One or Two-Step? Evaluating GMM Efficiency for Spatial Binary Probit Models

Gianfranco Piras*

Mauricio Sarrias †

July 6, 2023

Abstract

In this article we propose two-step generalized method of moment (GMM) procedure for a Spatial Binary Probit Model. In particular, we propose a series of two-step estimators based on different choices of the weighting matrix for the moments conditions in the first step, and different estimators for the variance-covariance matrix of the estimated coefficients. In the context of a Monte Carlo experiment, we compare the properties of these estimators, a linearized version of the one-step GMM and the recursive importance sampler (RIS). Our findings reveal that there are benefits related both to the choice of the weight matrix for the moment conditions and in adopting a two-step procedure.

JEL classification: C21, C31

Keywords: Spatial dependence, Probit model, Generalized moment estimation, Efficiency

*Catholic University of America Washington D.C. and University "G. d'Annunzio", Pescara, Italy

†Facultad de Economía y Negocios, Universidad de Talca, Talca, Chile. email:mauricio.sarrias@utalca.cl.

1 Introduction

Social agents are faced with decisions that are intrinsically discrete. There are several situations in which the variable that needs to be explained takes only two different values (whether a county decides to implement a certain policy, entering the labor force, dropping college, voting for democrats, etc.). In many of those situations, space plays a crucial role because economic agents are affected from the decisions taken by their neighbors. One simple example could relate to the decision to increase the police force in a given area (e.g., a county). Of course, reasonable candidates to explain such variable would be the level of crime in the area, the unemployment rate, the educational level, and the proportion housing units that are rented. However, it is also of interest to consider the decision made in neighboring areas about the commitment of increasing the police force. The explanation for this is that, *ceteris paribus*, if neighboring counties increase the police force, and a given county does not, that given county will become more attractive for criminals, with all the consequences that something like this bears.

Spatial limited dependent variable models have a quite long history in spatial econometrics. One of the first attempts to estimate the binary spatial error and the spatial autoregressive models is McMillen (1992) that proposes two categories of estimators, one based on the EM algorithm and one derived from the spatial expansion method (Casetti, 1972).[1]

The GMM approach to estimate spatial binary models was first proposed by Pinkse and Slade (1998) in the context of a spatial error probit model. They develop test for spatial error dependence and GMM estimators to account for spatial dependence when the restrictions are rejected. Their empirical application was focused on the evaluation of spatial patterns in retail-gasoline contracts.[2] One of the drawbacks of the GMM approach proposed by Pinkse and Slade (1998) is that it requires the inversion of an $n \times n$ matrix, which is still quite an obstacle in very large data. For this reason, Klier and McMillen (2008) propose a linearized version of the GMM put forth by Pinkse and Slade (1998). The linearization allows to estimate the model in two steps: the first step is a standard probit (or logit) model; in the second step the linearized model is estimated by two-stage least squares method.

For binomial discrete choice models, several estimators have been proposed in addition to the GMM and the EM.[3] LeSage (2000) suggests a Bayesian estimation of the limited dependent variable spatial autoregressive model.[4] Based on Vijverberg (1997), Beron and Vijverberg (2004) introduce the so-called recursive importance sampling (RIS) estimator and

[1]The paper also contained an empirical application based on the popular Columbus crime dataset (Anselin, 1988).

[2]For future reference, note that in their empirical application, they make use of a one-step GMM estimator where the weighting matrix of the moments is set to an identity matrix.

[3]For an exhaustive survey of methods see Fleming (2004).

[4]See also LeSage et al. (2011).

show how this estimator can be used to evaluate an n-dimensional normal probability. More recently, Billé and Leorato (2020), put forth a partial maximum likelihood estimator for a general spatial non-linear probit model, and perform a complete asymptotic analysis of their estimator. Finally, in the context of a Monte Carlo experiment, Calabrese and Elkink (2014) provide a comparison of the accuracy of some of these estimators. One of their conclusions was that only the linearized version of the GMM estimator proposed by Klier and McMillen (2008) can be used for very large samples, provided that the level of spatial correlation in the data is low.

In this paper we consider a set of two-step GMM estimators for a spatial binary probit model based on different choices of the weighting matrix for the moment conditions in the first step, and different estimators for the variance-covariance matrix of the estimated coefficients. To the best of our knowledge, this is the first attempt to consider two-step GMM. In the spirit of Calabrese and Elkink (2014), we report on a Monte Carlo experiment that compares between one- and two- step GMM estimators, the linearized GMM, and the RIS estimators.[5] Results from the Monte Carlo experiment reveal that there are benefits related both to the choice of the weight matrix for the moment conditions and in adopting a two-step procedure.

Finally, we provide an empirical application analyzing the decision to reopen after the Hurricane Katrina of a sample of firms located in the New Orleans area (LeSage et al., 2011). In particular, we compare results from the GMM estimators and those reported in the original paper. We observe that, in general, the statistical significance and the magnitude of the estimated coefficients are very similar.

The paper is organized as follow. In Section 2, we introduce the model specification. Section 3 deals with the description of the GMM estimators and the RIS. The results from the Monte Carlo experiment are presented in Section 4. Section 5 contains the empirical illustration, while Section 6 draws conclusions and gives indications for further research.

2 The model

As in Pinkse and Slade (1998), our point of departure is the spatial autoregressive specification of a binary probit model (SARB). The structural form of the SARB model can be written in the following way:

$$\begin{aligned} \mathbf{y}_n^* &= \rho \mathbf{W}_n \mathbf{y}_n^* + \mathbf{X}_n \boldsymbol{\beta} + \boldsymbol{\varepsilon}_n, \quad i = 1, ..., n \\ \mathbf{y}_n &= \mathbb{1}\left[\mathbf{y}_n^* > 0\right], \end{aligned} \tag{1}$$

where $\mathbf{y}_n^*$ is a $n \times 1$ vector of latent (unobserved) continuous variable, $\mathbf{y}_n$ is the vector of observed binary variable, and $\mathbb{1}\left[\cdot\right]$ is the indicator function. In other words, the binary variable $y_{i,n} = 1$ if $\mathbb{1}\left[y_{i,n}^* > 0\right]$, and zero otherwise. The matrix $\mathbf{X}_n$ is a $n \times k$ matrix of explanatory

[5]To keep the simulation feasible, we decided to exclude from the comparison the EM algorithm (McMillen, 1992), the Bayesian approach (LeSage, 2000), and the recent development in maximum likelihood (Billé and Leorato, 2020).

variables whose first column is the intercept, $\mathbf{W}_n$ is a non-stochastic $n \times n$ spatial weighting matrix with element $w_{ij,n}$, and $\mathbf{W}_n \mathbf{y}_n^*$ is the spatial lag of the continuous but unobserved variable $\mathbf{y}_n^*$ which introduces (unobserved) endogeneity; $\boldsymbol{\beta}$ is a $k \times 1$ vector of coefficients, ρ is the spatial autoregressive coefficient, and $\boldsymbol{\varepsilon}_n$ is the $n \times 1$ vector of error terms.

In line with the literature on spatial models, we make the following assumptions regarding the weighting matrix, the error terms and the spatial autoregressive parameter.

Assumption 1 *(a) All diagonal elements of $\mathbf{W}_n$ are zero. (b) $\rho \in (-1, 1)$. (c) The matrix $\mathbf{A}_n = (\mathbf{I}_n - \bar{\rho}\mathbf{W}_n)$ is nonsingular for all $\bar{\rho} \in (-1, 1)$.*

Assumption 2 *The innovations $\{\epsilon_{i,n} : 1 \leq i \leq n, n \geq 1\}$ are normally identically distributed and satisfy $\mathbb{E}(\epsilon_{i,n}) = 0$, $\mathbb{E}(\epsilon_{i,n}^2) = \sigma^2$, where $0 < \sigma^2 < b$. Furthermore, its moments $\mathbb{E}\left(\left|\epsilon_{i,n}^{4+\delta}\right|\right)$ exist for some $\delta > 0$.*

Assumption 3 *The row and column sums of the matrix $\mathbf{W}_n$ are bounded uniformly in absolute value by, respectively, one and some finite constant, and the row and column sums of the matrix $\mathbf{A}_n^{-1} = (\mathbf{I}_n - \bar{\rho}\mathbf{W}_n)^{-1}$ are bounded uniformly in absolute value by some constant.*

Under Assumption 1, the reduced form of Equation (1) can be expressed as the following latent process:

$$\begin{aligned} \mathbf{y}_n^* &= (\mathbf{I}_n - \rho\mathbf{W}_n)^{-1} (\mathbf{X}_n\boldsymbol{\beta} + \boldsymbol{\varepsilon}_n), \\ &= \mathbf{A}_n^{-1}\mathbf{X}_n\boldsymbol{\beta} + \mathbf{u}_n, \end{aligned} \tag{2}$$

where $\mathbf{A}_n = (\mathbf{I}_n - \rho\mathbf{W}_n)$ and $\mathbf{u}_n = \mathbf{A}_n^{-1}\boldsymbol{\varepsilon}_n$, so that, under Assumption 2, $\mathbf{u}_n \sim \mathrm{N}(\mathbf{0}, \boldsymbol{\Sigma}_u)$, with variance-covariance matrix given by:

$$\boldsymbol{\Sigma}_u = \mathbb{E}(\mathbf{u}_n\mathbf{u}_n^\top | \mathbf{X}_n, \mathbf{W}_n) = \sigma^2 (\mathbf{I}_n - \rho\mathbf{W}_n)^{-1} \left[(\mathbf{I}_n - \rho\mathbf{W}_n)^\top\right]^{-1} = \sigma^2(\mathbf{A}_n^\top\mathbf{A}_n)^{-1}. \tag{3}$$

Thus, the inclusion of $\mathbf{W}_n\mathbf{y}_n^*$ introduces endogeneity as well as heteroskedasticity. As standard in binary models, we need to fix σ^2 in the estimation procedure for identification of the parameters. Hereafter, we assume that $\sigma^2 = 1$. For further reference, note that $\mathbf{y}_n^* \sim \mathrm{N}(\mathbf{A}_n^{-1}\mathbf{X}_n\boldsymbol{\beta}, \boldsymbol{\Sigma}_u)$.

The SARB model in (1) has several distinctive peculiarities. First, since the spatial lag of the dependent variable $\mathbf{W}\mathbf{y}_n^*$ is correlated with the disturbances, the SARB models suffers from endogeneity which motives the use of instrumental variables. Following Kelejian and Prucha (1998), we use an approximation to the ideal instruments. In particular, let $\mathbf{H}_n$ be an $n \times p$ matrix of non-stochastic instruments where $p \geq k+1$. The assumptions on $\mathbf{X}_n$ and $\mathbf{H}_n$ are the following:

Assumption 4 *The regressor matrices $\mathbf{X}_n$ have full column rank for n large enough. Additionally, the elements of $\mathbf{X}_n$ are uniformly bounded in absolute value.*

Assumption 5 *The instrument matrices* $\mathbf{H}_n$ *have full column rank* $p \geq k+1$ *for all* n *large enough. Furthermore, the elements of the matrices* $\mathbf{H}_n$ *are uniformly bounded in absolute value. Additionally,* $\mathbf{H}_n$ *contain at least the linearly independent columns of* $(\mathbf{X}_n, \mathbf{W}_n\mathbf{X}_n, \mathbf{W}_n^2\mathbf{X}_n)$.

Second, standard Maximum Likelihood (ML) estimators are also inconsistent due to the heteroskedasticity induced by spatial dependence (Pinkse and Slade, 1998). In fact, from the reduced form in (2), we can obtain the conditional expectation of the observed outcome, for all $i = 1, ..., n$ as

$$\begin{aligned}\mathbb{E}(y_{i,n}|\mathbf{X}_n, \mathbf{W}_n) &= \Pr(y_{i,n} = 1|\mathbf{X}_n, \mathbf{W}_n) \\ &= \Pr\left(\{\mathbf{u}_n\}_i > -\left\{\mathbf{A}_n^{-1}\mathbf{X}_n\boldsymbol{\beta}\right\}_i|\mathbf{X}_n, \mathbf{W}_n)\right) \\ &= \Phi\left(\{\boldsymbol{\Sigma}_u\}_{ii}^{-1/2}\left\{\mathbf{A}_n^{-1}\mathbf{X}_n\boldsymbol{\beta}\right\}_i\right) \\ &= \Phi\left(a_{i,n}\right)\end{aligned} \tag{4}$$

where $\Phi(\cdot)$ is the normal cumulative distribution function (cdf), $\{\cdot\}_i$ is the ith element of the vector in brackets and $\{\cdot\}_{ii}$ is the ith diagonal element of the matrix in brackets. Equation (4) reveals that the probability of observing $y_{i,n} = 1$ is heteroskedastic and hence MLE is inconsistent and not fully efficient (Fleming, 2004).

3 GMM estimators and RIS

3.1 GMM estimators

Following Pinkse and Slade (1998) and Fleming (2004), this section describes the general GMM approach to estimate a SARB probit model.[6] We place particular emphasis on the difference between the one-step and the two-step GMM estimator (also known as optimal GMM estimator). Finally, we present the linearized version of the GMM estimator (Klier and McMillen, 2008), and the RIS estimator put forth by Beron and Vijverberg (2004).

3.1.1 Moment conditions

For the estimation of the SARB probit model using a GMM procedure we need population moment conditions based on the innovations. The main problem with the model in Equation (1) is that the error term $\boldsymbol{\varepsilon}_n$ is based on unobserved dependent variables, $\mathbf{y}_n^*$. For this reason, we need to rely on the concept of generalized residuals which are based on observed error terms (Cox and Snell, 1968; Chesher and Irish, 1987; Gourieroux et al., 1987).

To simplify the notation, let $a_{i,n}$ in Equation (4) be the ith element of the following $n \times 1$ vector:

[6] Although Pinkse and Slade (1998) derive a GMM procedure for the spatial error binary model (SEMB), it is not difficult to extend their procedure to the SARB model. See also Fleming (2004).

$$\mathbf{a}_n = \mathbf{D}_{n,\rho}^{-1}\mathbf{A}_n^{-1}\mathbf{X}_n\boldsymbol{\beta}, \tag{5}$$

where $\mathbf{D}_{n,\rho}$ is a $n \times n$ diagonal matrix with diagonal elements representing the square root of the diagonal elements of the conditional variance-covariance matrix of the error terms $\mathbf{u}_n$ given in Equation (3). Thus, the generalized residuals for spatial unit $i = 1, ..., n$ are (see also Pinkse and Slade, 1998):

$$\widetilde{u}_{i,n}(\boldsymbol{\theta}) = u_{i,n} \cdot \left[\frac{\phi(a_{i,n})}{\Phi(a_{i,n})(1-\Phi(a_{i,n}))}\right], \tag{6}$$

where $\phi(\cdot)$ is the standard normal density function, $u_{i,n} = y_{i,n} - \Phi(a_{i,n})$ and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \rho)^\top$ is the $(k+1) \times 1$ vector of population parameters.[7]

The $p \times 1$ population moment conditions are then:

$$\mathbb{E}\left[\mathbf{h}_{i,n}\widetilde{u}_{i,n}(\boldsymbol{\theta})\right] = \mathbf{0}, \tag{7}$$

where $\mathbf{h}_{i,n}$ is a $p \times 1$ vector of instruments such that $p \geq k+1$ for identification of the parameters. The sample analog in vector form of the population moment conditions in (7) is:

$$\mathbf{g}_n(\boldsymbol{\theta}) = n^{-1}\mathbf{H}_n^\top\widetilde{\mathbf{u}}_n, \tag{8}$$

where $\widetilde{\mathbf{u}}_n$ is the $n \times 1$ vector of generalized residuals. These sample moments reveal several aspects that deserve to be mentioned. First, unlike the traditional GMM methods for linear spatial models, the sample moments in (8) require to invert $\mathbf{A}_n$ to obtain $\mathbf{a}_n$ as shown in Equation (5).[8] Second, the sample moments do not consider the off-diagonal elements of $\boldsymbol{\Sigma}_u$.[9] Thus, GMM estimators are generally less efficient than techniques that account for the full variance-covariance structure of the error term (see also Calabrese and Elkink, 2014). Finally, the $n \times p$ matrix of instrument is given by the independent columns of $\mathbf{H} = (\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}, ..., \mathbf{W}^q\mathbf{X})$ for some given q (Kelejian and Prucha, 1998; Kelejian et al., 2004). The $p \times p$ variance matrix of the moment conditions is given by:[10]

$$\begin{aligned} \mathbf{S}_n(\boldsymbol{\theta}) &= \mathrm{Var}\left[\mathbf{g}_n(\boldsymbol{\theta})\right], \\ &= n^{-1}\mathbf{H}_n^\top\mathbf{T}_n\mathbf{H}_n, \end{aligned} \tag{9}$$

where $\mathbf{T}_n$ is a diagonal matrix whose elements are $\phi^2(a_{i,n})/\left[\Phi(a_{i,n})(1-\Phi(a_{i,n}))\right]$.

[7] If we let $q_{i,n} = 2y_{i,n} - 1$, the generalized residuals can be also written as:

$$\widetilde{u}_{i,n} = q_{i,n} \cdot \frac{\phi(q_{i,n} \cdot a_{i,n})}{\Phi(q_{i,n} \cdot a_{i,n})}.$$

We use this simplified expression for coding the procedure in the R programming language.

[8] Of course, this can be very time consuming in large datasets. However, this problem can be mitigated by using some matrix approximations (see, for example, Santos and Proença, 2019).

[9] One way to account for this problem would be to consider quadratic forms of the moment condition. While interesting, this is beyond the scope of our paper.

[10] See Appendix A.

3.1.2 General GMM estimator

Let $\widehat{\boldsymbol{\Psi}}_n$ be some $p \times p$ symmetric positive semidefinite moment-weighting matrix such that $\widehat{\boldsymbol{\Psi}}_n \xrightarrow{p} \boldsymbol{\Psi}$, then the corresponding GMM estimator is defined as:

$$\widehat{\boldsymbol{\theta}}_{n,GMM} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \;\; J_n(\boldsymbol{\theta}) = \mathbf{g}_n^\top(\boldsymbol{\theta}) \widehat{\boldsymbol{\Psi}}_n \mathbf{g}_n(\boldsymbol{\theta}), \tag{10}$$

where $\mathbf{g}_n$ is the $p \times 1$ vector of sample moments given in Equation (8), and $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \rho)^\top$ is the $k+1$ vector of parameters. Under certain regularity conditions, the GMM estimator is consistent and asymptotically normally distributed with estimated variance-covariance matrix as (see Pinkse and Slade, 1998, pag. 134):

$$\widehat{\mathbf{V}}_{n,GMM} = n \left[\left(\widehat{\mathbf{G}}_n^\top \mathbf{H}_n\right) \widehat{\boldsymbol{\Psi}}_n \left(\mathbf{H}_n^\top \widehat{\mathbf{G}}_n\right)\right]^{-1} \left[\left(\widehat{\mathbf{G}}_n^\top \mathbf{H}_n\right) \widehat{\boldsymbol{\Psi}}_n \widehat{\mathbf{S}}_n \widehat{\boldsymbol{\Psi}}_n \left(\mathbf{H}_n^\top \widehat{\mathbf{G}}_n\right)\right] \left[\left(\widehat{\mathbf{G}}_n^\top \mathbf{H}_n\right) \widehat{\boldsymbol{\Psi}}_n \left(\mathbf{H}_n^\top \widehat{\mathbf{G}}_n\right)\right]^{-1} \tag{11}$$

where $\widehat{\mathbf{G}}_n$ is $n \times (k+1)$ matrix of first derivatives of the generalized residuals (see Appendix B) such that:

$$\widehat{\mathbf{G}}_n = \left.\frac{\partial \widetilde{\mathbf{u}}_n}{\partial \boldsymbol{\theta}^\top}\right|_{\widehat{\boldsymbol{\theta}}_n}, \tag{12}$$

and the $p \times p$ matrix $\widehat{\mathbf{S}}_n$ is a consistent estimator of (9).

3.1.3 One-step GMM estimators

The one-step procedure estimates the model parameters based on an initial weight matrix $\widehat{\boldsymbol{\Psi}}_n$. Following Klier and McMillen (2008) and Pinkse and Slade (1998), we consider two types of one-step GMM estimators. The first estimator is obtained by setting $\widehat{\boldsymbol{\Psi}}_n = \left(n^{-1}\mathbf{H}_n^\top \mathbf{H}_n\right)^{-1}$, that is:

$$\widetilde{\boldsymbol{\theta}}_{\mathrm{OS},H} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \;\; J_n(\boldsymbol{\theta}) = \left(\frac{1}{n}\widetilde{\mathbf{u}}_n^\top \mathbf{H}_n\right) \left(n^{-1}\mathbf{H}_n^\top \mathbf{H}_n\right)^{-1} \left(\frac{1}{n}\mathbf{H}_n^\top \widetilde{\mathbf{u}}_n\right). \tag{13}$$

This one-step estimator is a natural adaptation of the estimator proposed by Klier and McMillen (2008) to estimate a spatial lag binary dependent model with a logistically distributed error term.[11] The variance-covariance matrix for $\widetilde{\boldsymbol{\theta}}_{\mathrm{OS},H}$ can be estimated as:

$$\begin{aligned}\widehat{\mathbf{V}}\left(\widetilde{\boldsymbol{\theta}}_{\mathrm{OS},H}\right) =& n\left[\widetilde{\mathbf{G}}_n^\top \mathbf{H}_n \left(\mathbf{H}_n^\top \mathbf{H}_n\right)^{-1} \mathbf{H}_n^\top \widetilde{\mathbf{G}}_n\right]^{-1} \left[\widetilde{\mathbf{G}}_n^\top \mathbf{H}_n \left(\mathbf{H}_n^\top \mathbf{H}_n\right)^{-1} \widetilde{\mathbf{S}}_n \left(\mathbf{H}_n^\top \mathbf{H}_n\right)^{-1} \mathbf{H}_n^\top \widetilde{\mathbf{G}}_n\right] \\ & \times \left[\widetilde{\mathbf{G}}_n^\top \mathbf{H}_n \left(\mathbf{H}_n^\top \mathbf{H}_n\right)^{-1} \mathbf{H}_n^\top \widetilde{\mathbf{G}}_n\right]^{-1},\end{aligned} \tag{14}$$

where the estimator for $\widetilde{\mathbf{S}}_n$ is:

[11] The generalized residuals for the Logit model are $\widetilde{u}_{i,n} = y_i - \exp(a_{i,n}) / \left(1 + \exp(a_{i,n})\right)$.

$$\widetilde{\mathbf{S}}_n\left(\widetilde{\boldsymbol{\theta}}_{\mathrm{OS},H}\right) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{h}_{i,n}\left[\frac{\phi^2(\widetilde{a}_{i,n})}{\Phi(\widetilde{a}_{i,n})(1-\Phi(\widetilde{a}_{i,n}))}\right]\mathbf{h}_{i,n}^{\top}, \tag{15}$$

where $\widetilde{a}_{i,n}$ is the ith element of Equation (5) evaluated at $\widetilde{\boldsymbol{\theta}}_{\mathrm{OS},H}$.

As in Pinkse and Slade (1998), the second one-step estimator sets $\widehat{\boldsymbol{\Psi}}_n = \mathbf{I}_p$ yielding:

$$\widetilde{\boldsymbol{\theta}}_{\mathrm{OS},\mathbf{I}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\ J_n(\boldsymbol{\theta}) = \mathbf{g}_n^{\top}\mathbf{g}_n. \tag{16}$$

This estimator can be viewed as an unweighted nonlinear least squares estimators in which $J_n(\boldsymbol{\theta})$ is the sum of p squared sample average of the moment conditions (Cameron and Trivedi, 2005). The estimator of the variance-covariance matrix in this case becomes:

$$\widehat{\mathbf{V}}\left(\widetilde{\boldsymbol{\theta}}_{\mathrm{OS},\mathbf{I}}\right) = n\left[\widehat{\mathbf{G}}_n^{\top}\mathbf{H}_n\mathbf{H}_n^{\top}\widehat{\mathbf{G}}_n\right]^{-1}\left[\widehat{\mathbf{G}}_n^{\top}\mathbf{H}_n\widehat{\mathbf{S}}_n\mathbf{H}_n^{\top}\widehat{\mathbf{G}}_n\right]\left[\widehat{\mathbf{G}}_n^{\top}\mathbf{H}_n\mathbf{H}_n^{\top}\widehat{\mathbf{G}}_n\right]^{-1} \tag{17}$$

Under the assumptions made, the choice of the weight matrix $\widehat{\boldsymbol{\Psi}}_n$ should not affect the consistency of the one-step estimators. However, we should observe differences in finite sample.

3.1.4 Two-step GMM estimator

It is widely known that one can gain efficiency by computing two-step estimators of the form:

$$\widehat{\boldsymbol{\theta}}_{\mathrm{TS}} = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\ J_n(\boldsymbol{\theta}) = \mathbf{g}_n^{\top}(\boldsymbol{\theta})\widehat{\boldsymbol{\Psi}}_n\mathbf{g}_n(\boldsymbol{\theta}). \tag{18}$$

where $\widehat{\boldsymbol{\Psi}}_n = \widetilde{\mathbf{S}}_n^{-1}$, and $\widetilde{\mathbf{S}}_n$ is an estimate of the variance-covariance matrix $\mathbf{S}_n$ based on either of the one-step estimator.

The step-wise procedure can be summarized as follows:

1. First, minimize the objective function (10) by choosing either $\widehat{\boldsymbol{\Psi}}_n = \mathbf{I}_p$ or $\widehat{\boldsymbol{\Psi}}_n = (n^{-1}\mathbf{H}_n^{\top}\mathbf{H}_n)^{-1}$ to obtain $\widetilde{\boldsymbol{\theta}}_{\mathrm{OS}}$. Note that in either case, $\widetilde{\boldsymbol{\theta}}_{\mathrm{OS}}$ is consistent as $n\to\infty$ but not fully efficient.

2. Second, use $\widetilde{\boldsymbol{\theta}}_{\mathrm{OS}}$ to obtain the residuals from the first step and calculate $\widetilde{\mathbf{S}}_n$ using Equation (15). Set $\widehat{\boldsymbol{\Psi}}_n = \widetilde{\mathbf{S}}_n^{-1}$ and minimize (18) to obtain the final round estimate $\widehat{\boldsymbol{\theta}}_{TS}$. The estimated asymptotic variance is given by:

$$\widehat{\mathbf{V}}_{EGMM} = n\left[\widehat{\mathbf{G}}_n^{\top}\mathbf{H}_n\widehat{\boldsymbol{\Psi}}\mathbf{H}_n^{\top}\widehat{\mathbf{G}}_n\right]^{-1}, \tag{19}$$

   where $\widehat{\boldsymbol{\Psi}}_n = \widetilde{\mathbf{S}}^{-1}$.[12] However, as pointed out by Cameron and Trivedi (2005) in finite samples the estimator in Equation (19) might be biased. In such case, it would be better to use Equation (11), where $\widehat{\boldsymbol{\Psi}}_n = \widetilde{\mathbf{S}}^{-1}$, and $\widehat{\mathbf{S}}_n$ is computed using Equation (15) and $\widehat{\boldsymbol{\theta}}_{\mathrm{TS}}$.

[12] Since any consistent estimate of $\boldsymbol{\theta}$ can be used, an alternative is to evaluate $\mathbf{S}$ at $\widehat{\boldsymbol{\theta}}_{\mathrm{TS}}$.

3.1.5 Lineralized GMM

One of the main drawback of the previous GMM estimators is that they require the inversion of the $n \times n$ matrix $\mathbf{A}_n = (\mathbf{I}_n - \rho \mathbf{W}_n)$ which can be very time consuming for large datasets. To overcome this problem, Klier and McMillen (2008) propose a linearized version of the one-step GMM estimator around the starting point $\rho = 0$. When $\rho = 0$, $\boldsymbol{\beta}$ is estimated consistently by standard probit model and $\mathbf{A}_n^{-1} = \mathbf{I}_n$ so that no matrices need to be inverted. Linearizing the generalized residuals around the initial estimates of $\boldsymbol{\theta}$ (i.e, $\boldsymbol{\theta}^0$), Klier and McMillen (2008) obtain $\widetilde{u}_{i,n} \approx \widetilde{u}_{i,n}^0 - \mathbf{G}_n(\boldsymbol{\theta}_n - \boldsymbol{\theta}_n^0)$. If we define $\upsilon_{i,n} = \widetilde{u}_{i,n}^0 + \mathbf{G}_n\boldsymbol{\theta}_n^0 - \mathbf{G}_n\boldsymbol{\theta}_n$ and letting $\boldsymbol{\Psi}_n = (\mathbf{H}_n^\top \mathbf{H}_n)$, the objective function becomes $\boldsymbol{\upsilon}_n^\top \mathbf{H}_n \left(\mathbf{H}_n^\top \mathbf{H}_n\right)^{-1} \mathbf{H}_n^\top \boldsymbol{\upsilon}_n$.

The steps for the linearized spatial Probit model are the following (Klier and McMillen, 2008, pag. 462):

1. Estimate the model by standard probit model, in which spatial autocorrelation and heteroskedasticity are ignored. The estimated values are $\widehat{\boldsymbol{\beta}}_0$. Calculate the generalized residuals in Equation (6) assuming that $\rho = 0$, and the gradient terms $\mathbf{G}_{\boldsymbol{\beta}} = -\partial \widetilde{\mathbf{u}}_n / \partial \boldsymbol{\beta}$ and $\mathbf{G}_\rho = -\partial \widetilde{\mathbf{u}}_n / \partial \rho$.

2. The second step is a two-stage least squares estimator of the linearized model. Thus, regress $\mathbf{G}_{\boldsymbol{\beta}}$ and $\mathbf{G}_\rho$ on $\mathbf{H}_n$. The predicted values are $\widehat{\mathbf{G}}_{\boldsymbol{\beta}}$ and $\widehat{\mathbf{G}}_\rho$. Then regress $u_0 + \mathbf{G}_{\boldsymbol{\beta}}^\top \widehat{\boldsymbol{\beta}}_0$ on $\widehat{\mathbf{G}} = \left[\widehat{\mathbf{G}}_{\boldsymbol{\beta}}, \widehat{\mathbf{G}}_\rho\right]$. The coefficients are the estimated values of $\boldsymbol{\beta}$ and ρ.

The variance-covariance matrix can be computed using the traditional White-corrected coefficient covariance matrix from the last two-stage least squares estimator of the linearized model.

3.2 The RIS estimator

The recursive importance sampling (RIS) was initially proposed by Beron and Vijverberg (2004) as a way to evaluate directly the n-dimensional integral implied by the model. Let $\boldsymbol{\nu}$ be distributed as a normal with mean zero and variance-covariance matrix $\boldsymbol{\Omega}$. The aim is to evaluate $\Pr[\boldsymbol{\nu} < V]$. Let $\mathbf{C}$ be an upper-triangular matrix such that $\mathbf{C}^\top\mathbf{C} = \boldsymbol{\Omega}^{-1}$, and let $\boldsymbol{\eta} = \mathbf{C}\boldsymbol{\nu}$. Note that $\boldsymbol{\eta}$ is $N(0,1)$. Additionally, we can define the matrix $\mathbf{B} = \mathbf{C}^{-1}$, where $\mathbf{B}$ an upper triangular matrix with $b_{ii} > 0 \;\; \forall i$. Then the upper limit of the inequality $\mathbf{B}\boldsymbol{\eta} < V$ can be written as

$$
\begin{aligned}
\boldsymbol{\eta}_n &< b_{nn}^{-1} V_n \equiv \boldsymbol{\eta}_{n0} \\
\boldsymbol{\eta}_j &< b_{ii}^{-1}\left[V_i - \sum_{j=i+1}^{n} b_{ij}\boldsymbol{\eta}_i\right] \equiv \boldsymbol{\eta}_{i0}.
\end{aligned}
$$

Let $h(\boldsymbol{\eta}_i)$ be a density function and let H be the associated cumulative distribution function, and define $h^c(\boldsymbol{\eta}_i) = h(\boldsymbol{\eta}_i)/H(\boldsymbol{\eta}_{i0})$ where $\boldsymbol{\eta}_i \leq \boldsymbol{\eta}_{i0}$. Then, we can compute the following probability

$$p = \Pr[\boldsymbol{\nu} < V] = \int_{-\infty}^{V} \phi_n(\boldsymbol{\nu}; 0, \boldsymbol{\Omega})\, d\boldsymbol{\nu} = \int_{-\infty}^{\boldsymbol{\eta}_{n0}} \cdots \int_{-\infty}^{\boldsymbol{\eta}_{n1,0}} \prod_{i=1}^{n} \phi(\boldsymbol{\eta}_i) d\boldsymbol{\eta}_1 \ldots d\boldsymbol{\eta}_n$$
$$= \int_{-\infty}^{\boldsymbol{\eta}_{n0}} \frac{\phi(\boldsymbol{\eta}_n)}{h^c(\boldsymbol{\eta}_n)} \left[\int_{-\infty}^{\boldsymbol{\eta}_{n-1,0}} \frac{\phi(\boldsymbol{\eta}_{n-1})}{h^c(\boldsymbol{\eta}_{n-1})} \cdots \left(\int_{-\infty}^{\boldsymbol{\eta}_{2,0}} \frac{\phi(\boldsymbol{\eta}_2)}{h^c(\boldsymbol{\eta}_2)} \Phi(\boldsymbol{\eta}_{1,0}) h^c(\boldsymbol{\eta}_2) d\boldsymbol{\eta}_2 \right) \cdots \right] h^c(\boldsymbol{\eta}_n) d\boldsymbol{\eta}_n$$

The RIS simulator can be implemented by drawing R random vectors of $\boldsymbol{\eta}$ from the distribution defined by h.[13] In the end, the simulated value for p can be obtained from the following equation:

$$\hat{p} = \frac{1}{R} \sum_{r=1}^{R} \left[\prod_{i=1}^{n} \Phi(\tilde{\boldsymbol{\eta}}_{i,0,r}) \right] \tag{20}$$

4 Monte Carlo

To assess the efficiency of the estimators presented in the previous section, we perform a Monte Carlo experiment. The set up of our Monte Carlo is similar to the one in Calabrese and Elkink (2014). In the next subsection we describe the design of our experiment highlighting the data generation process and the different estimators employed. Next, we focus on the results obtained from the Monte Carlo that give valuable information for applied researchers.

4.1 Monte Carlo design

The data generating process (DGP) is given by the following equation:

$$\begin{aligned} \mathbf{y}^* &= (\mathbf{I} - \rho \mathbf{W})^{-1} (\beta_0 \boldsymbol{\imath}_n + \beta_1 \mathbf{x} + \boldsymbol{\varepsilon}), \\ \mathbf{y} &= \mathbb{1}[\mathbf{y}_n^* > 0], \end{aligned} \tag{21}$$

where $\boldsymbol{\imath}_n$ in an $n \times 1$ vector of one and the elements of the vector $\mathbf{x}$ are normally distributed with mean 2 and standard deviation 4. The parameter β_0 is set to 4, while $\beta_1 = -2$. The error term is normally distributed with mean zero and standard deviation one. We account for five different values for the spatial parameter ρ, namely, 0, 0.2, 0.4, 0.6, 0.8. We consider two sample sizes $n = 50$ and $n = 500$ and generate $1,000$ replication for both sample sizes.[14]

The spatial weighting matrix $\mathbf{W}$ is computed as in Beron and Vijverberg (2004). We first generate uncorrelated random pair of coordinates from the uniform distributions for each $i = 1, ..., n$. Then, we set $w_{ij} = 1$ if $d_{ij} < d^{(n)}$ and 0 otherwise, where w_{ij} is the i, jth element of $\mathbf{W}$, d_{ij} is the euclidean distance between observations i and j, and $d^{(n)}$ is a threshold distance that depends on the sample size. For each sample size n, the threshold distances are set so as to obtain an average of five nearest neighbors for each observation. Finally,

[13] In line with the literature, in our Monte Carlo we drew samples from a normal distribution.

[14] A major concern with the simulation experiment is the amount of processing time, particularly for those estimator using simulated probabilities (Beron and Vijverberg, 2004). Similarly to Calabrese and Elkink (2014), the decision of generating $1,000$ Monte Carlo samples was taken to keep the experiment manageable in terms of computation time.

the row-standardized spatial weighting matrix is used to generate the matrix of instrument $\mathbf{H} = [\boldsymbol{\iota}, \mathbf{x}, \mathbf{W}\mathbf{x}, \mathbf{W}^2\mathbf{x}]$.

For each Monte Carlo sample, we compute the following estimators:[15]

1. $\breve{\boldsymbol{\theta}}_{\text{LGMM}}$: The linearized GMM estimator (LGMM).
2. $\widetilde{\boldsymbol{\theta}}_{\text{OS, I}}$: One-step GMM estimator setting the weighting-moment matrix $\widehat{\boldsymbol{\Psi}} = \mathbf{I}_p$ (OS-I).
3. $\widetilde{\boldsymbol{\theta}}_{\text{OS, H}}$: One-step GMM estimator setting the weighting-moment matrix to $\widehat{\boldsymbol{\Psi}} = \left(n^{-1}\mathbf{H}_n^\top\mathbf{H}_n\right)^{-1}$ (OS-H).
4. $\widehat{\boldsymbol{\theta}}_{\text{TS, I}}$: Two-step GMM estimator based on residuals from $\widetilde{\boldsymbol{\theta}}_{\text{OS, I}}$ to construct the second-step weighting matrix $\widehat{\boldsymbol{\Psi}}$ in Equation(18) (TS-I).
5. $\widehat{\boldsymbol{\theta}}_{\text{TS, H}}$: Two-step GMM estimator based on residuals from $\widetilde{\boldsymbol{\theta}}_{\text{OS, H}}$ to construct the second-step weighting matrix $\widehat{\boldsymbol{\Psi}}$ in Equation(18) (TS-H).
6. $\bar{\boldsymbol{\theta}}_{\text{RIS}}$: The RIS estimator (RIS)

It is important to note that the GMM and the RIS estimators use as initial values of the optimization procedure the estimates from a standard probit for the model parameters and the correlation between $\mathbf{W}\mathbf{y}$ and $\mathbf{y}$ as the initial value for ρ. All the GMM estimators are computed using an analytical gradient (see additional material associated with the paper in Appendix B), whereas the RIS estimator uses numerical derivatives. Furthermore, both GMM and RIS estimator use constrained optimization methods adopting the BFGS algorithm. Unfortunately, the LGMM procedure does not allow to constrain the values of ρ.[16]

4.2 Monte Carlo results

4.2.1 Bias and standard deviation

In presenting the Monte Carlo results we focus both on the spatial autoregressive parameter ρ as well as on β_1. In particular, the tables of the results contain the Mean Bias, the Median Bias, the standard deviation (SD), the root mean square error (RMSE) and the root mean square error with respect to the median (RMSE Median) for both sample sizes (Tables 1-4).[17]

[15]The GMM, LGMM and RIS estimators were coded by the authors using the R software environment. The functions used to estimate the models in this paper are part of an upcoming library named **spldv** which, at this time, is still under development. However, a preliminary version of the library is available on GitHub at the following link: `https://github.com/gpiras/spldv`.

[16]As suggested by a referee, one can impose the constraint $-1 < \rho < 1$ on the post estimation of ρ by applying the tilting approach of Hall and Huang (2002).

[17]Following Kelejian and Prucha (1998), the RMSE Median is computed as $\left[\text{bias}^2 + (\text{IQ}/1.35)^2\right]^{1/2}$ where bias is an absolute difference between the median of the empirical distribution and the true parameter value, and IQ is an interquantile range.

Finally, we also compare the size of the test on the two parameters for the estimators (Tables 5-8).

Figures 1 and 2 graph the boxplot for different values of ρ obtained from the Monte Carlo samples for the six estimators considered in the paper. In particular, Figure 1 refers to the smallest sample size, and Figure 2 refers to $n = 500$. A glance at Figure 1 reveals that the performance of the estimators varies with the value of the spatial parameter. When $\rho = 0$, all six estimator are unbiased with small variability (apart from few outliers). This is also evident looking at the statistics in the first panel of Table 1. Perhaps not surprisingly, the linearized GMM has the lowest mean and median bias of -0.009 and -0.000, respectively. At the same time the standard deviation of the LGMM is the highest among all estimators. Specifically, the SD for the LGMM is 0.335 while the RIS has standard deviation below 0.2. As the value of ρ increases, the number of outliers in the boxplot also increases. Note that, as we have mentioned before, all the GMM and the RIS estimator use constrained optimization, while the LGMM does not. Hence the outliers for the GMM and the RIS are values of ρ within the parameter space but the LGMM gives values outside of range. Interestingly, except for the LGMM, the variability is higher for the estimators that are based on the identity matrix as initial moment-weighting matrix (i.e., OS-I and TS-I). For $\rho = 0.2$ and $\rho = 0.4$, the two-step estimators are those with the lowest mean bias and SD. Finally, when the level of spatial autocorrelation is substantial ($\rho = 0.6$ and $\rho = 0.8$) the TS-H and the RIS estimator perform the best both in terms of bias and variability (SD and RMSE). One interesting point to note is that, whenever the level of spatial autocorrelation is large, the LGMM is, by far, the estimator that performs the worse.

(Insert Table 1 about here)

(Insert Figure 1 about here)

Moving to Figure 2 and Table 2 that relates to $n = 500$, it is evident again that for small values of the spatial parameter ($\rho = 0$ and $\rho = 0.2$) the performance of the six estimators is extremely similar. When $\rho = 0.4$, the LGMM consistently over estimate the spatial autocorrelation parameter and the variability is about twice as large as the other estimators. In contrast, the GMM and RIS estimator are unbiased and presents almost the same variability at the same true experimental value $\rho = 0.4$. For the highest value of $\rho = 0.8$, the GMM estimators TS-H and OS-H remain unbiased and efficient. However, the TS-I and OS-I are biased and inefficient. Clearly, this last point is very relevant and it was not observed in Calabrese and Elkink (2014) simply because they only considered the one-step GMM estimator based on the identity matrix.

(Insert Table 2 about here)

(Insert Figure 2 about here)

The results in Tables 3 and 4 show evidence of the bias and efficiency of the estimated β_1. In those two tables, a non-spatial probit model was added as a reference point. A quick look at Table 3 reveals that, no matter the level of spatial autocorrelation, all the estimators are biased for the small sample size. For the most part, the estimators present a downward bias. The only two exceptions are the positive values for the probit and LGMM when $\rho = 0.8$.[18] Interestingly, for all levels of spatial autocorrelation (less than $\rho = 0.8$), the two-step GMM estimators in columns five and six have the least bias. For what concern the standard deviation, the TS-I and the TS-H record the smallest values. Summarizing the results for $n = 50$, it seems almost prohibitive to estimate the model parameter β_1 in a precise and accurate way.

(Insert Table 3 about here)

The situation improves tremendously for $n = 500$. While—as expected—the probit and the LGMM are still biased, the GMM and the RIS presents very low figures for the bias ranging from -0.045 and 0.285 (when ρ is, at most, 0.6). When the spatial autocorrelation coefficient is 0.8, only the OS-H and TS-H remain unbiased. This means that using the identity matrix to weight the moment conditions is a sub-optimal choice that can lead to bias, particularly when the level of spatial correlation in the model is high. Interestingly, this is also true when the focus is on the efficiency. In fact, the SD of the OS-H and TS-H is always lower than that for OS-I and TS-I. It is important to note two additional points. First, the SD for both the probit and LGMM is consistently lower than the one obtained with other methods. Unfortunately, those two estimators are biased when there is presence of spatial autocorrelation. Secondly, the RIS and the OS-H and TS-H are comparable in terms of efficiency (with the RIS performing slightly better). At the same time though, RIS has a computational time that is way higher than the GMM, despite the fact that the GMM implies the inversion of a $n \times n$ matrix.

(Insert Table 4 about here)

4.2.2 Size of the test on ρ and β_1

Tables 5-8 report the size of a test on ρ and β_1 for the two sample sizes.[19] Note that those tables have nine columns. The first three columns are the LGMM, the OS-I and OS-H. The two-step GMM estimators differ in terms of the variance-covariance matrix employed. In particular, TS-I-A is based on the identity matrix for the first step and the variance-covariance matrix in Equation (11). The next two columns are based on the identity matrix for the first step, and the variance-covariance matrix in Equation (19), where $\mathbf{S}_n$ is evaluated

[18] This result has also been found by Calabrese and Elkink (2014).

[19] Note that the rejection rate in the tables are significantly different from the nominal level of 5% only if they are outside of the confidence interval $[0.0365; 0.0635]$

respectively, at $\widetilde{\boldsymbol{\theta}}_{OS,H}$ (TS-I-B) or $\hat{\boldsymbol{\theta}}_{TS}$ (TS-I-C). Analogously, the last three columns are based on OS-H for the first step and on different estimators for the variance-covariance. For the first column (TS-H-A) the variance-covariance is that in Equation (11). For the remaining two columns, the variance-covariance matrix is that in Equation (19), where $\mathbf{S}_n$ is evaluated respectively, at $\widetilde{\boldsymbol{\theta}}_{OS,H}$ (TS-H-B) or $\hat{\boldsymbol{\theta}}_{TS}$ (TS-H-C). This is done to test whether there are improvements in efficiency when the variance-covariance matrix is estimated with the final values of the parameter vector.

Let us first focus on Table 5 and 6, which refer to the spatial autocorrelation coefficient. A glance at Table 5 reveals that for small sample sizes there is actually a problem of the size and this is particularly true for higher values of ρ. Contrarily to the expectation, the two-step estimators TS-I-A and TS-H-A systematically over reject. Surprisingly, for lower-to-medium values of ρ, the one-step based on the identity matrix seems to be the most reliable. It is worth noting that the TS-I-B and the TS-H-B show sizes that, although outside of the confidence interval, are relatively close to the 5% nominal level. A final remark on Table 5 is that the rejection rates for the LGMM when there is substantial spatial dependence are way over the nominal level.

(Insert Table 5 about here)

Moving to Table 6 corresponding to the larger sample size, it is immediately evident that the estimators based on the "optimal" choice of weights in the first step (OS-H, TS-H-A, TS-H-B, and TS-H-C) are performing extremely well also for extreme values of the spatial parameter. Among those, the TS-H-B is the one whose sizes are all within the confidence interval (ranging from 0.042 to 0.052). Generally, the GMM estimators based on the identity matrix in the first step show a problem of the size. Finally, consistently with previous evidence, the figures for the size of the LGMM are way off also in larger sample sizes when the level of spatial autocorrelation is considerable.

(Insert Table 6 about here)

Table 7 ($n = 50$) and 8 ($n = 500$) have to do with the size of a test on β_1. Indeed similar consideration can be made also in this case. Focusing on Table 7, it is evident that inference based on LGMM cannot be done unless the spatial autocorrelation in the data is particularly moderate. The OS-H is still the one that behave best. Out of the remaining estimators (apart from TS-I-B, and TS-H-B), their performance is overall reasonable with some values of the size less than 5% indicating under rejection. For the larger sample size (Table 8), the two-step GMM estimators based on the "optimal" weighting matrix in the first step produce sizes that are the closest to the nominal level, and almost all of them are not statistically different from 5%. It is also remarkable that the one-step estimators fail even with small level of autocorrelation, while the two-step based on the identity matrix deteriorate only for high level of spatial dependence.

(Insert Table 7 about here)

(Insert Table 8 about here)

Summarizing, our Monte Carlo results show some interesting patterns that were not noted in previous studies. First of all, the LGMM is the estimator that presents the largest bias and higher standard deviation among all the considered estimators. This is particularly evident when the level of spatial dependence increases in the data. Second, for what concerns the one-step GMM estimators, the "optimal" GMM (OS-H) consistently outperform the estimator based on the identity matrix. To the best of our knowledge, all previous Monte Carlo experiments only considered the one-step GMM based on the identity matrix and hence GMM has always been deemed as worst compared to other, more efficient, estimators. Third, there is an advantage in considering two-step estimators based on the "optimal" weighting matrix in the first step compare to those that are based on the identity matrix. Fourth, given our Monte Carlo setup, efficiency not always improves when a two-step estimator is adopted, particularly for the small sample size. For what concerns the size of the test for β_1 when $n = 50$, our results indicates that the OS-H is the only estimator that has sizes within the confidence interval for all values of ρ. However, when the sample size increases the two-step GMM estimators based on a "optimal" weighting matrix in the first step are generally very close to the 5% nominal level (apart for a few values that are outside the confidence interval).

5 Empirical application

In this section, we provide an empirical application using LeSage et al. (2011)'s dataset. LeSage et al. (2011) analyze the decision to reopen for a sample of (slightly less than) seven hundred firms in the aftermath of Hurricane Katrina on major business thoroughfares in New Orleans. They assume that the decision to reopen is likely to depend on decisions made by neighboring firms. Thus, the model proposed has the following SAR structure:

$$y_i^* = \rho \sum_{j=1}^{n} w_{ij} y_j^* + \mathbf{x}_i^\top \boldsymbol{\beta} + \epsilon_i, \quad i = 1, ..., n \tag{22}$$

where y_i^* denotes the propensity of reopening the establishment i at same point in time after the disaster; $\sum_{j=1}^{n} w_{ij} y_j^*$ captures the propensity of reopening for the neighbors of unit i; $\mathbf{x}_i$ is a $k \times 1$ vector of explanatory variables; and ϵ_i is the error term assumed to be $\epsilon_i \sim \mathrm{N}(0, 1)$.

The propensity to be open for each firm depends on the difference in profits $y_i^* = \pi_{1i} - \pi_{0i}$, for $i = 1, ..., n$ where π_{1i} represents profits in the open state and π_{i0} in the closed state. The link between the observed choice made by the firm, y_i and the unobserved propensity to be open is

$$y_i = \begin{cases} 1 & \text{if } y_i^* = \pi_{1i} - \pi_{0i} \geq 0, \\ 0 & \text{if } y_i^* = \pi_{1i} - \pi_{0i} < 0. \end{cases} \tag{23}$$

LeSage et al. (2011) use different time horizons to create the dependent variable. Specifically, the dependent variable equals 1 if the firm reopened 3, 6 and 12 months after the hurricane. The spatial weight matrix $\mathbf{W}$ is constructed using the 15-nearest neighbors for each firm.[20] The main covariates are the flood depth (measured in feet) at the location of the firm, the logarithm of median income for the census block group in which the firm was located, two dummy variables indicating small and large size firms, with medium size firms representing the omitted category, two dummy variables indicating low and high socio-economic class of the store's buyers and two dummy variables for type of store ownership, one for sole proprietorship and the other for national chains, with regional chains being the omitted category.

Table 9 shows the results using as dependent variable whether the firm reopened within the 0-6 months time horizon. Each column presents the estimates for different estimation methods. For the GIBBS estimator, we show the posterior mean and the standard deviation in parentheses. For the rest of the estimators the standard errors are reported. The RIS estimator was computed using 1000 draws, whereas the GIBBS estimator was computed using 1000 MCMC. All GMM estimators use $\mathbf{H} = [\mathbf{X}, \mathbf{WX}, \mathbf{W}^2\mathbf{X}]$ as instruments.

(Insert Table 9 about here)

In terms of the covariates, we observe some differences in terms of magnitude of the point estimates and their significance across estimators. Flood depth is significant across all the estimator, except for LGMM and OS-I which are not significant at the traditional level. The logarithm of median income is only significant for the probit estimator. Other things equal, this would indicate that this variable is capturing some sort of spatial dependency of the propensity to reopen across firms. Small size is negative, but only slightly significant for the TS GMM estimators. Firms oriented to low status consumers are on average less likely to be opened than those oriented to medium status consumers for all the estimators. This variable is not significant only when using the OS-I estimator. The spatial autoregressive coefficient is positive and significant across all estimators suggesting a positive spatial dependence in firms' decisions regarding open-closed status. Thus, firms nearby exhibit similar decisions outcomes regarding reopening. As expected, the higher value is obtained when the LGMM is used, whereas the other estimators yield more similar values.

In general, our results are in line with those reported by Calabrese and Elkink (2014), but we show that the TS point-estimates are closer to those more efficient estimates (such as the RIS and GIBBS estimators) which can be beneficial if the sample size is large.

[20] The appropriate number of neighbors was found using the deviance information criterion (LeSage et al., 2011, pag. 1057).

6 Conclusions

This paper dealt with two-step GMM estimators for spatial binary probit model. These estimators were based on different choices of the weighting matrix for the moment conditions in the first step, and different estimators for the variance-covariance matrix of the estimated coefficients. The Monte Carlo experiment comparing one- and two- step GMM estimators, the linearized GMM, and the RIS estimators offered interesting results for potential applications in the area. In particular, the results highlighted that there are benefits related both to the choice of the weight matrix for the moment conditions and in adopting a two-step procedure. The empirical application based on the firm's decision to reopen after the event of the Hurricane Katrina in New Orleans showed that the results obtained from the two-step GMM estimator are in line with the RIS and GIBBS results. Finally, as a possible extension, one could consider GMM estimator based not only on linear but also on quadratic moments to exploit the entire structure of the variance-covariance matrix.

References

Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science. Springer Netherlands.

Beron, K. J. and Vijverberg, W. P. (2004). Probit in a spatial context: a Monte Carlo analysis. In *Advances in spatial econometrics*, pages 169–195. Springer.

Billé, A. G. and Leorato, S. (2020). Partial ml estimation for spatial autoregressive nonlinear probit model with autoregressive disturbances. *Econometric Reviews*, 39(5):437–475.

Calabrese, R. and Elkink, J. A. (2014). Estimators of binary spatial autoregressive models: A Monte Carlo study. *Journal of Regional Science*, 54(4):664–687.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Casetti, E. (1972). Generating Models by the Expansion method: Applications to Geographical Research. *Geographical Analysis*, 4:81–91.

Chesher, A. and Irish, M. (1987). Residual analysis in the grouped and censored normal linear model. *Journal of Econometrics*, 34(1-2):33–61.

Cox, D. R. and Snell, E. J. (1968). A general definition of residuals. *Journal of the Royal Statistical Society: Series B (Methodological)*, 30(2):248–265.

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

Gourieroux, C., Monfort, A., Renault, E., and Trognon, A. (1987). Generalised residuals. *Journal of econometrics*, 34(1-2):5–32.

Hall, P. and Huang, L.-S. (2002). Unimodal density estimation using kernel methods. *Statistica Sinica*, pages 965–990.

Kelejian, H. H. and Prucha, I. R. (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics*, 17(1):99–121.

Kelejian, H. H., Prucha, I. R., and Yuzefovich, Y. (2004). Instrumental variable estimation of a spatial autoregressive model with autoregressive disturbances: Large and small sample results. *Advances in Econometrics: Spatial and Spatio-Temporal econometrics*, pages 163–198.

Klier, T. and McMillen, D. P. (2008). Clustering of auto supplier plants in the united states: generalized method of moments spatial logit for large samples. *Journal of Business & Economic Statistics*, 26(4):460–471.

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

LeSage, J. P., Kelley Pace, R., Lam, N., Campanella, R., and Liu, X. (2011). New Orleans business recovery in the aftermath of Hurricane Katrina. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(4):1007–1027.

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

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

Santos, L. S. and Proença, I. (2019). The inversion of the spatial lag operator in binary choice models: Fast computation and a closed formula approximation. *Regional Science and Urban Economics*, 76:74–102.

Vijverberg, W. (1997). Monte Carlo evaluation of multivariate normal probabilities. *Journal of Econometrics*, 76:281–307.

Figures

Figure 1: Boxplot for $\widehat{\rho}$: $n = 50$

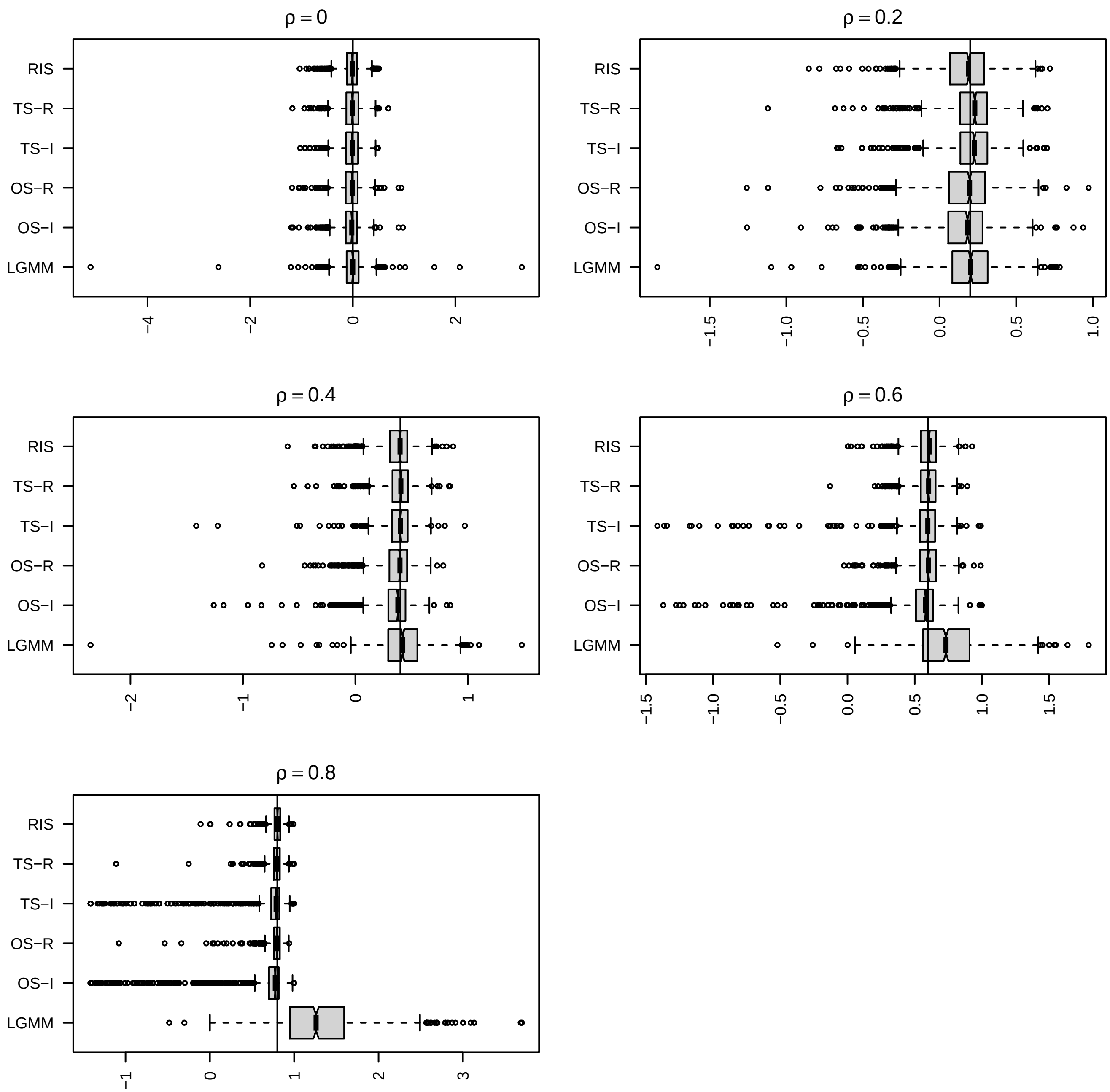
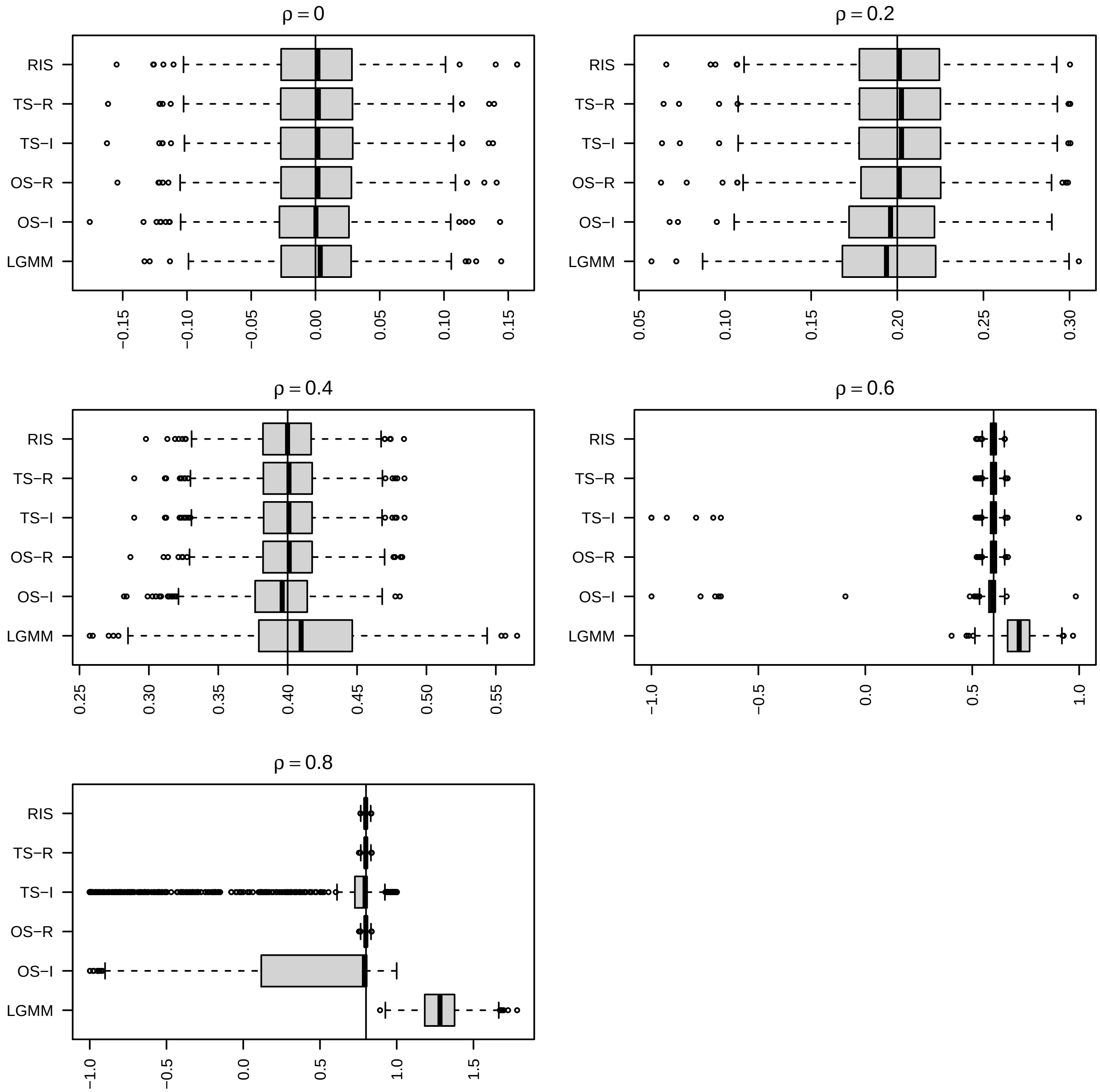


Figure 2: Boxplot for $\widehat{\rho}$: $n = 500$



Tables

Table 1: Simulation results for estimates of $\widehat{\rho}$: $n = 50$

	LGMM	OS-I	OS-H	TS-I	TS-H	RIS
			$\rho = 0$			
Mean Bias	-0.009	-0.035	-0.029	-0.027	-0.023	-0.025
Median Bias	-0.000	-0.019	-0.013	-0.012	-0.010	-0.008
SD	0.335	0.202	0.211	0.201	0.207	0.191
RMSE	0.335	0.205	0.213	0.203	0.208	0.192
RMSE Median	0.173	0.166	0.172	0.172	0.176	0.150
			$\rho = 0.2$			
Mean Bias	-0.008	-0.043	-0.032	0.007	0.006	-0.026
Median Bias	0.003	-0.018	-0.004	0.026	0.031	-0.010
SD	0.211	0.193	0.198	0.167	0.173	0.177
RMSE	0.212	0.197	0.201	0.167	0.173	0.179
RMSE Median	0.171	0.167	0.175	0.134	0.132	0.166
			$\rho = 0.4$			
Mean Bias	0.014	-0.056	-0.037	-0.015	-0.008	-0.029
Median Bias	0.021	-0.022	-0.004	0.001	0.005	-0.003
SD	0.230	0.183	0.166	0.154	0.130	0.157
RMSE	0.230	0.191	0.170	0.155	0.131	0.160
RMSE Median	0.193	0.115	0.116	0.102	0.101	0.115
			$\rho = 0.6$			
Mean Bias	0.137	-0.064	-0.008	-0.040	-0.005	-0.004
Median Bias	0.133	-0.019	0.002	-0.002	0.004	0.007
SD	0.264	0.239	0.109	0.228	0.095	0.106
RMSE	0.297	0.248	0.109	0.231	0.095	0.106
RMSE Median	0.288	0.096	0.089	0.083	0.081	0.083
			$\rho = 0.8$			
Mean Bias	0.499	-0.181	-0.022	-0.131	-0.016	-0.007
Median Bias	0.459	-0.025	-0.000	-0.013	-0.003	0.002
SD	0.511	0.456	0.121	0.405	0.100	0.090
RMSE	0.714	0.491	0.123	0.426	0.101	0.090
RMSE Median	0.663	0.089	0.052	0.072	0.053	0.051

Table 2: Simulation results for estimates of $\widehat{\rho}$: $n = 500$

	LGMM	OS-I	OS-H	TS-I	TS-H	RIS
			$\rho = 0$			
Mean Bias	0.001	-0.001	0.001	0.001	0.001	0.001
Median Bias	0.004	0.000	0.002	0.002	0.002	0.002
SD	0.040	0.041	0.040	0.041	0.041	0.040
RMSE	0.040	0.041	0.040	0.041	0.041	0.040
RMSE Median	0.040	0.040	0.040	0.041	0.041	0.041
			$\rho = 0.2$			
Mean Bias	-0.005	-0.004	0.001	0.001	0.001	0.001
Median Bias	-0.006	-0.004	0.001	0.002	0.002	0.001
SD	0.038	0.036	0.035	0.035	0.035	0.034
RMSE	0.039	0.036	0.035	0.035	0.035	0.034
RMSE Median	0.041	0.037	0.034	0.035	0.035	0.034
			$\rho = 0.4$			
Mean Bias	0.011	-0.006	-0.000	0.000	0.000	-0.000
Median Bias	0.010	-0.004	0.001	0.001	0.001	-0.000
SD	0.052	0.029	0.027	0.027	0.027	0.027
RMSE	0.053	0.030	0.027	0.027	0.027	0.027
RMSE Median	0.051	0.028	0.026	0.026	0.026	0.026
			$\rho = 0.6$			
Mean Bias	0.119	-0.014	-0.001	-0.010	-0.001	-0.002
Median Bias	0.120	-0.006	-0.001	-0.001	-0.000	-0.001
SD	0.077	0.102	0.020	0.115	0.020	0.020
RMSE	0.142	0.103	0.020	0.115	0.020	0.020
RMSE Median	0.142	0.023	0.020	0.020	0.019	0.019
			$\rho = 0.8$			
Mean Bias	0.482	-0.302	-0.001	-0.275	-0.001	-0.002
Median Bias	0.482	-0.010	-0.001	-0.006	-0.001	-0.002
SD	0.143	0.553	0.013	0.573	0.013	0.012
RMSE	0.503	0.630	0.013	0.636	0.013	0.012
RMSE Median	0.503	0.504	0.013	0.059	0.013	0.012

Table 3: Simulation results for estimates of $\widehat{\beta}_1$: $n = 50$

	Probit	LGMM	OS-I	OS-H	TS-I	TS-H	RIS
			$\rho = 0$				
Mean Bias	-51.838	-53.823	-51.994	-52.056	-2.596	-2.906	-56.515
Median Bias	-1.022	-1.473	-1.429	-1.738	-0.839	-0.801	-3.992
SD	491.727	524.796	491.711	491.706	6.479	5.201	492.402
RMSE	494.452	527.549	494.453	494.454	6.980	5.958	495.635
RMSE Median	13.018	16.047	12.958	12.938	3.250	3.141	21.713
			$\rho = 0.2$				
Mean Bias	-24.943	-26.679	-25.314	-25.553	-3.199	-3.240	-32.723
Median Bias	-0.208	0.003	-0.788	-1.034	-1.472	-1.321	-6.722
SD	149.922	161.306	149.861	149.864	4.417	5.361	150.683
RMSE	151.982	163.497	151.984	152.027	5.454	6.264	154.195
RMSE Median	3.413	6.446	3.581	4.312	4.169	4.093	20.596
			$\rho = 0.4$				
Mean Bias	-4.604	-2.972	-5.514	-5.693	-2.959	-3.005	-14.226
Median Bias	0.818	0.861	-0.247	-0.376	-2.114	-2.019	-4.977
SD	46.562	62.601	46.471	46.463	3.490	3.643	47.563
RMSE	46.789	62.671	46.797	46.810	4.575	4.722	49.645
RMSE Median	1.037	1.136	1.105	1.568	4.167	4.375	13.850
			$\rho = 0.6$				
Mean Bias	-3.772	-4.127	-5.244	-5.260	-2.670	-2.500	-14.516
Median Bias	1.319	1.303	0.059	-0.043	-2.007	-1.909	-4.727
SD	145.765	154.420	145.722	145.718	3.666	2.892	146.422
RMSE	145.814	154.475	145.817	145.813	4.536	3.823	147.140
RMSE Median	1.342	1.334	0.793	0.984	3.443	3.476	10.455
			$\rho = 0.8$				
Mean Bias	1.480	1.433	-1.017	-0.028	-1.639	-1.596	-8.467
Median Bias	1.621	1.604	0.205	0.201	-1.283	-1.219	-3.971
SD	1.664	1.822	3.966	1.784	21.073	2.114	13.106
RMSE	2.227	2.318	4.094	1.784	21.137	2.649	15.603
RMSE Median	1.626	1.611	1.042	0.806	2.583	2.359	8.376

Table 4: Simulation results for estimates of $\widehat{\beta}_1$: $n = 500$

	Probit	LGMM	OS-I	OS-H	TS-I	TS-H	RIS
			$\rho = 0$				
Mean Bias	-0.089	0.145	-0.070	-0.045	-0.107	-0.105	-0.119
Median Bias	-0.038	0.112	-0.035	-0.011	-0.053	-0.049	-0.063
SD	0.332	0.881	0.372	0.374	0.342	0.341	0.347
RMSE	0.344	0.892	0.379	0.377	0.359	0.357	0.367
RMSE Median	0.296	0.590	0.354	0.337	0.307	0.308	0.306
			$\rho = 0.2$				
Mean Bias	0.421	0.426	0.063	0.166	-0.121	-0.115	-0.114
Median Bias	0.447	0.434	0.094	0.201	-0.060	-0.056	-0.053
SD	0.211	0.390	0.310	0.316	0.362	0.360	0.356
RMSE	0.471	0.577	0.316	0.357	0.381	0.378	0.374
RMSE Median	0.490	0.534	0.304	0.345	0.316	0.314	0.320
			$\rho = 0.4$				
Mean Bias	1.058	0.976	0.193	0.200	-0.103	-0.098	-0.016
Median Bias	1.068	0.978	0.217	0.236	-0.052	-0.045	0.036
SD	0.104	0.149	0.303	0.356	0.369	0.365	0.358
RMSE	1.063	0.988	0.359	0.408	0.383	0.378	0.358
RMSE Median	1.072	0.987	0.359	0.394	0.328	0.320	0.300
			$\rho = 0.6$				
Mean Bias	1.432	1.377	0.285	0.236	-0.217	-0.143	0.136
Median Bias	1.437	1.385	0.327	0.288	-0.086	-0.085	0.175
SD	0.057	0.072	0.501	0.335	1.412	0.404	0.324
RMSE	1.433	1.379	0.576	0.410	1.429	0.428	0.352
RMSE Median	1.438	1.387	0.423	0.419	0.363	0.357	0.337
			$\rho = 0.8$				
Mean Bias	1.682	1.654	-3.129	0.340	-3.346	-0.198	0.565
Median Bias	1.684	1.658	0.263	0.383	-0.394	-0.049	0.604
SD	0.031	0.037	6.335	0.327	217.823	0.666	0.283
RMSE	1.682	1.655	7.066	0.472	217.849	0.694	0.632
RMSE Median	1.684	1.658	5.286	0.483	7.541	0.490	0.661

Table 5: Rejection rates for $\widehat{\rho}$: $n = 50$

	LGMM	OS-I	OS-H	TS-I-A	TS-I-B	TS-I-C	TS-H-A	TS-H-B	TS-H-C
$\rho = 0$	0.026	0.078	0.091	0.163	0.111	0.187	0.164	0.105	0.181
$\rho = 0.2$	0.035	0.060	0.080	0.143	0.078	0.159	0.138	0.079	0.163
$\rho = 0.4$	0.079	0.035	0.049	0.104	0.057	0.133	0.104	0.064	0.123
$\rho = 0.6$	0.151	0.056	0.035	0.113	0.048	0.150	0.093	0.042	0.117
$\rho = 0.8$	0.420	0.140	0.027	0.124	0.063	0.195	0.065	0.031	0.132

Table 6: Rejection rates for $\widehat{\rho}$: $n = 500$

	LGMM	OS-I	OS-H	TS-I-A	TS-I-B	TS-I-C	TS-H-A	TS-H-B	TS-H-C
$\rho = 0$	0.048	0.046	0.040	0.056	0.052	0.056	0.056	0.052	0.056
$\rho = 0.2$	0.046	0.039	0.038	0.062	0.051	0.062	0.063	0.046	0.063
$\rho = 0.4$	0.081	0.028	0.033	0.056	0.040	0.056	0.057	0.044	0.057
$\rho = 0.6$	0.518	0.033	0.040	0.076	0.048	0.079	0.068	0.050	0.068
$\rho = 0.8$	0.992	0.379	0.037	0.356	0.322	0.449	0.082	0.042	0.083

Table 7: Rejection rates for $\widehat{\beta}_1$: $n = 50$

	LGMM	OS-I	OS-H	TS-I-A	TS-I-B	TS-I-C	TS-H-A	TS-H-B	TS-H-C
$\rho = 0$	0.014	0.024	0.043	0.020	0.031	0.139	0.028	0.052	0.154
$\rho = 0.2$	0.053	0.042	0.052	0.025	0.046	0.102	0.031	0.052	0.138
$\rho = 0.4$	0.269	0.045	0.052	0.020	0.029	0.135	0.020	0.030	0.137
$\rho = 0.6$	0.762	0.070	0.042	0.026	0.019	0.121	0.015	0.019	0.122
$\rho = 0.8$	0.966	0.159	0.057	0.065	0.060	0.117	0.030	0.032	0.104

Table 8: Rejection rates for $\widehat{\beta}_1$: $n = 500$

	LGMM	OS-I	OS-R	TS-I-R	TS-I-E	TS-I-E2	TS-R-R	TS-R-E	TS-R-E2
$\rho = 0$	0.056	0.025	0.020	0.047	0.050	0.047	0.047	0.040	0.047
$\rho = 0.2$	0.266	0.044	0.083	0.037	0.019	0.038	0.036	0.019	0.036
$\rho = 0.4$	0.993	0.122	0.171	0.047	0.029	0.047	0.046	0.031	0.046
$\rho = 0.6$	1.000	0.192	0.160	0.056	0.030	0.069	0.048	0.029	0.050
$\rho = 0.8$	1.000	0.493	0.184	0.314	0.293	0.358	0.046	0.038	0.046

Table 9: SARB estimates 0-6 months time horizon. Data from LeSage et al. (2011)

	Probit	RIS	GIBBS	LGMM	OS-I	OS-H	TS-I-A	TS-I-B	TS-H-A	TS-H-B
Constant	-8.327^{**}	-3.198	-2.877	2.177	-8.331	-1.346	-1.122	-1.122	-1.294	-1.294
	(2.792)	(2.716)	(2.341)	(4.528)	(6.636)	(1.213)	(0.957)	(0.901)	(1.119)	(1.121)
Flood depth	-0.261^{***}	-0.106^{**}	-0.106^{***}	0.026	-0.084	-0.077^{**}	-0.059^{**}	-0.059^{**}	-0.069^{**}	-0.069^{**}
	(0.034)	(0.037)	(0.032)	(0.105)	(0.056)	(0.031)	(0.021)	(0.021)	(0.027)	(0.027)
Log(Median income)	0.886^{**}	0.335	0.302	-0.226	0.840	0.159	0.137	0.137	0.153	0.153
	(0.273)	(0.268)	(0.230)	(0.469)	(0.664)	(0.122)	(0.095)	(0.090)	(0.112)	(0.113)
Small size	-0.128	-0.134	-0.094	-0.161	-0.215	-0.212	-0.242^{*}	-0.242^{*}	-0.245^{*}	-0.245^{*}
	(0.144)	(0.157)	(0.154)	(0.121)	(0.142)	(0.130)	(0.131)	(0.129)	(0.129)	(0.129)
Large size	-0.456	-0.509	-0.403	-0.410^{*}	-0.447	-0.411	-0.393	-0.393	-0.415	-0.415
	(0.295)	(0.463)	(0.315)	(0.243)	(0.333)	(0.298)	(0.298)	(0.289)	(0.295)	(0.296)
Low status customers	-0.513^{**}	-0.342^{**}	-0.336^{**}	-0.311^{**}	-0.202	-0.351^{**}	-0.291^{**}	-0.291^{**}	-0.313^{**}	-0.313^{**}
	(0.156)	(0.171)	(0.154)	(0.155)	(0.190)	(0.129)	(0.099)	(0.097)	(0.118)	(0.117)
High status customers	0.086	0.035	0.039	0.058	0.020	-0.001	-0.034	-0.034	0.008	0.008
	(0.150)	(0.163)	(0.150)	(0.124)	(0.137)	(0.123)	(0.118)	(0.116)	(0.120)	(0.120)
Sole proprietorship	0.312^{*}	0.363^{*}	0.333^{*}	0.302^{*}	0.373^{**}	0.238	0.207	0.207	0.239	0.239
	(0.182)	(0.189)	(0.175)	(0.162)	(0.177)	(0.158)	(0.151)	(0.147)	(0.155)	(0.155)
National chain	0.154	0.275	0.294	0.213	0.178	-0.231	-0.560	-0.560	-0.341	-0.341
	(0.347)	(0.574)	(0.393)	(0.267)	(0.414)	(0.389)	(0.399)	(0.379)	(0.389)	(0.388)
ρ		0.624^{***}	0.586^{***}	1.028^{**}	0.584^{**}	0.752^{***}	0.843^{***}	0.843^{***}	0.782^{***}	0.782^{***}
		(0.126)	(0.076)	(0.369)	(0.283)	(0.131)	(0.099)	(0.097)	(0.120)	(0.120)
N	673	673	673	673	673	673	673	673	673	673

Significance: $* \equiv p < 0.1$; $** \equiv p < 0.05$; $*** \equiv p < 0.001$

A Appendix: Derivation of S

Since $\mathbb{E}(u_{i,n}) = 0$ and $\text{Var}(y_{i,n} - \Phi(a_{i,n})) = \Phi(a_{i,n})(1 - \Phi(a_{i,n}))$, the variance for the generalized residuals in Equation (6) is:

$$\begin{aligned}
\text{Var}(\widetilde{u}_{i,n}) &= \mathbb{E}(\widetilde{u}_{i,n}^2) \\
&= \mathbb{E}\left(\left[\phi(a_{i,n})^2 \cdot \left[\frac{y_{i,n} - \Phi(a_{i,n})}{\Phi(a_{i,n})(1 - \Phi(a_{i,n}))}\right]^2\right]\right) \\
&= \mathbb{E}\left(\left[\frac{\phi(a_{i,n})}{\Phi(a_{i,n})(1 - \Phi(a_{i,n}))}\right]^2 \mathbb{E}_{y_{i,n}|\mathbf{x}_{i,n}}\left(\{y_{i,n} - \Phi(a_{i,n})\}^2 |\mathbf{x}_{i,n}\right)\right) \\
&= \mathbb{E}\left(\left[\frac{\phi(a_{i,n})}{\Phi(a_{i,n})(1 - \Phi(a_{i,n}))}\right]^2 \Phi(a_{i,n})\left(1 - \Phi(a_{i,n})\right)\right) \\
&= \frac{\phi(a_{i,n})^2}{\Phi(a_{i,n})(1 - \Phi(a_{i,n}))}
\end{aligned} \tag{A.1}$$

Then,

$$\begin{aligned}
\text{Var}\left(\mathbf{H}_n^\top \widetilde{\mathbf{u}}_n\right) &= \mathbf{H}_n^\top \text{Var}(\widetilde{\mathbf{u}}_n)\mathbf{H}_n \\
&= \mathbf{H}_n^\top \mathbf{T}_n \mathbf{H}_n
\end{aligned}$$

where $\mathbf{T}_n$ is a diagonal matrix whose elements are given by (A.1).

B Appendix: Derivation of G

In this section we drop the subscript n to simplify the notation. Let the generalized residuals given by:

$$\widetilde{\mathbf{u}} = \mathbf{q} \cdot \frac{\phi(\mathbf{q} \cdot \mathbf{a})}{\Phi(\mathbf{q} \cdot \mathbf{a})} \tag{B.1}$$

where $\mathbf{a}$ is given in Equation (5) and $\mathbf{q} = 2 \cdot \mathbf{y} - \boldsymbol{\imath}_n$ where $\boldsymbol{\imath}_n$ is an n-vector of ones. Taking the derivative of Equation (B.1) with respect to $\boldsymbol{\beta}$, we obtain:

$$\frac{\partial \widetilde{\mathbf{u}}}{\partial \boldsymbol{\beta}^\top} = \mathbf{q}^2 \cdot \left[\frac{\phi'(\mathbf{q} \cdot \mathbf{a}) \cdot \Phi(\mathbf{q} \cdot \mathbf{a}) - \phi(\mathbf{q} \cdot \mathbf{a})^2}{\Phi(\mathbf{q} \cdot \mathbf{a})^2}\right] \cdot \mathbf{D}_\rho^{-1}\mathbf{A}_\rho^{-1}\mathbf{X} \tag{B.2}$$

where $\phi'(z) = -z\phi(z)$. Taking the derivative of Equation (B.1) with respect to ρ, we obtain:

$$\frac{\partial \widetilde{\mathbf{u}}}{\partial \rho} = \mathbf{q}^2 \cdot \left[\frac{\phi'(\mathbf{q} \cdot \mathbf{a}) \cdot \Phi(\mathbf{q} \cdot \mathbf{a}) - \phi(\mathbf{q} \cdot \mathbf{a})^2}{\Phi(\mathbf{q} \cdot \mathbf{a})^2}\right] \cdot \frac{\partial \mathbf{a}}{\partial \rho}$$

where:

$$\frac{\partial \mathbf{a}}{\partial \rho} = \left[\mathbf{A}_\rho^{-1}\mathbf{W}\mathbf{a} - \mathbf{D}_\rho^{-1}\left[\frac{\partial \mathbf{D}}{\partial \rho}\right]\mathbf{a}\right] \tag{B.3}$$

and:

$$
\begin{aligned}
\frac{\partial \mathbf{D}_\rho}{\partial \rho} &= \frac{\partial \operatorname{diag}(\boldsymbol{\sigma}_{\mathbf{u}})}{\partial \rho} \\
&= \frac{1}{2}\mathbf{D}_\rho^{-1} \operatorname{diag}\left[\left((\mathbf{A}_\rho^\top \mathbf{A}_\rho)^{-1}\left[\mathbf{A}_\rho + \mathbf{A}_\rho^\top\right]\mathbf{W}(\mathbf{A}_\rho^\top \mathbf{A}_\rho)^{-1}\right)\right]
\end{aligned}
\tag{B.4}
$$

If $\mathbf{A}_\rho$ is symmetric, then:

$$
\frac{\partial \mathbf{D}_\rho}{\partial \rho} = \mathbf{D}_\rho^{-1} \operatorname{diag}(\mathbf{A}_\rho^{-1}\mathbf{W}\mathbf{A}_\rho^{-1}\mathbf{A}_\rho^{-1}) \tag{B.5}
$$