

# Estimation of Sample Selection Models with Spatial Dependence\*

Alfonso Flores-Lagunes<sup>†</sup>      Kurt Erik Schnier<sup>‡</sup>

December 2004

## Abstract

We consider the estimation of sample selection (type II Tobit) models that exhibit spatial dependence. Attention is focused mainly on the spatial error dependence model (or spatial autoregressive error model, SAE), but our method can also be used to estimate the spatial lag dependence model (or spatial autoregressive model, SAL).

The method considered is motivated by a two-step strategy analogous to the popular heckit model. The first step of estimation is based on a spatial probit model following a methodology proposed by Pinkse and Slade (1998) that yields consistent estimates. The consistent estimates of the selection equation are used to estimate the inverse Mills ratio (IMR) to be included as a regressor in the estimation of the outcome equation (second step). Since the appropriate IMR turns out to depend on a parameter from the second step under SAE, we propose to estimate the two steps jointly within a generalized method of moments (GMM) framework.

We explore the large sample properties of the proposed estimator and undertake a Monte Carlo experiment to assess its performance in finite samples. Finally, we discuss the importance of the spatial sample selection model in applied work and illustrate its application by estimating the relative spatial efficiency of the production process within a fishery, using catch-per-unit-effort (CPUE) as a measure of efficiency.

Key words and phrases: sample selection, spatial dependence, generalized method of moments.

JEL classification: C13, C15, C24, C49

---

\*We want to thank Dan Akerberg, Carlos A. Flores and Kei Hirano for useful discussions, and seminar participants at the 2004 EEA-ESEM meeting in Madrid, the department of Agricultural & Resource Economics at the University of Arizona, and the 2003 Midwest Econometrics Group meetings. All errors are our own.

<sup>†</sup>Department of Economics, University of Arizona. alfonso@eller.arizona.edu

<sup>‡</sup>Department of Environmental and Natural Resource Economics, University of Rhode Island. schnier@uri.edu

# 1 Introduction

Econometric models taking into account spatial interactions among economic units have been increasingly used by economists over the last several years.<sup>1</sup> The different approaches for undertaking estimation and inference in linear regression models with spatial effects are well developed and have been summarized in the work by Anselin (1988, 2001), Anselin and Bera (1998), and other researchers.

The estimation of nonlinear models that include spatial interactions, in particular limited dependent variable models, is not as well developed as that of linear models. In fact, only recently have methods for estimating and conducting statistical inference in spatial models with limited dependent variables been proposed. This literature has concentrated mainly on the probit model with spatial effects, as in Case (1992), Beron and Vijverberg (1999), Fleming (2002), LeSage (2000), McMillen (1992), and Pinkse and Slade (1998). In this paper we contribute to this literature by introducing a sample selection model with spatial dependence and proposing a method for its estimation. The type of sample selection model considered is the widely used heckit model (Heckman, 1976, 1979), also known as the Tobit type II model in the terminology of Amemiya (1985). As for spatial dependence, we focus our attention mainly on the spatial autoregressive error (SAE) model, but we point out that our method can also be applied to another type of spatial dependence, the spatial autoregressive lag (SAL) model.

Our estimation strategy can be thought of as a two-step procedure analogous to the popular heckit model that is estimated jointly as a "pseudo" sequential estimator using GMM (Newey, 1984). The first step of estimation is based on a spatial probit model following a methodology by Pinkse and Slade (1998), which yields consistent, although not fully efficient, estimates of the selection equation. As in the heckit procedure, the consistent estimates of the selection equation are used to estimate the inverse Mills ratio (IMR) to be included in the estimation of the outcome equation to correct for selectivity bias. Since in the presence of spatial error dependence the IMR depends upon unknown parameters from the outcome equation, we propose to estimate the model jointly within a GMM framework.

The estimation of a probit model with spatial dependence introduces a non-spherical variance-covariance matrix that renders the simple probit estimator inconsistent. In turn, to obtain consistent and fully efficient estimates, one has to deal with multidimensional

---

<sup>1</sup>Some examples are Case (1991), Fishback, Horrace and Kantor (2002), Topa (2001), among many others.

integrals. To obtain consistent and efficient parameter estimates of the spatial probit model, LeSage (2000) and Beron and Vijverberg (1999) employ simulation methods to approximate these multidimensional integrals. Unfortunately, the simulation of multidimensional integrals is computationally intensive, restricting estimation to moderate sample sizes. This same limitation applies to the estimation of sample selection models with spatial dependence using simulation methods to approximate the multidimensional integrals in the likelihood function.<sup>2</sup>

In an attempt to avoid approximating multidimensional integrals but still achieve consistency of the probit estimates (at the expense of efficiency), some authors propose to ignore the off-diagonal elements of the variance-covariance matrix and focus on the heteroskedasticity induced by the spatial dependence (Case (1992), McMillen (1992) and Pinkse and Slade (1998)). We use Pinkse and Slade’s estimator in the first step of the sample selection model for the following reasons. First, it yields consistent estimates of the selection equation that are necessary to obtain consistent estimates of the parameters in the outcome equation. Second, it is computationally simpler than both of the other estimators that approximate multidimensional integrals. Third, it has been developed in the framework of GMM, the same framework we employ in the joint estimation of our sample selection models.

The consistent estimates obtained in the first step are used to construct the IMR used in the outcome equation to correct for selectivity bias (Heckman, 1979). In practice, both the selection and outcome equations are likely to exhibit spatial dependence, and generally the spatial autoregressive parameters will be different in each equation. In this case, the IMR under SAE turns out to be a function of the unknown spatial parameter in the outcome equation (second step). In order to increase the efficiency of the estimator and to obtain its variance-covariance matrix directly, we propose to estimate all parameters of the model simultaneously. For this, we employ the sequential estimation framework proposed by Newey (1984) to jointly estimate the sample selection model with SAE. In addition, building on Pinkse and Slade’s (1998) asymptotic results for their spatial probit model and standard GMM theory, we derive the asymptotic properties of our estimator.

As for the sample selection model with SAL dependence, we show that the corresponding IMR depends only upon parameters from the selection equation and thus can be estimated in two steps, similar to the traditional heckit model. However, we point out that the estimation

---

<sup>2</sup>For an example of a likelihood function for a sample selection model under independent observations see Heckman (1979) or Maddala (1983).

of this model using the sequential estimation GMM framework has the advantage of directly yielding the appropriate variance-covariance matrix of the estimated parameters.

Our estimators have two potential disadvantages. First, they have lower efficiency as compared to the computationally intensive simulation methods. However, given the fact that the multidimensional integration is in the order of the number of observations, the computational simplicity of our method is preferable in many relevant instances when the amount of data available to researchers is large. Second, our estimators require the availability of at least one instrumental variable in addition to the common "exclusion restriction" desirable in the heckit model. Nevertheless, the instrumental variables proposed by Kelejian and Prucha (1998) can be employed. These instruments are always available to the practitioner since they are constructed from polynomials of available variables. In our Monte Carlo simulations, we find that the Kelejian and Prucha instruments work well in practice.

To our knowledge, there is only one other paper attempting to specify and estimate a sample selection model with spatial effects. McMillen (1995) first specifies a model similar to the heckit model with spatial effects presented in our paper and proposes an estimator based on the EM algorithm. However, this estimator is not actually employed since it is impractical due to its computational intensity. Next, McMillen (1995) specifies and estimates an extension of the spatial expansion model of Casetti (1972) that is used in geography. However, this model is not explicitly spatial since additional variables are required to control for the spatial effects, and the consistency of the proposed estimator depends heavily on correctly assuming the functional form of the underlying heteroskedasticity induced by the spatial dependence. Compared to McMillen (1995), we propose a feasible estimator for the heckit model with spatial effects.

The paper is organized as follows. Section 2 presents the sample selection model with spatial autoregressive errors (SAE), which is the main focus of our paper, and also outlines the spatial autoregressive lag (SAL) model with sample selection. Section 3 introduces our proposed method of estimation (the "spatial heckit") and states its large-sample properties. Particular attention is devoted to the model with SAE, since the model with SAL can be analyzed with more standard asymptotic methods. Section 4 presents the results of a Monte Carlo experiment for the SAE model. Section 5 discusses the practical importance of both spatial sample selection models and presents an empirical application of the SAE model with sample selection. Concluding remarks are provided in the last section of the paper, and

proofs of the asymptotic properties of the estimators are presented in the Appendix.

## 2 The Sample Selection Models with Spatial Dependence

Our main focus is the estimation of a sample selection (Tobit type II) model with spatial autoregressive errors (SAE), since this model is slightly more challenging than the model exhibiting spatial autoregressive lags (SAL). We present here both models, which include spatial effects in both the selection and the outcome equations separately.

The spatial autoregressive error (SAE) model specifies spatially autocorrelated disturbances:

$$y_{1i}^* = \alpha_0 + x'_{1i}\alpha_1 + u_{1i}, \quad u_{1i} = \delta \sum_{j \neq i} c_{ij}u_{1j} + \varepsilon_{1i} \quad (1)$$

$$y_{2i}^* = \beta_0 + x'_{2i}\beta_1 + u_{2i}, \quad u_{2i} = \gamma \sum_{j \neq i} c_{ij}u_{2j} + \varepsilon_{2i} \quad (2)$$

where  $y_{1i}^*$  and  $y_{2i}^*$  are latent variables with the following relationship with the observed variables:  $y_{1i} = 1$  if  $y_{1i}^* > 0$  and  $y_{1i} = 0$  otherwise, and  $y_{2i} = y_{2i}^* * y_{1i}$ . Therefore, (1) is the selection equation while (2) is the outcome equation. Note that each of these equations exhibit spatial dependence, as  $u_{1i}$  and  $u_{2i}$  depend on other  $u_{1j}$  and  $u_{2j}$  through their location in space, as given by the spatial weights  $c_{ij}$  and the spatial autoregressive parameters  $\delta$  and  $\gamma$ . Typically, the spatial weights are specified by the econometrician based on some function of contiguity or (economic) distance (Anselin, 1988; Anselin and Bera, 1998). Note also that, in general, one will specify different spatial autoregressive parameters for the selection and outcome equations.<sup>3</sup> It is assumed that:

**Assumption A** The errors  $\varepsilon_{1i}$  and  $\varepsilon_{2i}$ ,  $i = 1, \dots, N$ , are *iid*  $N(\mathbf{0}, \Sigma)$  with

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}.$$

The model in (1)-(2) can also be presented in a reduced form:

$$y_{1i}^* = \alpha_0 + x'_{1i}\alpha_1 + \sum_j \omega_{ij}^1 \varepsilon_{1j} \quad (3)$$

$$y_{2i}^* = \beta_0 + x'_{2i}\beta_1 + \sum_j \omega_{ij}^2 \varepsilon_{2j} \quad (4)$$

---

<sup>3</sup>Without loss of generality, we specify the same spatial weights in each of the two equations.

where the weights  $\omega_{ij}^1$  and  $\omega_{ij}^2$  are the  $(i, j)$  elements of the inverse matrices  $(I - \delta C)^{-1}$  and  $(I - \gamma C)^{-1}$ , respectively, with  $C$  the matrix of spatial weights  $c_{ij}$ . Note that both sets of weights,  $\omega_{ij}^1$  and  $\omega_{ij}^2$ , depend upon the unknown parameters  $\delta$  and  $\gamma$ , respectively.

The spatial autoregressive lag (SAL) model specifies that the dependent variable of each unit is spatially dependent on the value of the dependent variable in all (or some) of the other units:

$$y_{1i}^* = \alpha_0 + x'_{1i}\alpha_1 + \delta \sum_{j \neq i} c_{ij}y_{1j}^* + \varepsilon_{1i} \quad (5)$$

$$y_{2i}^* = \beta_0 + x'_{2i}\beta_1 + \gamma \sum_{j \neq i} c_{ij}y_{2j}^* + \varepsilon_{2i} \quad (6)$$

where  $y_{1i}^*$  and  $y_{2i}^*$  are latent variables with the following relationship with the observed variables:  $y_{1i} = 1$  if  $y_{1i}^* > 0$  and  $y_{1i} = 0$  otherwise, and  $y_{2i} = y_{2i}^* * y_{1i}$ . Assumption A is applied here as well to the error terms  $\varepsilon_{1i}$  and  $\varepsilon_{2i}$ . The spatial dependence implied by the SAL model is given by the spatial weights  $c_{ij}$  and the spatial autoregressive lag parameters  $\delta$  and  $\gamma$ . A reduced form of the SAL model similar to equations (3) and (4) is:

$$y_{1i}^* = \alpha_0 \sum_j \omega_{ij}^1 + \alpha_1 \sum_j \omega_{ij}^1 x'_{1j} + \sum_j \omega_{ij}^1 \varepsilon_{1j} \quad (7)$$

$$y_{2i}^* = \beta_0 \sum_j \omega_{ij}^2 + \beta_1 \sum_j \omega_{ij}^2 x'_{2j} + \sum_j \omega_{ij}^2 \varepsilon_{2j} \quad (8)$$

where the weights  $\omega_{ij}^1$  and  $\omega_{ij}^2$  are as defined in the reduced form of the SAE model, and we note again that these weights are functions of the unknown spatial parameters  $\delta$  and  $\gamma$ , respectively, in both the SAL and SAE models.

### 3 Estimation of the Sample Selection Models with Spatial Dependence

We now describe our proposed estimation method for the sample selection models with spatial dependence presented in the previous section. As previously described, we follow a two-step procedure in the spirit of Heckman (1976, 1979) that is estimated jointly in a GMM framework. The selection equation is estimated using Pinkse and Slade's (1998) GMM estimator for the spatial probit model, while the outcome equation is estimated with the spatial methods for linear models developed by Kelejian and Prucha (1998), although other methods can be employed as well. An estimate of the inverse Mills ratio is included in the outcome

equation to correct for selectivity bias. To estimate these two parts simultaneously, the corresponding moment conditions are stacked, and a GMM criterion function is minimized with respect to all parameters in the model. In what follows, we consider each type of spatial dependence separately.

### 3.1 Estimation of the SAE sample selection model

To motivate the estimation of the SAE model in (1)-(2), we start with the following calculations:

$$\text{var}(u_{1i}) = \sigma_1^2 \sum_j (\omega_{ij}^1)^2 \quad (9)$$

$$\text{var}(u_{2i}) = \sigma_2^2 \sum_j (\omega_{ij}^2)^2 \quad (10)$$

$$E(u_{1i}, u_{2i}) = \sigma_{12} \sum_j \omega_{ij}^1 \omega_{ij}^2. \quad (11)$$

In the typical heckit model, a probit model is employed in the first step to estimate the probability of each observation being included in the observed sample. The presence of SAE errors, however, induces heteroskedasticity in the error terms in (9), resulting in inconsistent probit estimates. Pinkse and Slade (1998) propose a consistent estimator for this spatial probit model by taking into account the known form of the induced heteroskedasticity.

Define  $\theta_1 = \{\alpha_0, \alpha'_1, \delta\}$  as the parameters to be estimated in the spatial probit model, and  $\psi_i(\theta_1) = \frac{\alpha_0 + x'_{1i} \alpha_1}{\sqrt{\text{var}(u_{1i})}}$  the index function of the probit model weighted by the standard deviation of the residual. The corresponding generalized residuals of this model are:

$$\tilde{u}_{1i}(\theta_1) = \{y_{1i} - \Phi[\psi_i(\theta_1)]\} \cdot \frac{\phi[\psi_i(\theta_1)]}{\Phi[\psi_i(\theta_1)] \{1 - \Phi[\psi_i(\theta_1)]\}}. \quad (12)$$

The GMM estimates for  $\theta_1$  can be obtained as follows:

$$\hat{\theta}_{1,GMM} = \arg \min_{\theta_1 \in \Theta_1} S_N(\theta_1)' M_N S_N(\theta_1) \quad (13)$$

where  $S_N = \frac{1}{N} z'_N \tilde{u}_{1N}(\theta_1)$ ,  $z_N$  is a matrix of regressors plus at least one instrument (to identify the extra parameter  $\delta$ ), and  $M_N$  is a positive definite matrix such that  $M_N \xrightarrow{p} M$ . Pinkse and Slade (1998) show that this estimator is consistent and asymptotically normal.

The consistent estimates of  $\theta_1$  will be used in the construction of the inverse Mills ratio (IMR) to correct for sample selection bias. Note that the conditional regression function for

the outcome equation (2) has the following form:

$$\begin{aligned}
E[y_{2i}|y_{1i} > 0] &= \beta_0 + x'_{2i}\beta_1 + E[u_{2i}|u_{1i} > -(\alpha_0 + x'_{1i}\alpha_1)] \\
&= \beta_0 + x'_{2i}\beta_1 + \frac{E(u_{1i}, u_{2i})}{\sqrt{\text{var}(u_{1i})}} \cdot \frac{\phi[-\psi_i(\theta_1)]}{\{1 - \Phi[-\psi_i(\theta_1)]\}} \\
&= \beta_0 + x'_{2i}\beta_1 + \frac{\sigma_{12} \sum_j \omega_{ij}^1 \omega_{ij}^2}{\sqrt{\sigma_1^2 \sum_j (\omega_{ij}^1)^2}} \cdot \frac{\phi[-\psi_i(\theta_1)]}{\{1 - \Phi[-\psi_i(\theta_1)]\}} \\
&= \beta_0 + x'_{2i}\beta_1 + \frac{\sigma_{12}}{\sigma_1} \cdot \frac{\sum_j \omega_{ij}^1 \omega_{ij}^2}{\sqrt{\sum_j (\omega_{ij}^1)^2}} \cdot \frac{\phi[-\psi_i(\theta_1)]}{\{1 - \Phi[-\psi_i(\theta_1)]\}}
\end{aligned}$$

Therefore, the selectivity correction implies the following "adjusted" IMR:

$$\lambda_i \equiv \frac{\sum_j \omega_{ij}^1 \omega_{ij}^2}{\sqrt{\sum_j (\omega_{ij}^1)^2}} \cdot \frac{\phi[-\psi_i(\theta_1)]}{\{1 - \Phi[-\psi_i(\theta_1)]\}}. \quad (14)$$

Once estimated, the "adjusted" IMR may be included as an additional regressor in the outcome equation, which in turn could be estimated with any of the spatial methods developed for this linear equation.<sup>4</sup> We illustrate our estimator in this section by employing feasible generalized least squares (FGLS) to estimate the augmented outcome equation:  $y_{2i} = \beta_0 + x'_{2i}\beta_1 + \mu\hat{\lambda}_i + v_{2i}$ .

However, note that the "adjusted" IMR in (14) depends on a parameter that is not estimated in the first step:  $\gamma$ , which is included in the weights  $\omega_{ij}^2$ . In order to increase the efficiency of the estimator and directly obtain its variance covariance matrix, we propose using GMM to estimate simultaneously all parameters of the sample selection model by rewriting it as a sequential estimator (Newey, 1984) composed of the Pinkse and Slade (1998) and FGLS estimators. More specifically, we stack their corresponding moment conditions:

$$g(z, \theta) = [s(z_{1N}, \theta)', m(z_{2N}, \theta)']', \quad \theta = \{\alpha_0, \alpha'_1, \delta, \beta_0, \beta'_1, \mu, \gamma\}$$

with

$$\begin{aligned}
s(z_{1N}, \theta) &= z'_{1N} \tilde{u}_{1N}(\theta), \quad \tilde{u}_{1N}(\theta) \text{ as in (12)}, \\
m(z_{2N}, \theta) &= [y_{1N} \cdot z_{2N}]' \tilde{u}_{2N}(\theta), \quad \tilde{u}_{2N}(\theta) = y_{2N} - \beta_0 - x'_{2N}\beta_1 - \mu\hat{\lambda}(\delta, \gamma)
\end{aligned}$$

---

<sup>4</sup>The linear SAE model could be estimated, for instance, using the methods by Kelejian and Prucha (1998), Lee (2001a), or maximum likelihood (see Anselin 1988).



and  $z_{1N}$  includes the regressors of the selection equation plus at least one instrument, and  $z_{2N}$  includes the regressors of the outcome equation, the estimated "adjusted" IMR, and at least one instrument.<sup>5</sup> Both instruments could in principle be the same. We note that the instruments proposed by Kelejian and Prucha (1998) can be used if no other instruments are available.<sup>6</sup>

Define  $z'_N = (z'_{1N}, [y_{1N} \cdot z_{2N}])'$  and  $\tilde{u}_N(\theta) = (\tilde{u}'_{1N}(\theta), \tilde{u}'_{2N}(\theta))'$ . Then, all parameters of the SAE sample selection model can be estimated as:

$$\hat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} g_N(\theta)' M_N g_N(\theta) \quad (15)$$

where  $g_N = N^{-1} z'_N \tilde{u}_N(\theta)$ , for a positive definite  $M_N$  such that  $M_N \xrightarrow{p} M$ .

We show in the appendix that, under conditions similar to those in Pinkse and Slade (1998),  $\hat{\theta}_{GMM}$  is consistent and asymptotically normal. We call it the "spatial heckit" estimator for the SAE model.

**Proposition 1** Under assumptions A1 to A8 (stated in the appendix),  $\hat{\theta}_{GMM} \xrightarrow{p} \theta_0$ .

The following two propositions concern the asymptotic distribution and the estimation of the variance-covariance matrix of the spatial heckit estimator for the SAE model.

**Proposition 2** Under assumptions A1 to A11 (stated in the appendix),

$$\sqrt{N}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(\mathbf{0}, [\Psi_2(\theta_0)]^{-1} [\partial g'(\theta_0) / \partial \theta] M \Psi_1(\theta_0) M [\partial g(\theta_0) / \partial \theta'] [\Psi_2(\theta_0)]^{-1})$$

where  $\Psi_1(\theta_0) = \lim_{N \rightarrow \infty} E\{N g_N(\theta_0) g_N(\theta_0)'\}$ , and  $\Psi_2(\theta_0) = [\partial g'(\theta_0) / \partial \theta] M [\partial g(\theta_0) / \partial \theta']$ .

**Proposition 3** Under assumptions A1 to A14 (stated in the appendix), then

$$\begin{aligned} \Psi_{1N}(\hat{\theta}_{GMM}) &\xrightarrow{p} \Psi_1(\theta_0) \quad \text{and} \quad \Psi_{2N}(\hat{\theta}_{GMM}) \xrightarrow{p} \Psi_2(\theta_0), \quad \text{where} \\ \Psi_{1N}(\hat{\theta}_{GMM}) &= NE\{g_N(\hat{\theta}_{GMM}) g_N(\hat{\theta}_{GMM})'\} \quad \text{and} \\ \Psi_{2N}(\hat{\theta}_{GMM}) &= [\partial g'_N(\hat{\theta}_{GMM}) / \partial \theta] M_N [\partial g_N(\hat{\theta}_{GMM}) / \partial \theta']. \end{aligned}$$

---

<sup>5</sup>Kelejian and Prucha (1999) propose a set of three moment conditions for the estimation of the spatial autoregressive parameter in a SAE model, which are based on second-order moments of the residuals. These moments are included in both  $s(z_{1N}, \theta)$  and  $m(z_{2N}, \theta)$ , although they are not explicitly shown here to simplify the exposition.

<sup>6</sup>In the context of a linear SAL model, Kelejian and Prucha (1998) show that the optimal set of instruments in a linear equation is approximated by the linearly independent columns of  $[x, Cx, C^2x, \dots]$ , where  $x$  is the set of exogenous independent variables and  $C$  is the spatial weighting matrix. These instruments can also be used in our context because they are exogenous.

Finally, the following corollary gives the (efficient) asymptotic variance of our estimator when the optimal GMM weighting matrix is used.

**Corollary 1** If  $M_N = [\Psi_{1N}(\hat{\theta}_{GMM})]^{-1}$  is chosen, such that  $M_N \xrightarrow{p} [\Psi_1(\theta_0)]^{-1}$ , then the asymptotic distribution of  $\hat{\theta}_{GMM}$  simplifies to:

$$\sqrt{N}(\hat{\theta}_{GMM} - \theta_0) \xrightarrow{d} N(\mathbf{0}, [\Psi_2(\theta_0)]^{-1}).$$

To end this subsection, we point out that the SAE sample selection model can also be estimated using a two-step procedure. In the first step of this procedure, consistent estimates of the parameters in  $\theta_1$  are obtained from (13). In the second step, nonlinear least squares (NLLS) is employed to estimate the parameters in  $y_{2i} = \beta_0 + x'_{2i}\beta_1 + \mu\hat{\lambda}_i + v_{2i}$ , in which the parameter  $\gamma$  enters nonlinearly in the "adjusted" IMR ( $\hat{\lambda}_i$ ). This procedure is attractive since it preserves the two-step intuition of the heckit model, however, to estimate the correct standard errors for the second-step estimates one must adjust for the fact that the parameters in  $\theta_1$  are estimated in a first step, and also employ a heteroskedasticity-consistent variance-covariance estimator for NLLS since  $v_{2i}$  is non-spherical. We avoid the extra steps to obtain correct standard errors by employing the sequential GMM estimator in (15).

### 3.2 Estimation of the SAL sample selection model

We start by noting that in the sample selection model with SAL dependence the IMR does not depend on parameters from the outcome equation (second step). To see this, note that from (5) and (6):

$$\text{var}(\varepsilon_{1i}) = 1, \quad \text{var}(\varepsilon_{2i}) = \sigma_2^2, \quad \text{and} \quad E(\varepsilon_{1i}, \varepsilon_{2i}) = \sigma_{12}. \quad (16)$$

and the conditional regression function for the outcome equation (6) becomes:

$$\begin{aligned} E[y_{2i}|y_{1i} > 0] &= \beta_0 + x'_{2i}\beta_1 + \gamma \sum_{j \neq i} c_{ij}y_{2j} + E[\varepsilon_{2i}|\varepsilon_{1i} > -(\alpha_0 + x'_{1i}\alpha_1 + \delta \sum_{j \neq i} c_{ij}y_{1j})] \\ &= \beta_0 + x'_{2i}\beta_1 + \gamma \sum_{j \neq i} c_{ij}y_{2j} + \frac{E(\varepsilon_{1i}, \varepsilon_{2i})}{\sqrt{\text{var}(\varepsilon_{1i})}} \cdot \frac{\phi[-\varphi_i(\theta_1)]}{\{1 - \Phi[-\varphi_i(\theta_1)]\}} \\ &= \beta_0 + x'_{2i}\beta_1 + \gamma \sum_{j \neq i} c_{ij}y_{2j} + \frac{\sigma_{12}}{\sigma_1} \cdot \frac{\phi[-\varphi_i(\theta_1)]}{\{1 - \Phi[-\varphi_i(\theta_1)]\}} \end{aligned}$$

where  $\varphi_i(\theta_1) \equiv \alpha_0 + x'_{1i}\alpha_1 + \delta \sum_{j \neq i} c_{ij}y_{1j}$ . The second multiplicand in the last term is the usual IMR that only depends on parameters from the selection equation. We suggest the

use of the same sequential estimation framework as with the SAE model to directly obtain the appropriate variance-covariance matrix of the estimated parameters (Newey, 1984). In addition, we note that under the assumptions maintained for the SAL model and regularity conditions on the spatial dependence process (such as  $A6$  and  $A9$  in the appendix), standard GMM asymptotic results apply, and thus the spatial heckit estimator for the SAL model is consistent and asymptotically normal.<sup>7</sup>

For completeness, we outline the moment conditions upon which the spatial heckit estimator for the SAL model is based. First, consider the selection equation in (5) that implies a binary choice model for the observed variable  $y_{1i}$ , which under assumption A can be rewritten as:

$$P(y_{1i} = 1) = \Phi(\alpha_0 + x'_{1i}\alpha_1 + \delta \sum_{j \neq i} c_{ij}y_{1j}) = \Phi(\varphi_i). \quad (17)$$

Note that this probit model contains both endogenous and exogenous variables, and thus a nonlinear two-stage least-squares (NL2SLS) procedure (Amemiya, 1974) can be used to estimate the parameters of interest. Let  $z_{1i}$  be a vector of variables that contains a constant,  $x'_{1i}$ , plus at least one instrument.<sup>8</sup> Replacing  $\varphi_i$  for  $\psi_i$  in the definition of the generalized residuals in (12), the NL2SLS estimator can be obtained by solving the following moment conditions:

$$s(z_1, \theta_1) = z'_{1N} \tilde{u}_{1N}(\theta_1) = 0$$

where  $\theta_1 = \{\alpha_0, \alpha'_1, \delta\}$ .

The outcome equation, now expanded by the inclusion of the IMR, can be estimated with any of the spatial methods developed for this linear equation.<sup>9</sup> We illustrate our estimator employing 2SLS. Define

$$m(z_{2N}, \theta) = [y_{1N} \cdot z_{2N}]' \tilde{u}_{2N}(\theta), \quad \tilde{u}_{2N}(\theta) = y_{2N} - \beta_0 - x'_{2N}\beta_1 - \mu\hat{\lambda}(\theta_1) - \gamma C y_{2N}$$

where  $z_{2N}$  includes the regressors of the outcome equation, the estimated IMR, and at least one instrument, which could be in principle the same instrument(s) contained in  $z_1$ . Defining  $z'_N = (z'_{1N}, [y_{1N} \cdot z_{2N}]')'$  and  $\tilde{u}_N(\theta) = (\tilde{u}'_{1N}(\theta_1), \tilde{u}'_{2N}(\theta))'$ . Then, all parameters of the sample selection model with SAL dependence can be estimated as:

$$\hat{\theta}_{GMM} = \arg \min_{\theta \in \Theta} g_N(\theta)' M_N g_N(\theta)$$

<sup>7</sup>For a review of the asymptotic properties of GMM sequential estimators see Newey (1984).

<sup>8</sup>In particular, the same instruments proposed by Kelejian and Prucha (1998) can be used.

<sup>9</sup>The linear SAL model could be estimated, for instance, with 2SLS, using the methods by Lee (2001b), or maximum likelihood (see Anselin 1988).

where  $g_N = N^{-1}z'_N\tilde{u}_N(\theta)$ , for a positive definite  $M_N$  such that  $M_N \xrightarrow{p} M$ .

## 4 Monte Carlo Experiment

We conduct a Monte Carlo experiment to explore the finite-sample performance of the spatial heckit estimator for the sample selection model with spatially autoregressive errors (SAE). The spatial heckit estimator is compared to three other estimators: the Kelejian and Prucha (1998) estimator for the SAE model that ignores sample selection but accounts for spatial dependence; the simple heckit estimator that accounts for sample selection but ignores spatial dependence; and finally the ordinary least squares (OLS) estimator that ignores both sample selection and spatial dependence. We compare these estimators in terms of their finite sample bias and root-mean square error.

Given that we combine features of a sample selection model with a spatial dependence specification, we pay close attention to previous simulation studies in specifying each of the two features of our models, such as Cosslett (1991) and Leung and Yu (1996) for the sample selection model, and Beron and Vijverberg (1999) and Kelejian and Prucha (1999) for spatially dependent models.

Our data generating process (DGP) is as follows :

$$y_{1i}^* = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_{1i}, \quad u_{1i} = \delta \sum_{j \neq i} c_{ij} u_{1j} + \varepsilon_{1i} \quad (18)$$

$$y_{2i}^* = \beta_0 + \beta_1 x_{3i} + \beta_2 x_{1i} + u_{2i}, \quad u_{2i} = \gamma \sum_{j \neq i} c_{ij} u_{2j} + \varepsilon_{2i}. \quad (19)$$

Each of our models consists of three uncorrelated exogenous variables, one of which is common to both equations, as in Cosslett's (1991) experimental design. These exogenous variables are generated as  $x_k \sim U(0, 1)$ ,  $k = 1, 2, 3$ . The innovations  $\varepsilon_{1i}$  and  $\varepsilon_{2i}$  are generated bivariate normal as follows:

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \sim N \left( \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \quad (20)$$

where we set  $\rho = 0.5$  for the correlation between the innovations in each of the two equations. The parameters of the model that are not related to the spatial dependence feature are set at  $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 1$  and  $\beta_0 = 0$ . The parameter  $\alpha_0$  is used to control the amount of sample selection, for which we consider two cases: 25% censoring ( $\alpha_0 = -0.3$ ) and 40% censoring ( $\alpha_0 = -.77$ ). We consider in the experiment three different sample sizes:  $N = 100$ ,

225, and 400 observations. Importantly, these sample sizes refer to the uncensored sample, that is, the average number of observations available for estimation of the outcome equation is 75, 168, and 300, respectively, for the case of 25% sample selection; and 60, 135, and 240, respectively, for the case of 40% sample selection.

Regarding the spatial autoregressive parameter, we consider three different values: 0.25, 0.5, and 0.75. In addition, we consider two different types of instruments in our simulations to gauge their relative performance: the Kelejian and Prucha (KP) instruments, which are always available to the researcher;<sup>10</sup> and two artificially generated exogenous instruments that mirror a situation in which the researcher has available additional instruments. These generated instruments are specified to have a correlation of about 0.2 with the exogenous included variables in the model. Due to computational constraints, 250 replications are undertaken for the models with  $N = 100$ , 150 replications for the models with  $N = 225$ , whereas 50 replications are undertaken for the models with  $N = 400$ .

The matrix of spatial weights has to be specified. For this, we create three grids of 10 by 10, 15 by 15 and 20 by 20 for the observation matrices of 100, 225 and 400, respectively. Each grid is assigned an X and Y coordinate centered on the grid such that the bottom left corner of the grid had a value of (0.5, 0.5). We use these grids to create a weighting matrix that is based on the square of the inverse Euclidean distance between any two points. After creating the location specific weights for each grid, the matrix is row standardized so that the diagonal elements of the weighting matrix are all zeros and the sum of any one row is equal to 1. Finally, a band is used to determine the number of observations that may influence a centered observation. Such band is set with a lower bound of 0 and an upper bound equal to  $\sqrt{5}$ . As a result, there exists spatial dependence for all the observations, and, on average, each observation has 4 neighbors.<sup>11</sup> This way of specifying the spatial weighting matrix is widely used within the literature, see, for instance, Anselin (1988).

Tables 1 through 3 present simulation results for the outcome equation of the SAE model for samples of size 100, 225 and 400, respectively. In addition to presenting simulation results for OLS, heckit and Kelejian and Prucha's (1998) estimator for the SAE model (KP-SAE), Table 1 also shows simulation results for the two different versions of our spatial heckit estimator: one in which the generated instruments discussed above are used (Speck Gen)

---

<sup>10</sup>More specifically, in the simulations we use the independent columns of  $[x, Cx, C^2x]$ .

<sup>11</sup>We note that Kelejian and Prucha (1999) find that controlling or not for the number of neighbors per unit when specifying a weighting matrix does not lead to significantly different results in their simulation study.

and another in which the Kelejian and Prucha instruments are employed (Speck KP).<sup>12</sup> The first column in each table indicates the extent of sample selection (sel) and spatial dependence ( $\delta = \gamma$ ) for the models considered. For each of these models, up to four parameters of interest are reported, which are listed in the second column of each table. The remaining columns in the table are arranged in two blocks that correspond to the average bias (BIAS) and the root-mean squared error (RMSE) of the different estimators.

The first estimator reported is OLS, which ignores both features of the data: sample selection and spatial dependence. As a result, OLS is inconsistent, which is reflected in the fact that it has large bias and RMSE that increase as the amount of sample selection or spatial dependence increases. In addition, neither the bias nor the RMSE decrease as the sample size increases. In general, though, OLS is able to estimate  $\beta_1$  (the coefficient on the variable that does not appear in the selection equation) with relatively small bias compared to the other coefficients. This is due to the fact that the variables  $x_k$  ( $k = 1, 2, 3$ ) are generated independently, and thus there is little effect of the sample selection on the coefficient on  $x_3$  ( $\beta_1$ ).

The second estimator reported is the heckit estimator, which accounts for sample selection but ignores spatial dependence. The consequence of spatial dependence on this estimator is inconsistency, as explained above, since the probit model that is estimated in the first step is heteroskedastic due to the spatial dependence. Therefore, we might expect that the bias and RMSE of the heckit will increase as the amount of spatial dependence increases. This is typically the case in Table 1, except for the bias of the parameter  $\beta_1$ . In addition, the bias also frequently increases as the amount of sample selection increases, except for the same parameter  $\beta_1$ . Compared to OLS, the heckit estimator shows a great improvement, even though it is also inconsistent in theory. While the average bias and RMSE of  $\beta_1$  is very small and comparable to that of OLS, the other two coefficients ( $\beta_0$  and  $\beta_2$ ) have larger bias: in the case of  $\beta_2$ , the bias has the interpretation of percentage and ranges from 3.9-13.3, 1.5-9.7 and 2.2-4.2 percent for sample sizes 100, 225 and 400, respectively. In the case of  $\beta_0$  the bias (not interpreted as percentage) ranges from -1.32 to -0.22, -0.16 to 0.003 and -0.12 to -0.07 for sample sizes 100, 225 and 400, respectively. Thus, the range of the bias on these coefficients decreases across model specifications as the sample size increases. Interestingly,

---

<sup>12</sup>The spatial heckit estimators require starting values. Both in the simulations and in the empirical application presented below, we employ starting values that are available in practice. In particular, the starting values employed for all parameters except  $\delta$  and  $\gamma$  are the heckit parameter estimates. The starting values employed for  $\delta$  and  $\gamma$  are equal to the KP-SAE estimate of  $\gamma$ .

the bias of these two coefficients is of opposite sign compared to the bias in OLS, with only a few exceptions within the models with 100 observations. Finally, the RMSE of the heckit estimator for the coefficients  $\beta_0$  and  $\beta_2$  is much larger than that of OLS.

The third estimator reported is KP-SAE that accounts for spatial dependence but ignores the sample selection feature of the data, which would result in inconsistent parameter estimates. In agreement with this notion, the bias and RMSE increase as the amount of sample selection increases, but they also typically increase as the amount of spatial dependence increases. Compared to the previous two estimators, with the exception of  $\beta_1$ , the KP-SAE shows a bias in the coefficients  $\beta_0$  and  $\beta_2$  of about the same magnitude as that of the OLS estimator, which is significantly larger than that of the heckit estimator. In the same way, the RMSE of the KP-SAE estimator is of about the same magnitude as that of OLS and thus smaller than that of heckit. The similarity between the results of OLS and KP-SAE might not be surprising since in theory the latter offers only a gain in efficiency over OLS in the absence of sample selection.

KP-SAE is the first estimator that produces an estimate of the SAE parameter  $\gamma$ , which shows bias and RMSE that increase with both the amount of sample selection and spatial dependence. The bias of  $\gamma$  ranges from 21.6 to 53.2 percent of the true  $\gamma$  in the models with sample size of 100, while in the models with sample size of 225 and 400 the bias ranges from 17.6 to 73 and 36.4 to 79.7 percent of the true  $\gamma$ , respectively. Therefore, the bias in the KP-SAE estimate of  $\gamma$  appears not to vanish as the sample size increases. Finally the RMSE does not become smaller as more observations are available, as it ranges from 0.075 to 0.407, 0.051 to 0.548 and 0.091 to 0.599 for sample sizes 100, 225 and 400 respectively.

The last two estimators are the two versions of the spatial heckit estimator, Spheck Gen and Spheck KP, which are consistent for all parameters in the model. Reporting these two versions allows the comparison of the spatial heckit using potentially available external instruments and using the always available instruments proposed by Kelejian and Prucha (KP).

In Table 1 ( $N = 100$ ), the Spheck Gen estimator of  $\beta_0$  and  $\beta_2$  possesses smaller bias than OLS and KP-SAE, but typically larger bias than heckit, especially when the sample selection is 25%. In fact, Spheck Gen performs much better than all other three estimators (including heckit) in the model with the highest sample selection (40%) and spatial dependence (0.75). In terms of the RMSE of  $\beta_0$  and  $\beta_2$ , however, Spheck Gen has typically the highest among

all estimators across all models.<sup>13</sup> In terms of  $\beta_1$ , Speck Gen has low bias and RMSE, comparable to that of other estimators. Comparing the KP-SAE estimate of  $\gamma$  with that of Speck Gen, the latter estimator has smaller bias and RMSE than the former in all models, except those with low spatial dependence ( $\gamma = 0.25$ ). Finally, comparing the two versions of Speck using  $N = 100$ , the version that employs generated instruments clearly outperforms in most models the version that uses the KP instruments. Nevertheless, as we show below, this difference in performance vanishes as the sample size increases.

As the sample size is increased to 225 observations in Table 2, the performance of the Speck Gen estimates of  $\beta_0$  and  $\beta_2$  improve in terms of bias, possessing now a smaller bias than the estimates of all other estimators including hekit. At the same time, however, the bias of the Speck Gen estimator of  $\beta_1$  is typically slightly higher than that of the other three estimators (although this bias is small in absolute terms), and the RMSE for this parameter is small and comparable to that of the other estimators. In terms of RMSE of  $\beta_0$  and  $\beta_2$ , the Speck Gen estimator shows a great improvement over the smaller sample size, now having a comparable RMSE to the hekit estimator, but still larger than the other two estimators (OLS and KP-SAE). Regarding the estimate of  $\gamma$ , the comparison between KP-SAE and Speck Gen is similar to the one with the smaller sample size: Speck Gen has smaller bias and RMSE in all models, except those with low spatial dependence ( $\gamma = 0.25$ ). Nevertheless, while the RMSE of Speck Gen generally decreases as we move from 100 to 225 observations, that of KP-SAE actually increases. Finally, the results using  $N = 225$  show that, while Speck Gen still outperforms Speck KP, the two estimators are now more comparable in terms of bias and RMSE.

Table 3 presents simulation results for a sample size of  $N = 400$ . In terms of bias of  $\beta_0$  and  $\beta_2$ , Speck Gen still outperforms all other estimators in all models (except for the hekit in two cases in which the two estimators are very close). The bias of  $\beta_1$  using the Speck Gen estimator, however, is typically higher than that of the other estimators. In terms of RMSE and in contrast to the simulations with smaller sample sizes, the Speck Gen estimator now has similar RMSE to the other estimators considered (most notably OLS and hekit) for all three parameters  $\beta_1$ ,  $\beta_2$ , and  $\beta_3$ . This may be attributed to the larger sample size ( $N = 400$ ), which indicates that in this case the finite sample properties of the spatial

---

<sup>13</sup>Recall that even though  $N = 100$  in these models, only about 75 (25% selection) or 60 (40% selection) observations are available to estimate the outcome equation. The small number of observations likely causes the Speck estimators to have higher variance compared to the other estimators.



heckit estimator are reasonably good for a sample as small as 400 observations. In terms of the spatial autoregressive parameter  $\gamma$ , Speck Gen outperforms KP-SAE in terms of both bias and RMSE, except again in the models with low spatial dependence ( $\gamma = 0.25$ ). As before, we note that the RMSE of  $\gamma$  for Speck Gen further decreases with the larger sample size, while that for KP-SAE increases. Finally, comparing the performance of Speck Gen with that of Speck KP in the models with  $N = 400$  corroborates the previously noted trend that the two estimators become more similar in terms of bias and RMSE as the sample size increases. This lends support to the notion that as more observations are available, the version of the spatial heckit estimator that employs KP instruments yields similar results to employing exogenous instruments (if available).<sup>14</sup>

Table 4 presents simulation results for the selection equation. This equation is only estimated in the heckit (using a probit model) and the two versions of Speck. However, given their similar performance and to save space, only the results for Speck Gen are presented.<sup>15</sup> Even though we present results for the two cases of 25% and 40% censoring, the same number of observations are employed in each case since it is the selection equation. However, the value of the constant term is different in each case.

It is evident from Table 4 that the spatial heckit estimator (Speck Gen) typically outperforms the heckit estimator in terms of bias, while the reverse is true in terms of RMSE. However, the RMSE of the Speck Gen estimator is comparable to that of the heckit model even in a sample as small as 100 observations. In addition, the improvement in RMSE of the Speck Gen estimator as the sample size increases is larger than the one observed in the heckit model. As a result, the Speck Gen estimator achieves a RMSE essentially identical to that of the heckit estimator in the models with 400 observations. Finally, it should be pointed out that the estimate of the spatial parameter  $\delta$  by the Speck Gen estimator shows considerable bias across model specifications, with a relatively large bias for models with lower spatial dependence. Importantly, however, the bias and RMSE of this parameter estimate decreases substantially as more observations are available.

---

<sup>14</sup>Recall that the current comparison uses exogenously generated instrumental variables with a correlation of 0.2 with the included regressors. Depending on the application, such correlation could be considered high, which lends further support to the notion that the KP instruments are a good alternative to use in practice when the quality of other instruments is doubtful.

<sup>15</sup>The results for the selection equation for the Speck KP estimator are available upon request. In summary, the results for Speck Gen are only slightly better than for the Speck KP model. It is worth noting that, contrary to the outcome equation, the two versions of the spatial heckit estimator are very similar, even in the models with the smallest sample size.

In summary, we regard the simulation results as encouraging with respect to the finite sample properties of the spatial heckit estimator. In particular, the fact that the advantages of our estimator are evident in the simulations despite using relatively small sample sizes is worth pointing out. Finally, the simulations also show that as more observations become available the two different sets of instruments considered for the spatial heckit estimator produce very similar results in terms of both bias and RMSE in all four parameters. This finding is important for in practice researchers might find it hard to find quality instrumental variables, whereas the KP instruments are always available.

## 5 The Sample Selection Models with Spatial Dependence in Practice

In this section we discuss the empirical importance of taking into account sample selection bias when estimating models that exhibit spatial dependence. Furthermore, we illustrate the application of the spatial heckit estimator for the sample selection model with SAE dependence using a data set from a fishery, which is censored for confidentiality reasons.

McMillen (1995) motivates the pervasiveness of sample selection problems in spatial data, in particular in urban economics and regional science. His main example deals with data on land use and values in the city of Chicago during the 1920s (see references in McMillen, 1995). In this case, unobserved variables that make a parcel more likely to receive residential zoning may increase the value of residential land (McMillen, 1995). Other applications of sample selection models with spatial data discussed in McMillen (1995) include models of housing prices, rent and tenure choice (Goodman, 1988), office rents and lease provisions (Benjamin et al., 1992), and home improvement choice (Montgomery, 1992) in urban economics; the choice between central city and suburban employment (McMillen, 1993), and analysis of earnings and migration (Borjas et al., 1992) in labor economics. In fact, the increasing availability of geo-coded data makes even more relevant the availability of methods to deal with sample selection when spatial dependence is present.

Our application in this section is in the area of natural resource economics, in particular fisheries economics. We employ a sample selection model with SAE dependence to estimate the spatial efficiency of production within a fishery, using a data set from the Pacific cod fishery within the Eastern Bering Sea that is censored for reasons of confidentiality. As in the previous section, we compare the performance of the spatial heckit estimator for the

SAE with OLS, the traditional heckit model and Kelejian and Prucha's SAE estimator.

Within fisheries management there has been a strong push to expand the suite of management regimes implemented and economic models used to evaluate them, to incorporate the spatial and temporal structure of the bioeconomic model. This has even lead some to draw the conclusion that future spatial management regimes will be defined not only over time, technology and location but also over depth and degree of implementation (Wilens, 2004). Therefore, the challenge for fisheries economists is to expand their models, both empirical and theoretical, to reflect this frontier in fisheries management. An initial interest is to investigate the production process within fisheries over the spatial region defined by the distribution of the metapopulation harvested. The catch-per-unit-effort (CPUE) has been traditionally implemented to analyze the productivity and efficiency of the production process within fisheries, where CPUE is defined as the catch per a "haul" executed. A "haul" represents the technology used such as a trawl device, pot vessel, hook-and-line, jig, etc.

Previous empirical work has been focused on investigating the technical efficiency present within a fishery in an effort to determine the factors that explain deviations from the production frontier (Kirkley et al. 1995, 1998; Sharma and Leung 1998; Squires and Kirkley 1999; Pascoe and Coglán 2002; Viswanathan et al. 2002; Garcia del Hoyo et al. 2004). Applying the results of the previous investigations to the current front line of spatial fisheries management would be inappropriate because the spatial processes present are not incorporated into the model. Therefore, a spatial efficiency model is needed to determine the fleet's level of efficiency across space and time. Ignoring the spatial distribution of fleet efficiency may have an adverse impact on the welfare of fishermen. For instance, should a managing body decide to close a given spatial region within the fishery with a low level of spatial technical efficiency this will displace fishing effort into the surrounding areas. If these areas possess a higher level of spatial technical efficiency, it will force them to more exhaustively push the frontier of their production capabilities to capture the same amount of the target species. This will invariably yield a higher cost of harvesting and lower rents for the fishermen, more so than if a high efficiency area is closed instead.

The primary method utilized in previous studies of technical efficiency within a fishery is the stochastic frontier production function model developed by Aigner, Lovell and Schmidt (1977). This method is conducted on a vessel-by-vessel basis and specifies an error term with two components: the first is a symmetric (normally distributed) disturbance that allows the

production frontier to be stochastic; while the second error component is one-sided and reflects the inefficiency of individual vessels. The model we employ in this application is similar to Zellner, Kmenta and Dreze (1966) in that we allow the production frontier to be stochastic but ignore the potential inefficiency with respect to the frontier. We choose this specification for two reasons. First, we are not interested in the vessel specific characteristics over space but the total fleet's level of spatial technical efficiency over space, and thus it seems more natural to focus on the average production function. Secondly, an extension of our methods to a model such as that of Aigner, Lovell and Schmidt (1977) is beyond the scope of the current paper. Our main purpose with this empirical application is to illustrate the use of the spatial heckit estimator.

Determining the spatial rates of technical efficiency requires a very fine resolution of data revelation. Often times this data is screened for confidentiality reasons to preserve the privacy of the fishermen within the fleet. The current publicly available data set on fishing effort within the Eastern Bering Sea of Alaska is compiled by the National Oceanic and Atmospheric Administration (NOAA) from the observer and log-book data collected by the National Marine Fisheries Service (NMFS). This data set is censored by not reporting the CPUE within a location unless 4 or more vessels, possessing similar characteristics, fish within that region. Therefore, the use of this data by researchers is limited unless they employ an empirical method that can control for this censorship, which justifies the use of the spatial sample selection model developed within this paper. Within our data set, there are 320 observations, of which we observe 208, which equates to a sample selection rate of 35%.

Although the data set does not contain vessel identifiers, it is still possible to determine the overall level of the fleet's spatial efficiency using this data set. Should one decide to refine the analysis by focusing on inter and intra vessel differences in the spatial distribution of technical efficiency, vessel specific data would be necessary. This ultimately may appear to be a more interesting question. However, given that a researcher will invest a substantial amount of time and effort prior to obtaining this information, it may be beneficial to investigate the fleet performance to test for spatial heterogeneity in the fleets' spatial rate of technical efficiency before investigating the vessel specific model.

Our analysis is conducted on the Pacific cod fishery within the Eastern Bering Sea of Alaska for the year 1997 using the NOAA data and the spatial heckit model outlined within

this paper. Pacific cod (*Gadus macrocephalus*) is a demersal species targeted in the Alaskan groundfish fishery. Estimating the fleets' spatial efficiency with regard to this species is beneficial due to its broad distribution within the Eastern Bering Sea which makes it susceptible to recent regulations targeted to protect the Stellar sea lion rookeries and the even more recent concerns of essential fish habitat (EFH) management.

To conduct the spatial efficiency estimation, the spatial resolution of what is deemed a "location" must be defined. The spatial resolution utilized are the Alaska Department of Fish and Game's (ADF&G) statistical reporting units. This unit of measure divides the Eastern Bering Sea into a grid with each cell being one-half degree latitude by one degree longitude. For the year analyzed this divides the fishery into 90 spatially different locations within the fishery. This resolution is a finer resolution than has been used by many studies within the location choice literature (for example, Bockstael and Opaluch 1983; Eales and Wilen 1986; Dupont 1993; Holland and Sutinen 1999, 2000; Mistiaen and Strand 2000; Smith 2000).

The CPUE for the Pacific cod fishery is defined as the metric tons of fish caught during the year within the ADF&G statistical reporting regions. This measure is the average of all vessels that fished within this region of like vessel characteristics. Vessels were grouped according the size of vessel, gear utilized and type of vessel (catcher-processor vs. catcher-vessel) by NOAA. Therefore, each observation represents a relatively homogeneous micro-fleet within the Pacific cod fishery that fished within the ADF&G region. Given that these observations are spatially defined, it is plausible that they are spatially correlated, and therefore a spatial econometric method must be utilized to obtain appropriate estimates of the spatial efficiency. Indeed, two different Moran-I tests employed soundly reject the null hypothesis of zero spatial autocorrelation on the data. The first Moran-I test uses the residuals resulting from an OLS regression on the observed data, while the second test employs the residuals from the heckit model (Kelejian and Prucha 2001). The tests' statistics and p-values are reported in Table 5.

To allow comparison among the four estimators employed in the Monte Carlo experiment in the previous section, the relative spatial efficiency is estimated with each of them. We note that given the documented spatial error dependence and the sample selection in the data, a sample selection model with SAE dependence is likely the appropriate model. The OLS and SAE models are estimated using only the observations that are not censored,

while the heckit model accounts for the sample selection in the data but ignores the spatial dependence. These features most likely render the estimates of these three models inconsistent. The sample selection model with spatial dependence in (1)-(2) is estimated with  $y_2$  as the natural logarithm of CPUE and  $x_2$  containing the log-transformed bathymetric measurements corresponding with the maximum and minimum depth within a ADF&G statistical reporting area, the stock assessment data resulting from the NMFS annual biomass trawl survey, and dummy variables for the following vessel characteristics that determine the homogenized unit observed: catcher-vessel (CV), hook-and-line gear (HAL), non-pelagic trawl gear (NPT), and vessel at least 125 feet long (Large). As for the selection equation,  $x_1$  contains the same variables as  $x_2$  along with an additional variable, a one-year lagged biomass trawl survey observation, under the assumption that this lagged variable influences the probability that four or more vessels will fish in a given statistical reporting location but not the amount of "hauls" that will be conducted. A lower stock signal in the previous year will presumably reduce the attractiveness of revisiting this location in the upcoming year and therefore reduce the total number of vessels that fish within the location.

In order to determine the spatial weighting matrix, we use the following common specification to assign spatial weights among the statistical reporting units we use as locations:  $c_{ij} = \frac{1}{d_{ij}^2}$ , where  $c_{ij}$  is the spatial weight assigned to the distance between location  $i$  and location  $j$ , and  $d_{ij}$  is the Euclidian distance between locations  $i$  and  $j$ . The spatial weights,  $c_{ij}$ , are row standardized such that the diagonal elements of the spatial weighting matrix are all zero and the sum of any one row is one.

The results from the four models are presented in Table 5.<sup>16</sup> These coefficients can be used to determine the relative spatial technical efficiency by conducting a corrected residuals procedure (Richmond 1974, Greene 1980). To implement this method the largest residual from the estimation procedure is added to the constant term of the regression to obtain corrected residuals that provide a benchmark for the level of efficiency present relative to the observation with the largest positive residual. To estimate efficiency, the following transformation of the modified errors is utilized (Russel and Young 1983):  $Eff_i = \exp\{-\varepsilon_i\}$ , where  $\varepsilon_i$  represents the transformed residuals. Following the estimation of the efficiency levels for the different statistical reporting areas, they are ordered to determine the relative rates of

---

<sup>16</sup>We note that the numerical optimizations needed to estimate the spatial heckit make it computationally intensive relative to the other three estimators. In the current application, it takes 134 minutes to compute the spatial heckit estimator in a computer with a 1.69 GHz processor.

efficiency that each of the four models predicts. Table 7 provides a list of the top ten statistical reporting areas in terms of their relative spatial technical efficiency rates. Because there are more observations than statistical reporting areas, we often observe more than one class of vessels fishing within a geographic region. To account for this, the technical efficiency levels for each of the 90 locations are averaged across each class of vessels.

Analyzing the results in Table 5 for the outcome equation, the estimated coefficients from each of the models are somewhat similar in sign and magnitude. The primary exceptions are the "Max. Depth" and "Dum HAL" coefficient in the heckit model and the "Dum Large" coefficient in the spatial heckit model. Within the heckit model, the coefficient on "Max. Depth" is negative and statistically insignificant in contrast to the positive and statistically significant estimates of the other models, whereas the coefficient on "Dum HAL" is negative and statistically significant, contrary to the positive and statistically significant estimates of the other three models. Regarding the latter coefficient, given that the hook-and-line (HAL) technology is one of the most effective technologies within the Pacific cod fishery, it is likely that this estimate in the heckit model is misleading.<sup>17</sup> The coefficient on "Dum Large" within the spatial heckit model is not statistically significant, contrary to the positive and statistically significant estimates from the other three models. This might be due to the spatial heckit model correctly accounting for the nature of the data, although it could also be due to other factors. In addition, the estimated spatial autocorrelation parameter ( $\gamma$ ) is substantially different in the KP-SAE and spatial heckit models. The magnitude of the estimate of  $\gamma$  in the KP-SAE (0.09) is smaller than that within the spatial heckit model (0.79). This may be attributed to the KP-SAE's model inability to correctly capture the spatial autocorrelation in the data due to the sample selection.

Table 6 presents the estimated coefficients for the selection equation using the heckit and the spatial heckit estimators. The estimates obtained from each of these two estimators are fairly similar with the exception of the coefficient on "Dum Large", which is negative and statistically significant in the heckit model, but positive and statistically insignificant in the spatial heckit model. This is consistent with the result for the spatial heckit model in the outcome equation. It is also worth noting that the spatial heckit model yields smaller standard errors for all of the coefficients. Perhaps surprisingly, even though the estimate of

---

<sup>17</sup>The three primary production technologies used to harvest Pacific cod in the Bering Sea in 1997 are HAL, NPT, and pelagic trawl gear (PTR). They account for approximately 91% of the "unique hauls" deployed. Among these three primary production technologies, HAL is more productive per haul executed.

the spatial autoregressive parameter ( $\delta$ ) of the selection equation is positive (0.1827), it is statistically insignificant.

Finally, with regard to the ranking of the statistical reporting areas in terms of their relative spatial technical efficiency rates, reported in Table 7, the results are substantially different across the four models considered. The OLS, heckit, and KP-SAE models yield a similar ranking of spatial efficiency, yet they include only 4 of the 10 regions ranked using the spatial heckit estimator. Given the documented nature of the data set, it is likely that the spatial heckit estimator is more appropriate to use than the other estimators, which would suggest that inaccurate inferences may be drawn regarding spatial efficiencies using the alternative estimation techniques. This said, all four models do possess the same highest ranked location. Therefore, the benchmark utilized for the spatial efficiency calculations is identical.

Although a more complete analysis would be required before any concrete policy recommendations are made from this exercise, they do suggest that there exists heterogeneity in the spatial efficiency rates within the Pacific cod fishery. This may be attributed to a number of different factors such as climatic conditions, skipper ability, and interactions with other fisheries (to name a few) which are all lumped into the corrected residual used to estimate efficiency.

More importantly for the purposes of the present paper, these results are indicative of the potential benefits of accounting simultaneously for both sample selection and spatial dependence. Failing to account simultaneously for both of these features can result in inaccurate inferences and thus potentially misleading policy recommendations.

## 6 Conclusion

This paper proposes a method of estimation for two sample selection models with spatial dependence that differ in terms of the type of spatial correlation present. The first, a model with spatial autoregressive errors (SAE), is slightly more challenging than the second, a model with spatial autoregressive lags (SAL), and thus more attention is devoted to the former model.

The method of estimation for both models is analogous to the popular heckit model (and thus we call our estimator the "spatial heckit"), in which consistent estimates of the probability of observing a particular unit (selection equation) are estimated using a modification



of the probit model (Pinkse and Slade 1998). Then, the odds of observing each unit are calculated (the inverse Mills ratio) and used as an additional regressor that controls for the selection bias in the equation of interest (outcome equation). Importantly, in the method of estimation we propose for the sample selection model with SAE dependence, the appropriate inverse Mills ratio depends on the SAE parameter of the outcome equation. Therefore, to increase efficiency of the resulting estimator and to obtain directly its variance-covariance matrix, we propose to estimate the model jointly by nesting the two equations into a sequential GMM framework (Newey, 1984).

We explore the properties of the spatial heckit for the model with SAE dependence by deriving its asymptotic properties, conducting simulations, and applying it to actual data. The estimator is consistent and asymptotically normally distributed. The simulations show the potential biases incurred by other estimators that ignore sample selection, spatial dependence, or both, and also show that our estimator is valuable when the data exhibits both of these characteristics. Importantly, the finite sample properties of our estimator are shown to be acceptable even for relatively small sample sizes. Finally, the empirical application section illustrates that sample selection is a common occurrence in spatial data sets typically available to researchers, and shows that our estimator is both feasible and valuable to use in practice.

To our knowledge, the proposed estimator is among the first to account for sample selection and spatial dependence simultaneously. Nevertheless, some shortcomings are worth mentioning. First, our estimator relies on a distributional assumption (joint normality) of the error terms in selection and outcome equations, just as the heckit estimator does. This shortcoming indicates an area for future research. Second, is the relatively greater computational intensity of our estimator compared to the available methods for spatial models without sample selection. However, our estimator still compares favorably in this respect with other estimation methods that would require approximation of multidimensional integrals.

## References

- [1] Aigner, D., C.A.K. Lovell and P. Schmidt. 1977. Formulation and Estimation of Stochastic Frontier Production Function Models. *Journal of Econometrics* 6:21-37.
- [2] Amemiya, T. 1974. The Nonlinear Two-Stage Least-Squares Estimator. *Journal of Econometrics*, 2: 105-10.
- [3] Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press.
- [4] Andrews, D.W.K. 1992. Generic Uniform Convergence. *Econometric Theory*, 8: 241-57.
- [5] Anselin, L. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.
- [6] Anselin, L. 2001. Spatial Econometrics. In B. H. Baltagi (ed.), *A Companion to Theoretical Econometrics*, Oxford: Blackwell.
- [7] Anselin, L. and A. Bera. 1998. Spatial Dependence in Linear Regression Models with an Introduction to Spatial Econometrics. In A. Ullah and D.E.A. Giles (eds.), *Handbook of Applied Economic Statistics*. New York: Marcel Dekker.
- [8] Benjamin, J.D., J. Sa-Aadu and J.D. Shilling. 1992. Influence of Rent Differentials on the Choice Between Office Rent Contracts with and without Relocation Provisions. *Journal of American Real Estate and Urban Economics Association*. 20: 289-302.
- [9] Bernstein, S. 1927. Sur l'Extension du Theoreme du Calcul des Probabilites aux Sommes de Quantities Dependantes. *Mathematische Annalen*. 97, 1-59.
- [10] Beron, K. and W. Vijverberg. 1999. Probit in a Spatial Context: A Monte Carlo Analysis. Forthcoming in L. Anselin, R. Florax and S. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*. New York: Springer.
- [11] Bockstael, N.E. and J. Opaluch. 1983. Discrete modeling of supply responses under uncertainty: the case of the fishery. *Journal of Environmental Economics and Management* 10:125-137.
- [12] Borjas, G.J., S.G. Bronars, and S. Trejo. 1992. Self-Selection and Internal Migration in the United States. *Journal of Urban Economics*. 32: 159-85.
- [13] Case, A.C. 1991. Spatial Patterns in Household Demand. *Econometrica*. 59: 953-965.
- [14] Case, A.C. 1992. Neighborhood Influence and Technology Change. *Regional Sciences and Urban Economics*. 22: 491-508.
- [15] Casetti, E. 1972. Generating Models by the Expansion Method: Applications to Geographical Research. *Geographical Analysis*. 4: 81-91.
- [16] Cosslett, S.R. 1991. Semiparametric Estimation of a Regression Model with Sample Selectivity. In W.A. Barnett, J. Powell and G. Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*. Cambridge: Cambridge University Press.
- [17] Davidson, J. 1994. *Stochastic Limit Theory*. Oxford University Press: Oxford.
- [18] Dupont, D.P. 1993. Price uncertainty, expectations formation, and fishers' location choices. *Marine Resource Economics* 8:219-247.

- [19] Eales, J. and J.E. Wilen. 1986. An examination of fishing location choice in the pink shrimp fishery. *Marine Resource Economics* 2:331-351.
- [20] Fishback, P.V., W.C. Horrace and S. Kantor. 2002. Do Federal Programs Affect Internal Migration? The Impact of New Deal Expenditures on Mobility During the Great Depression. Working Paper Series, NBER, Number 8283.
- [21] Fleming, Mark. 2002. Techniques for Estimating Spatially Dependent Discrete Choice Models. Forthcoming in L. Anselin, R. Florax and S. Rey (eds.), *Advances in Spatial Econometrics: Methodology, Tools and Applications*. New York: Springer.
- [22] Garcia del Hoyo, J.J., D.C. Espino and R.J. Toribio. 2004. Determination of technical efficiency of fisheries by stochastic frontier models: a case on the Gulf of Cadiz (Spain). *ICES Journal of Marine Science* 61:417-422.
- [23] Goodman, A.C. 1988. An Econometric Model of Housing Price, Permanent Income, Tenure Choice, and Housing Demand. *Journal of Urban Economics*. 23: 327-53.
- [24] Greene, W.H. 1980. Maximum Likelihood Estimates of Econometric Production Frontier Functions. *Journal of Econometrics* 13:27-56.
- [25] Heckman, J. 1976. The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for such Models. *The Annals of Economic and Social Measurement*. 5: 475-492.
- [26] Heckman, J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*. 47: 153-61.
- [27] Holland, D. and J. Sutinen. 1999. An empirical model of fleet dynamics. *Can. J. Fish. Aquat. Sci.* 56:253-264.
- [28] Holland, D. and J. Sutinen. 2000. Location choice in New England trawl fisheries: old habits die hard. *Land Economics* 76:133-149.
- [29] Kelejian, H.H. and I.R. Prucha. 1998. A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances. *Journal of Real Estate Finance and Economics*, 17: 99-121.
- [30] Kelejian, H.H. and I.R. Prucha. 1999. A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model. *International Economic Review*, 40: 509-33.
- [31] Kelejian, H.H. and I.R. Prucha. 2001. On the Asymptotic Distribution of the Moran  $I$  Test Statistic with Applications. *Journal of Econometrics*. 104: 219-57.
- [32] Kirkley, J., D. Squires and I.E. Strand. 1995. Assessing Technical Efficiency in Commercial Fisheries: The Mid-Atlantic Sea Scallop Fishery. *American Journal of Agricultural Economics* 77:689-697.
- [33] Kirkley, J., D. Squires and I.E. Strand. 1998. Characterizing Managerial Skill and Technical Efficiency. *Journal of Productivity Analysis* 9:145-160.
- [34] Lee, L.F. 2001a. Generalized Method of Moments Estimation of Spatial Autoregressive Processes. Working paper, Ohio State University.
- [35] Lee, L.F. 2001b. GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models. Working paper, Ohio State University.

- [36] LeSage, J.P. 2000. Bayesian Estimation of Limited Dependent Variable Spatial Autoregressive Models. *Geographical Analysis*. 32: 19-35.
- [37] Leung, S.F and S. Yu. 1996. On the Choice Between Sample Selection and Two-Part Models. *Journal of Econometrics*. 72: 197-229.
- [38] Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.
- [39] McLeish, D.L. 1974. Dependent Central-Limit Theorems and Invariance Principles. *Annals of Probability*. 2, 620-628.
- [40] McMillen, D.P. 1992. Probit with Spatial Autocorrelation. *Journal of Regional Science*. 32: 335-48.
- [41] McMillen, D.P. 1993. Can Blacks Earn More in the Suburbs? Racial Differences in Intrametropolitan Earnings Variation. *Journal of Urban Economics*. 33: 167-88.
- [42] McMillen, D.P. 1995. Selection Bias in Spatial Econometric Models. *Journal of Regional Science*. 35: 417-36.
- [43] Mistean, J. and I. Strand. 2000. Location choice of commercial fishermen with heterogeneous risk preferences. *American Journal of Agricultural Economics* 82:1184-1190.
- [44] Montgomery, C. 1992. Explaining Home Improvement in the Context of Household Investment in Residential Housing. *Journal of Urban Economics*. 32: 326-50.
- [45] Newey, W. 1984. A Method of Moments Interpretation of Sequential Estimators. *Economics Letters*. 14: 201-206.
- [46] Pascoe, S. and L. Coglan. 2002. The Contribution of Unmeasurable Inputs to Fisheries Production: An Analysis of Technical Efficiency of Fishing Vessels in the English Channel. *American Journal of Agricultural Economics* 84:585-597.
- [47] Pinkse, J. and M.E. Slade. 1998. Contracting in Space. An Application of Spatial Statistics to Discrete-Choice Models. *Journal of Econometrics*. 85: 125-54.
- [48] Richmond, J. 1974. Estimating the Efficiency of Production. *International Economic Review* 15:515-521.
- [49] Russel, M. and T. Young. 1983. Frontier Production Functions and the Measurement of Technical Efficiency. *Journal of Agricultural Economics* 34(2):139-150.
- [50] Sharma, K.R. and P. Leung. 1998. Technical Efficiency of Hawaii's Longline Fishery. *Marine Resource Economics* 13:259-274.
- [51] Smith, M.D. 2000. Spatial search and fishing location choice: methodological challenges and empirical modeling. *American Journal of Agricultural Economics* 82: 1198-1206.
- [52] Squires, D. and J. Kirkley. 1999. Skipper skill and panel data in fishing industries. *Can. J. Fish. Aquat. Sci.* 56:2011-2018.
- [53] Topa, G. 2001. Social Interactions, Local Spillovers, and Unemployment. *Review of Economic Studies*. 2: 133-138.
- [54] Viswanathan, K.K., I.H. Omar, Y. Jeon, J. Kirkley, D. Squires and I. Susilowati 2002. Fishing Skill in Developing Country Fisheries: The Kedah, Malaysia Trawl Fishery. *Marine Resource Economics* 16:293-314.

- [55] Wilen, J. 2004. Spatial Management of Fisheries. *Marine Resource Economics* 19:7-20.
- [56] Zellner, A., J. Kmenta and J. Dreze 1966. Specification and Estimation of Cobb-Douglas Production Functions. *Econometrica* 34:784-795.

Table 1. Simulation Results for N=100

| (sel, $\delta=\gamma$ ) |           | BIAS   |        |        |            |           | RMSE  |        |        |            |           |
|-------------------------|-----------|--------|--------|--------|------------|-----------|-------|--------|--------|------------|-----------|
|                         |           | OLS    | Heckit | KP-SAE | Spheck Gen | Spheck KP | OLS   | Heckit | KP-SAE | Spheck Gen | Spheck KP |
| (25%,0.25)              | $\beta_0$ | 0.280  | 0.038  | 0.290  | 0.134      | -0.900    | 0.447 | 1.204  | 0.464  | 2.682      | 4.674     |
|                         | $\beta_1$ | 0.024  | 0.026  | 0.028  | 0.028      | 0.005     | 0.418 | 0.417  | 0.426  | 0.436      | 0.457     |
|                         | $\beta_2$ | -0.198 | -0.039 | -0.212 | -0.159     | 0.489     | 0.469 | 0.848  | 0.488  | 1.807      | 3.221     |
|                         | $\gamma$  |        |        | 0.054  | 0.115      | 0.101     |       |        | 0.075  | 0.180      | 0.178     |
| (25%,0.5)               | $\beta_0$ | 0.295  | -0.015 | 0.311  | 0.239      | -0.988    | 0.475 | 1.959  | 0.566  | 3.035      | 3.978     |
|                         | $\beta_1$ | 0.023  | 0.027  | 0.034  | 0.033      | 0.010     | 0.421 | 0.421  | 0.520  | 0.446      | 0.468     |
|                         | $\beta_2$ | -0.204 | -0.073 | -0.231 | -0.107     | 0.527     | 0.485 | 1.044  | 0.621  | 1.754      | 2.694     |
|                         | $\gamma$  |        |        | -0.188 | -0.007     | -0.020    |       |        | 0.196  | 0.135      | 0.140     |
| (25%,0.75)              | $\beta_0$ | 0.370  | 0.220  | 0.415  | 0.217      | -1.254    | 0.641 | 2.267  | 1.175  | 3.250      | 5.237     |
|                         | $\beta_1$ | -0.014 | -0.009 | 0.013  | 0.004      | -0.013    | 0.467 | 0.469  | 1.231  | 0.501      | 0.535     |
|                         | $\beta_2$ | -0.225 | -0.077 | -0.308 | -0.138     | 0.526     | 0.577 | 1.298  | 1.389  | 1.870      | 2.757     |
|                         | $\gamma$  |        |        | -0.399 | -0.105     | -0.128    |       |        | 0.407  | 0.163      | 0.180     |
| (40%,0.25)              | $\beta_0$ | 0.423  | -0.107 | 0.435  | 0.223      | -1.016    | 0.567 | 2.397  | 0.582  | 2.304      | 6.295     |
|                         | $\beta_1$ | 0.007  | 0.017  | 0.000  | -0.088     | -0.150    | 0.448 | 0.453  | 0.471  | 0.456      | 0.666     |
|                         | $\beta_2$ | -0.239 | 0.127  | -0.244 | -0.117     | 0.498     | 0.525 | 1.316  | 0.530  | 1.308      | 3.139     |
|                         | $\gamma$  |        |        | 0.084  | 0.255      | -0.084    |       |        | 0.107  | 0.280      | 0.149     |
| (40%,0.5)               | $\beta_0$ | 0.448  | -0.193 | 0.447  | 0.232      | -0.738    | 0.609 | 2.107  | 0.689  | 2.382      | 4.223     |
|                         | $\beta_1$ | -0.007 | 0.004  | -0.004 | -0.077     | -0.116    | 0.449 | 0.453  | 0.604  | 0.472      | 0.530     |
|                         | $\beta_2$ | -0.245 | 0.066  | -0.219 | -0.071     | 0.288     | 0.567 | 1.193  | 0.660  | 1.063      | 2.139     |
|                         | $\gamma$  |        |        | -0.138 | 0.086      | 0.091     |       |        | 0.152  | 0.136      | 0.135     |
| (40%,0.75)              | $\beta_0$ | 0.516  | -1.321 | 0.546  | -0.033     | -0.961    | 0.768 | 16.853 | 1.221  | 3.243      | 6.246     |
|                         | $\beta_1$ | -0.034 | -0.028 | -0.001 | -0.089     | -0.158    | 0.504 | 0.506  | 1.225  | 0.544      | 0.670     |
|                         | $\beta_2$ | -0.259 | 0.133  | -0.266 | 0.005      | 0.492     | 0.637 | 1.554  | 1.306  | 1.790      | 3.136     |
|                         | $\gamma$  |        |        | -0.360 | -0.077     | -0.089    |       |        | 0.373  | 0.141      | 0.152     |

Note: Simulation results are based on 250 replications.

Table 2. Simulation Results for N=225

| (sel, $\delta=\gamma$ ) |           | BIAS   |        |        |            |           | RMSE  |        |        |            |           |
|-------------------------|-----------|--------|--------|--------|------------|-----------|-------|--------|--------|------------|-----------|
|                         |           | OLS    | Heckit | KP-SAE | Spheck Gen | Spheck KP | OLS   | Heckit | KP-SAE | Spheck Gen | Spheck KP |
| (25%,0.25)              | $\beta_0$ | 0.296  | -0.056 | 0.297  | -0.010     | -0.187    | 0.373 | 0.718  | 0.378  | 0.769      | 1.350     |
|                         | $\beta_1$ | 0.005  | 0.007  | -0.008 | -0.042     | -0.069    | 0.279 | 0.280  | 0.283  | 0.274      | 0.290     |
|                         | $\beta_2$ | -0.206 | 0.037  | -0.203 | 0.010      | 0.117     | 0.328 | 0.562  | 0.328  | 0.587      | 1.076     |
|                         | $\gamma$  |        |        | -0.065 | 0.141      | 0.125     |       |        | 0.068  | 0.168      | 0.157     |
| (25%,0.5)               | $\beta_0$ | 0.304  | -0.058 | 0.312  | 0.003      | -0.323    | 0.396 | 0.828  | 0.432  | 0.899      | 1.888     |
|                         | $\beta_1$ | 0.008  | 0.010  | 0.000  | -0.019     | -0.056    | 0.284 | 0.285  | 0.357  | 0.291      | 0.295     |
|                         | $\beta_2$ | -0.209 | 0.038  | -0.217 | 0.000      | 0.054     | 0.331 | 0.624  | 0.387  | 0.684      | 1.012     |
|                         | $\gamma$  |        |        | -0.313 | 0.023      | 0.009     |       |        | 0.313  | 0.084      | 0.085     |
| (25%,0.75)              | $\beta_0$ | 0.357  | -0.156 | 0.362  | -0.065     | -0.398    | 0.504 | 1.419  | 0.841  | 1.376      | 2.406     |
|                         | $\beta_1$ | 0.006  | 0.007  | 0.063  | -0.029     | -0.065    | 0.332 | 0.332  | 0.960  | 0.337      | 0.345     |
|                         | $\beta_2$ | -0.226 | 0.097  | -0.282 | 0.023      | 0.057     | 0.371 | 0.891  | 0.954  | 0.920      | 0.983     |
|                         | $\gamma$  |        |        | -0.547 | -0.060     | -0.075    |       |        | 0.548  | 0.093      | 0.105     |
| (40%,0.25)              | $\beta_0$ | 0.435  | -0.032 | 0.430  | 0.065      | 0.205     | 0.511 | 1.191  | 0.508  | 1.441      | 1.095     |
|                         | $\beta_1$ | 0.004  | 0.011  | 0.003  | -0.144     | -0.182    | 0.321 | 0.324  | 0.320  | 0.338      | 0.351     |
|                         | $\beta_2$ | -0.260 | 0.015  | -0.251 | -0.058     | -0.104    | 0.383 | 0.749  | 0.390  | 0.753      | 0.606     |
|                         | $\gamma$  |        |        | -0.044 | 0.279      | 0.264     |       |        | 0.051  | 0.288      | 0.273     |
| (40%,0.5)               | $\beta_0$ | 0.458  | -0.071 | 0.443  | 0.063      | -0.028    | 0.546 | 1.154  | 0.553  | 1.244      | 1.107     |
|                         | $\beta_1$ | 0.004  | 0.017  | 0.007  | -0.105     | -0.159    | 0.333 | 0.329  | 0.399  | 0.339      | 0.365     |
|                         | $\beta_2$ | -0.268 | 0.040  | -0.255 | -0.044     | 0.012     | 0.409 | 0.779  | 0.450  | 0.840      | 0.751     |
|                         | $\gamma$  |        |        | -0.288 | 0.108      | 0.097     |       |        | 0.289  | 0.125      | 0.115     |
| (40%,0.75)              | $\beta_0$ | 0.503  | 0.003  | 0.459  | -0.006     | -0.316    | 0.620 | 2.273  | 0.812  | 1.379      | 3.055     |
|                         | $\beta_1$ | 0.012  | 0.027  | 0.114  | -0.050     | -0.138    | 0.338 | 0.367  | 0.860  | 0.363      | 0.391     |
|                         | $\beta_2$ | -0.246 | 0.058  | -0.254 | -0.033     | 0.100     | 0.423 | 0.952  | 0.793  | 0.794      | 1.305     |
|                         | $\gamma$  |        |        | -0.525 | -0.035     | -0.046    |       |        | 0.527  | 0.071      | 0.077     |

Note: Simulation results are based on 150 replications.

Table 3. Simulation Results for N=400

| (sel, $\delta=\gamma$ ) |           | BIAS   |        |        |            |           | RMSE  |        |        |            |           |
|-------------------------|-----------|--------|--------|--------|------------|-----------|-------|--------|--------|------------|-----------|
|                         |           | OLS    | Heckit | KP-SAE | Spheck Gen | Spheck KP | OLS   | Heckit | KP-SAE | Spheck Gen | Spheck KP |
| (25%,0.25)              | $\beta_0$ | 0.327  | -0.065 | 0.327  | -0.057     | -0.220    | 0.358 | 0.441  | 0.359  | 0.439      | 0.492     |
|                         | $\beta_1$ | -0.004 | -0.007 | 0.023  | -0.048     | -0.033    | 0.189 | 0.187  | 0.207  | 0.196      | 0.167     |
|                         | $\beta_2$ | -0.250 | 0.022  | -0.269 | 0.015      | 0.104     | 0.309 | 0.320  | 0.323  | 0.320      | 0.346     |
|                         | $\gamma$  |        |        | -0.105 | 0.129      | 0.129     |       |        | 0.106  | 0.143      | 0.143     |
| (25%,0.5)               | $\beta_0$ | 0.333  | -0.068 | 0.356  | -0.062     | -0.120    | 0.369 | 0.458  | 0.398  | 0.466      | 0.486     |
|                         | $\beta_1$ | -0.006 | -0.009 | 0.003  | -0.040     | -0.040    | 0.199 | 0.198  | 0.283  | 0.211      | 0.210     |
|                         | $\beta_2$ | -0.242 | 0.030  | -0.290 | 0.025      | 0.054     | 0.300 | 0.309  | 0.353  | 0.313      | 0.320     |
|                         | $\gamma$  |        |        | -0.356 | 0.022      | 0.019     |       |        | 0.356  | 0.054      | 0.053     |
| (25%,0.75)              | $\beta_0$ | 0.365  | -0.111 | 0.490  | -0.119     | -0.162    | 0.430 | 0.702  | 0.623  | 0.733      | 0.728     |
|                         | $\beta_1$ | 0.029  | -0.017 | -0.011 | -0.046     | -0.056    | 0.227 | 0.242  | 0.713  | 0.253      | 0.250     |
|                         | $\beta_2$ | -0.268 | 0.042  | -0.429 | 0.032      | 0.035     | 0.333 | 0.377  | 0.594  | 0.374      | 0.354     |
|                         | $\gamma$  |        |        | -0.598 | -0.048     | -0.051    |       |        | 0.599  | 0.059      | 0.063     |
| (40%,0.25)              | $\beta_0$ | 0.459  | -0.086 | 0.453  | 0.075      | 0.013     | 0.484 | 0.542  | 0.476  | 0.520      | 0.548     |
|                         | $\beta_1$ | -0.004 | -0.001 | 0.018  | -0.156     | -0.153    | 0.213 | 0.212  | 0.220  | 0.264      | 0.260     |
|                         | $\beta_2$ | -0.287 | 0.037  | -0.301 | -0.037     | 0.003     | 0.344 | 0.352  | 0.360  | 0.343      | 0.355     |
|                         | $\gamma$  |        |        | -0.091 | 0.278      | 0.276     |       |        | 0.091  | 0.281      | 0.279     |
| (40%,0.5)               | $\beta_0$ | 0.479  | -0.103 | 0.495  | 0.079      | 0.071     | 0.511 | 0.612  | 0.531  | 0.592      | 0.673     |
|                         | $\beta_1$ | 0.003  | 0.006  | -0.017 | -0.125     | -0.131    | 0.244 | 0.242  | 0.309  | 0.265      | 0.278     |
|                         | $\beta_2$ | -0.302 | 0.040  | -0.310 | -0.042     | -0.020    | 0.363 | 0.377  | 0.394  | 0.377      | 0.423     |
|                         | $\gamma$  |        |        | -0.342 | 0.113      | 0.110     |       |        | 0.342  | 0.121      | 0.119     |
| (40%,0.75)              | $\beta_0$ | 0.541  | -0.121 | 0.636  | -0.050     | -0.061    | 0.599 | 0.777  | 0.728  | 0.766      | 0.816     |
|                         | $\beta_1$ | 0.001  | -0.001 | -0.005 | -0.089     | -0.130    | 0.296 | 0.315  | 0.628  | 0.317      | 0.350     |
|                         | $\beta_2$ | -0.301 | 0.033  | -0.455 | 0.002      | 0.010     | 0.375 | 0.399  | 0.690  | 0.384      | 0.387     |
|                         | $\gamma$  |        |        | -0.584 | -0.025     | -0.031    |       |        | 0.584  | 0.045      | 0.052     |

Note: Simulation results are based on 50 replications.



Table 4. Simulation Results for the Selection Equation

|                    |            | N=100  |        |        |        | N=225  |        |        |        | N=400  |        |        |        |
|--------------------|------------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|--------|
|                    |            | BIAS   |        | RMSE   |        | BIAS   |        | RMSE   |        | BIAS   |        | RMSE   |        |
| (sel, $\delta=y$ ) |            | Heckit | Spheck | Heckit | Spheck | Heckit | Spheck | Heckit | Spheck | Heckit | Spheck | Heckit | Spheck |
| (25%,0.25)         | $\alpha_0$ | 0.015  | 0.010  | 0.431  | 0.482  | 0.018  | -0.046 | 0.292  | 0.319  | 0.007  | -0.031 | 0.180  | 0.189  |
|                    | $\alpha_1$ | 0.059  | 0.082  | 0.538  | 0.596  | 0.017  | 0.035  | 0.363  | 0.394  | 0.039  | 0.053  | 0.218  | 0.225  |
|                    | $\alpha_2$ | -0.060 | -0.022 | 0.562  | 0.619  | 0.006  | 0.059  | 0.377  | 0.407  | -0.053 | -0.019 | 0.214  | 0.218  |
|                    | $\delta$   |        | -0.204 |        | 0.313  |        | -0.169 |        | 0.240  |        | -0.143 |        | 0.193  |
| (25%,0.5)          | $\alpha_0$ | 0.011  | 0.021  | 0.459  | 0.524  | 0.022  | -0.029 | 0.312  | 0.334  | 0.012  | -0.012 | 0.198  | 0.208  |
|                    | $\alpha_1$ | 0.011  | 0.041  | 0.533  | 0.598  | -0.020 | 0.005  | 0.366  | 0.388  | 0.000  | 0.018  | 0.219  | 0.235  |
|                    | $\alpha_2$ | -0.070 | -0.043 | 0.565  | 0.622  | -0.030 | 0.017  | 0.367  | 0.384  | -0.081 | -0.051 | 0.211  | 0.211  |
|                    | $\delta$   |        | -0.281 |        | 0.356  |        | -0.237 |        | 0.288  |        | -0.196 |        | 0.223  |
| (25%,0.75)         | $\alpha_0$ | 0.020  | -0.032 | 0.539  | 0.601  | 0.046  | -0.021 | 0.359  | 0.387  | 0.031  | 0.002  | 0.264  | 0.288  |
|                    | $\alpha_1$ | -0.112 | 0.008  | 0.534  | 0.620  | -0.124 | -0.066 | 0.383  | 0.410  | -0.120 | -0.070 | 0.271  | 0.277  |
|                    | $\alpha_2$ | -0.170 | -0.065 | 0.560  | 0.596  | -0.156 | -0.074 | 0.389  | 0.380  | -0.197 | -0.141 | 0.285  | 0.276  |
|                    | $\delta$   |        | -0.280 |        | 0.337  |        | -0.235 |        | 0.270  |        | -0.199 |        | 0.223  |
| (40%,0.25)         | $\alpha_0$ | -0.045 | -0.168 | 0.545  | 0.512  | -0.018 | -0.115 | 0.309  | 0.351  | -0.014 | -0.105 | 0.171  | 0.203  |
|                    | $\alpha_1$ | 0.069  | 0.138  | 0.538  | 0.575  | 0.011  | 0.046  | 0.351  | 0.388  | 0.020  | 0.029  | 0.221  | 0.240  |
|                    | $\alpha_2$ | 0.005  | 0.124  | 0.572  | 0.577  | 0.003  | 0.083  | 0.371  | 0.403  | 0.007  | 0.075  | 0.214  | 0.232  |
|                    | $\delta$   |        | -0.188 |        | 0.298  |        | -0.143 |        | 0.214  |        | -0.111 |        | 0.159  |
| (40%,0.5)          | $\alpha_0$ | 0.007  | -0.105 | 0.458  | 0.522  | -0.009 | -0.080 | 0.320  | 0.354  | 0.031  | -0.052 | 0.179  | 0.188  |
|                    | $\alpha_1$ | 0.010  | 0.095  | 0.517  | 0.585  | -0.013 | 0.022  | 0.357  | 0.390  | -0.024 | 0.006  | 0.232  | 0.255  |
|                    | $\alpha_2$ | -0.067 | 0.065  | 0.526  | 0.565  | -0.018 | 0.054  | 0.374  | 0.405  | -0.068 | 0.004  | 0.209  | 0.205  |
|                    | $\delta$   |        | -0.238 |        | 0.320  |        | -0.204 |        | 0.245  |        | -0.159 |        | 0.194  |
| (40%,0.75)         | $\alpha_0$ | 0.109  | -0.009 | 0.545  | 0.620  | 0.079  | -0.011 | 0.371  | 0.383  | 0.112  | 0.031  | 0.257  | 0.260  |
|                    | $\alpha_1$ | -0.126 | 0.022  | 0.538  | 0.733  | -0.136 | -0.069 | 0.388  | 0.402  | -0.140 | -0.077 | 0.247  | 0.239  |
|                    | $\alpha_2$ | -0.208 | -0.030 | 0.572  | 0.601  | -0.141 | -0.026 | 0.379  | 0.374  | -0.187 | -0.106 | 0.274  | 0.252  |
|                    | $\delta$   |        | -0.248 |        | 0.307  |        | -0.218 |        | 0.250  |        | -0.166 |        | 0.184  |

Notes: Simulation results are based on 250 replications for N=100, 150 for N=225, and 50 for N=400. The spatial heckit estimator (Spheck) is computed using generated instruments. See text for details.

**Table 5. Estimated Coefficients for the Outcome Equation**

| Variable                   | OLS                | Heckit             | KP-SAE              | Spatial Heckit <sup>1</sup> |
|----------------------------|--------------------|--------------------|---------------------|-----------------------------|
| Constant                   | 7.5491<br>(0.196)  | 8.8629<br>(0.514)  | 7.559<br>(0.129)    | 6.4013<br>(0.047)           |
| Max. Depth                 | 0.0788<br>(0.004)  | -0.0017<br>(0.074) | 0.0801<br>(0.003)   | 0.1324<br>(0.009)           |
| Min. Depth                 | 0.0101<br>(0.003)  | 0.0328<br>(0.049)  | 0.0028<br>(0.002)   | 0.0149<br>(0.037)           |
| Biomass                    | 0.1959<br>(0.003)  | 0.1779<br>(0.057)  | 0.1977<br>(0.002)   | 0.0495<br>(0.003)           |
| Dum CV                     | 1.309<br>(0.044)   | 1.549<br>(0.216)   | 1.331<br>(0.030)    | 1.442<br>(0.102)            |
| Dum HAL                    | 0.1081<br>(0.041)  | -0.4813<br>(0.236) | 0.1163<br>(0.027)   | 0.1585<br>(0.101)           |
| Dum NPT                    | -0.5448<br>(0.054) | -0.8001<br>(0.270) | -0.5326<br>(0.037)  | -0.5297<br>(0.082)          |
| Dum Large                  | 0.5898<br>(0.016)  | 0.5422<br>(0.157)  | 0.5890<br>(0.011)   | -0.2168<br>(0.179)          |
| IMR                        | -----              | -0.9636<br>(0.247) | -----               | 1.6881<br>(0.137)           |
| SAE parameter ( $\gamma$ ) | -----              | -----              | 0.0991 <sup>2</sup> | 0.7909<br>(0.001)           |
| Moran-I Test (p-value)     | 13.26 (0.00)       | 10.51 (0.00)       | -----               | -----                       |

Notes: Dependent variable is catch-per-unit-effort (CPUE). Standard errors in parentheses.

<sup>1</sup> The spatial heckit is estimated using the Kelejian and Prucha (KP) instruments.

<sup>2</sup> The KP-SAE method does not provide a standard error for the SAE parameter since it is treated as a nuisance parameter. For details, see Kelejian and Prucha (1998, 1999).

**Table 6. Estimated Coefficients for the Selection Equation**

| Variable                   | Heckit             | Spatial Heckit      |
|----------------------------|--------------------|---------------------|
| Constant                   | 0.9015<br>(0.236)  | 0.4699<br>(0.091)   |
| Max. Depth                 | 0.0020<br>(0.002)  | 0.0024<br>(0.0001)  |
| Min. Depth                 | -0.0024<br>(0.002) | -0.0029<br>(0.002)  |
| Biomass                    | -0.0025<br>(0.003) | -0.0042<br>(0.002)  |
| Lag Biomass                | -0.0032<br>(0.002) | -0.0009<br>(0.0001) |
| Dum CV                     | -0.8356<br>(0.179) | -0.3121<br>(0.167)  |
| Dum HAL                    | 0.5765<br>(0.195)  | 0.2701<br>(0.114)   |
| Dum NPT                    | 0.0442<br>(0.259)  | -0.1126<br>(0.238)  |
| Dum Large                  | -0.4313<br>(0.163) | 0.2288<br>(0.143)   |
| SAE parameter ( $\delta$ ) | -----              | 0.1827<br>(0.503)   |

Notes: Dependent variable is whether or not CPUE is observed for that observation. Standard errors in parentheses.

**Table 7. Top Ten Statistical Reporting Areas in Terms of Their Relative Spatial Technical Efficiency**

| Rank |  | OLS       | KP-SAE    | Heckit    | Spatial Heckit |
|------|--|-----------|-----------|-----------|----------------|
| 1    |  | <b>58</b> | <b>58</b> | <b>58</b> | <b>58</b>      |
| 2    |  | <b>19</b> | <b>19</b> | 62        | <b>54</b>      |
| 3    |  | <b>9</b>  | <b>9</b>  | <b>28</b> | <b>51</b>      |
| 4    |  | <b>28</b> | <b>28</b> | <b>9</b>  | <b>48</b>      |
| 5    |  | 79        | 62        | <b>19</b> | <b>46</b>      |
| 6    |  | 62        | 79        | 24        | <b>19</b>      |
| 7    |  | 71        | 71        | 71        | <b>57</b>      |
| 8    |  | 24        | 24        | 6         | <b>53</b>      |
| 9    |  | 3         | 3         | 3         | <b>28</b>      |
| 10   |  | 6         | 6         | 59        | <b>9</b>       |

Notes: Relative spatial technical efficiency is estimated using the coefficient from Table 5 and a corrected residuals procedure. See text for details. The entries are the individual identifiers of the 90 statistical reporting areas.

## 7 Appendix

The parameter estimate of the sample selection model with SAE dependence is obtained from the solution to (15), where  $g_N(\theta) = \frac{1}{N} z_N' \tilde{u}_N(\theta)$ . Denote  $g(\theta) \equiv \lim_{n \rightarrow \infty} E[g_n(\theta)]$ . Then, the unknown parameter vector  $\theta_0$  satisfies  $\lim_{n \rightarrow \infty} E[g_N(\theta_0)] = 0$ . Further, define the objective function as  $Q_N = g_N'(\theta) M_N g_N(\theta)$ , where  $M_N \xrightarrow{p} M$ , and  $Q = g'(\theta) M g(\theta)$ .

The proof of all three propositions follow closely Pinkse and Slade (1998). The main difference is that some extra conditions have to be verified in for the additional moments stacked in  $g_N(\theta)$  and the estimated inverse-Mills ratio (IMR).

### Assumptions

**A1**  $\theta_0$  is in the interior of the parameter space  $\Theta$ , which is a compact set.

**A2**  $Q$  is uniquely minimized at  $\theta_0$

**A3** The vector valued function  $g(\theta)$  is continuous.

**A4** The density of observations in any region whose area exceeds a fixed minimum is bounded.

**A5** The elements of  $z_N$  are uniformly bounded.<sup>18</sup>

**A6** Let  $d_{ij}$  denote the distance between location  $i$  and  $j$ , then

$$\begin{aligned} \sup_{Nij} |cov(y_{1i}, y_{1j})| &\leq \alpha(d_{ij}), \\ \sup_{Nij} |cov(y_{2i}, y_{2j})| &\leq \alpha(d_{ij}) \text{ and} \\ \alpha(d) &\rightarrow 0 \text{ as } d \rightarrow \infty. \end{aligned}$$

**A7** The moments  $var(u_{ij}) = \sigma_1^2 \sum_j [\omega_{ij}^1]^2$ ,  $var(u_{2i}) = \sigma_2^2 \sum_j [\omega_{ij}^2]^2$ , and  $E[u_{1i}, u_{2i}] = \sigma_{12} \sum_j \omega_{ij}^1 \omega_{ij}^2$  are uniformly bounded, bounded away from zero, and boundedly differentiable (with respect to  $\theta$ ).

**A8** As  $N \rightarrow \infty$ ,  $M_N \xrightarrow{p} M$  for some positive definite matrix  $M$ .

**A9** As  $d \rightarrow \infty$ ,  $d^2 \alpha(d) / \alpha(d^*) \rightarrow 0$ , for all fixed  $d^* > 0$ .

**A10** The area in which the observations are located grows at a rate of  $\sqrt{N}$  in both directions.

**A11**  $\Psi_1(\theta_0) = \lim_{N \rightarrow \infty} E\{N g_N(\theta_0) g_N'(\theta_0)\}$  and  $\Psi_2(\theta_0) = [\partial g'(\theta_0) / \partial \theta] M [\partial g(\theta_0) / \partial \theta']$  are positive definite matrices.

**A12** Let  $\rho_{Nij}(\theta)$  be the covariance between  $\frac{u_{Ni}}{\sqrt{var(u_{Ni})}}$  and  $\frac{u_{Nj}}{\sqrt{var(u_{Nj})}}$ . For some fixed  $N^* > 0$ ,  $\rho_{Nij}(\theta)$  is boundedly differentiable, uniformly in  $\theta \in \Theta$ ,  $N > L$ , and  $i \neq j$ .

---

<sup>18</sup> $\lambda(\delta, \gamma)$  is an element of  $z_N$  which will be shown below to be uniformly bounded, as well as its first derivative.

**A13**  $|\rho_{Nij}(\theta)|$  is boundedly away from one from below, uniformly in  $\theta \in H(\theta_0)$ ,  $N > N^*$ , and  $i \neq j$ , with  $H(\theta_0)$  some neighborhood of  $\theta_0$ .

**A14** As  $N \rightarrow \infty$ ,  $\frac{1}{N} \sum_{ij} |\rho_{Nij}(\theta)|$  is uniformly bounded in  $\theta \in H(\theta_0)$ .

A discussion of the implications of these assumptions can be found in Pinkse and Slade (1998).

### Proof of Proposition 1

By assumption A2,  $Q$  is uniquely minimized at  $\theta_0$ . Thus, we only need to establish that  $Q_N$  converges uniformly to  $Q$  over the parameter space  $\Theta$ .

To show that  $Q_N$  converges uniformly to  $Q$  over  $\Theta$ , it suffices to show that

(a)  $Q_N \xrightarrow{p} Q$  at all  $\theta \in \Theta \iff g_N(\theta) \xrightarrow{p} g(\theta)$  at all  $\theta \in \Theta$ .

(b)  $Q_N$  is stochastically equicontinuous and  $Q$  is continuous on  $\Theta \iff g_N(\theta) \xrightarrow{p} g(\theta)$  is stochastically equicontinuous.

For (a), note that  $g(\theta) \equiv \lim_{N \rightarrow \infty} E[g_N(\theta)]$ , that is,  $E[g_N(\theta)] \xrightarrow{p} g(\theta)$ . If  $g_N(\theta) \xrightarrow{p} E[g_N(\theta)]$ , then  $g_N(\theta) \xrightarrow{p} g(\theta)$ , so we show the former.

Define the functions

$$\tau_{1i}(\psi_i(\theta)) \equiv \phi(\psi_i(\theta)) / \{\Phi(\psi_i(\theta)) [1 - \Phi(\psi_i(\theta))]\}$$

such that  $\tilde{u}_{1i}(\theta) = \tau_{1i}(\psi_i(\theta)) (y_i - \Phi(\psi_i(\theta)))$ .

Let  $\tau_i(\psi_i(\theta)) \equiv (\tau_{1i}(\psi_i(\theta)))'$ ,  $\mathbf{1}$  is a conformable vector of ones such that  $\tilde{u}_N(\theta) = ((\tau_1(\psi(\theta)) (y - \Phi(\psi(\theta))))', \mathbf{1}u'_{2N}(\theta))'$ .

Then,

$$\begin{aligned} & \lim_{n \rightarrow \infty} E |g_N(\theta) - E[g_N(\theta)]|^2 \\ &= \lim_{n \rightarrow \infty} \frac{1}{N^2} \sum_{ij} z'_i \tau_{Ni} \tau_{Nj} z_j \text{cov}[y_i, y_j] \\ &\leq \lim_{N \rightarrow \infty} \frac{1}{N^2} C \sum_{ij} \alpha(d_{ij}) \quad \text{by assumptions A5-A7} \\ &= 0 \end{aligned}$$

since  $\alpha(d_{ij}) \rightarrow 0$  as  $d \rightarrow \infty$  by assumption A6. Therefore, since  $g_N(\theta) \xrightarrow{p} E[g_N(\theta)]$  then  $g_N(\theta) \xrightarrow{p} g(\theta)$ .

For (b), given that by assumption A3  $g(\theta)$  is continuous, we need to show that  $g_N(\theta)$  is stochastically equicontinuous.

Using the mean value theorem for  $\theta^*$  between  $\theta$  and  $\tilde{\theta}$  we rewrite

$$g_N(\theta) - g_N(\tilde{\theta}) = \frac{1}{N} z'_N \{ \tilde{u}_N - \tilde{u}_N(\tilde{\theta}) \} = \frac{1}{N} z'_i \frac{\partial \tilde{u}_{Ni}(\theta^*)}{\partial \theta'} (\theta - \tilde{\theta}).$$

Following Andrews(1992), stochastic equicontinuity is implied by

$$\sup_{\theta \in \Theta} \left| \frac{1}{N} \sum_i z_i' \frac{\partial \tilde{u}_{Ni}(\theta)}{\partial \theta'} \right| = Op(1).$$

Recall that  $\tau_i(\psi(\theta)) \equiv (\tau_{1i}(\psi(\theta))', \mathbf{1})'$  with  $\tau_{1i}(\psi_i(\theta)) \equiv \phi[\psi_i(\theta)]/\Phi[\psi_i(\theta)]\{1 - \Phi(\psi_i(\theta))\}$ . Similarly, define  $h_i(\theta) \equiv ((y_i - \Phi(\psi_i(\theta)))', \tilde{u}_{2i}(\theta)')'$ , such that  $u_{Ni}(\theta) = \tau_i(\psi_i(\theta)) h_i(\theta)$  and the derivative becomes:

$$\frac{\partial \tilde{u}_{Ni}(\theta)}{\partial \theta'} = \left[ \frac{\partial \tau_i(\psi_i(\theta))}{\partial \theta} h_i(\theta) - \frac{\partial h_i(\theta)}{\partial \theta} \tau_i(\psi_i(\theta)) \right] \frac{\partial \psi_i(\theta)}{\partial \theta}$$

where

$$\begin{aligned} \frac{\partial \tau_i(\psi_i(\theta))}{\partial \theta} &= \frac{\partial}{\partial \theta} \begin{bmatrix} \tau_{1i}(\psi_i(\theta))' \\ \mathbf{1} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial \theta} \left[ \frac{\Phi[\psi_i(\theta)]}{\Phi[\psi_i(\theta)]\{1 - \Phi(\psi_i(\theta))\}} \right]' \\ \mathbf{0} \end{bmatrix} \\ \frac{\partial h_i(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} \begin{bmatrix} [y_1 - \Phi[\psi_i(\theta)]]' \\ [y_2 - \beta_0 - x'_{2i}\beta_1 - \mu\lambda_i(\delta, \gamma)]' \end{bmatrix} = \begin{bmatrix} [-\Phi[\psi_i(\theta)]]' \\ - \left[ 1 - x'_{2i} - \mu \frac{\partial \lambda_i(\delta, \gamma)}{\partial \theta} - \lambda_i(\delta, \gamma) \right]' \end{bmatrix} \\ \frac{\partial \psi_i(\theta)}{\partial \theta} &= \frac{\partial}{\partial \theta} \left[ \frac{\alpha_o + x'_{1i}\alpha_1}{\sqrt{\sum_j [\omega_{ij}^1(\theta)]^2}} \right] \\ &= \left[ \sqrt{\sum_j [\omega_{ij}^1(\theta)]^2} (1 + x'_{1i}) - (\alpha_o + x_{1i}\alpha_1) \frac{\partial \sqrt{\sum_j [\omega_{ij}^1(\theta)]^2}}{\partial \theta} \right] \cdot \left[ \sum_j [\omega_{ij}^1(\theta)]^2 \right] \end{aligned}$$

Then the task is to show that each of the parts of  $\frac{\partial \tilde{u}_{Ni}(\theta)}{\partial \theta'}$  is bounded uniformly. Following Pinkse and Slade (1998), we first establish

$$\sup_{y_i \in 0,1; t \in \mathbb{R}, y_2 \in \mathbb{R}} \left| \frac{\partial \tau(t)}{\partial t} h(t) - \frac{\partial h(t)}{\partial t} \tau(t) \right| < \infty$$

for which it is enough to show (i)  $\frac{\partial \tau(\theta)}{\partial \theta}$  and (ii)  $\frac{\partial h(\theta)}{\partial \theta} \tau(\theta)$  are bounded uniformly in  $t$ .

For (i), given that the only non-zero components of  $\frac{\partial \tau(\theta)}{\partial \theta}$  are those of  $\frac{\partial \tau_1(t)}{\partial \theta}$ , we concentrate on those. These components are the same as in Pinkse and Slade (1998) setup and thus their same arguments apply. Note that:

$$\frac{\partial \tau_1(\theta)}{\partial \theta} = \frac{1}{\Phi(t)} \left[ \frac{\phi(t)}{1 - \Phi(t)} \left\{ \frac{\phi(t)}{1 - \Phi(t)} - t \right\} \right] - \frac{\phi^2(t)}{\Phi^2(t)(1 - \Phi(t))}$$

and the only places it can be unbounded are at  $\pm\infty$ . Since the expression is an even function, it suffices to check  $t \rightarrow \infty$ . Define  $\Upsilon(t) = \frac{\phi(t)}{1 - \Phi(t)}$  and rewrite the above expression as:

$$\frac{\partial \tau_1(t)}{\partial t} = \frac{(\Upsilon(t) - t)^2 + t(\Upsilon(t) - t) - \phi(t)(\Upsilon(t) - t)/\Phi(t) - t\phi(t)/\Phi(t)}{\Phi(t)}$$

and noting that as  $t \rightarrow \infty$ ,  $\Phi(t) \rightarrow 1$  and  $t\phi(t) \rightarrow 0$ , and that

$$1 - \Phi(w) = \int_w^\infty \phi(t) dt = \int_w^\infty \frac{t\phi(t)}{t} dt = \frac{\phi(w)}{w} \left\{ 1 + \frac{1}{w^2} + O(w^{-4}) \right\}$$

as  $w \rightarrow \infty$ , the remaining term is

$$\Upsilon(t) - t = \frac{t}{1 + 1/t^2 + O(t^{-4})} - t = \frac{1}{t} + O(t^{-3})$$

as  $t \rightarrow \infty$ .

Thus, analyzing

$$\frac{\partial \tau_1(t)}{\partial t} = \frac{[\frac{1}{t} + O(t^{-3})]^2 + t[\frac{1}{t} + O(t^{-3})] - \phi(t)[\frac{1}{t} + O(t^{-3})]/\Phi(t) - t\phi(t)/\Phi(t)}{\Phi(t)} \rightarrow 1$$

and hence  $\frac{\partial \tau_1(\theta)}{\partial \theta}$  is bounded.

For (ii), that is,  $\frac{\partial h(t)}{\partial \theta} \tau(t)$ , recall that

$$\frac{\partial h(t)}{\partial t} = \begin{bmatrix} [-\phi(t)]' \\ -\left[1 - x_2' - \mu \frac{\partial \lambda(\delta, \gamma)}{\partial \theta} - \lambda(\delta, \gamma)\right]' \end{bmatrix} \text{ and } \tau(t) = \begin{bmatrix} \tau_1(t)' \\ \mathbf{1} \end{bmatrix}$$

and note it can be written as

$$\frac{\partial h(t)}{\partial t} \tau(t) = \begin{bmatrix} -\phi(t) \tau_1(t) \\ -\left[1 - x_2' - \mu \frac{\partial \lambda(t)}{\partial t} - \lambda(t)\right] \end{bmatrix} = \begin{bmatrix} -\phi(t) \frac{1}{\Phi(t)} [\Upsilon(t) - t + t] \\ -\left[1 - x_2' - \mu \frac{\partial \lambda(t)}{\partial t} - \lambda(t)\right] \end{bmatrix}.$$

The first component,  $\phi(t) \frac{1}{\Phi(t)} [\Upsilon(t) - t + t]$  is bounded by the same arguments above:  $\Phi(t) \rightarrow 1$ ,  $(\Upsilon(t) - t) = t^{-1} + O(t^{-3})$  and  $\phi(t) \rightarrow 0$ . So that it remains to check that the second component is bounded:  $\left[1 - x_2' - \mu \frac{\partial \lambda(t)}{\partial t} - \lambda(t)\right]$ .

$$\text{Take } \lambda(t) = \frac{\sum_j \omega_{ij}^1 \omega_{ij}^2}{\sum_j [\omega_{ij}^1]^2} \cdot \frac{\phi(t)}{\{1 - \Phi(t)\}} = \frac{\sum_j \omega_{ij}^1 \omega_{ij}^2}{\sum_j [\omega_{ij}^1]^2} \Upsilon(t).$$

Since  $\Upsilon(t) = \frac{1}{1/t + 1/t^3 + (1/t)O(t^{-4})}$  as  $t \rightarrow \infty$  and by assumption A7,  $\lambda(t)$  is bounded.

Taking  $\frac{\partial \lambda(t)}{\partial t} = \frac{\sum_j \omega_{ij}^1 \omega_{ij}^2}{\sum_j [\omega_{ij}^1]^2} \cdot \frac{\partial \Upsilon(t)}{\partial t}$ , by assumption A7 the first term is bounded, while using results above, the second term:

$$\begin{aligned} \frac{\partial}{\partial t} \left[ \frac{\phi(t)}{1 - \Phi(t)} \right] &= \frac{\{1 - \Phi(t)\} \phi(t) (-t) - \phi(t) (-\phi(t))}{[\{1 - \Phi(t)\}]^2} = \frac{\phi^2(t) - \phi(t)t(1 - \Phi(t))}{[\{1 - \Phi(t)\}]^2} \\ &= \frac{\phi^2(t)}{\phi^2(t) [1/t + 1/t^3 + (1/t)0 \cdot (t^{-4})]^2} - \frac{[1/t + 1/t^2 + 0 \cdot (t^{-4})]}{[1/t + 1/t^3 + (1/t)0 \cdot (t^{-4})]^2} \end{aligned}$$

and thus  $\frac{\partial}{\partial t} \Upsilon(t)$  is bounded and therefore  $\frac{\partial \lambda(t)}{\partial t}$  is bounded and the expression  $\frac{\partial h(t)}{\partial t}$  is also bounded.

Note also that the term  $h(t) = \left[ \begin{array}{c} y_1 - \Phi(t) \\ [y_2 - \beta_o - x'_2 \beta_1 - \mu \lambda(t)] \end{array} \right]$  is also bounded as  $\Phi(t) \rightarrow 1$  and  $\lambda(t)$  is bounded.

By assumption A5 the elements of  $z_N$  are bounded and thus for some constant  $C$ :

$$\sup_{\theta \in \Theta} \left| \frac{1}{N} \sum_i z_i' \frac{\partial \tilde{u}_{Ni}(\theta)}{\partial \theta'} \right| \leq C \sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^N \left\| \frac{\partial \psi_i(\theta)}{\partial \theta} \right\|.$$

Finally, checking  $\frac{\partial \psi_i(\theta)}{\partial \theta}$  with previous results,

$$\frac{\partial \psi_i(\theta)}{\partial \theta} = \left[ \sqrt{\sum_j [\omega_{ij}^1(\theta)]^2} (1 + x'_{2i}) - (\alpha_0 + x'_{1i} \alpha_1) \frac{\partial \sqrt{\sum_j [\omega_{ij}^1(\theta)]^2}}{\partial \theta} \right] \cdot \left[ \sum_j [\omega_{ij}^1(\theta)]^2 \right].$$

Since  $x_{1i}$  is bounded by A5 and the terms in the sums are also bounded by A7, then  $\sup_{\theta \in \Theta} \frac{1}{N} \sum_{i=1}^N \left\| \frac{\partial \psi_i(\theta)}{\partial \theta} \right\|$  is bounded by A6 and A1. *QED*

## Proof of Proposition 2

The first order conditions based on the objective function  $Q_N = g_{N'}(\theta) M_N g_N(\theta)$  are given by  $\frac{\partial Q_N(\hat{\theta})}{\partial \theta} = 0$ . Using the mean value theorem for  $\theta^*$  between  $\hat{\theta}$  and  $\theta_0$  we can write:

$$\left( \hat{\theta}_{GMM} - \theta_0 \right) = \left[ \frac{\partial^2 Q_N(\theta^*)}{\partial \theta \partial \theta'} \right]^{-1} \frac{\partial Q_N(\theta_0)}{\partial \theta} \quad (21)$$

The expression that corresponds to the second derivative of the objective function  $Q_N$  can be written as:

$$\frac{\partial^2 Q_N(\theta)}{\partial \theta \partial \theta'} = \frac{2}{N^2} \left[ \sum_{i=1}^N \frac{\partial^2 \tilde{u}_{Ni}(\theta)}{\partial \theta \partial \theta'} z_i' M_N z_j \tilde{u}_{Nj}(\theta) + \sum_{i=1}^N \frac{\partial \tilde{u}_{Ni}(\theta)}{\partial \theta} z_i' M_N z_j \frac{\partial \tilde{u}_{Nj}(\theta)}{\partial \theta'} \right].$$

We start by analyzing the convergence properties of this second derivative of the objective function. The following lemma will be useful.

**Lemma 1** (Pinkse and Slade (1998) Lemma A3)

For any  $\tilde{\theta}$  consistent for  $\theta_0$ , i.e.  $\tilde{\theta} \xrightarrow{p} \theta_0$ :

$$(a) \quad \frac{\partial g_N(\tilde{\theta})}{\partial \theta'} \xrightarrow{p} \frac{\partial g(\theta_0)}{\partial \theta'}$$

$$(b) \quad g_N(\tilde{\theta}) \xrightarrow{p} g_N(\theta_0)$$



Proof:

(a) Need to show that for  $\omega$  such that  $\|\omega\| = 1$ ,

$$\omega' \left[ \frac{\partial g_N(\tilde{\theta})}{\partial \theta'} - \frac{\partial g_N(\theta_0)}{\partial \theta'} \right] \xrightarrow{p} 0$$

as  $\lim_{N \rightarrow \infty} \frac{\partial g_N(\theta_0)}{\partial \theta'} = \frac{\partial g(\theta_0)}{\partial \theta'}$  follows from  $g_N(\theta_0) \xrightarrow{p} g(\theta_0)$  and A3.

Setting  $\bar{z}_i = \omega' z_i$  and using the mean value theorem, the above expression can be written as

$$= \frac{1}{N} \sum_{i=1}^N \bar{z}_i \left[ \frac{\partial \tilde{u}_{Ni}(\tilde{\theta})}{\partial \theta'} - \frac{\partial \tilde{u}_{Ni}(\theta_0)}{\partial \theta'} \right] = (\tilde{\theta} - \theta_0)' \frac{1}{N} \sum_{i=1}^N \bar{z}_i \frac{\partial^2 \tilde{u}_{Ni}(\theta^*)}{\partial \theta \partial \theta'}$$

where  $\theta^*$  is between  $\tilde{\theta}$  and  $\theta_0$ .

Analyzing the last expression,  $\frac{1}{N} \sum_{i=1}^N \bar{z}_i \frac{\partial^2 \tilde{u}_{Ni}(\theta^*)}{\partial \theta \partial \theta'}$  is bounded since by A5  $\bar{z}_i$  is uniformly bounded (including  $\lambda(\delta, \gamma)$ , which was shown in the proof of proposition 1) and  $\frac{\partial^2 \tilde{u}_{Ni}(\theta^*)}{\partial \theta \partial \theta'}$  is also bounded. Therefore, since  $\tilde{\theta} \xrightarrow{p} \theta_0$ ,  $(\tilde{\theta} - \theta_0) \xrightarrow{p} 0$  and thus  $\omega' \left[ \frac{\partial g_N(\tilde{\theta})}{\partial \theta'} - \frac{\partial g_N(\theta_0)}{\partial \theta'} \right] \xrightarrow{p} 0$  ■

(b) The proof is analogous to part (a).

Now returning to analyzing the second derivative of  $Q_N(\theta)$ , note that  $\sum_{j=1}^N z_j \tilde{u}_{Nj}(\theta) = op(N)$  from (a) in the proof of proposition 1. Furthermore, by lemma 1  $g_N(\theta^*) \xrightarrow{p} g_N(\theta_0)$  and also  $\frac{1}{N} \sum_{i=1}^N \frac{\partial^2 \tilde{u}_{Ni}(\theta^*)}{\partial \theta \partial \theta'} z_i' \omega$  is bounded in probability in probability  $\forall \|\omega\| = 1$ . Then, the first term inside the square brackets will vanish asymptotically.

Considering the remaining term of  $\frac{\partial^2 Q_N(\theta)}{\partial \theta \partial \theta'}$ , looking at  $\frac{1}{N} \sum_{i=1}^N z_i \frac{\partial \tilde{u}_{Ni}(\tilde{\theta})}{\partial \theta'}$ , it will converge in probability to  $\frac{\partial g(\theta_0)}{\partial \theta'}$  by lemma 1. Finally, since  $M_N \xrightarrow{p} M$  by A8,  $\frac{\partial^2 Q_N(\theta^*)}{\partial \theta \partial \theta'} \xrightarrow{p} \left[ \frac{\partial g'(\theta_0)}{\partial \theta} \right] M \left[ \frac{\partial g(\theta_0)}{\partial \theta'} \right] \equiv \Psi_2(\theta_0)$ .

Now we turn to the term  $\frac{\partial Q_N(\theta_0)}{\partial \theta}$  in equation (21), which is equal to  $\frac{\partial Q_N(\theta_0)}{\partial \theta} = 2 \frac{\partial g'_N(\theta_0)}{\partial \theta} M_N g_N(\theta_0)$  and it follows from previous results that  $\frac{\partial g'_N(\theta_0)}{\partial \theta} \xrightarrow{p} \frac{\partial g'(\theta_0)}{\partial \theta}$ .

The remaining task is to show that  $g_N(\theta_0) \rightarrow N(0, \Psi_1(\theta_0))$ . To do this, we follow again closely Pinkse and Slade (1998)'s strategy and employ Berenstein (1927) blocking method using McLeish (1974) central limit theorem for dependent processes in Davison's (1994, chapter 24).

Start by defining  $Y_{0N} = \omega' \{E[Ng_N(\theta_0)g'_N(\theta_0)]\}^{-\frac{1}{2}} \sqrt{N}g_N(\theta_0) = \frac{1}{\sqrt{N}} \sum_{t=1}^N A_{Nt}$  for implicitly defined  $A_{Nt}$  and  $\forall \|\omega\| = 1$ , and thus the task is to show  $Y_{0N} \xrightarrow{d} N(0, 1)$ .

Following Davison (1994, chapter 24), split the region in which the observations are located into  $a_N$  areas of size  $c_1\sqrt{b_N} \times c_2\sqrt{b_N}$  each, where  $a_N$  and  $b_N$  are integers such that  $a_N b_N = N$ .

Without loss of generality, set  $c_1 = c_2 = 1$  and let  $a_N$  and  $b_N$  be such that  $\alpha(b_N)a_N \rightarrow 0$  and  $N^{l-(\frac{1}{2})}b_N < 1$  uniformly in  $N$  for some fixed  $0 < l < \frac{1}{2}$ .

Define the set of indices  $\Lambda_{Nj}$  that correspond to observations in area  $j$ . By assumption, a number  $c > 0$  exists such that  $\max_j |\Lambda_{Nj}| < Cb_N$ , where  $|\cdot|$  applied to sets denotes the

cardinality of that set. Define  $D_{Nj} = \frac{1}{\sqrt{N}} \sum_{t \in \Lambda_{Nj}} A_{Nt}$  such that  $Y_{0N} = \sum_{j=1}^{a_N} D_{Nj}$ .

Following Davison's (1994) theorem 24.1, McLeish's (1974) CLT requires that the following conditions hold.

- (a)  $T_{Na_N} = \prod_{j=1}^{a_N} (1 + i\lambda D_{Nj})$  is uniformly integrable in  $N > N^*$  for some fixed  $N^*$  and  $\lambda > 0$ .
- (b)  $E[T_{Na_N}] - 1 \rightarrow 0$
- (c)  $\sum_{j=1}^{a_N} D_{Nj}^2 - 1 \xrightarrow{p} 0$
- (d)  $\max_j |D_{Nj}| \xrightarrow{p} 0$

Before establishing each of these conditions, we note that for sufficiently large  $N$ ,  $A_{Nt}$  is bounded, since  $\Psi_1(\theta_0)$  is positive definite (p.d.) and thus  $E[Ng_N(\theta_0)g'_N(\theta_0)]$  is p.d. and its inverse is bounded. We establish each of the above conditions in turn.

- (a) It needs to be shown that for some fixed  $N^*$ ,

$$\sup_{N > N^*} E|T_{Na_N} I\{|T_{Na_N}| > K\}| \rightarrow 0 \text{ as } K \rightarrow \infty$$

where  $I$  is an indicator function.

Following Pinkse and Slade (1998), we begin by showing that  $P \left[ \sup_{N > N^*} |T_{Na_N}| > K \right] = 0$  for some  $K > 0$  which will imply the above condition.

Note that

$$\begin{aligned} P \left[ \sup_{N > N^*} |T_{Na_N}| > K \right] &= P \left[ \sup_{N > N^*} \left| \prod_{j=1}^{a_N} (1 + i\lambda D_{Nj}) \right| > K \right] \\ &\leq P \left[ \sup_{N > N^*} \prod_{j=1}^{a_N} (1 + \lambda^2 D_{Nj}^2)^{\frac{1}{2}} > K \right] \end{aligned}$$

$$\begin{aligned}
&= P \left[ \sup_{N > N^*} \prod_{j=1}^{a_N} (1 + \lambda^2 D_{Nj}^2)^{\frac{1}{2}} > K \mid \sup_{N > N_j^*} N^l |D_{Nj}| \leq C \right] \cdot P \left[ \max_j N^l |D_{Nj}| \leq C \right] + \\
&P \left[ \sup_{N > N^*} \prod_{j=1}^{a_N} (1 + \lambda^2 D_{Nj}^2)^{\frac{1}{2}} > K \mid \sup_{N > N_j^*} N^l |D_{Nj}| > C \right] \cdot P \left[ \sup_{N > N_j^*} N^l |D_{Nj}| > C \right] \\
&\leq P \left[ \sup_{N > N^*} \prod_{j=1}^{a_N} (1 + \lambda^2 D_{Nj}^2)^{\frac{1}{2}} > K \mid \sup_{N > N_j^*} N^l |D_{Nj}| \leq C \right] + P \left[ \sup_{N > N_j^*} N^l |D_{Nj}| > C \right]
\end{aligned}$$

with  $C$  a uniform upper bound to the  $A_{Nt's}$ . Nothing that the first summand is bounded by  $\sup_{N > N^*} I \{ (1 + \lambda^2 C N^{-2l})^{a_{n/2}} > K \}$  which is zero for sufficiently large  $K$ , and that the

second summand is bounded by  $P \left[ \sup_{N > N^*, l} N^{l - (\frac{1}{2})} b_N |A_{Nt}| > C \right] = 0$  since  $N^{l - (\frac{1}{2})} b_N < 1$  by

construction. Therefore,  $P \left[ \sup_{N > N^*} |T_{Na_N}| > K \right] = 0$  for some  $K > 0$ , which in turn implies

$\sup_{N > N^*} E |T_{Na_N} I \{ |T_{Na_N}| > K \}| \rightarrow 0$  as  $K \rightarrow \infty$  and thus condition (a) holds.

(b) From Davison (1994) we can write  $T_{Na_N} = (i\lambda) \sum_{j=1}^{a_N} D_{Nj} T_{N,j-1}$ . Then, it is enough to show that the  $\max_j |E [D_{Nj} T_{N,j-1}]| = o(a_N^{-1})$ .

Writing  $T_{N,j-1} = \prod_{k \in \Xi_{Nj1}} (1 + i\lambda D_{Nk}) \cdot \prod_{k \notin \Xi_{Nj1}} (1 + i\lambda D_{Nk}) = \prod_{k \in \Xi_{Nj1}} (1 + i\lambda D_{Nk}) T_{RNj}$  where  $T_{RNj}$  is implicitly defined and  $\Xi_{Nj1}$  is the set of blocks adjacent to block  $j$ .

Thus,  $T_{N,j-1} = (i\lambda) \sum_{\gamma \in \Gamma_{Nj}} \left( \prod_{k \in \Xi_{Nj1}} D_{Nk}^{\gamma_k} \right) T_{RNj}$  with  $\Gamma_{Nj}$  the set of vectors of size equal to  $|\Xi_{Nj1}|$  whose elements are all either zero or one. Because the number of elements in  $\Gamma_{Nj}$  is finite, it is enough to show that  $\max_{j, \gamma} \left| E \left[ D_{Nj} T_{RNj} \prod_{k \in \Xi_{Nj1}} D_{Nk}^{\gamma_k} \right] \right| = o(a_N^{-1})$ . To show this, we show that (i)  $\max_j |E [D_{Nj} T_{RNj}]| = o(a_N^{-1})$  and (ii)  $\max_{j \neq k} |E [D_{Nj} D_{Nk} T_{RNj}]| = o(a_N^{-1})$ .

For (i), note that the observations in  $T_{RNj}$  are located at least a distance  $b_N^{\frac{1}{2}}$  away from those in  $D_{Nj}$  by construction. Therefore,  $\max_j |E [D_{Nj} T_{RNj}]|$  is bounded by  $C_1 \max_j E |D_{Nj} T_{RNj}| \alpha(\sqrt{b_N})$  for large  $C_1 > 0$  given the conditions on the covariances and that  $E [D_{Nj}] = 0$ . Using the properties of the constructed  $a_N$ ,  $b_N$  and the properties of  $\alpha$  in A6:

$C_1 \max_j E |D_{Nj} T_{RNj}| \alpha(\sqrt{b_N}) = o\left(N^{-\frac{1}{2}} b_N \alpha(\sqrt{b_N})\right) = o\left(a_N^{-\frac{1}{2}} b_N^{-\frac{1}{2}} \alpha(\sqrt{b_N})\right) = o(a_N^{-1})$  for large  $N$ , as required.

For (ii),  $\max_{j \neq k} |E [D_{Nj} D_{Nk} T_{RNj}]| \leq \max_{j \neq k} N^{-\frac{1}{2}} b_N |E [D_{Nj} T_{RNj}]|$  by the boundedness property of the  $A_{Nt's}$ . The maximum is thus of order  $O\left(N^{-\frac{1}{2}} b_N a_N^{-1}\right)$  which is  $o(a_N^{-1})$ . Repeating the argument for the remaining elements in  $\Gamma_{Nj}$  completes the proof.

Therefore, condition (b)  $E [T_{Na_N}] \rightarrow 1$  is demonstrated.

(c) We start by showing that  $\sum_{j=1}^{a_N} (D_{Nj}^2 - E[D_{Nj}^2]) \xrightarrow{p} 0$

Take  $\sum_{i,j=1}^{a_N} E \{ (D_{Nj}^2 - E[D_{Nj}^2]) (D_{Ni}^2 - E[D_{Ni}^2]) \} \leq C_2 \sum_{l=0}^{C_1\sqrt{a_N}} (l+1) \alpha(\sqrt{b_N}l) \max_i E[D_{Ni}^4]$

where  $C_1, C_2 > 0$  are sufficiently large constants, and the inequality is a result of the conditions on the covariances and locations in assumptions A6 and A9. The right hand side of the inequality is of order  $0(N^{-2}b_N^3a_N)$  since it follows from the conditions in A6 and A9 that  $\max_i E[D_{Ni}^4]$  is bounded by

$$\begin{aligned} & \frac{1}{N^2} \max_i \sum_{t_1, t_2, t_3, t_4 \in \Lambda_{Ni}} |E[A_{Nt_1}A_{Nt_2}A_{Nt_3}A_{Nt_4}]| \leq \\ & C_1 \frac{1}{N^2} \max_i \sum_{t_1, t_2, t_3, t_4 \in \Lambda_{Ni}} \{ \alpha(d_{t_1, t_2}) + \dots + \alpha(d_{t_3, t_4}) \} \leq C_2 \frac{1}{N^2} b_N^2 \max_i \sum_{t_1, t_2 \in \Lambda_{Ni}} \alpha(d_{t_1, t_2}) \leq \\ & C_4 \frac{1}{N^2} b_N^2 \max_i \sum_{t_1 \in \Lambda_{Ni}} \sum_{l=0}^{C_3 t_N^{1/2}} l \alpha(l) \text{ for some } C_1, C_2, C_3, C_4 > 0, \text{ and since } \sum_{l=0}^{\infty} l \alpha(l) \text{ is bounded} \end{aligned}$$

by A6, the last term is of order  $O(N^{-2}b_N^3a_N)$ .

Finally,  $C_2 \sum_{l=0}^{C_1\sqrt{a_N}} (l+1) \alpha(\sqrt{b_N}l) \max_i E[D_{Ni}^4] = O(N^{-2}b_N^3a_N) = o(1)$  as  $n \rightarrow \infty$  by the properties of  $a_N$  and  $b_N$ . Thus we have shown that  $\sum_{j=1}^{a_N} (D_{Nj}^2 - E[D_{Nj}^2]) \xrightarrow{p} 0$  and condition

(c) can be written as:  $\sum_{j=1}^{a_N} D_{Nj}^2 - 1 = \sum_{j=1}^{a_N} E[D_{Nj}^2] - 1 + op(1)$  which can be further rewritten as  $\sum_{j=1}^{a_N} E[D_{Nj}^2] - 1 + op(1) = E[Y_{0N}^2] - 1 - \sum_{i \neq j} E[D_{Ni}D_{Nj}] + op(1)$ .

To check the order of convergence we need only to analyze the term  $\sum_{i \neq j} E[D_{Ni}D_{Nj}]$ . For this it is enough to analyze  $\max_i \sum_{i \neq j} |E[D_{Ni}D_{Nj}]|$  since each of the summations over  $i$  and  $j$  contain terms with  $a_N$  or  $a_{N-1}$ . Take  $\Xi_{Ni}$  as previously defined, then, for some  $C_1 > 0$ ,  $\max_i \sum_{i \neq j} |E[D_{Ni}D_{Nj}]|$  can be bounded by

$$\max_i \sum_{l=1}^{C_1\sqrt{a_N}} \sum_{j \in \Xi_{Ni}} |E[D_{Ni}D_{Nj}]| \leq \max_i \sum_{j \in \Xi_{Ni}} |E[D_{Ni}D_{Nj}]| + \max_i \sum_{l=2}^{C_1\sqrt{a_N}} \sum_{j \in \Xi_{Ni}} |E[D_{Ni}D_{Nj}]| \quad (22)$$

and we analyze each of the terms in the right-hand side .

For the first term, note that  $\max_{i \neq j} |E[D_{Ni}D_{Nj}]| = \max_{i \neq j} \left| \frac{1}{N} \sum_{t \in \Lambda_{Ni}, s \in \Lambda_{Ni}} E[A_{Nt}A_{Ns}] \right|$   
 $\leq \max_{i \neq j} C_1 \frac{1}{N} \sum_{t \in \Lambda_{Ni}, s \in \Lambda_{Ni}} \alpha(d_{ts})$  for some large  $C_1 > 0$ , by the boundedness of  $A_{Nt}$ 's and A6.

Consider adjacent blocks for which dependence will be typically stronger, then, by A9 and A6, the number of  $(t, s)$  combinations within distance  $d$  is bounded by  $C_2 b_N^{1/2} d^2$  for some  $C_2 > 0$ .

Letting  $C_3 = C_2 C_1$  the expression is bounded by  $C_3 \max \frac{1}{N} b_N^{1/2} \sum_{d=0}^{C_4\sqrt{b_N}} d^2 \alpha(d)$  for some  $C_4 > 0$

and since by A9  $d^2 \alpha(d) \rightarrow 0$  the expression is  $o\left(\frac{1}{N}b_N\right)$  and thus  $\max_{i \neq j} |E[D_{Ni}D_{Nj}]| = o\left(\frac{1}{N}b_N\right)$  or  $o\left(a_N^{-1}\right)$  (since  $n = a_N b_N$ ), and so is the first term in (22).

For the second term, first note that by the boundedness of  $A_{Nt}$  and  $E[A_{Nt}] = 0$ :  $\max_{j \in \Xi_{Nil}} \max_{t \in \Lambda_{Ni}} \max_{s \in \Lambda_{Nj}} |E[A_{Nt}A_{Ns}]| = O\left(\alpha\left(\sqrt{b_N}(l-1)\right)\right)$  uniformly in  $l$ . Therefore, the second term is

$$\text{bounded by } C_2 \max_i \sum_{l=2}^{C_1 \sqrt{a_N}} \frac{1}{N} |\Xi_{Nil}| |\Lambda_{Ni}| |\Lambda_{Nj}| \alpha\left(\sqrt{b_N}(l-1)\right) \leq$$

$$C_3 \frac{1}{N} b_N^2 \sum_{l=1}^{C_1 \sqrt{a_N}} l \alpha\left(\sqrt{b_N}l\right) = o\left(\frac{1}{N} b_N \sum_{l=1}^{C_1 \sqrt{a_N}} l \alpha(l)\right) = o\left(a_N^{-1}\right). \text{ The second-to-last equality}$$

is due to  $\frac{\alpha(ts)}{\alpha(s)} = o(t^2)$  as  $t \rightarrow \infty$  and the last one due to the boundedness of the sum

$$\sum_{l=1}^{C_1 \sqrt{a_N}} l \alpha(l).$$

Finally, since  $\max_i \sum_{i \neq j} |E[D_{Ni}D_{Nj}]|$  is  $o\left(a_N^{-1}\right)$  and thus  $\sum_{i \neq j} E[D_{Ni}D_{Nj}]$  is  $o(1)$ , condition (c) is demonstrated.

(d) Using the boundedness of  $A_{Nt}$ ,

$$\max_j |D_{Nj}| = \max_j \left| \frac{1}{\sqrt{N}} \sum_{t \in \Lambda_{Nj}} A_{Nt} \right| \leq \frac{1}{\sqrt{N}} |\Lambda_{Nj}| \max_t |A_{Nt}| = Op\left(\frac{1}{\sqrt{N}} b_N\right) = op(1) \text{ by the construction of } b_N.$$

Since conditions (a)-(d) are satisfied under the current assumptions,  $Y_{0N} \xrightarrow{d} N(0, 1) \iff g_N(\theta_0) \xrightarrow{d} N(0, \Psi_1(\theta_0))$  which concludes the proof of proposition 2. *QED*

### Proof of Proposition 3

$\Psi_{2N}(\hat{\theta}_{GMM}) \xrightarrow{p} \Psi_2(\theta_0)$  follows from the fact that  $\frac{\partial g_N(\hat{\theta}_{GMM})}{\partial \theta'} \xrightarrow{p} \frac{\partial g(\theta_0)}{\partial \theta'}$  by part (a) of Lemma 1, by the consistency of  $\theta_{GMM} \xrightarrow{p} \theta_0$ ; and by A8 which assumes  $M_N \xrightarrow{p} M$

To show that  $\Psi_{1N}(\hat{\theta}_{GMM}) \xrightarrow{p} \Psi_1(\theta_0)$  we can show that  $\Psi_{1N}(\hat{\theta}_{GMM}) \xrightarrow{p} \Psi_{1N}(\theta_0)$  as we did in Lemma 1.

Note that

$$\begin{aligned} \Psi_{1N}(\theta_0) &= \frac{1}{N} \sum_{ij} \tau_{Ni}(\theta_0) \tau_{Nj}(\theta_0) z_i z_j' \left\{ \Phi_2(\psi_i(\theta_0), \psi_j(\theta_0), \rho_{Nij}(\theta_0)) - \Phi_2(\psi_i(\theta_0), \psi_j(\theta_0), 0) \right\} \\ &= \frac{1}{N} \sum_{ij} \tau_{Ni}(\theta_0) \tau_{Nj}(\theta_0) z_i z_j' \rho_{Nij}(\theta_0) \frac{\partial \Phi_2(\psi_i(\theta_0), \psi_j(\theta_0), \rho_{Nij}(\theta_0^*))}{\partial \rho} \end{aligned} \quad (23)$$

where  $\tau(\cdot)$  is as defined in the proof of proposition 1,  $\Phi_2$  stands for the bivariate normal distribution, and the second equality follows from the mean value theorem for  $\theta^*$  between  $\hat{\theta}_{GMM}$  and  $\theta_0$ .

First consider the partial derivative term and show that it is bounded. Take the following bivariate normal distribution.

$$\Phi_2(a, b, \rho) = \int_{-\infty}^b \int_{-\infty}^a \frac{1}{2\Pi(1-\rho)^{1/2}} \exp\left\{-\frac{1}{2(1-\rho^2)}(t^2 - 2\rho ts + s^2)\right\} dt ds$$

and make the following change of variable from  $t$  to  $u = (t - \rho s)(1 - \rho^2)^{-1/2}$ . Integrating over  $u$  and differentiating with respect to  $\rho$  yields:

$$\frac{\partial \Phi(a, b, \rho)}{\partial \rho} = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^b \left\{ -s(1 - \rho^2)^{-1/2} + (a - \rho s)\rho(1 - \rho^2)^{-3/2} \right\}$$

which, by the conditions on  $\rho$  assumed in A12 - A14, is bounded.

Repeating the application of the mean value theorem with respect to  $\theta$  in (23), using the boundedness of the above partial derivative and the fact that  $\hat{\theta}_{GMM} \xrightarrow{p} \theta_0$  yields the result that  $\Psi_{1N}(\hat{\theta}_{GMM}) \xrightarrow{p} \Psi_{1N}(\theta_0)$  and thus  $\Psi_{1N}(\hat{\theta}_{GMM}) \xrightarrow{p} \Psi_1(\theta_0)$ . *QED*