Estimation of Sample Selection Models with Spatial Dependence*

Alfonso Flores-Lagunes† Kurt Erik Schnier‡

This version: June 2006

Abstract

We consider the estimation of sample selection (type II Tobit) models that exhibit spatial error dependence or spatial autoregressive errors (SAE). 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 finite sample properties of the proposed estimator using a Monte Carlo experiment; discuss the importance of the spatial sample selection model in applied work, and illustrate the application of our method by estimating the spatial production within a fishery with data that is censored for reasons of confidentiality.

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

JEL classification: C13, C15, C24, C49

*We want to thank Dan Ackerberg, Anil Bera, Carlos A. Flores, Kei Hirano and Dan McMillen for useful comments; and seminar participants at the Universities of Arizona (AREC), Illinois, Ohio State, Purdue (AE), Syracuse, the 2004 EEA-ESEM meeting, and the 2003 Midwest Econometrics Group meetings. Very competent research assistance was provided by Alex Shcherbakov. Remaining errors are our own.

†Corresponding author. Department of Economics, Eller College of Management, University of Arizona. McClelland Hall 401, Tucson, AZ, 85721-0108. Phone: (520) 626-3165. Email: alfonso@eller.arizona.edu

‡Department of Environmental and Natural Resource Economics, University of Rhode Island. Email: 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.[1] 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 Beron and Vijverberg (2004), Case (1992), Fleming (2004), 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 error dependence (or spatial autoregressive errors, SAE) 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).

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 generalized method of moments (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 generalized method of moments (GMM) framework. In order to identify the spatial autoregressive error parameters, the moment conditions suggested by Kelejian and Prucha (1999) are added to the set of moments used in estimation.

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

[1] Some examples are Case (1991), Fishback, Horrace and Kantor (2006), Topa (2001), among many others.

integrals. To obtain consistent and efficient parameter estimates of the spatial probit model, LeSage (2000) and Beron and Vijverberg (2004) 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.[2]

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 error parameters will be different in each equation. In this case, the IMR turns out to be a function of the unknown spatial parameter in the outcome equation. 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 spatial autoregressive errors. Building on Pinkse and Slade's (1998) asymptotic results for their spatial probit model and standard GMM theory, our proposed estimator is consistent, asymptotically normally distributed, and its covariance matrix can be estimated.

We note that our estimator has lower efficiency compared to a maximum likelihood estimator that requires computationally-intensive simulation methods for multidimensional

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

integration. However, given the fact that the multidimensional integration is in the order of the number of observations, the relative computational simplicity of our method is preferable in many relevant instances when the amount of data available to researchers is large.

To our knowledge, despite the pervasiveness of selection bias in the economics literature and the serious consequences of spatial dependence in the typical selection models, there is only one other paper that attempts specifying and estimating a sample selection model with spatial dependence. McMillen (1995) first specifies a model similar to the heckit model with spatial effects presented in the present paper and outlines an estimator based on the EM algorithm. However, this estimator requires knowledge of the true spatial autoregressive error parameters and it is fairly computationally intensive. As a result, McMillen (1995) specifies and estimates an extension of the spatial expansion model of Casetti (1972) that is used in geography. Nevertheless, this model is not explicitly spatial since additional variables are required to indirectly control for the spatial dependence, and the consistency of the proposed estimator depends heavily on correctly assuming the functional form of the underlying heteroskedasticity induced by the spatial dependence. Therefore, compared to the pioneering work by McMillen (1995), we propose a feasible estimator for the heckit model that explicitly accounts for and estimates the spatial error dependence parameters.

The paper is organized as follows. Section 2 presents the sample selection model with spatial autoregressive errors (SAE) and discusses some of the alternative estimators for linear and non-linear models with spatial dependence to posit our method in context. Section 3 introduces our proposed method of estimation (the "spatial heckit"), states its large-sample properties, and discusses some aspects of its implementation. Section 4 presents the results of a Monte Carlo experiment to analyze the finite sample properties of our estimator; while section 5 discusses the practical importance of the spatial sample selection model and presents an empirical example to illustrate the application of our technique. Concluding remarks are provided in the last section of the paper.

2 The Sample Selection Models with Spatial Dependence

Our focus is the estimation of a sample selection (Tobit type II) model with spatial autoregressive errors (SAE) in both the selection and the outcome equations. The spatial autoregressive error (SAE) model specifies spatially autocorrelated disturbances:

$$y_{1i}^* = \alpha_0 + x_{1i}'\alpha_1 + u_{1i}, \qquad u_{1i} = \delta \sum_{j \neq i} c_{ij} u_{1j} + \varepsilon_{1i} \tag{1}$$

$$y_{2i}^* = \beta_0 + x_{2i}'\beta_1 + u_{2i}, \qquad u_{2i} = \gamma \sum_{j \neq i} c_{ij} u_{2j} + \varepsilon_{2i} \tag{2}$$

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 δ and γ. 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.[3] It is assumed that:

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

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

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

$$y_{1i}^* = \alpha_0 + x_{1i}'\alpha_1 + \sum_j \omega_{ij}^1 \varepsilon_{1j} \tag{3}$$

$$y_{2i}^* = \beta_0 + x_{2i}'\beta_1 + \sum_j \omega_{ij}^2 \varepsilon_{2j} \tag{4}$$

where the weights ω_{ij}^1 and ω_{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, ω_{ij}^1 and ω_{ij}^2, depend upon the unknown parameters δ and γ, respectively.

Note that in the absence of any sample selection, equation (2) is just a linear model with SAE, for which a number of estimation methods exist. We briefly review some of those methods here. Ordinary least squares (OLS) ignoring the SAE results in consistent but inefficient estimates, and does not produce an estimate of the SAE parameter. The use of

[3] Without loss of generality, we specify the same spatial weights in each of the two equations.

maximum likelihood (ML) to obtain asymptotically efficient estimates, including the SAE parameter, was suggested by Ord (1975) and rigorously analyzed by Lee (2004). This method relies on normality and its computational demands increase with the sample size since the Jacobian of the likelihood function entails obtaining the determinant of a full matrix of dimension equal to the sample size (Anselin and Bera, 1998; Kelejian and Prucha, 1999). Another approach to obtain consistent estimates of all parameters in the linear SAE model is the three-step FGLS procedure of Kelejian and Prucha (1998) that we refer to as KP-SAE. First, residuals are obtained using OLS, which are used to estimate the SAE parameter using the "generalized moments" estimator in Kelejian and Prucha (1999); finally, OLS is applied to a model transformed with a Cochrane-Orcutt type procedure.[4] Compared to ML, KP-SAE is asymptotically less efficient, although has been found to be "virtually as efficient" in simulations (Kelejian and Prucha, 1999), while being computationally simpler and not relying on normality.

Note also that equation (1) is a probit model with spatial autoregressive errors. In this case, SAE introduces a fully non-spherical variance-covariance matrix that renders the simple probit estimator inconsistent. As a result, to obtain ML estimates that are consistent and asymptotically efficient (relying on normality), multidimensional integration on the order of the sample size is necessary. This is possible using simulation algorithms (LeSage, 2000; Beron and Vijverberg, 2004), but becomes infeasible even for moderate-size sample sizes. Alternatively, less efficient methods that obtain consistent estimates of the parameters in equation (1) that account for the induced heteroskedasticity while ignoring the off-diagonal elements of the variance-covariance matrix have been proposed. For instance, Case (1992) transforms a particular model structure to obtain homoskedastic errors;[5] McMillen (1992) uses the EM algorithm to account for the heteroskedasticity induced by the SAE process; while Pinkse and Slade (1998) propose a GMM estimator that also takes into account the heteroskedasticity induced by the SAE process. In our method below, we use Pinkse and Slade's (1998) estimator, which allows us to obtain consistent estimates of all parameters in equation (1) including the SAE parameter, is computationally simple, and allows us to use the GMM framework of estimation when we consider estimation of the full model in (1)-(2).

[4] Note that the original procedure in Kelejian and Prucha (1998) is more general since it allows for spatial lag dependence as well. For a related approach based on GMM for this linear model with SAE, see Lee (2001).

[5] Unfortunately, Case's (1992) method is constrained to situations in which the population can be partitioned into groups (e.g. "districts") whose errors can be assumed independent.

Before introducing our estimator in the next section, we finally note that in the absence of SAE the model in (1)-(2) reduces to the standard Heckman's (1979) sample selection model. This model can be estimated with a two-step procedure (heckit) or a ML method, which are asymptotically equivalent under standard conditions. In the presence of spatial error dependence, however, neither of these two methods result in consistent estimates of the parameters since (1)-(2) is a limited-dependent variable model and will be affected by the same problems discussed in the context of the probit model. As a result, consistent and fully efficient estimates can only be obtained with methods that account for the full structure of the non-spherical variance-covariance matrix, such as ML with multidimensional integrals that are on the order of the sample size. Unfortunately, just as with the probit model, the available simulation methods to approximate multidimensional integrals become infeasible when the sample size at hand is relatively large. Therefore, we propose below a feasible estimator for relatively large samples that achieves consistency by accounting for the heteroskedasticity induced by the spatial error dependence while sacrificing efficiency, as the off-diagonal elements of the variance-covariance matrix are ignored.

3 Estimation of the Sample Selection Models with Spatial Dependence

We now describe our proposed estimation method for the sample selection model with spatial dependence presented in the previous section. 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.

To motivate the estimation of the SAE model in (1)-(2), we start with the following calculations (McMillen, 1995):

$$var(u_{1i}) = \sigma_1^2 \sum_j (\omega_{ij}^1)^2 \tag{5}$$

$$var(u_{2i}) = \sigma_2^2 \sum_j (\omega_{ij}^2)^2 \tag{6}$$

$$E(u_{1i}, u_{2i}) = \sigma_{12} \sum_j \omega_{ij}^1 \omega_{ij}^2. \tag{7}$$

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 (5), 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 $\psi_i(\theta_1) = \frac{\alpha_0 + x_{1i}'\alpha_1}{\sqrt{var(u_{1i})}}$ the index function of the probit model weighted by the standard deviation of the residual. The corresponding "generalized residuals" of this model are:

$$\tilde{u}_{1i}(\theta_1) = \{y_{1i} - \Phi[\psi_i(\theta_1)]\} \cdot \frac{\phi[\psi_i(\theta_1)]}{\Phi[\psi_i(\theta_1)]\{1 - \Phi[\psi_i(\theta_1)]\}}. \tag{8}$$

The GMM estimates for θ_1 can be obtained as follows:

$$\hat{\theta}_{1,GMM} = \underset{\theta_1 \in \Theta_1}{\arg\min} \quad S_N(\theta_1)' M_N S_N(\theta_1) \tag{9}$$

where $S_N(\theta_1) = \frac{1}{N} z_N' \tilde{u}_{1N}(\theta_1)$, z_N is a data matrix of regressors plus at least one instrument (to identify the extra parameter δ),[6] $\tilde{u}_{1N}(\theta_1)$ is the vector of generalized residuals with elements as shown in (8), 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 θ_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

[6] The original formulation of Pinkse and Slade's (1998) estimator uses an instrument to identify δ. In our estimator below we do not require such an instrument since we use specific moment conditions to estimate δ.

the outcome equation (2) has the following form (McMillen, 1995):

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

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

$$\lambda_i \equiv \frac{\sum\limits_j \omega_{ij}^1\omega_{ij}^2}{\sqrt{\sum\limits_j[\omega_{ij}^1]^2}} \cdot \frac{\phi\left[-\psi_i(\theta_1)\right]}{\{1 - \Phi\left[-\psi_i(\theta_1)\right]\}}. \tag{10}$$

Once estimated ($\hat{\lambda}_i$), 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, such as those described in the last section. We illustrate our estimator in this section by employing KP-SAE to estimate the augmented outcome equation: $y_{2i} = \beta_0 + x_{2i}'\beta_1 + \mu\widehat{\lambda}_i + v_{2i}$.

However, note that the "adjusted" IMR in (10) depends on a parameter that is not estimated in the first step: γ, which is included in the weights ω_{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 KP-SAE estimators. More specifically, we stack their corresponding moment conditions:

$$g(z_N, \theta) = \left[s(z_{1N}, \theta)', m(z_{2N}, \theta)'\right]', \quad \theta = \{\alpha_0, \alpha_1', \delta, \beta_0, \beta_1', \mu, \gamma\}$$

with

$$\begin{aligned}
s(z_{1N}, \theta) &= z_{1N}'\tilde{u}_{1N}(\theta), \quad \tilde{u}_{1N}(\theta) \text{ as in (8)}, \\
m(z_{2N}, \theta) &= [y_{1N} \cdot z_{2N}]'\tilde{u}_{2N}(\theta), \quad \tilde{u}_{2N}(\theta) = y_{2N} - \beta_0 - x_{2N}'\beta_1 - \mu\widehat{\lambda}_N(\delta, \gamma)
\end{aligned}$$

where the subscript N denotes the corresponding vector or matrix of data, we have let $z'_N = (z'_{1N}, [y_{1N} \cdot z_{2N}]')'$, z_{1N} includes the regressors of the selection equation, and z_{2N} includes the regressors of the outcome equation plus the estimated "adjusted" IMR, which is represented as $\widehat{\lambda}_N(\delta, \gamma)$ to make explicit its dependence on both SAE parameters.[7]

Defining $\tilde{u}_N(\theta) = (\tilde{u}'_{1N}(\theta), \tilde{u}'_{2N}(\theta))'$ then all parameters of the SAE sample selection model can be estimated as:

$$\hat{\theta}_{GMM} = \underset{\theta \in \Theta}{\arg\min} \;\; g_N(\theta)' M_N g_N(\theta) \tag{11}$$

where $g_N(\theta) = \frac{1}{N} z'_N \tilde{u}_N(\theta)$, for a conformable positive definite M_N such that $M_N \xrightarrow{p} M$. We call $\hat{\theta}_{GMM}$ the "spatial heckit" estimator for the sample selection model with SAE.

Denote $g(\theta) \equiv \lim_{N \to \infty} E[g_N(\theta)]$ and let θ_0 be the true parameter vector. Under conditions similar to those in Pinkse and Slade (1998), $\hat{\theta}_{GMM}$ is consistent ($\hat{\theta}_{GMM} \xrightarrow{p} \theta_0$) and asymptotically normal:[8]

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

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

Furthermore, the asymptotic variance of $\hat{\theta}_{GMM}$ can be estimated:

$$\Psi_{1N}(\hat{\theta}_{GMM}) \xrightarrow{p} \Psi_1(\theta_0) \;\text{ and }\; \Psi_{2N}(\hat{\theta}_{GMM}) \xrightarrow{p} \Psi_2(\theta_0), \;\text{ where}$$

$$\begin{aligned} \Psi_{1N}(\hat{\theta}_{GMM}) &= N E\{g_N(\hat{\theta}_{GMM}) g_N(\hat{\theta}_{GMM})'\} \text{ and} \\ \Psi_{2N}(\hat{\theta}_{GMM}) &= [\partial g'_N(\hat{\theta}_{GMM})/\partial\theta] M_N [\partial g_N(\hat{\theta}_{GMM})/\partial\theta']. \end{aligned}$$

The efficient $\hat{\theta}_{GMM}$ within the class of GMM estimators is obtained when the optimal GMM weighting matrix is used. That is, 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 the optimal $\hat{\theta}_{GMM}$ simplifies to:

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

We also 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

[7] For the estimation of the SAE parameters (δ and γ) we use the "generalized moments" procedure of Kelejian and Prucha (1999) that employs the following three moment conditions based on second-order moments of the residuals: $E[\varepsilon'_r \varepsilon_r] = N\sigma_r^2$; $E[\varepsilon'_r C'C\varepsilon_r] = \sigma_r^2 tr(C'C)$; and $E[\varepsilon'_r C\varepsilon_r] = 0$; where $r = 1, 2$ refers to each equation in the model. In practice, these moment conditions are included in both $s(z_{1N}, \theta)$ and $m(z_{2N}, \theta)$, avoiding the need of an instrument; although they are not explicitly shown here to simplify the exposition. We discuss in more detail the implementation of our estimator below.

[8] An appendix with the derivation of the formal asymptotic properties of $\hat{\theta}_{GMM}$ under "high-level" assumptions similar to those in Pinkse and Slade (1998) is available upon request from the authors.

in θ_1 are obtained from (9). In the second step, nonlinear least squares (NLLS) is employed to estimate the parameters in $y_{2i} = \beta_0 + x'_{2i}\beta_1 + \mu\widehat{\lambda}_i + v_{2i}$, in which the parameter γ enters nonlinearly in the "adjusted" IMR ($\widehat{\lambda}_i$). 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 θ_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 (11).

We now discuss other aspects of the implementation of the spatial heckit estimator. Recall that the original spatial probit estimator by Pinkse and Slade (1998) makes use of an instrumental variable (IV) to identify the SAE parameter δ. In this context, where the IV is needed to identify a parameter related to the spatial error dependence and not to instrument for an endogenous variable, it stands to reason that such IV should be correlated with the underlying spatial process. A practical difficulty in general IV estimation is finding relevant IVs. However, in the case of spatial dependence, there are two related sets of available instruments that have been proposed in the context of a linear spatial autoregressive lag (SAL) model in which the dependent variable is spatially lagged and used as a regressor (thus becoming endogenous). The two sets of instruments are optimal in the sense that they approximate the expected value of the endogenous variables. Even though these sets of IVs are for a different model than the SAE, they are related to the spatial process and are exogenous; making them reasonable candidates for use in the present context.

An alternative to the use of IVs to identify the SAE parameter δ is to apply the set of moment conditions proposed in Kelejian and Prucha (1999) to the generalized residuals (8). Recall that these moments are employed in the estimation of γ in the outcome equation using KP-SAE. Given that these moment conditions explicitly involve the SAE parameters, it seems more desirable to employ them over the use of IVs to identify δ and γ. Consequently, one version of the spatial heckit estimator employs such moment conditions and no additional instrumental variables to identify δ and γ.

Noting that the instrumental variables suggested in the context of the SAL model are exogenous, related to the spatial process, and always available, there is the possibility that they can be used to construct additional moment conditions to attain higher asymptotic efficiency. This could be the case if additional useful information is contained in those moment conditions. An open question, of course, is whether the use of such additional instruments results in finite sample improvements in the estimator, given that there may be a tradeoff between the smaller variance that can be achieved and the finite sample bias introduced if the additional instruments are not strong. We explore this issue in the Monte Carlo experiment below by computing two additional versions of the spatial heckit estimator with different sets of instruments: the Kelejian and Prucha (1998) and the Lee (2003) instruments.

The two sets of instruments are motivated as follows. In the context of the SAL model, Kelejian and Prucha (1998) show that the optimal instruments depend on the unknown spatial autoregressive parameter, and thus they propose to approximate them with the linearly independent columns of $[x, Cx, C^2x, ...]$, where x is the set of exogenous independent variables and C is the spatial weighting matrix. Alternatively, Lee (2003) proposes to use the estimated optimal instruments, which, in the context of the selection equation, are given by $(I - \widehat{\delta}C)x$. Kelejian et al. (2004) present simulation evidence about the finite sample performance of these two sets of IVs in the context of the SAL model, concluding that their performance is similar.[9] In our present context, however, it is uncertain which of the two sets of IVs could perform better.

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 (KP-SAE); the traditional heckit estimator that accounts for sample selection but ignores spatial dependence (heckit); 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 (2004) and Kelejian and Prucha (1999)

[9] Kelejian et al. (2004) also introduce series-type IVs in the context of the SAL model, which we do not consider here.

for spatially dependent models.

Our data generating process (DGP) is as follows :

$$y_{1i}^* = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + u_{1i}, \quad u_{1i} = \delta \sum_{j \neq i} c_{ij} u_{1j} + \varepsilon_{1i} \tag{12}$$

$$y_{2i}^* = \beta_0 + \beta_1 x_{3i} + \beta_2 x_{1i} + u_{2i}, \quad u_{2i} = \gamma \sum_{j \neq i} c_{ij} u_{2j} + \varepsilon_{2i}. \tag{13}$$

Each of our models consists of three independent 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 ε_{1i} and ε_{2i} are generated bivariate normal as follows:

$$\begin{bmatrix} \varepsilon_{1i} \\ \varepsilon_{2i} \end{bmatrix} \backsim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right) \tag{14}$$

where we set $\rho = 0.5$ for the correlation between the innovations in each of the two equations.[10] 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 α_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, therefore, 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 four different values: 0, 0.25, 0.5, and 0.75. We gauge the relative performance of three versions of the spatial heckit estimator that differ in the IVs employed: no additional instruments ("No-Inst"), Kelejian and Prucha (1998) instruments ("KP-Inst"),[11] and Lee (2003) instruments ("Lee-Inst"); while the three versions employ the moment conditions from Kelejian and Prucha (1999) to identify the SAE parameters. For the models with $N = 100$ and with $N = 225$ we undertake $1,000$ replications, whereas 500 replications are undertaken for the models with $N = 400$.

[10] The single choice of $\rho = 0.5$ is admitedly arbitrary, but it allows focusing on other important features of the experiment while keeping it manageable. In our design, the coefficient on the (adjusted) IMR is equal to ρ, and if $\rho = 0$ sample selection will not have any consequence, thus we want to avoid this choice of ρ. In a similar model to the sample selection model without SAE, Hartman (1991) found no effect of ρ on the performance of a similar two-step procedure.

[11] More specifically, in the simulations we use the independent columns of $[x, Cx, C^2x]$.

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-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 $\sqrt{5}$.[12] 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 sample selection 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), these tables also show simulation results for the three different versions of our spatial heckit estimator: Spheck No-Inst, Spheck KP-Inst, and Spheck Lee-Inst.[13] 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, both of which typically increase as the amount of sample selection or spatial dependence increases. In addition, the bias of OLS does not decrease as the sample size increases. In general, though, OLS is able to estimate β_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

[12] The number of neighbors varies between 10 and 12, depending on the sample size.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.

[13] The spatial heckit estimators require starting values. Both in the simulations and in the empirical illustration below, we employ starting values that are available in practice. In particular, the staring values employed for all parameters except δ and γ are the heckit parameter estimates. The starting values employed for δ and γ are equal to the KP-SAE estimate of γ.

($k = 1, 2, 3$) are generated independently, and thus there is little effect of the sample selection on the coefficient on x_3 (β_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. The only exceptions are the two models without spatial dependence ($\delta = \gamma = 0$), for which heckit is in fact the correct estimator and it is expected to perform best. Therefore, we expect that the bias and RMSE of the heckit estimator will increase as the amount of spatial dependence increases. This is typically the case with the RMSE of heckit for all three sample sizes. With respect to the bias of heckit, it does not always increase with the amount of spatial dependence in the sample sizes of 100 and 225. However, it typically does increase in the sample size of 400 (Table 3). In addition, the bias also frequently increases as the amount of sample selection increases for all three sample sizes.

Compared to OLS, the heckit estimator shows a great improvement, even though it is also inconsistent in theory (except when $\delta = \gamma = 0$). While the average bias and RMSE of β_1 is very small and fairly comparable to that of OLS, the other two coefficients (β_0 and β_2) have smaller bias: in the case of β_2, the bias has the interpretation of percentage and ranges from 0.5 to 9, 1.8 to 4.4 and 1.3 to 3.2 percent for sample sizes 100, 225 and 400, respectively. In the case of β_0 the bias (not interpreted as percentage) ranges from -0.08 to 0.09, -0.076 to 0.028 and -0.019 to 0.088 for sample sizes 100, 225 and 400, respectively. Thus, the range of the bias on β_2 decreases while that of β_0 initially decreases but then appears to increase across model specifications as the sample size increases. Interestingly, the bias of these two coefficients is of opposite sign compared to the bias in OLS, with only a few exceptions when spatial dependence is highest. Finally, the RMSE of the heckit estimator for the coefficients β_0 and β_2 is larger than that of OLS (even when $\delta = \gamma = 0$), except in the models with the largest sample size. This is expected as the non-linear heckit model is a more demanding estimation technique. This will also be true of the spatial heckit estimator below.

The third estimator reported is KP-SAE that accounts for spatial dependence but ignores the sample selection feature of the data, which results in inconsistent parameter estimates in all model specifications. In agreement with this notion, the bias and RMSE increase as the amount of sample selection increases, but they also increase substantially as the amount of spatial dependence increases. This perhaps reflects that spatial dependence is not entirely accounted for in KP-SAE due to sample selection. Compared to the previous two estimators, KP-SAE shows a bias in all of the coefficients that is slightly higher than that of the OLS estimator, which is significantly larger than that of the heckit estimator (with the exception of the bias on β_1). With respect to RMSE, KP-SAE is fairly comparable to OLS when spatial dependence is 0 or 0.25. However, the RMSE of KP-SAE deteriorates substantially when spatial dependence increases to 0.5 and 0.75. As a result, compared to the RMSE of heckit, the RMSE of KP-SAE is typically smaller in the two cases of low spatial dependence (except when $N = 400$), but larger in the two cases of high spatial dependence.

KP-SAE is the first estimator that produces an estimate of the SAE parameter γ. The bias of γ in KP-SAE typically decreases in absolute terms with the amount of spatial dependence and sample selection (except when $N = 100$); while the RMSE tends to decrease with the level of spatial dependence and no clear pattern emerges with respect to the amount of sample selection. In relative terms, the bias of γ ranges from 16.2 to 76.8 percent of the true γ in the models with sample size of 100 (average of -0.135 in the models with $\delta = \gamma = 0$), while in the models with sample size of 225 and 400 the bias ranges from 0.93 to 33.2 (average of -0.93 in the models with $\delta = \gamma = 0$) and 2 to 19.6 (average of -0.45 in the models with $\delta = \gamma = 0$) percent of the true γ, respectively. Therefore, the bias in the KP-SAE estimate of γ appears to decrease as the sample size increases. Finally, the RMSE of γ becomes smaller as more observations are available, as it ranges from 0.252 to 0.414, 0.141 to 0.327 and 0.118 to 0.251 for sample sizes 100, 225 and 400 respectively. Similar patterns are also found for γ in the spatial heckit estimators below.

The last three estimators in Tables 1 through 3 are the versions of the spatial heckit estimator we consider: Spheck No-Inst, Spheck KP-Inst and Spheck Lee-Inst, which are consistent for all parameters across model specifications. Reporting these three different versions allows the comparison of spatial heckit estimators using different sets of moment conditions (instruments), which were discussed in Section 3.

We start by analyzing the smallest sample size ($N = 100$) shown in Table 1. All three Spheck estimators for the parameters β_0, β_1 and β_2 possess significantly smaller bias than OLS and KP-SAE, but typically larger bias than heckit;[14] although the bias of the Spheck estimators is fairly similar to that of heckit in the models with the highest spatial dependence

[14] Recall that the heckit estimator is the correct model under no spatial dependence ($\delta = \gamma = 0$).

(0.75). In terms of the RMSE of β_0 and β_2, however, all three Spheck estimators typically have the highest among all estimators across all models. Recall, however, that just as was pointed out for heckit, Spheck is a more demanding estimation technique, and in Table 1, even though $N = 100$, only about 75 (25% selection) or 60 (40% selection) observations are available to estimate the outcome equation. This small number of observations likely causes the Spheck estimators to have higher variance compared to the other estimators. Finally, in terms of β_1, the Spheck estimators have low bias and RMSE, comparable to that of other estimators.

Comparing the KP-SAE estimate of γ with that of the three Spheck estimators, the latter estimators have smaller bias and RMSE than the former in all models, except in those with no spatial dependence ($\gamma = 0$), in which their RMSE is only slightly larger. Finally, comparing the three versions of Spheck using $N = 100$, there is no clear favorite in terms of bias as the three are very similar in this regard. In terms of RMSE, however, Spheck KP-Inst typically dominates the other two, albeit by a slight margin.

As the sample size is increased to 225 observations in Table 2, the performance of the Spheck estimates of β_0 and β_2 improve in terms of bias, as expected, possessing smaller bias than OLS and KP-SAE and comparable bias to that of heckit, especially in models with low sample selection and high spatial dependence, although in some of the other models the bias of the Spheck estimators can be substantially larger than that of heckit (especially when $\delta = \gamma = 0$). At the same time, the bias on β_1 of the Spheck estimators is typically smaller than that of the other estimators. In terms of RMSE of β_0 and β_2, the Spheck estimators experience a great improvement over the smaller sample size, now having more comparable RMSE to the other estimators, especially when spatial dependence is an important feature of the model; however, in a few instances the RMSE of the Spheck estimators is still sizable.

Regarding the estimate of γ in this intermediate sample size, both the KP-SAE and Spheck estimators show similar bias and RMSE although, as we discussed above, the Spheck estimators outperform KP-SAE with respect to the bias and RMSE on all other coefficient estimates. Finally, the results using $N = 225$ suggest that Spheck Lee-Inst performs slightly better than the other two Spheck estimators in terms of both bias and RMSE, followed by Spheck KP-Inst. The trend in terms of RMSE may not be surprising since, as discussed in Section 3, both Spheck Lee-Inst and Spheck KP-Inst use potentially useful information in the form of instruments that can result in smaller (asymptotic) variance. In any case, the results of the three estimators are typically very close to each other.

Table 3 presents simulation results for a sample size of $N = 400$. In terms of bias of β_0 and β_2, the Spheck estimators outperform OLS and KP-SAE by far, plus they become fairly comparable to the heckit estimator, outperforming it in many cases. Similarly, in terms of RMSE, and in contrast to the simulations with smaller sample sizes, the Speck estimators now have similar RMSE to the other estimators considered for all three parameters β_0, β_1, and β_2; even performing best in this respect in a few cases. This improvement in the Spheck estimators may be attributed to the larger sample size ($N = 400$), which indicates that in this case the finite sample properties of the spatial heckit estimator are reasonably good for a sample as small as 400 observations, even when only 75 or 60 percent of those observations are available to estimate the outcome equation.

In terms of the spatial autoregressive parameter γ, the Spheck and KP-SAE estimates show again very similar bias and RMSE, with typically slightly higher values for the Spheck estimators. Finally, comparing the performance of the three Spheck estimators in the models with $N = 400$ corroborates the previously noted trend that the three estimators perform fairly similarly in terms of bias and RMSE. If at all, there is only a slight improvement in bias from using Spheck No-Inst, and a slight improvement in RMSE from using either Spheck KP-Inst or Spheck Lee-Inst.

Table 4 presents simulation results for the selection equation. This equation is only estimated in the heckit (using a probit model) and the three versions of Spheck. However, given their similar performance and to save space, only the results for Spheck KP-Inst are presented.[15] 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. Nonetheless, the value of the constant term is different in each case.

It is evident from Table 4 that the heckit estimator typically outperforms the Spheck KP-Inst estimator in terms of both bias and RMSE. However, the RMSE of the Spheck KP-Inst estimator is comparable to that of the heckit model even in a sample as small as 225 observations, and even in the models with no spatial dependence, where the heckit estimator is the correct model. For the sample of $N = 400$, the performance in terms of both bias and RMSE becomes very similar between Spheck KP-Inst and heckit. In addition,

[15] The results for the selection equation for the Spheck No-Inst and Spheck Lee-Inst estimators are available upon request. In summary, Spheck KP-Inst outperforms the other two estimators, albeit by a small margin; however, the relatively better performance by Spheck KP-Inst erodes as the sample size increases. For $N = 400$, the performance of the three Spheck estimators is almost identical in the selection equation.

the improvement of the Spheck KP-Inst estimator as the sample size increases is larger than the one observed in the heckit model. Finally, it should be noted that the estimate of the spatial parameter δ by the Spheck KP-Inst estimator is estimated with relatively low bias and RMSE across model specifications, especially in the models with sample size 225 and 400, and when there is substantial spatial dependence.

In summary, we regard the simulation results as encouraging with respect to the finite sample properties of the three versions of the spatial heckit estimator we consider here. 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 the three versions considered of the spatial heckit estimator yield similar results with only slight improvements in RMSE in the parameters that correspond to the outcome equation when using additional instruments (Spheck KP-Inst or Spheck Lee-Inst); while for the parameters in the selection equation the difference among the three estimators becomes negligible as the sample size increases from 100 to 400.

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 error 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 error 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 spatial 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 (Wilen, 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 non-spatially defined production within a fishery in an effort to determine the factors that explain deviations from the production frontier (Kirkley et al. 1995, 1998; Squires and Kirkley 1999; Pascoe and Coglan 2002). 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. 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. Estimating spatial efficiency is beyond the scope of this application, but investigating spatial production is a necessary first step in the process of facilitating fisheries policy.

Determining the spatial rates of production requires a very fine spatial resolution of data. 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 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 207, 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 production using this data set. Should one decide to refine the analysis by focusing on inter and intra vessel differences in the spatial distribution of production, 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 production 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 NMFS data. Pacific cod is targeted in the Alaskan groundfish fishery. Estimating the fleets' spatial production 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 production 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.

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.[16] This measure is the average of all vessels that fished within this region of like vessel characteristics. Vessels were grouped according to the size of vessel, gear utilized and type of vessel (catcher-processor vs. catcher-vessel). 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 their spatial production. Indeed, the Moran-I test statistic using the OLS residuals soundly rejects the null hypothesis of zero spatial autocorrelation on the data with a value of 16.1 (p-value of 0.00).[17]

To allow comparison among the four estimators employed in the Monte Carlo experiment in the previous section, the spatial production model 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 estimator with SAE dependence is likely more appropriate than the other estimators. 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

[16] For simplicity, we concentrate on the catch of Pacific cod only and ignore all other species caught.

[17] We also computed a Moran I test statistic using the heckit residuals following Kelejian and Prucha (2001). Surprisingly, in this case the null hypothesis of zero spatial autocorrelation cannot be rejected with a p-value of 0.9. On the contrary, a Moran I test statistic using the probit residuals (selection equation) strongly rejects the null of zero spatial autocorrelation with a p-value of 0.001.

reporting location but not the amount of "hauls" that will be conducted.

In order to determine the spatial weighting matrix, we use the following common specification (also used in the previous section) to assign spatial weights among the statistical reporting units we use as locations: $c_{ij} = \frac{1}{d_{ij}^{f}}$, where c_{ij} is the spatial weight assigned to the distance between location i and location j, d_{ij} is the Euclidian distance between locations i and j, and f is a "friction" parameter.[18] To control the number of neighbors per statistical reporting unit a band is chosen. Finally, 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. We use a band of 7 and a friction parameter of 2 in the estimations below. The results from all estimators for the outcome equation are presented in Table 5.[19]

The estimated coefficients from each of the models are somewhat similar in sign and magnitude, although there are a number of important differences. First, the OLS and KP-SAE estimators often differ in magnitude, sign and statistical significance compared to the estimates yield by the Spheck estimators. Second, the Spheck estimators and heckit agree to some extent in the magnitude of the estimates in most coefficients, although not always in their statistical significance. Third, the three Spheck estimators (No-Inst, KP-Inst, and Lee-Inst) are very similar in the magnitude of their estimated coefficients and for the most part in their statistical significance; although in general Spheck KP-Inst and Spheck Lee-Inst seem to achieve smaller standard errors than Spheck No-Inst. Fourth, the estimate of the SAE parameter (γ) is high and agrees across the KP-SAE and Spheck estimators. All estimates of γ are statistically significant but more so the Spheck estimates. Importantly, all these features of the outcome equation are largely in agreement with the simulation results described in the previous section.

Interestingly, despite the relatively high amount of selection in the sample (35%), the IMR is not statistically significant in all but one of the estimators (Spheck KP-Inst), although it is estimated to be positive in all four estimators; and is of about the same magnitude in the Spheck estimators but considerably larger in heckit. Comparing some of the Spheck and heckit estimates, even though they are similar in magnitude, the statistical significance of

[18] The spatial weighting matrix was constructed by superimposing the ADF&G on to the $X-Y$ coordinate plane.

[19] We note that the numerical optimizations needed to estimate the spatial heckit make it computationally intensive relative to the other three estimators. In the currrent application, it takes about 35 minutes to compute the spatial heckit estimators in a computer with a Pentium M processor at 1.6 GHz with 496 MB of RAM.

some of the Spheck coefficients relative to heckit agrees with expectations for these data. For instance, in the case of "Max. Depth", it is expected that vessels fishing in deeper areas would obtain a larger CPUE. Similarly, it is also expected that areas with high "Biomass" signal would attain more productivity per haul executed, if such signal is accurate. Finally, "Dum HAL" is one of the more productive technologies (per haul executed) used to harvest Pacific cod in the Bering Sea (in 1997). Summarizing the results for the outcome equation in this empirical illustration, the Spheck estimators yield more sensible results than OLS, heckit, and KP-SAE; while at the same time they achieve smaller standard errors, especially Spheck KP-Inst and Spheck Lee-Inst.

Table 6 presents the estimated coefficients for the selection equation using the heckit and the spatial heckit estimators. Once again, the estimators yield somewhat similar parameter estimates albeit with some important differences. For instance, the estimated coefficient on "Max. Depth" in heckit is smaller than that of the Spheck estimators and is marginally statistically significant; while the estimated coefficient on "Min. Depth" is smaller in absolute value in the Spheck estimators and marginally statistically significant in one instance (Spheck KP-Inst) compared to heckit. Similarly, the coefficient on "Dum HAL" is larger in the Spheck estimators while being statistically significant in both Spheck (except No-Inst) and heckit.

Comparing the Spheck estimators in the selection equation, all three yield very similar results, except in a couple of coefficients in which Spheck No-Inst comes up with relatively high standard errors that result in statistically insignificant coefficients: "Dum CV" and "Dum HAL", although the magnitude is always similar. This may suggest that, in practice, using extra instruments may be more important for the precision of the estimated coefficients in the selection equation. Finally, all three Spheck estimators yield values of the SAE parameter (δ) that are high (0.82 to 0.91) and highly statistically significant, as expected.

Although a more complete analysis would be required before any concrete policy recommendations are made from this exercise, they do suggest that there exists some degree of heterogeneity in the spatial production 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).

In summary, the results from this empirical illustration of our methodology 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 a sample selection model with spatial autoregressive errors (SAE). The method of estimation 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, 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). It is also noted the availability of instrumental variables that can be used to develop additional moment conditions that can potentially result in higher efficiency of the spatial heckit estimator.

We explore the properties of the spatial heckit for the model with SAE dependence by stating 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 linear spatial models without sample selection. However, our estimator still compares favorably in this respect with other estimation methods for spatial sample selection models that would require approximation of multidimensional integrals.

References

[1] Amemiya, T. 1985. *Advanced Econometrics*. Cambridge, Mass.: Harvard University Press.

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

[3] Anselin, L. 2001. Spatial Econometrics. In B. H. Baltagi (ed.), *A Companion to Theoretical Econometrics,* Oxford: Blackwell.

[4] 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.

[5] 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.

[6] Beron, K. and W. Vijverberg. 2004. 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*. Berlin: Springer-Verlag.

[7] 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.

[8] Case, A.C. 1991. Spatial Patterns in Household Demand. *Econometrica*. 59: 953-965.

[9] Case, A.C. 1992. Neighborhood Influence and Technology Change. *Regional Sciences and Urban Economics*. 22: 491-508.

[10] Casetti, E. 1972. Generating Models by the Expansion Method: Applications to Geographical Research. *Geographical Analysis*. 4: 81-91.

[11] 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.

[12] Fishback, P.V., W.C. Horrace and S. Kantor. 2006. The Impact of New Deal Expenditures on Mobility During the Great Depression. *Explorations in Economic History*, 43:179-222.

[13] Fleming, Mark. 2004. 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*. Berlin: Springer-Verlag.

[14] Goodman, A.C. 1988. An Econometric Model of Housing Price, Permanent Income, Tenure Choice, and Housing Demand. *Journal of Urban Economics*. 23: 327-53.

[15] Hartman, R.S. 1991. A Monte Carlo Analysis of Alternative Estimators in Models Involving Selectivity. *Journal of Business and Economic Statistics*, 9: 41-49.

[16] 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.

[17] Heckman, J. 1979. Sample Selection Bias as a Specification Error. *Econometrica*. 47: 153-61.

[18] 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.

[19] 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.

[20] 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.

[21] Kelejian, H.H., I.R. Prucha and Y. Yuzefovich. 2004. Instrumental Variable Estimation of a Spatial Autoregressive Model with Autoregressive Disturbances: Large and Small Sample Results. In LeSage, J.P. and R.K. Pace (eds.), *Advances in Econometrics: Spatial and Spatio-Temporal Econometrics*. Oxford: Elsevier.

[22] 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.

[23] Kirkley, J., D. Squires and I.E. Strand. 1998. Characterizing Managerial Skill and Technical Efficiency. *Journal of Productivity Analysis* 9:145-160.

[24] Lee, L.F. 2001. Generalized Method of Moments Estimation of Spatial Autoregressive Processes. Working paper, Ohio State University.

[25] Lee, L.F. 2003. Best Spatial Two-Stage Least Squares Estimation for a Spatial Autoregressive Model with Autoregressive Disturbances. *Econometric Reviews*, 22: 307-335.

[26] Lee, L.F. 2004. Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. *Econometrica*, 72: 1899-1925.

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

[28] Leung, S.F and S. Yu. 1996. On the Choice Between Sample Selection and Two-Part Models. *Journal of Econometrics*. 72: 197-229.

[29] Maddala, G.S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press.

[30] McMillen, D.P. 1992. Probit with Spatial Autocorrelation. *Journal of Regional Science*. 32: 335-48.

[31] McMillen, D.P. 1993. Can Blacks Earn More in the Suburbs? Racial Differences in Intrametropolitan Earnings Variation. *Journal of Urban Economics*. 33: 167-88.

[32] McMillen, D.P. 1995. Selection Bias in Spatial Econometric Models. *Journal of Regional Science*. 35: 417-36.

[33] Montgomery, C. 1992. Explaining Home Improvement in the Context of Household Investment in Residential Housing. *Journal of Urban Economics*. 32: 326-50.

[34] Newey, W. 1984. A Method of Moments Interpretation of Sequential Estimators. *Economics Letters*. 14: 201-206.

[35] Ord, J.K. 1975. Estimation Methods for Models of Spatial Interaction. *Journal of the American Statistical Association* 70:120-126.

[36] Pascoe, S. and L. Coglan. 2002. The Contribution of Unmeasureable Inputs to Fisheries Production: An Analysis of Technical Efficiency of Fishing Vessels in the English Channel. *American Journal of Agricultural Economics* 84:585-597.

[37] 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.

[38] Squires, D. and J. Kirkley. 1999. Skipper skill and panel data in fishing industries. *Can. J. Fish. Aquat. Sci.* 56:2011-2018.

[39] Topa, G. 2001. Social Interactions, Local Spillovers, and Unemployment. *Review of Economic Studies*. 2: 133-138.

[40] Wilen, J. 2004. Spatial Management of Fisheries. *Marine Resource Economics* 19:7-20.

Table 1. Simulation Results for N=100

(sel, δ=γ)		BIAS						RMSE					
		OLS	Heckit	KP-SAE	Spheck No-Inst	Spheck KP-Inst	Spheck Lee-Inst	OLS	Heckit	KP-SAE	Spheck No-Inst	Spheck KP-Inst	Spheck Lee-Inst
(25%,0)	β_0	0.324	-0.042	0.321	-0.167	-0.194	-0.148	0.459	1.155	0.467	3.080	2.984	2.989
	β_1	0.034	-0.003	0.034	-0.005	-0.001	-0.004	0.398	0.401	0.408	0.432	0.428	0.432
	β_2	-0.264	0.026	-0.258	0.123	0.147	0.123	0.493	0.921	0.514	2.296	2.187	2.257
	γ			-0.136	-0.161	-0.159	-0.164			0.395	0.462	0.458	0.464
(25%,0.25)	β_0	0.315	-0.057	0.308	-0.145	-0.119	-0.095	0.461	1.041	0.492	3.016	2.741	2.886
	β_1	0.040	0.002	0.046	0.002	0.004	0.002	0.395	0.396	0.440	0.422	0.419	0.421
	β_2	-0.253	0.043	-0.245	0.127	0.107	0.093	0.491	0.895	0.562	2.037	1.837	1.852
	γ			-0.123	-0.113	-0.115	-0.115			0.387	0.439	0.435	0.439
(25%,0.5)	β_0	0.333	-0.040	0.350	-0.296	-0.280	-0.365	0.500	1.028	0.704	2.653	2.512	3.249
	β_1	0.035	-0.002	0.041	0.004	0.005	0.003	0.400	0.403	0.670	0.409	0.407	0.409
	β_2	-0.261	0.023	-0.292	0.234	0.217	0.267	0.519	0.860	0.911	1.957	1.826	2.115
	γ			-0.081	-0.047	-0.054	-0.051			0.332	0.365	0.361	0.363
(25%,0.75)	β_0	0.401	0.099	0.490	-0.109	-0.162	-0.162	0.660	1.242	1.536	2.941	2.718	3.098
	β_1	0.030	-0.002	0.031	0.009	0.013	0.012	0.448	0.450	1.826	0.426	0.422	0.425
	β_2	-0.300	-0.090	-0.448	0.111	0.133	0.144	0.599	0.953	2.185	2.046	1.982	2.127
	γ			-0.072	-0.030	-0.038	-0.035			0.252	0.264	0.259	0.261
(40%,0)	β_0	0.466	-0.075	0.471	-0.202	-0.275	-0.199	0.601	1.389	0.626	3.390	3.766	3.343
	β_1	0.037	-0.010	0.031	-0.008	-0.006	-0.008	0.450	0.454	0.487	0.484	0.484	0.486
	β_2	-0.331	0.060	-0.332	0.191	0.232	0.187	0.577	1.086	0.612	2.373	2.505	2.278
	γ			-0.135	-0.171	-0.170	-0.171			0.396	0.485	0.481	0.484
(40%,0.25)	β_0	0.468	-0.080	0.468	-0.099	-0.045	-0.079	0.606	1.367	0.643	3.229	2.275	3.031
	β_1	0.037	-0.009	0.038	-0.013	-0.013	-0.012	0.446	0.452	0.504	0.477	0.474	0.477
	β_2	-0.327	0.061	-0.326	0.100	0.070	0.085	0.578	1.078	0.645	2.106	1.697	1.994
	γ			-0.192	-0.190	-0.191	-0.193			0.414	0.477	0.477	0.478
(40%,0.5)	β_0	0.482	-0.055	0.494	-0.312	-0.318	-0.322	0.641	1.574	0.792	3.200	3.356	3.295
	β_1	0.041	-0.006	0.046	-0.003	0.000	-0.002	0.459	0.465	0.661	0.476	0.474	0.476
	β_2	-0.327	0.054	-0.346	0.278	0.267	0.279	0.596	1.174	0.901	2.400	2.378	2.411
	γ			-0.178	-0.142	-0.146	-0.145			0.378	0.416	0.414	0.415
(40%,0.75)	β_0	0.549	0.052	0.645	-0.077	-0.103	-0.087	0.798	1.468	1.568	2.772	2.722	2.746
	β_1	0.030	-0.014	-0.012	0.000	0.005	0.003	0.509	0.516	1.834	0.507	0.502	0.504
	β_2	-0.312	0.005	-0.430	0.112	0.124	0.121	0.665	1.083	1.930	1.948	1.940	1.976
	γ			-0.137	-0.077	-0.085	-0.083			0.306	0.310	0.310	0.311

Note: Simulation results are based on 1000 replications.

Table 2. Simulation Results for N=225

(sel, δ=γ)		BIAS OLS	BIAS Heckit	BIAS KP-SAE	BIAS Spheck No-Inst	BIAS Spheck KP-Inst	BIAS Spheck Lee-Inst	RMSE OLS	RMSE Heckit	RMSE KP-SAE	RMSE Spheck No-Inst	RMSE Spheck KP-Inst	RMSE Spheck Lee-Inst
(25%,0)	β_0	0.327	-0.052	0.324	-0.083	-0.081	-0.064	0.382	0.549	0.383	0.919	0.862	0.730
	β_1	0.003	-0.008	0.004	-0.008	-0.008	-0.008	0.257	0.258	0.261	0.275	0.275	0.274
	β_2	-0.244	0.041	-0.241	0.092	0.090	0.083	0.347	0.469	0.350	0.715	0.697	0.655
	γ			-0.102	-0.113	-0.113	-0.113			0.327	0.362	0.360	0.361
(25%,0.25)	β_0	0.330	-0.036	0.331	-0.033	-0.027	-0.027	0.390	0.530	0.407	0.586	0.584	0.583
	β_1	0.002	-0.009	0.005	-0.008	-0.008	-0.008	0.260	0.261	0.300	0.272	0.271	0.271
	β_2	-0.246	0.028	-0.250	0.061	0.056	0.057	0.352	0.454	0.386	0.546	0.549	0.547
	γ			-0.030	-0.021	-0.023	-0.023			0.273	0.292	0.290	0.290
(25%,0.5)	β_0	0.347	-0.033	0.370	-0.066	-0.056	-0.033	0.415	0.566	0.563	1.139	1.083	0.812
	β_1	0.003	-0.008	0.010	-0.002	-0.002	-0.002	0.272	0.274	0.583	0.270	0.269	0.269
	β_2	-0.250	0.025	-0.293	0.089	0.084	0.070	0.363	0.475	0.626	0.807	0.795	0.661
	γ			0.043	0.063	0.058	0.060			0.215	0.230	0.226	0.226
(25%,0.75)	β_0	0.404	0.028	0.604	-0.041	-0.035	-0.027	0.518	0.660	1.710	0.932	0.942	0.905
	β_1	0.001	-0.009	-0.056	0.000	0.000	0.000	0.311	0.311	2.084	0.278	0.279	0.278
	β_2	-0.261	-0.019	-0.555	0.080	0.074	0.070	0.400	0.486	2.264	0.664	0.672	0.648
	γ			0.066	0.095	0.089	0.090			0.141	0.158	0.154	0.154
(40%,0)	β_0	0.475	-0.066	0.475	-0.173	-0.175	-0.161	0.526	0.729	0.527	1.401	1.495	1.292
	β_1	0.007	-0.008	0.007	-0.007	-0.007	-0.007	0.287	0.288	0.291	0.307	0.307	0.307
	β_2	-0.316	0.042	-0.316	0.140	0.140	0.136	0.423	0.537	0.427	1.062	1.079	0.980
	γ			-0.084	-0.097	-0.096	-0.097			0.295	0.337	0.335	0.337
(40%,0.25)	β_0	0.478	-0.076	0.481	-0.149	-0.119	-0.151	0.534	0.743	0.547	1.269	0.996	1.089
	β_1	0.010	-0.005	0.016	-0.005	-0.005	-0.005	0.290	0.290	0.327	0.304	0.303	0.304
	β_2	-0.318	0.044	-0.329	0.125	0.107	0.128	0.429	0.543	0.462	0.929	0.773	0.808
	γ			-0.083	-0.074	-0.075	-0.075			0.274	0.299	0.297	0.297
(40%,0.5)	β_0	0.492	-0.069	0.516	-0.142	-0.108	-0.104	0.556	0.740	0.672	1.293	1.183	0.946
	β_1	0.010	-0.004	-0.005	-0.002	-0.003	-0.002	0.299	0.299	0.949	0.302	0.301	0.301
	β_2	-0.313	0.043	-0.344	0.136	0.111	0.115	0.438	0.549	0.730	0.886	0.792	0.731
	γ			-0.044	-0.015	-0.019	-0.018			0.224	0.236	0.234	0.235
(40%,0.75)	β_0	0.559	-0.022	0.733	-0.093	-0.114	-0.068	0.666	0.888	1.645	1.399	1.419	1.356
	β_1	-0.004	-0.018	-0.078	-0.012	-0.015	-0.013	0.338	0.339	1.916	0.310	0.310	0.309
	β_2	-0.321	0.018	-0.545	0.122	0.131	0.111	0.474	0.601	2.047	0.912	0.918	0.909
	γ			0.007	0.050	0.044	0.045			0.151	0.161	0.158	0.159

Note: Simulation results are based on 1000 replications.

Table 3. Simulation Results for N=400

(sel, δ=γ)		OLS	Heckit	KP-SAE	Spheck No-Inst	Spheck KP-Inst	Spheck Lee-Inst	OLS	Heckit	KP-SAE	Spheck No-Inst	Spheck KP-Inst	Spheck Lee-Inst
(25%,0)	β_0	0.318	-0.025	0.316	-0.016	-0.021	-0.018	0.351	0.350	0.351	0.336	0.339	0.340
	β_1	0.011	0.001	0.012	-0.004	-0.004	-0.004	0.193	0.192	0.196	0.205	0.205	0.205
	β_2	-0.232	0.018	-0.229	0.037	0.042	0.041	0.299	0.304	0.299	0.318	0.322	0.322
	γ			-0.047	-0.050	-0.050	-0.050			0.251	0.259	0.258	0.257
(25%,0.25)	β_0	0.328	-0.023	0.329	-0.008	-0.009	-0.012	0.361	0.364	0.371	0.344	0.345	0.347
	β_1	0.006	-0.004	0.011	-0.002	-0.003	-0.003	0.195	0.195	0.227	0.201	0.201	0.201
	β_2	-0.242	0.013	-0.246	0.031	0.034	0.036	0.307	0.311	0.325	0.326	0.327	0.328
	γ			0.025	0.036	0.034	0.034			0.202	0.210	0.209	0.208
(25%,0.5)	β_0	0.346	-0.024	0.380	-0.008	-0.013	-0.009	0.385	0.381	0.517	0.444	0.499	0.447
	β_1	0.000	-0.009	0.014	-0.007	-0.008	-0.007	0.201	0.201	0.495	0.202	0.202	0.202
	β_2	-0.241	0.017	-0.309	0.045	0.048	0.045	0.312	0.314	0.521	0.381	0.411	0.387
	γ			0.091	0.103	0.104	0.108			0.181	0.191	0.192	0.194
(25%,0.75)	β_0	0.393	0.040	0.477	0.015	-0.009	-0.004	0.465	0.468	2.976	0.506	0.554	0.528
	β_1	-0.002	-0.011	0.079	-0.005	-0.003	-0.004	0.225	0.225	4.399	0.209	0.209	0.209
	β_2	-0.243	-0.023	-0.422	0.041	0.055	0.053	0.333	0.333	4.694	0.355	0.395	0.377
	γ			0.123	0.138	0.134	0.133			0.154	0.162	0.158	0.158
(40%,0)	β_0	0.456	-0.036	0.454	-0.135	-0.119	-0.119	0.484	0.421	0.483	0.573	0.488	0.492
	β_1	0.017	0.006	0.020	-0.002	-0.003	-0.002	0.214	0.213	0.215	0.218	0.219	0.219
	β_2	-0.292	0.023	-0.291	0.108	0.096	0.099	0.361	0.341	0.363	0.451	0.395	0.400
	γ			-0.043	-0.042	-0.042	-0.042			0.232	0.242	0.240	0.240
(40%,0.25)	β_0	0.462	-0.046	0.462	-0.108	-0.104	-0.119	0.491	0.441	0.496	0.491	0.488	0.506
	β_1	0.012	0.001	0.018	-0.003	-0.003	-0.004	0.215	0.214	0.236	0.221	0.221	0.220
	β_2	-0.291	0.030	-0.293	0.099	0.096	0.108	0.361	0.345	0.376	0.400	0.397	0.408
	γ			-0.048	-0.038	-0.038	-0.038			0.213	0.219	0.219	0.218
(40%,0.5)	β_0	0.480	-0.043	0.508	-0.069	-0.067	-0.084	0.513	0.479	0.591	0.492	0.481	0.502
	β_1	0.010	-0.002	0.017	-0.001	-0.001	-0.001	0.222	0.220	0.403	0.221	0.221	0.222
	β_2	-0.293	0.029	-0.342	0.086	0.084	0.096	0.366	0.363	0.515	0.394	0.387	0.400
	γ			-0.004	0.018	0.015	0.015			0.172	0.177	0.176	0.176
(40%,0.75)	β_0	0.527	-0.063	0.789	-0.047	-0.047	-0.056	0.595	0.610	1.829	0.612	0.661	0.661
	β_1	0.017	0.005	-0.069	0.009	0.009	0.009	0.262	0.260	2.253	0.238	0.238	0.238
	β_2	-0.293	0.045	-0.667	0.097	0.097	0.104	0.389	0.400	2.196	0.418	0.458	0.455
	γ			0.054	0.094	0.089	0.084			0.121	0.144	0.141	0.160

Note: Simulation results are based on 500 replications.

Table 4. Simulation Results for the Selection Equation

		N=100				N=225				N=400			
		BIAS		RMSE		BIAS		RMSE		BIAS		RMSE	
(sel, δ=γ)		**Heckit**	**Spheck KP-Inst**	**Heckit**	**Spheck KP-Inst**	**Heckit**	**Spheck KP-Inst**	**Heckit**	**Spheck KP-Inst**	**Heckit**	**Spheck KP-Inst**	**Heckit**	**Spheck KP-Inst**
(25%,0)	α_0	-0.020	-0.020	0.349	0.419	-0.021	-0.039	0.240	0.282	-0.009	-0.013	0.174	0.199
	α_1	0.030	0.193	0.561	0.690	0.019	0.192	0.362	0.460	0.004	0.166	0.254	0.342
	α_2	0.051	0.199	0.523	0.682	0.030	0.198	0.332	0.445	0.013	0.166	0.256	0.331
	δ		-0.220		0.533		-0.128		0.394		-0.023		0.267
(25%,0.25)	α_0	-0.017	-0.007	0.342	0.417	-0.018	-0.016	0.243	0.282	-0.010	0.002	0.174	0.209
	α_1	0.022	0.188	0.537	0.660	0.012	0.196	0.368	0.467	0.006	0.181	0.252	0.345
	α_2	0.045	0.222	0.508	0.673	0.028	0.213	0.327	0.444	0.007	0.189	0.250	0.346
	δ		-0.200		0.485		-0.091		0.324		0.002		0.202
(25%,0.5)	α_0	-0.013	-0.004	0.361	0.462	-0.008	0.000	0.252	0.310	0.007	0.024	0.188	0.242
	α_1	-0.012	0.199	0.561	0.706	-0.039	0.196	0.363	0.474	-0.043	0.188	0.248	0.349
	α_2	0.032	0.256	0.502	0.701	-0.003	0.231	0.320	0.464	-0.032	0.202	0.248	0.369
	δ		-0.148		0.383		-0.053		0.230		0.008		0.143
(25%,0.75)	α_0	-0.020	-0.075	0.439	0.729	-0.004	-0.059	0.299	0.451	0.011	-0.022	0.236	0.343
	α_1	-0.109	0.251	0.565	0.914	-0.166	0.200	0.366	0.527	-0.152	0.189	0.285	0.369
	α_2	-0.042	0.294	0.499	0.887	-0.093	0.251	0.322	0.567	-0.111	0.194	0.264	0.403
	δ		-0.124		0.261		-0.042		0.129		-0.010		0.081
(40%,0)	α_0	-0.034	-0.113	0.337	0.413	-0.022	-0.094	0.227	0.278	-0.010	-0.065	0.169	0.198
	α_1	0.045	0.174	0.518	0.625	0.016	0.137	0.314	0.378	0.003	0.104	0.237	0.284
	α_2	0.042	0.163	0.477	0.586	0.022	0.139	0.305	0.378	0.014	0.111	0.227	0.277
	δ		-0.279		0.568		-0.152		0.409		-0.084		0.303
(40%,0.25)	α_0	-0.019	-0.112	0.343	0.440	-0.012	-0.091	0.234	0.285	0.003	-0.066	0.167	0.204
	α_1	0.032	0.173	0.518	0.633	0.007	0.149	0.320	0.391	-0.008	0.123	0.237	0.292
	α_2	0.032	0.197	0.467	0.700	0.010	0.152	0.309	0.390	-0.006	0.127	0.231	0.290
	δ		-0.194		0.508		-0.055		0.321		0.014		0.242
(40%,0.5)	α_0	0.006	-0.152	0.364	0.531	0.022	-0.117	0.252	0.343	0.027	-0.107	0.182	0.256
	α_1	-0.008	0.216	0.512	0.675	-0.033	0.196	0.321	0.434	-0.040	0.181	0.237	0.332
	α_2	0.007	0.259	0.454	0.743	-0.034	0.193	0.307	0.428	-0.033	0.196	0.232	0.340
	δ		-0.097		0.381		0.025		0.215		0.076		0.168
(40%,0.75)	α_0	0.122	-0.173	0.451	0.831	0.102	-0.177	0.316	0.524	0.098	-0.181	0.251	0.401
	α_1	-0.116	0.298	0.545	1.102	-0.155	0.239	0.345	0.521	-0.126	0.262	0.259	0.406
	α_2	-0.092	0.317	0.448	0.870	-0.127	0.239	0.326	0.550	-0.141	0.231	0.270	0.393
	δ		-0.072		0.238		0.010		0.106		0.035		0.073

Note: Simulation results are based on 1000 replications for N=100 and N=225, and 500 for N=400.

Table 5. Estimated Coefficients for the Outcome Equation

	OLS	Heckit	KP-SAE[1]	Spheck No-Inst	Spheck KP-Inst	Spheck Lee-Inst
Constant	7.562 ***	5.003 *	7.342 ***	5.421 ***	5.243 ***	5.314 ***
	(0.451)	(2.592)	(0.332)	(1.583)	(0.969)	(1.170)
Max. Depth	0.071	0.316	0.025	0.270 *	0.284 ***	0.268 **
	(0.068)	(0.284)	(0.052)	(0.162)	(0.107)	(0.130)
Min. Depth	0.011	-0.108	-0.058	-0.066	-0.081	-0.042
	(0.050)	(0.160)	(0.044)	(0.117)	(0.071)	(0.075)
Biomass	0.201 ***	0.181	0.202 ***	0.195 **	0.170 *	0.186 **
	(0.055)	(0.114)	(0.052)	(0.088)	(0.089)	(0.079)
Dum CV	1.316 ***	0.013	2.514 ***	0.014	0.016	0.017
	(0.214)	(1.244)	(0.207)	(0.925)	(0.248)	(0.436)
Dum HAL	0.102	1.074	0.793	1.114	1.309 **	1.133 **
	(0.205)	(0.966)	(0.179)	(0.763)	(0.547)	(0.561)
Dum NPT	-0.542 **	-0.339	0.210 **	-0.180	-0.061	-0.310
	(0.237)	(0.466)	(0.262)	(0.461)	(0.472)	(0.452)
Dum Large	0.600 ***	0.470	0.665 ***	0.480 **	0.454 *	0.471 **
	(0.131)	(0.303)	(0.165)	(0.239)	(0.237)	(0.231)
IMR		2.909		1.753	1.872 **	1.113
		(2.676)		(1.439)	(0.912)	(0.780)
SAE parameter *(γ)*			0.912 *	0.900 ***	0.909 ***	0.973 ***
			(0.509)	(0.163)	(0.084)	(0.028)

Notes: Dependent variable is average catch-per-unit-effort (CPUE) for statistical reporting regions in the Eastern Bering Sea. Sample size is 320 with 35% selection. Standar errors in parentheses; *, **, *** significant at the 10%, 5%, and 1% level, respectively.

[1] The standard errors for KP-SAE are computed following Kelejian and Prucha (2005).

Table 6. Estimated Coefficients for the Selection Equation

	Heckit	Spheck No-Inst	Spheck KP-Inst	Spheck Lee-Inst
Constant	-0.104	0.037	0.052	-0.070
	(0.648)	(1.104)	(0.776)	(0.687)
Max. Depth	0.179 *	0.317	0.240	0.208
	(0.092)	(0.348)	(0.159)	(0.130)
Min. Depth	-0.093	-0.194	-0.147 *	-0.100
	(0.068)	(0.219)	(0.088)	(0.068)
Biomass	0.005	0.002	0.008	0.004
	(0.078)	(0.123)	(0.087)	(0.077)
Lag Biomass	-0.043	-0.074	-0.084	-0.046
	(0.080)	(0.145)	(0.067)	(0.067)
Dum CV	-0.739 ***	-1.252	-0.995 **	-0.912 ***
	(0.183)	(1.227)	(0.441)	(0.256)
Dum HAL	0.650 ***	1.069	1.089 **	0.945 ***
	(0.202)	(1.103)	(0.430)	(0.290)
Dum NPT	0.073	0.034	0.257	0.101
	(0.261)	(0.464)	(0.377)	(0.340)
Dum Large	-0.078	-0.056	-0.094	-0.079
	(0.176)	(0.290)	(0.240)	(0.225)
SAE parameter (δ)		0.908 ***	0.866 ***	0.817 ***
		(0.211)	(0.114)	(0.006)

Notes: Dependent variable is whether or not CPUE is observed for that unit. Standar errors in parentheses; *, **, *** significant at the 10%, 5%, and 1% level, respectively.