently estimated. This is a crucial result in the present paper, which fills a gap in the literature on the distribution-free estimation of binary choice SAR models with unobserved heterogeneity.

4.2 Asymptotic Normality

In what follows, we give sufficient conditions for the asymptotic normality of the estimator in (9). We maintain the previous assumptions, which imply consistency. Recall ω is defined in Assumption 9.

Assumption 15. *The data-dependent bandwidth parameter* $h_n = cn^{-\gamma}$*, with* $c > 0$ *and* $[4\omega]^{-1} < \gamma < min\{\iota/[(2\iota+d)(k+q)+d], (2\iota-d)/[4k(2\iota+d)+4d]\}$*. In addition,* $q > (p_w+1)/2$ *for* $p_w = dim\, M_{in}$ *and* $M_{in} = (x_{it}, y_{it})$*.*

Assumption 15 strengthens the bandwidth condition we needed for consistency. See Jenish (2014) for an exhaustive discussion.

Assumption 16. *(a)* θ_0 *is in the interior of* Θ_0*. Assumption 8 holds with* $\hat{\alpha}(r)$ *s.t.* $\sum_{r=1}^{\infty} r^{d(c+1)-1}\hat{\alpha}^{-1/2} < \infty$*,*

(b)

$$S = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{s=1}^{n} E\Delta_0(S_s, x)' \left\{[\Delta y_s - \Phi_s(\theta_0, \sigma_0)]\right\} \left\{[\Delta y_i - \Phi_i(\theta_0, \sigma_0)]\right\}' \Delta_0(S_i, x),$$

exists and is positive definite,

$$H = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} E\Delta_0(S_i, x)'\Delta_0(S_i, x)$$

exists uniformly on $(\theta, \sigma) \in \Theta \times \mathfrak{S}$ *and is nonsingular.*

This assumption is used to identify the asymptotic variance–covariance matrix of the SSNLLS estimator. It ensures the convergence of the Jacobian and covariance matrices of the sample moments. The assumption on the mixing number is used to apply the central limit theorem (CLT) to the $G_i(\theta_0, \sigma_0)$ used in the asymptotic normality proof, which exploits empirical process theory.

Theorem 2. *Under Assumptions 1–16, the NLLS estimator* $\hat{\theta}$ *satisfies* $\sqrt{n}(\hat{\theta}-\theta_0) \xrightarrow{d} N(0, H^{-1}SH^{-1})$*.*

The estimation of the asymptotic variance–covariance matrix should be straightforward given the definition of H and S and substituting the matrices with the estimated sample counterparts. Observe that we don't provide a limiting distribution for the choice probability functions. This estimator will only converge at the nonparametric rate of the scale function estimator. We leave these topics for future research.

5 Monte Carlo Simulations

In order to evaluate the finite sample performance of the proposed estimators, we conducted a Monte Carlo simulation. The data generating process (DGP) is

$$y_{it}^* = \lambda_0 \sum_{j=1}^{n} w_{ij} y_{it}^* + \beta_0' x_{it} + \alpha_i + \epsilon_{it}, \quad t = 1, 2, \tag{12}$$

where $y_{it} = 1$ if $y^*_{it} > 0$ and zero otherwise. We set $\beta_0 = 1$ in order to allow identification and $\alpha_i \sim N(0, 1)$. In this experiment, we consider three designs, corresponding to three different distributions of ϵ_{it}:

1. $\epsilon_{it} \sim$ normal, median 0, variance 1.

2. $\epsilon_{it} \sim$ uniform, median 0, variance 1.

3. $\epsilon_{it} = 0.25(1 + x^2_{1it})v_{it}, \quad v_{it} \sim$ logistic, median 0, variance 1.

Borrowing from Qu and Lee (2012), we construct the row-stochastic nearest neighbor spatial weight matrix W_n using Lesage's econometrics toolbox. In practice, we generate two random vectors of coordinates which serves as the geographic location for each observation. Then, we find the l nearest neighbors according to the spatial distance and we set $w_{ij} = 1$ for these individuals and zero otherwise. Finally, we row-normalize W_n.

In the simulation, we compare three different estimators which are commonly used to estimate binary choice SAR models with our semiparametric spatial NLLS estimator given in (9). We refer to these four methods as Spatial Maximum Score (SMS), Spatial Smoothed Maximum Score (SSMS), Spatial local NLLS (SLNLLS) and SSNLLS. The SMS is the spatial version of the Maximum score proposed by Manski (1987). The SSMS is its smoothed version as in Lei (2013). The SLNLLS is the spatial version of the Local Non-linear Least Squares proposed in Khan (2013) and Blevins and Khan (2010). For the SSMS, SLNLLS and SSNLLS, the bandwidth is selected according to Silverman's rule of thumb: $h_n = 1.06 \cdot \hat{s} \cdot n^{-1/5}$, where $\hat{s}$ is the sample standard deviation of y_{it}. We also allow the strength of the spatial interactions to differ, setting λ= 0.2 and 0.5 to investigate how the endogenous effects parameter affects the estimates.

The setup of our simulations is as follows. The population size is set to 200, 500 and 1000 observations. We estimate model (12) 1,000 times with SMS, SSMS, SLNLLS and SSNLLS. The control variable x_{it} and the error terms are randomly generated for each replication.

Tables 1–3 report the empirical mean (Avg Point Estimation) and the root mean squared errors (RMSE) of each estimator. For each value of the sample size n (n=200, 500 and 1000), we show the performance of each estimator for low and medium strength spatial interactions (λ= 0.2 and 0.5). Tables 1–3 are given in Appendix C.

Table 1 reports the performance of each estimator when the distribution of errors is normal.

Table 2 reports the performance of each estimator when the distribution of errors is uniform. Table 3 reports the perfomance of each estimator when the distribution of errors is logistic and the errors are heteroskedastic.

The simulation results are summarized as follows.

1. The bias of the parameter estimators SSMS and SSNLLS is very small when the errors have a normal or uniform distribution (designs 1 and 2). The magnitude of this

bias slightly increases when the distribution of the errors is logistic and exhibits heterosckedasticity (design 3).

2. The bias of the estimators does not vary with the magnitude of the spatial interactions, λ. It decreases with the sample size, n.

3. The root mean-squared error of the estimators SLNLLS and SSNLLS are systematically greater than those of SMS and SSMS. While SLNLLS has a slow rate of convergence (see, e.g. Khan, 2013), the variance of SSNLLS depends on the estimation of the first-step parameters.

4. Table 1 shows that the RMSE of those estimators decreases as n increases. In contrast, it increases as λ increases.

6 Conclusions

In this paper we proposed a new distribution-free estimation procedure for heteroskedastic-autocorrelated spatial binary response models in panel data settings. The SSNLLS estimator permits estimating the choice probabilities. Conditions to prove the limiting distribution of the regression coefficients estimator have been given. A simulation study showed that the estimator performs adequately well in finite samples.

We leave for future research the limiting distribution theory for the (pointwise) choice probability, and marginal effects estimators. Also it would be useful to explore other nonparametric estimators for the scale function, such us series estimators, since they have the virtue of being easy to implement.

To conclude, the SSNLLS estimator fills a gap in the literature of semiparametric estimation of SAR binary choice models, being a procedure which allows estimating the response probability, something which is useful for most policy purposes.

References

Andrews, D. W. (1994). Asymptotics for semiparametric econometric models via stochastic equicontinuity, *Econometrica: Journal of the Econometric Society* pp. 43–72.

Andrews, D. W. (1994b). Empirical process methods in econometrics, *Handbook of econometrics* **4**: 2247–2294.

Beron, K. J. and Vijverberg, W. P. (2000). Probit in a spatial context: A monte carlo aproach, *Advances in Spatial Econometrics* .

Blevins, J. R. and Khan, S. (2010). Distribution-free estimation of heteroskedastic binary response models in stata, *The Stata Journal* p. 588.

Chamberlain, G. (1984). Panel data, in z. griliches and md intriligator,(eds.), handbook of econometrics 2.

Charlier, E., Melenberg, B. and Soest, A. v. (1995). A smoothed maximum score estimator for the binary choice panel data model with an application to labour force participation, *Statistica Neerlandica* **49**(3): 324–342.

Chen, S. and Khan, S. (2003). Rates of convergence for estimating regression coefficients in heteroskedastic discrete response models, *Journal of Econometrics* **117**(2): 245–278.

Davidson, J. (1994). *Stochastic Limit Theory: An Introduction for Econometricians: An Introduction for Econometricians*, Oxford University Press, UK.

Horowitz, J. L. (1992). A smoothed maximum score estimator for the binary response model, *Econometrica: journal of the Econometric Society* pp. 505–531.

Jenish, N. (2014). Spatial semiparametric model with endogenous regressors, *Econometric Theory* pp. 1–26.

Jenish, N. and Prucha, I. R. (2009). Central limit theorems and uniform laws of large numbers for arrays of random fields, *Journal of econometrics* **150**(1): 86–98.

Jenish, N. and Prucha, I. R. (2012). On spatial processes and asymptotic inference under near-epoch dependence, *Journal of econometrics* **170**(1): 178–190.

Khan, S. (2013). Distribution free estimation of heteroskedastic binary response models using probit/logit criterion functions, *Journal of Econometrics* **172**(1): 168–182.

Lei, J. (2013). Smoothed spatial maximum score estimation of spatial autoregressive binary choice panel models.

LeSage, J. P. (2000). Bayesian estimation of limited dependent variable spatial autoregressive models, *Geographical Analysis* **32**(1): 19–35.

Manski, C. F. (1975). Maximum score estimation of the stochastic utility model of choice, *Journal of econometrics* **3**(3): 205–228.

Manski, C. F. (1987). Semiparametric analysis of random effects linear models from binary panel data, *Econometrica: Journal of the Econometric Society* pp. 357–362.

McMillen, D. P. (1992). Probit with spatial autocorrelation, *Journal of Regional Science* **32**(3): 335–348.

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

Pinkse, J. and Slade, M. E. (1998). Contracting in space: An application of spatial statistics to discrete-choice models, *Journal of Econometrics* **85**(1): 125–154.

Pinkse, J. and Slade, M. E. (2010). The future of spatial econometrics*, *Journal of Regional Science* **50**(1): 103–117.

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

Qu, X. and Lee, L.-f. (2012). Lm tests for spatial correlation in spatial models with limited dependent variables, *Regional Science and Urban Economics* **42**(3): 430–445.

Wang, H., Iglesias, E. M. and Wooldridge, J. M. (2013). Partial maximum likelihood estimation of spatial probit models, *Journal of Econometrics* **172**(1): 77–89.

Xu, X. and Lee, L.-f. (2015). Maximum likelihood estimation of a spatial autoregressive tobit model, *Journal of Econometrics* **188**(1): 264 – 280.

Appendix A: Stochastic equicontinuity and functional central limit results

In the following proofs we establish a stochastic equicontinuity criterion and a functional CLT for NED random fields. These results are derived in Jenish (2014). Let $G(\cdot,\cdot) : \mathcal{M} \times \mathcal{S} \to \mathcal{R}$ be a real function indexed by an infinite-dimensional metric space $\mathcal{S}$. We assume that for each $\sigma \in \mathcal{S}$, $G(m,\sigma)$ is Borel measurable in m and that the family $\{G(\cdot,\sigma)\}$ belongs to the Sobolev space $\mathcal{T}^{q,2}(\mathcal{M})$ equipped with norm (10). Following Andrews (1991), we take the pseudo-metric ρ on $\mathcal{S}$ to be

$$\rho_{\mathcal{S}} = ||G(\cdot,\sigma_1) - G(\cdot,\sigma_2)||_{\mathcal{M}} = \left(\int_{\mathcal{M}} |G(\cdot,\sigma_1) - G(\cdot,\sigma_2)|^2 dm\right)^{1/2}$$

and consider the empirical process $v_n(\cdot)$ defined as

$$v_n(\sigma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[G_i(m_{i,n},\sigma) - E(G_i(m_{i,n},\sigma))\right].$$

Definition 2. *$\{v_n(\cdot),\, n \geq 1\}$ is uniformly stochastically equicontinuous iff for every $\epsilon > 0$ and $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$\limsup_{n\to\infty} P^* \left(\sup_{\sigma_1,\sigma_2 \in \mathcal{S}: \rho_{\mathcal{S}}(\sigma_1,\sigma_2) < \delta} |v_n(\sigma_1) - v_n(\sigma_2)| > \varepsilon \right) < \epsilon,$$

where P^* denotes $P-$outer probability.

Claim A.1. *Suppose that the following conditions are met:*

(a) $\mathcal{M}$ is an open bounded subset of $\mathcal{R}^p$ with minimally smooth boundary;

(b) writing p for $Dim(m_{i,n})$, $\sup_{\sigma\in\mathcal{S}} ||G(\cdot,\sigma)||_{q,2,\mathcal{M}} < \infty$ for some integer $q > (p+1)/2$;

(c) $\{M_{in}\}$is a $\mathcal{M}-$valued random field that is L_2-NED of size $-d$ on $\{\eta_{i,t}, i \in D_n\}$, where D_n satisfies Assumption 1, and the mixing coefficient of $\{\eta_{i,t}\}$ satisfies $\alpha(k,l,r) \leq (k+l)^c \hat{\alpha}(r)$, $c > 0$, $\hat{\alpha}(r)$ s.t.$\sum_{r=1}^{\infty} r^{d(c+1)-1} \hat{\alpha}(r)^{1/2} < \infty$.

Then $\{v_n(\cdot),\, n \geq 1\}$ is uniformly stochastically equicontinuous and $(\mathcal{S},\rho)$ is totally bounded.

Claim A.2. *If the conditions of Claim A.1 and the following conditions are met,*

(d) $(G(m,\sigma))$ satisfies for any m_1, $m_2 \in \mathcal{M}$: $|G(m_1,\sigma) - G(m_2,\sigma)| \leq C|m_1 - m_2|$ for some $C < \infty$;

(e) For any $\sigma = (\sigma_1, \ldots, \sigma_p)' \in \mathcal{S}^p$ *and* $p \geq 1$, $S_p(\sigma) = \lim_{n\to\infty} E(v_n(\sigma)v_n(\sigma)')$ *exists and is positive definite, where* $v_n(\sigma) = (v_n(\sigma_1), \ldots, v_n(\sigma_p))'$,

then $v_n(\cdot)$ *converges weakly to a zero-mean Gaussian process with covariance function* $S(\cdot,\cdot)$ *whose sample paths are uniformly continuous on* $(\mathcal{S}, \rho)$ *almost surely.*

Appendix B: Proofs

Lemma B.1. *Under Assumptions 3–5(b),* $\Gamma \equiv |\lambda_0| \sup_n ||W_n||_1 < \infty$ *and* $||\lambda_0^l |W_n^l|||_1 \leq max(lN, 1)\Gamma\zeta^{l-1}$*, where* $|W_n| \equiv (|w_{ij,n}|)_{n\times n}$.

Proof. See Lemma 1 in Xu and Lee (2015). □

Proof of Proposition 2. (a) $||a_i(\beta' x_{nt} + \epsilon_{nt})||_p \leq \sum_{j=1}^n a_{ij,n} ||\beta' x_{jnt} + \epsilon_{jnt}||_p < \infty$ by Assumption 4 which implies Lemma B.1. Thus, $\{y_{i,n}^*\}_{i=1}^n$ is bounded.

Even though $1(\cdot)$ is not differentiable, we can apply the mean value theorem of a convex function (Wegge, 1974) since the indicator function of a convex set is a convex function. Thus, we have $y(\cdot) : y_1 - y_2 = f_{D_n}[a_{1n}(\beta' x_{1nt} + \epsilon_{1nt}) - a_{2n}(\beta' x_{2nt} + \epsilon_{2nt})]$, where f_{D_n} is a diagonal matrix whose ith diagonal element is some subgradient of $y(\cdot)$ at some points between $a_{1n}(\beta' x_{1nt} + \sigma_1(\cdot)\epsilon_{1nt})$ and $a_{2n}(\beta' x_{2nt} + \epsilon_{2nt})$. Observe that the subgradients of $y(\cdot)$ are always between 0 and 1 and therefore, given the L_p boundness of $\{y_{i,n}^*\}_{i=1}^n$, we have that the R.H.S of the mean value equality is always L_p bounded.

(b) Following the proof of Proposition 1 in Jenish and Prucha (2012), we have

$$||y_{it} - E(y_{it}|\Im_{i,n}(m))||_2 \leq \sup_{j,n} ||\beta' x_{jnt} + \epsilon_{jnt}||_2 \sup_{i,n} \sum_{j:d(i,j)>s} a_{ij,n}.$$

Under Assumption 5(a), by applying Proposition 1 (i) in Xu and Lee (2015), we have

$$\sup_{i,n} \sum_{j:d(i,j)>s} a_{ij,n} \leq \sup_{i,n} \sum_{l=[s/d_c]+1} \zeta^l \geq (1-\zeta)^{-1}\zeta^{s/d_c},$$

which implies

$$||y_{it} - E(y_{it}|\Im_{i,n}(s))||_2 \leq C_0(1-\zeta)^{-1}\zeta^{s/d_c}.$$

Under Assumption 5(b), by applying Proposition 1 (ii) in Xu and Lee (2015), we have

$$\sup_{i,n} \sum_{j:d(i,j)>s} a_{ij,n} \leq C_1 C_2 2^\alpha (\alpha - d)^{-1} s^{d-\alpha}.$$

This implies

$$||y_{it} - E(y_{it}|\Im_{i,n}(s))||_2 \leq C_0 C_1 C_2 2^\alpha (\alpha - d)^{-1}/s^{\alpha - d}.$$

Finally, the NED properties $\{y_{i,n}^*\}_{i=1}^n$ are obvious from $\{y_{i,n}\}_{i=1}^n$.

□

Proof of Proposition 3. The NED of the process $\{\Delta y_{it}\}$ follows directly from Theorem 17.8 in Davidson (1994), which can also be applied under spatial dependence. We follow the

strategy of Jenish (2014) to prove the NED of the process $\left\{\Phi\left(S_i^{-1}(\lambda)\beta'\Delta x_i \exp(l(S_i,x))\right)\right\}$. by Assumption 13, we have that $E(\Delta y_i|x) = \left\{\Phi\left(S_i^{-1}(\lambda)\beta'\Delta x_i \exp(l(S_i,x))\right)\right\}$. Thus, given Assumptions 1–7, $E(\Delta y_i|x)$ is continuously differentiable in its arguments, and hence satisfies a Lipschitz condition in its arguments with a bounded Lipschitz coefficient. Then, by Proposition 2 of Jenish and Prucha (2012), $E(\Delta y_i|x)$ is L_2-NED of the same size as $\{y^*_{i,n}\}_{i=1}^n$. Similarly, $\left\{\left(\frac{\partial \Phi_i(\theta,\sigma)}{\theta}\right)\right\}$, is also L_2-NED of the same size $\{y^*_{i,n}\}_{i=1}^n$

□

Proof of Proposition 4. See Theorem 1 of Jenish (2014)

□

Lemma B.2. *Under Assumptions 1–3, let us define our limiting objective function as*

$$G(\theta) = -\frac{1}{n}EG_i(\theta) = -\frac{1}{n}E\left\{\left[\Delta y_i - \Phi\left(\frac{S_i^{-1}(\lambda)\beta'\Delta x_i}{\sigma_n(S,x_i)}\right)\right]^2\right\}.$$

Then $G(\theta_0) > G(\theta)$ for all $\theta = (\lambda,\beta) \in \Lambda \times \mathcal{R}^q$, where $\beta \neq \beta_0$ when $\lambda = \lambda_0$.

Proof. Lei (2013) shows that the smoothed maximum score estimator for the SAR binary choice model are identified in the Manski (1987) sense and that θ_0 is the unique maximizer.

So, following Khan (2013), we observe that every maximum of this objective function, say $\tilde{\theta}$ and $\tilde{\sigma}$, must satisfy $\frac{S_i^{-1}(\lambda)\tilde{\beta}'\Delta x_i}{\tilde{\sigma_n}(S,x_i)} = \frac{S_i^{-1}(\lambda_0)\beta_0'\Delta x_i}{\sigma_0(S,x_i)}$ by the strict monotonicity of $\Phi(\cdot)$. Suppose $\tilde{\theta},\tilde{\sigma} \neq \theta_0,\sigma_0$. Then by Assumptions 1 and 2, we have on a set of positive probability that $sign(S_i^{-1}(\lambda_0)\beta_0'\Delta x_i) \neq sign(S_i^{-1}(\lambda)\tilde{\beta}'\Delta x_i)$. Therefore $\frac{S_i^{-1}(\lambda)\tilde{\beta}'\Delta x_i}{\tilde{\sigma_n}(S,x_i)} \neq \frac{S_i^{-1}(\lambda_0)\beta_0'\Delta x_i}{\sigma_0(S,x_i)}$, given that both scale functions are positive everywhere. Observe that if $\tilde{\theta} = \theta_0$, we have that $\tilde{\sigma} = \sigma_0$ since $P(S_i^{-1}(\lambda)\beta'\Delta x_i = 0) = 0$ by Assumption 2. □

Proof of Theorem 1. We have to satisfy all the conditions C of Theorem A.1 of Andrews (1994). In particular:

(a) There exists a function $G(\theta,\sigma)$ such that $\bar{G}_i \xrightarrow{p} G(\theta,\sigma)$, where

$$\bar{G}_i = \frac{1}{n}\sum_{i=1}^n \Delta(x)'\left\{[\Delta y_i - \Phi_i(\theta,\sigma)]\right\}$$

uniformly over $(\theta,\sigma) \in \mathcal{S}$.

Assumption (a) holds if

$$G(\theta,\sigma) = \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^n E\Delta(x)'\left\{[\Delta y_i - \Phi_i(\theta,\sigma)]\right\},$$

where $\Delta(x) = \left(\frac{\partial \Phi_i(\theta,\sigma)}{\theta}\right)$ and $\Phi_i(\theta,\sigma) = \left\{\left[\Phi\left(\frac{S_i^{-1}(\lambda)\beta'\Delta x_i}{\sigma_n(S,x_i)}\right)\right]\right\}$. By Proposition 3, $\{\bar{G}_i\}$ is L_2-NED of the same size as y. By Theorem 1 of Jenish and Prucha (2012),$\{\bar{G}_i\}$ satisfies an LLN, that is for each (θ,σ) we have

$$\left|\bar{G}_i(\theta,\sigma) - E\bar{G}_i(\theta,\sigma)\right| \xrightarrow{p} 0.$$

Next we have to show the this convergence holds uniformly over $\Theta \times \mathcal{S}$. In order to do this, it suffices, by the ULLN of Jenish and Prucha (2009), to show that (i) $\{\bar{G}_i(\theta,\sigma)\}$ is stochastically equicontinuous on $\Theta \times \mathcal{S}$ w.r.t. the pseudometric ρ_1, and $\Theta \times \mathcal{S}, \rho_1$ is totally bounded, where

$$\rho_1((\theta_a,\sigma_a),(\theta_b,\sigma_b)) = ||\theta_a - \theta_b|| + \sup_{\chi}|\sigma_a - \sigma_b|.$$

To this end, consider the following inequality.

$$\begin{aligned} |(y_i - \Phi_{i1})^2 - (y_i - \Phi_{i2})^2| = |(\Phi_{i1} - \Phi_{i2})(\Phi_{i1} + \Phi_{i2} - 2y_i)| &\leq \\ |(\Phi_{i1} - \Phi_{i2})(\Phi_{i1} + \Phi_{i2})| &\leq \quad (13) \\ |(\Phi_{i1} - \Phi_{i2})^2| & \quad (14) \\ & \quad (15) \end{aligned}$$

Under Assumptions 1–7, we have by Proposition 3 that both $\left\{\Phi\left(S_i^{-1}(\lambda)\beta'\Delta x_i \exp(l(S_i,x))\right)\right\}$, and $\left\{\left(\frac{\partial \Phi_i(\theta,\sigma)}{\theta}\right)\right\}$, are L_2 NED on the input process, and thus by the mean value expansion we can write

$$|(y_i - \Phi_{i1})^2 - (y_i - \Phi_{i2})^2| \leq C\rho_1((\theta_a,\sigma_a),(\theta_b,\sigma_b)),$$

for some $C < \infty$, which proves the stochastic ρ_1-equicontinuity of $\{(y_i - \Phi_{i1})^2\}$ on $\Theta \times \mathcal{S}$. Moreover, $\mathcal{S}$ is uniformly ρ_2-continuous on χ, where $\rho_2 = \sup_{\chi}|\sigma_a(z) - \sigma_b(z)|$, and $\mathcal{S}$ is equibounded on $\Theta \times \mathcal{S}$. Then, by the Arzela–Ascoli theorem, $(\mathcal{S},\rho_2)$ is totally bounded, and hence $(\mathcal{S} \times \Theta, \rho_1)$ is also totally bounded given the compactness of Θ. Thus, by the ULLN of Jenish and Prucha (2009),

$$\sup_{\Theta\times\mathcal{S}}\left|\bar{G}_i(\theta,\sigma) - E\bar{G}_i(\theta,\sigma)\right| \xrightarrow{p} 0.$$

Then, by Assumption , we have that

$$\sup_{\Theta\times\mathcal{S}}\left|E\bar{G}_i(\theta,\sigma)-G_i(\theta,\sigma)\right|\xrightarrow{p}0,$$

and we also have

$$\sup_{\Theta\times\mathcal{S}}\left|\bar{G}_i(\theta,\sigma)-G_i(\theta,\sigma)\right|\xrightarrow{p}0.$$

It also follows that $G_i(\theta,\sigma)$ is uniformly ρ_1-continuous on $\Theta\times\mathcal{S}$, and by the total boundedness of $(\mathcal{S}\times\Theta,\rho_1)$, that $\sup_{\Theta\times\mathcal{S}}|G_i(\theta,\sigma)|<\infty$.

(b) $\sup_{\theta\in\Theta}||G(\theta,\hat{\sigma})-G(\theta,\sigma_0)||\xrightarrow{p}0$ for some $\sigma_0\in\mathcal{S}$, and $P(\hat{\sigma}\in\mathcal{S})\to 1$.

The first part of Assumption (b) holds by Proposition 4, because

$$\begin{aligned}\sup_{\theta\in\Theta}||G(\theta,\hat{\sigma})-G(\theta,\sigma_0)|| &\leq\\ \sup_{\theta\in\Theta}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int||(\hat{\Delta}(x)-\Delta_0(x))||||\Phi_i(\theta,\sigma_0)-\Phi_i(\theta,\hat{\sigma})||dF_i(x) &\leq\\ C\left[\sup_{\theta\in\Theta}E||\Phi_i(\theta,\sigma_0)-\Phi_i(\theta,\hat{\sigma})||^2\right]^{\frac{1}{2}}\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int\left[||(\hat{\Delta}(x)-\Delta_0(x))||^2dF_i(x)\right]^{\frac{1}{2}}&\xrightarrow{p}0.\end{aligned}$$

We next show that $P(\hat{\sigma}\in\mathcal{S})\to 1$. By Proposition 4, we have that

$$||\hat{\sigma}-\sigma_0||_{q,\chi}\leq\sum_{|v|\leq q}\left(\int_{\chi}|D^v\hat{\sigma}-D^v\sigma_0|^2dx\right)^{1/2}\leq C\sum_{|v|\leq q}\sup_{\chi}|D^v\hat{\sigma}-D^v\sigma_0|^2\xrightarrow{p}0.$$

It then follows that

$$||\hat{\sigma}||_{q,\chi}\leq||\hat{\sigma}-\sigma_0||_{q,\chi}+||\sigma_0||_{q,\chi}\leq o_p(1)+B.$$

Hence, $P(||\hat{\sigma}||_{q,\chi}\leq B)\to 1$ and the conclusion follows.

(c) d(G) is uniformly continuous on $G\in\{G\in\Re: G=G(\theta,\sigma)\ for\ some\ \theta\in\Theta,\sigma\in\mathcal{S}\}$, where $d(G)=G'G/2$ (see, Andrews, 1994).

Assumption C holds with $d(G)=G'G/2$ and a) given that $\sup_{\Theta\times\mathcal{S}}|G_i(\theta,\sigma)|<\infty$. Observe that by the uniform continuity of $|G(\theta,\sigma)|$ in (θ,σ), $d(G(\theta,\sigma_0))$ is continuous on Θ.

(d) For every neighborhood $\Theta_0(\subset\Theta)$ of θ_0, $\inf_{\theta\in\Theta/\Theta_0}d(G(\theta,\sigma_0))>d(G(\theta_0,\sigma_0))$.

Assumption D holds with Θ compact, $d(G(\theta,\sigma_0))$ continuous on Θ and Lemma B.2.

□

Proof of theorem 2. Let us define $z_i(\theta) = S_i^{-1}(\lambda)\beta'\Delta x_i l_i(S, x_i)$. The estimator $\hat{\theta}$ satisfies the following first order conditions:

$$\bar{G}_i = \frac{1}{n}\sum_{i=1}^{n}\left\{\left[\Delta y_i - \Phi_i(\hat{\theta}, \hat{\sigma})\right]\right\}\phi(\hat{\theta}, \hat{\sigma})\frac{\partial z_i(\theta)}{\partial \theta} = o_p(1). \tag{16}$$

Note that by assumption, $\bar{G}_i(\theta, \sigma)$ is continuously differentiable in the interior of Θ, and $\theta_0 \in$ the interior of Θ. Taking the mean value expansion of $\bar{G}_i(\hat{\theta}, \hat{\sigma})$ about θ_0 yields

$$\bar{G}_i(\hat{\theta}, \hat{\sigma}) = \bar{G}_i(\theta_0, \hat{\sigma}) + \frac{\partial \bar{G}_i(\bar{\theta}, \hat{\sigma})}{\partial \theta}(\hat{\theta} - \theta_0), \tag{17}$$

where $\bar{\theta}$ is between $\hat{\theta}$ and θ_0. Plugging equation (17) into (16) gives

$$\sqrt{n}(\hat{\theta} - \theta_0) = -\bar{G}_i(\theta_0, \hat{\sigma})\left[\frac{\partial \bar{G}_i(\bar{\theta}, \hat{\sigma})}{\partial \theta}\right]^{-1} + o_P(1). \tag{18}$$

From Assumption 9(b), and arguments analogous to those in Theorem 1 (a) (by applying Theorem 1 of Jenish and Prucha, 2012), the Hessian term $\frac{\partial \bar{G}_i(\bar{\theta}, \hat{\sigma})}{\partial \theta}$ satisfies an ULLN. Therefore, $\frac{\partial \bar{G}_i(\bar{\theta}, \hat{\sigma})}{\partial \theta} \xrightarrow{p} H$. By applying Slutzky's theorem, we have

$$\sqrt{n}(\hat{\theta} - \theta_0) = -H'^{-1}\sqrt{n}\bar{G}_i(\theta_0, \hat{\sigma}) + o_P(1). \tag{19}$$

Recall $G_i(\theta) = \left\{\left[\Delta y_i - \Phi\left(\frac{S_i^{-1}(\lambda)\beta'\Delta x_i}{\sigma_n(S, x_i)}\right)\right]\right\}\phi(\hat{\theta}, \hat{\sigma})\frac{\partial z_i(\theta)}{\partial \theta}$.
Let

$$v_n(\sigma) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[G_i(\theta_0, \sigma) - E(G_i(\theta_0, \sigma))\right].$$

Then,

$$\sqrt{n}(\hat{\theta} - \theta_0) = -H^{-1'}\hat{v}_n - H^{-1'}\sqrt{n}E\bar{G}_i(\theta_0, \hat{\sigma}) + o_P(1).$$

We now show that $\sqrt{n}E\bar{G}_i(\theta_0, \hat{\sigma}) = o_p(1)$.

From the definition of our moment condition, we have $\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} E[G_i(\theta_0, \hat{\sigma})]\right| = 0\,\forall\sigma$, for large n. Thus, the result follows.

Following Andrews (1994), we need

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[G_i(\theta_0, \hat{\sigma}) - E(G_i(\theta_0, \sigma_0))\right] = o_p(1). \tag{20}$$

Given that, we have

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[G_i(\theta_0, \hat{\sigma}) - E(G_i(\theta_0, \sigma_0))\right] = v_n(\hat{\sigma}) - v_n(\sigma_0) - \sqrt{n}\bar{G}_i(\theta_0, \hat{\sigma}). \tag{21}$$

Therefore, Equation (20) holds if and only if the R.H.S. of Equation (21) is $o_p(1)$.

Thus, it remains to show that $v_n(\hat{\sigma}) - v_n(\sigma_0) \xrightarrow{p} 0$. To this end, it suffices to show: (i) $P(\hat{\sigma} \in \mathcal{S}) \to 1$ (already proven), (ii) $\rho_n(\hat{\sigma}, \sigma_0) \xrightarrow{p} 0$, and (iii) $\{v_n(\cdot), n \geq 1\}$ is stochastically equicontinuous at σ_0. The sufficiency of showing (i), (ii), and (iii) is because given any η and $\epsilon > 0$, there exists a $\delta > 0$ such that

$$\begin{aligned} \bar{\lim_{n\to\infty}} P(|v_n(\hat{\sigma}) - v_n(\sigma_0)\ | > \eta) &\leq \quad (22) \\ \bar{\lim_{n\to\infty}} P(|v_n(\hat{\sigma}) - v_n(\sigma_0)\ | > \eta,\ \hat{\sigma} \in \mathcal{S},\ \rho_n(\hat{\sigma}, \sigma_0) \leq \delta) + \bar{\lim_{n\to\infty}} P(\hat{\sigma} \notin \mathcal{S}\ or\ \rho_n(\hat{\sigma}, \sigma_0) > \delta) &\leq \\ \bar{\lim_{n\to\infty}} P^*(\sup_{\sigma\in\mathcal{S},\ \rho_n(\hat{\sigma},\sigma_0)\leq\delta} |v_n(\hat{\sigma}) - v_n(\sigma_0)\ | > \eta) &\leq \epsilon. \end{aligned}$$

Here P* is the outer measure (see, e.g. Andrews, 1994b). To show (ii), observe that by Assumption 4, $G(\cdot, \theta_0, \sigma)$ belongs to the Sobolev space $\mathcal{T}^{q,2}(\mathcal{M})$ with norm of order $q > (K_M + 1)/2$, and by Theorem 1 in Jenish (2014),

$$\rho_n(\hat{\sigma}, \sigma_0) \leq Leb^{1/2}(\mathcal{M}) \sup_{m\in\mathcal{M}} |G(m, \theta_0, \hat{\sigma}) - G(m, \theta_0, \sigma_0)| \xrightarrow{p} 0.$$

To establish (iii), we verify the assumptions of Claim A.1. Assumptions (a) and (c) of Claim A.1 hold by Assumptions 3–5(a) or (b), 6, and Assumption 16. Now, by Assumption 7,

$$||G(\cdot, \theta_0, \sigma)||_{q,2,\mathcal{M}} = \sup_{\sigma\in\mathcal{S}} \sum_{|\mu|\leq q} ||D_m^\mu G(m, \theta_0, \sigma)||_{L_2(\mathcal{M})} < \infty.$$

Thus, $\{v_n(\cdot), n \geq 1\}$ is stochastically equicontinuous at σ_0. Lastly, we show $v_n(\sigma_0) \xrightarrow{d} N(0, S)$ by verifying the assumptions of Claim A.2. Assumptions (a)–(d) of Claim A.2 have been verified above. Assumption (e) of Claim A.2 for $p = 1$ holds by Assumption 16.

Therefore, our (element by element) mean value expansion becomes

$$\sqrt{n}(\hat{\theta} - \theta_0) = -H^{-1'}(\hat{v}_n - v_n(\sigma_0) + v_n(\sigma_0)) - H^{-1'}\sqrt{n}E\bar{G}_i(\theta_0, \hat{\sigma}) + o_P(1) =$$

$$\sqrt{n}(\hat{\theta} - \theta_0) = -H^{-1'}o_p(1) - H^{-1'}v_n(\sigma_0) + o_P(1) \xrightarrow{d} N(0, H^{-1'}S(H^{-}1')'),$$

by the functional CLT in Claim A.2. □

Appendix C: Tables

Table 1: Design 1: normal distribution.

		SMS	SSMS	SLNLLS	SSNLLS
n=200					
$\lambda_0 = 0.2$	$\hat{\lambda}$	0.123	0.202	0.215	0.217
	RMSE	0.679	0.317	0.418	0.615
$\lambda_0 = 0.4$	$\hat{\lambda}$	0.470	0.420	0.417	0.436
	RMSE	0.399	0.293	0.419	0.682
n=500					
$\lambda_0 = 0.2$	$\hat{\lambda}$	0.130	0.199	0.206	0.197
	RMSE	0.356	0.300	0.205	0.398
$\lambda_0 = 0.4$	$\hat{\lambda}$	0.350	0.440	0.402	0.412
	RMSE	0.388	0.207	0.292	0.447
n=1000					
$\lambda_0 = 0.2$	$\hat{\lambda}$	0.143	0.193	0.178	0.164
	RMSE	0.363	0.262	0.319	0.311
$\lambda_0 = 0.4$	$\hat{\lambda}$	0.368	0.426	0.403	0.390
	RMSE	0.277	0.160	0.226	0.342

Table 2: Design 2: uniform distribution

		SMS	SSMS	SLNLLS	SSNLLS
n=200					
$\lambda_0 = 0.2$	$\hat{\lambda}$	0.165	0.190	0.170	0.213
	RMSE	0.380	0.368	0.451	0.558
$\lambda_0 = 0.4$	$\hat{\lambda}$	0.599	0.439	0.466	0.454
	RMSE	0.569	0.364	0.818	0.293
n=500					
$\lambda_0 = 0.2$	$\hat{\lambda}$	0.150	0.194	0.176	0.190
	RMSE	0.332	0.208	0.355	0.351
$\lambda_0 = 0.4$	$\hat{\lambda}$	0.596	0.428	0.431	0.400
	RMSE	0.388	0.214	0.306	0.433
n=1000					
$\lambda_0 = 0.2$	$\hat{\lambda}$	0.230	0.201	0.171	0.198
	RMSE	0.198	0.165	0.229	0.254
$\lambda_0 = 0.4$	$\hat{\lambda}$	0.580	0.421	0.390	0.400
	RMSE	0.277	0.166	0.228	0.356

Table 3: Design 3: logistic distribution with heteroskedastic errors

		SMS	SSMS	SLNLLS	SSNLLS
n=200					
$\lambda_0 = 0.2$	$\hat{\lambda}$	0.233	0.238	0.240	0.305
	RMSE	0.679	0.355	0.449	0.620
$\lambda_0 = 0.4$	$\hat{\lambda}$	0.633	0.534	0.491	0.568
	RMSE	0.669	0.357	0.620	1.257
n=500					
$\lambda_0 = 0.2$	$\hat{\lambda}$	0.245	0.267	0.256	0.235
	RMSE	0.356	0.358	0.216	0.412
$\lambda_0 = 0.4$	$\hat{\lambda}$	0.350	0.525	0.488	0.483
	RMSE	0.588	0.237	0.363	0.488
n=1000					
$\lambda_0 = 0.2$	$\hat{\lambda}$	0.266	0.261	0.243	0.211
	RMSE	0.363	0.228	0.164	0.328
$\lambda_0 = 0.4$	$\hat{\lambda}$	0.543	0.491	0.432	0.381
	RMSE	0.390	0.158	0.491	0.482

Alma Mater Studiorum - Università di Bologna
DEPARTMENT OF ECONOMICS

Strada Maggiore 45
40125 Bologna - Italy
Tel. +39 051 2092604
Fax +39 051 2092664
http://www.dse.unibo.it