Empir Econ
DOI 10.1007/s00181-017-1330-6

Spatial econometric Monte Carlo studies: raising the bar

James P. LeSage[1] · R. Kelley Pace[2]

Received: 2 November 2016 / Accepted: 5 July 2017
© Springer-Verlag GmbH Germany 2017

Abstract We discuss Monte Carlo methodology that can be used to explore alternative approaches to estimating spatial regression models. Our focus is on models that include spatial lags of the dependent variable, e.g., the SAR specification. A major point is that practitioners rely on scalar summary measures of *direct and indirect effects* estimates to interpret the impact of changes in explanatory variables on the dependent variable of interest. We argue that these should be the focus of Monte Carlo experiments. Since *effects estimates* reflect a nonlinear function of both β and ρ, past studies' focus exclusively on β and ρ parameter estimates may not provide useful information regarding statistical properties of *effects estimates* produced by alternative estimators. Since effects estimates have recently become the focus of inference regarding the significance of (scalar summary) direct and indirect impacts arising from changes in the explanatory variables, empirical measures of dispersion produced by simulating draws from the (estimated) variance–covariance matrix of the parameters β and ρ should be part of the Monte Carlo study. An implication is that differences in the quality of estimated variance–covariance matrices arising from alternative estimators also plays a role in determining the accuracy of inference. An applied illustration is used to demonstrate how these issues can impact conclusions regarding the performance of alternative estimators.

This research was partially supported by a grant from the National Science Foundation (Award No. 1212112).

✉ James P. LeSage
james.lesage@txstate.edu

1 Finance and Economics, Texas State University, San Marcos, TX, USA

2 Department of Finance, Louisiana State University, Baton Rouge, LA, USA

Published online: 31 August 2017

Keywords Spatial autoregressive models · Simulation of scalar summary direct and indirect effects · Estimated variance–covariance matrices

JEL Classification C11 · C18

1 Introduction

According to Elhorst (2010), the period 2007–2009 ushered in a "sea change" in spatial econometrics, with focus on models that included more than one spatial interaction effect. For example, Elhorst notes that Kelejian in his 2007 keynote address to the first World Conference of the Spatial Econometrics Association advocated models that include both a spatially lagged dependent variable plus a spatial lag of the error term. In contrast, LeSage in his 2007 presidential address at the 54th North American Meeting of the Regional Science Association International advocated models that included spatial lags of the dependent and explanatory variables. Such models have often been incorrectly interpreted in the past. Elhorst (2010) in his review of the 2009 book *Introduction to Spatial Econometrics* by LeSage and Pace entitled: "Applied spatial econometrics: raising the bar" provides a lucid presentation of the issues and the interpretive approach followed by LeSage and Pace.[1]

LeSage and Pace (2009) proposed summary estimates for the impacts arising from changes in explanatory variables of spatial regression models containing spatial lags of the dependent (and possibly spatial lags of independent) variables. These would include the spatial autoregessive (SAR) specification: $y = \rho W y + X\beta + \varepsilon$, or spatial Durbin model (SDM) specification: $y = \rho W y + X\beta + W X\theta + \varepsilon$, [see LeSage and Pace (2009)]. These reflect scalar summary measures of the own- and cross-partial derivatives from these models, showing how the dependent variable in (the typical) region i responds to changes in characteristics of (the typical) region i (own-partials), and how neighboring regions j respond to the same change in characteristics of region i (cross-partials).

They labeled these direct and indirect effects estimates, respectively, and these are calculated by most spatial regression software (e.g., the Spatial Econometrics Toolbox for MATLAB software by LeSage,[2] the R-language package spdep by Roger Bivand,[3] and the Stata code described in Drukker et al. (2013a, b) all report these estimates).[4] In addition to reporting scalar summary estimates, these software provide empirically derived measures of dispersion used for inference regarding the statistical significance of direct and indirect (spillover) effects. Because of the widely available software for spatial regression models, it is now almost universal practice to properly interpret estimates from spatial regression models, rather than rely on past practice where β *coefficients* were incorrectly interpreted *as if they represented partial derivative responses* of the dependent variable to changes in the explanatory variables.

[1] The title of Elhorst's review inspired the title of this manuscript.

[2] http://www.spatial-econometrics.com.

[3] https://r-forge.r-project.org/projects/spdep/.

[4] We provide more details regarding effects estimates in the next section.

In light of these recent changes in applied practice regarding spatial regression models, Monte Carlo studies should assess performance of alternative estimators using a comparison of bias and mean squared error (MSE) of the parameter estimates of β and ρ as well as the effects estimates on which practitioners base their inferences about the impact of explanatory variables. Since the effects estimates involve nonlinear transformations of possibly covarying underlying parameter estimates of (β and ρ), assessing bias or MSE for the parameters estimates of β and ρ may not provide useful information regarding the statistical operating characteristics of alternative estimators when it comes to the effects estimates used by practitioners. An implication is that Monte Carlo practice should examine the distribution of the effects estimates.

A related point is that empirical estimates of dispersion for the *effects estimates* used for inference regarding the significance of (scalar summary) direct and indirect impacts arising from changes in the explanatory variables are produced by simulating from the (estimated) variance–covariance matrix of the parameter estimates of β, ρ and σ_ε. An implication is that differences in the quality of estimated variance–covariance matrices arising from alternative estimators also plays a role in the accuracy of inference, and should be part of the Monte Carlo study.

Section 2 provides background information on the relationship between underlying parameters β, θ, ρ and the effects estimates for the SAR and SDM spatial regression models. Section 3 focuses on issues pertaining to Monte Carlo approaches that can be used to produce evidence regarding empirical measures of dispersion calculated by alternative spatial regression estimators for the effects estimates that are used for inference. Since these measures of dispersion are empirically determined using simulated values for the underlying parameters β, θ, ρ, σ_ε from the (estimated) variance–covariance matrix, alternative variance–covariance expressions can have a large impact on the empirical estimates of dispersion arising from alternative estimation procedures. Section 4 carries out an applied illustration based on a recent Monte Carlo study by Dogan and Taspinar (2014) to demonstrate how these issues impact conclusions regarding the performance of alternative estimators.

2 Background

For concreteness in our discussion we let a spatial regression (SAR) model take the form:

$$
\begin{aligned}
y &= \alpha\iota_n + \rho W y + X\beta + \varepsilon \\
\varepsilon &\sim N\left(0, \sigma_\varepsilon^2 I_n\right)
\end{aligned}
\tag{1}
$$

where y is an $n \times 1$ vector of observations on the dependent variable and X is an $n \times k$ matrix of observations on the explanatory variables. Each of these observations on the dependent and explanatory variables comes from regions or points in space. Also, β is a $k \times 1$ vector of parameters associated with the explanatory variables, ι_n is a vector of ones and α is the associated intercept parameter. The n disturbances ε are distributed normally with constant variance σ_ε^2 and zero covariance across observations. The spatial Durbin model (SDM) replaces the matrix X with $(\, X W X\,)$. The matrix W is the spatial weight matrix that contains positive elements w_{ij} if observations j and i are *neighbors* and zero otherwise. Typically, the matrix W is normalized to have row sums of unity, so the $n \times 1$ *spatial lag vector* Wy contains values constructed from a linear combination of values from *neighboring observations*. Elements of the weight matrix W are assumed to be fixed or nonstochastic, which means the elements are not random variables.

LeSage and Pace (2009) propose cumulative scalar summary measures of the impacts or partial derivative responses of the dependent variable to changes in the explanatory variables in spatial lag regression models. The partial derivatives take the form of $n \times n$ matrices shown in (2) for the kth explanatory variable in the SAR specification and (3) for the SDM.

$$\partial y / \partial X_k' = (I_n - \rho W)^{-1} I_n \beta_k \tag{2}$$

$$\partial y / \partial X_k' = (I_n - \rho W)^{-1} (I_n \beta_k + W \theta_k) \tag{3}$$

To produce manageable estimates that mirror standard regression practice, LeSage and Pace (2009) summarize the $n \times n$ matrix of partial derivatives using an average (or median) of the main diagonal elements reflecting own-partial derivatives: $\partial y_i / \partial X_{i,k}$ which they label *direct effects*. The cumulative sum of off-diagonal elements are also averaged to produce a scalar summary of the cross-partial derivatives: $\partial y_j / \partial X_{i,k}$, which they label *indirect effects*.[5]

We note that the SAR and SDM models allow for spillovers from neighbors to neighbors of neighbors, and so on, which accounts for the popularity of these models in regional science applications. To see this, consider the infinite series expansion of the matrix inverse: $(I_n - \rho W)^{-1}$ shown in (4).

$$(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 \ldots \tag{4}$$

The matrix W reflects neighboring regions, W^2 neighbors to the neighboring regions, W^3 neighbors to the neighbors of W^2, and so on. Since the spatial dependence parameter is bounded by $\rho < 1$, the matrix inverse assigns declining influence to higher-order neighbors.

An implication of the matrix inverse is that the scalar summary measures of direct and indirect effects arising from changes in the kth explanatory variable are nonlinear functions of the parameters ρ and β_k (and θ_k in the case of the SDM specification). Monte Carlo studies that focus on bias and MSE assessments for the parameter estimates β, ρ (in the case of the SAR model), might not adequately reflect bias and MSE in the effects estimates used by practitioners. The goal of most applied spatial regression analysis is to determine how the dependent variable responds to changes in the explanatory variables, information provided by the (scalar summary) effects estimates. Since the parameter estimates β are not useful for this purpose, some recent studies using spatial regression models do not report estimates of the parameters β (e.g., LeSage and Sheng 2014). For studies that do report these estimates, there is typically no discussion of the signs or statistical significance of the estimates for β or θ, since these may not reflect the signs or significance of the effects estimates.[6]

[5] One can also use the median rather than average of the cumulative sum of off-diagonal elements.

[6] This is true for the SDM specification, but not for the SAR specification where the signs of the coefficients β will have the same sign as both the direct and indirect effects. However even in the case of the SAR

Given this disconnect between the object of inference represented by the (scalar summary) effects estimates, and the parameter estimates of β, ρ or in the case of SDM β, ρ, θ, we propose that Monte Carlo studies focus their comparative assessment of alternative estimation methods on effects estimates rather than the underlying parameter estimates of β and ρ.

We note that our focus here is on Monte Carlo practice as it relates to cross-sectional spatial regression models involving continuous dependent variables. However, the proposals set forth also apply to scalar summary measures of effects estimates that have been proposed for the case of spatial regression variants of probit models, LeSage et al. (2011), as well as spatial regression versions of spatial interaction (gravity) models, LeSage and Thomas-Agnan (2015), static spatial regression panel models, Elhorst (2014), and space-time dynamic spatial panel models, Debarsy et al. (2012).

Our focus here is on the issue of Monte Carlo methodology as it applies to examination of effects estimates and associated measures of dispersion for these estimates. Therefore, we avoid discussion of other potential issues that arise when designing a Monte Carlo study. For example, issues pertaining to design of spatial weight matrix specifications, use of spatially dependent versus independent explanatory variables, controlling for signal-to-noise, etc.

3 Proper statistical summaries of Monte Carlo outcomes

Practitioners motivation for use of spatial regression models that include spatial lags of the dependent variable is often inference regarding the magnitude and significance of spatial spillovers. Because of this, indirect effects estimates as well as measures of dispersion for these should be a focus of Monte Carlo studies.

Let V represent the variance–covariance matrix pertaining to estimates from the SDM model in (5), and let τ represent the total effect, which is the sum of direct and indirect effects.

$$y = X\beta + WX\theta + \rho Wy + \varepsilon \tag{5}$$

$$\tau = \frac{\beta - \theta}{1 - \rho} \tag{6}$$

$$V = \begin{pmatrix} \mathrm{var}(\tilde{\beta}) & \mathrm{cov}(\tilde{\beta}, \tilde{\theta}) & \mathrm{cov}(\tilde{\beta}, \tilde{\rho}) \\ \mathrm{cov}(\tilde{\theta}, \tilde{\beta}) & \mathrm{var}(\tilde{\theta}) & \mathrm{cov}(\tilde{\theta}, \tilde{\rho}) \\ \mathrm{cov}(\tilde{\rho}, \tilde{\beta}) & \mathrm{cov}(\tilde{\rho}, \tilde{\theta}) & \mathrm{var}(\tilde{\rho}) \end{pmatrix} \tag{7}$$

A Monte Carlo study could examine the distributions of the estimates $\tilde{\beta}, \tilde{\theta}\ \tilde{\rho}$ for different estimators. This information provides an incomplete picture when the desire is to examine the distribution of statistics such as $\tilde{\tau}$ for different estimators, since τ is a nonlinear function of the parameters β, θ, ρ. Conclusions regarding bias and mean squared error of parameters β, θ, ρ do not necessarily reflect bias and mean squared

Footnote 6 continued
specification, the statistical significance of the coefficients β will not necessarily relate to the statistical significance of the direct and indirect effects associated with a particular explanatory variable.

error of (scalar summary) effects estimates like the total effect τ (or the direct and indirect effects component parts).

In general, the covariance of estimates will play a role. For example, when the explanatory variables in the matrix X are spatially dependent, they will be positively correlated with those in WX, leading to nonzero covariance between $\tilde{\beta}$ and $\tilde{\theta}$. It is also the case that $\tilde{\rho}$ may not vary independently from $\tilde{\beta}$ and $\tilde{\theta}$.

For example, in the case of the SAR specification, negative covariance between the parameter β^r for the rth explanatory variable and ρ implies that combining these two parameters in the matrix expression for the effects associated with the rth explanatory variable: $\partial y/\partial X^r = (I_n - \rho W)^{-1}\beta^r$ allows upward bias in the parameter β^r to be offset by downward bias in ρ (and vice versa) when calculating effects estimates.

In addition, measures of dispersion related to parameters β, ρ will not reflect the empirically derived measures of dispersion for the effects estimates. There is an important role played by the estimated variance–covariance matrix used to simulate draws of the parameters β, ρ that are in turn used to determine empirical measures of dispersion for the effects. Ultimately, inference regarding the significance of spatial spillovers may depend heavily on alternative estimated variance–covariance matrices used by different estimators.

Even for a given estimation method such as maximum likelihood, different approaches can be used to construct an estimate of the variance–covariance matrix for the parameter estimates of $\beta, \rho, \sigma_\varepsilon$. LeSage and Pace (2009, Chapter 3) point out that various procedures appear in the literature regarding use of an analytical or numerical Hessian, or the related information matrix to produce an estimate of the variance–covariance matrix conventionally used to produce t-statistics associated with the estimates of β, ρ. For maximum likelihood, the Hessian is the matrix of second-derivatives of the log-likelihood with respect to the parameters.

For large numbers of observations, calculation of the dense matrix inverse: $(I_n - \rho W)^{-1}$ that is part of the analytical Hessian may become computationally difficult.[7] This has lead to use of a numerical procedure to calculate second-derivatives of the log-likelihood with respect to the parameters for the Hessian.[8] There is also a mixed analytical and numerical procedure proposed in LeSage and Pace (2009, Chapter 3).

When we compare alternative estimation approaches such as maximum likelihood (ML) or generalized method of moments (GMM), differences in variance–covariance matrix calculation may have an important impact on bias and mean squared error of the effects estimates. This is because the empirical measures of dispersion for the effects estimates are constructed using draws for the parameter estimates of β, ρ based on the estimated variance–covariance matrix. These draws are then used in the nonlinear matrix expression: $\partial y/\partial X^r = (I_n - \rho W)^{-1}\tilde{\beta}$ to simulate (say) m different $n \times n$ matrices. Empirical measures of dispersion for the scalar summary direct effect

[7] The most difficult term to compute is trace$(W(I_n - \rho W)^{-1}W(I_n - \rho W)^{-1})$. However, this can be expanded as a weighted sum of trace(W^j) for $j = 1 \ldots n$ and these traces can be easily estimated as discussed in LeSage and Pace (2009, p. 98).

[8] The Spatial Econometrics Toolbox uses numerical derivatives for problems involving more than 500 observations. This approach can of course be changed by users, at the cost of additional time (and computer memory) required to produce estimates based on the analytical Hessian for large values of n.

estimates are based on an average (or median) of the main diagonal of these m different $n \times n$ matrices. Empirical measures of dispersion for the (cumulative) scalar summary indirect effect estimates are based on the sum of off-diagonal elements (for each row) using the average (or median) of the rows.[9] Despite the obvious importance of different approaches to calculating estimates of the variance–covariance matrices used by different estimators, there has been little attention given to this issue in the spatial econometrics Monte Carlo literature.

Practitioners may question the accuracy of numerical Hessian estimates calculated by maximum likelihood estimation. A comparison of t-statistics from maximum likelihood to those from Bayesian MCMC estimation of the same model using uninformative priors for the parameters in Bayesian MCMC estimation should produce estimates quite similar to maximum likelihood. The variance–covariance matrix does not need to be constructed during Bayesian MCMC estimation to produce simulated draws for calculating empirical measures of dispersion for the effects estimates. Instead, (retained) MCMC draws made during estimation of β, ρ are used to produce empirical measures of dispersion for the effects estimates (as well as for the parameters β, ρ). Large differences between t-statistics from MCMC versus maximum likelihood estimation likely point to a problem with the numerical estimate of the variance–covariance matrix. Gelfand et al. (1990) point out that draws from MCMC sampling can be used to produce valid posterior distributions for nonlinear functions of the parameters that are of interest, such as the direct and indirect effects estimates.

We propose an approach that can be used in Monte Carlo studies to explore the impact of alternative variance–covariance estimates on empirical measures of dispersion for the effects estimates. In terms of alternative variance–covariance estimates, we can rely on numerical or analytic Hessian matrices, as well as different asymptotic analytical expressions employed by GMM, or Bayesian Markov Chain Monte Carlo (MCMC) draws.[10]

For a Monte Carlo experiment involving $m = 1, \ldots, M$ iterations/trials, where we are interested in calculating bias and mean squared error of the direct and indirect effects estimates:

Step (1) For each Monte Carlo iteration m, calculate **mean/median** direct and indirect effects estimates based on the last q MCMC draws in the case of Bayesian MCMC estimation, or q simulated draws of the parameter estimates from the asymptotic variance–covariance matrix from GMM, maximum likelihood, or some other method of interest where q is a large number such as 1000.

Step (2) For each Monte Carlo iteration m, calculate **empirical standard deviations** (measures of dispersion) for the direct and indirect effects estimates based on the q simulated draws of the parameter estimates or MCMC draws of the parameters.

[9] Of course, the row and column sums are the same, so either could be used.

[10] The retained MCMC draws (those collected after a "burn-in" set of draws) have been shown by Gelfand and Smith (1990) to reflect draws from the joint posterior distribution of the parameters β and ρ. As such, they should embody any nonzero covariance between these sets of parameters.

Step (3) Store these mean/median *and* standard deviation estimates for the direct and indirect effects, producing a sample of size M.

We then suggest calculating measures of bias and mean squared error based on the set of M point estimates of the direct and indirect effects, *as well as* the set of M empirical estimates of the standard deviations.

Calculation of bias and mean squared error would involve:

- Analyze bias of the estimator using the difference between the true direct and indirect effects and each of the m point estimates for these effects. An average (mean) of the M different measures of bias should be used as a scalar summary measure of bias.
- Analyze dispersion of the estimator using the *mean* of the M *empirical standard deviation* estimates from each of the m Monte Carlo iterations.
- Mean squared error (MSE) measures for the estimator should be calculated using mean of the M different biases squared plus the mean of the M different empirical standard deviations squared.

The intuitive motivation for this approach is that practitioners will use the mean/median and *empirical standard deviation* constructed from the q simulated effects estimates to draw inferences. We care about the quality of inferences on average arising from different *estimators*. The quality of inference will depend on the process used to produce empirical standard deviations based on simulated effects estimates by different estimators. This of course will be a function of alternative approaches to estimating the variance–covariance matrix. This is because empirical measures of dispersion are based on draws made using the (estimated) variance–covariance matrix.

In applied practice we use numerical or analytical Hessians in maximum likelihood, Bayesian MCMC draws, or a GMM asymptotic variance–covariance matrix to calculate standard deviations of the effects estimates for practitioners. Our proposed Monte Carlo approach includes an examination of variance–covariance estimates produced by these various estimators.

By way of conclusion, we should not ignore the fact that draws from MCMC or an estimated variance–covariance matrix are used to produce measures of dispersion for estimated β, ρ that play an important role in inference. Further, alternative approaches to calculating dispersion for these parameters may play an important role when it comes to producing draws used in the nonlinear relationship between these parameters and our empirical measures of dispersion for the effect estimates.

4 An illustrative example

To illustrate how these issues impact conclusions drawn from Monte Carlo studies, we compare three estimation procedures, maximum likelihood (ML), robust Bayesian MCMC sampling (RBayes) and a robust GMM (RGMM) method set forth in a Monte Carlo study by Dogan and Taspinar (2014).

We carry out two Monte Carlo experiments, one in Sect. 4.1 where the three methods are applied to a data generating process (DGP) that exhibits the typical normally distributed constant variance disturbances, with explanatory variables that are independent normally distributed. From the spatial econometric literature we known that all three estimators ML, RBayes, RGMM produce estimates of β, ρ that are asymptotically equivalent (Lee 2007). This means we would expect similar bias and MSE for β, ρ from all three estimation procedures as the Monte Carlo experiment sample size increases. It should be noted that we rely on the best moment functions for the i.i.d normal case from Lee (2007) that produce GMM estimators with the same asymptotic efficiency as ML. Differences in outcomes from various approaches to determining measures of dispersion should become small as sample size increases, but asymptotic properties may manifest themselves at very large sample sizes. As a practical matter, the efficiency of GMM can depend on correlation between Wy and the instruments $W(I_n - \rho W)^{-1}X\beta$. For example, Pace et al. (2012) show that IV estimation methods produce poor performance for some spatial regression models even for sample sizes of 250,000–500,000 in the face of spatial dependence in the explanatory variables.

A second experiment in Sect. 4.2 follows the Monte Carlo study by Dogan and Taspinar (2014) that relies on a DGP involving a special type of heteroscedasticity that relates to the number of neighboring observations (more generally, aspects of the W-matrix). As a practical matter, Griffiths (2007) and Lee (2010) argue that this type of heteroscedasticity arises when spatial data are created by averaging regions with a different number of observations/subregions. In the presence of heteroscedastic disturbances of this type both ML and GMM estimators are inconsistent. Kelejian and Prucha (2010) modify the moment functions from their two-step GMM estimation approach to accommodate unknown forms of heteroscedasticity, and Lin and Lee (2010) propose a one-step robust GMM estimator for the heteroscedastic SAR specification, which they show is consistent and asymptotically normally distributed. In practice, a consistent estimate of the asymptotic covariance matrix is used to construct empirical confidence regions. In this case of heteroscedasticity, there is no set of best moment functions, so Lin and Lee (2010) suggest use of the moment conditions from the i.i.d case (without normality). These were used in Dogan and Taspinar (2014) and in this study as well. This second experiment should produce systematic bias in ML estimates, but other estimators such as RBayes and RGMM that do not exhibit bias may lead to increased dispersion, resulting in the classic trade-off between bias and precision, which we examine using mean squared error.

4.1 A comparison of ML, RBayes, RGMM when disturbances are normal and homoscedastic

We use the spatial coordinates added to the Boston house price data from Harrison and Rubinfeld (1978) by Gilley and Pace (1996) to create a 506 observation dataset. A larger sample of 3107 observations was based on spatial coordinates for US counties from Pace and Barry (1997).

The SAR model DGP was used: $y = \rho Wy + X\beta + \varepsilon$, where $\beta = \begin{pmatrix}1 & 0 & -1\end{pmatrix}'$. Normal, independent and constant variance disturbances were used, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$. Three explanatory variables were generated from an independent normal distribution with a mean of zero and variance of unity. The same set of explanatory variables,

Table 1 Results for $n = 506$, SNR = 0.8972, $R^2 = 0.8993$

	Bias			MSE		
	ML	RBayes	RGMM	ML	RBayes	RGMM
β_1	0.0047	0.0035*	0.0040	0.0005*	0.0005*	0.0005*
β_2	0.0019*	0.0037	0.0019*	0.0005*	0.0005*	0.0005*
β_3	0.0042*	0.0050	0.0051	0.0005*	0.0005*	0.0005*
ρ	−0.0019	−0.0023	−0.0001*	0.0001*	0.0003*	0.0003*
Direct						
x_1	−0.0398	0.0013*	−0.0399	0.0022	0.0008*	0.0008*
x_2	−0.0001	−0.0000*	0.0003	0.0006*	0.0007	0.0007
x_3	0.0400	−0.0011*	0.0402	0.0022	0.0007*	0.0007*
Indirect						
x_1	0.0111*	−0.0366	0.0287	0.0351*	0.0369	0.0368
x_2	0.0005	0.0004*	0.0012	0.0025	0.0024*	0.0025
x_3	−0.0102*	0.0374	−0.0271	0.0343*	0.0367	0.0360
	Standard deviation					
Indirect						
x_1	0.1285*	0.1916	0.1912			
x_2	0.0504	0.0500*	0.0512			
x_3	0.1267*	0.1901	0.1878			

parameters for β, ρ were used in all experiments with $n = 506$, with changes in the noise variance σ_ε^2 made to explore the impact of changes in the signal-to-noise ratio on estimation performance. The same was true for the case of $n = 3,107$, explanatory variables were fixed across all experiments.

It should be noted that we are only interested in providing an illustrative example of the implications of adopting our proposed Monte Carlo methodology. We do not intend an exhaustive study of the alternative estimators over the full range of parameter values, signal-to-noise ratios, etc.

Table 1 shows results averaged over 1000 trials based on the proposed procedure set forth in Sect. 3. The signal-to-nose ratio (SNR) was controlled using a noise variance σ^2 calculated using the relationships in (8).[11]

$$\begin{aligned} \text{SNR} &= \frac{A'A}{A'A + \sigma^2 tr(B'B)} \\ A &= BX\beta \\ B &= (I_n - \rho W)^{-1} \\ \sigma^2 &= \frac{A'A(1 - \text{SNR})}{tr(B'B)(\text{SNR})} \end{aligned} \tag{8}$$

Of course, the SNR in (8) will lie between 0 and 1, similar to the R^2-statistic. To impose a value of (say) SNR = 0.7, we solve expression (8) for

[11] This approach was proposed by Pace et al. (2012).

$\sigma^2 = 0.3(A'A)/0.7tr(B'B)$. An R-squared statistic calculated using: $\hat{y} = (I_n - \hat{\rho}W)^{-1}X\hat{\beta}$ in the usual least-squares formula tends to values similar to the SNR.

The lowest bias and MSE across the three alternative estimators are indicated with an asterisk ($*$) symbol in the table. We see similar performance in terms of MSE for all three methods in terms of the parameters β, ρ.

Empirical estimates of means and standard deviations for the direct and indirect effects were constructed based on 1000 draws used in the nonlinear expressions for calculating scalar summary measures of the effects estimates (LeSage and Pace, p 115). These draws were constructed using: (1) for RBayes, the last 1000 draws of β, ρ from a total of 2500 MCMC draws used for estimation,[12] (2) for ML, 1000 draws for β, ρ made using a normal distribution with a variance–covariance matrix based on the analytical Hessian, and (3) for RGMM, 1000 draws based on the asymptotic variance–covariance matrix associated with the best moment functions for the i.i.d case (without normality) from Lin and Lee (2010).[13]

The MSE reported in the table show roughly equal performance for direct effects across all three methods for the three explanatory variables. The MSE associated with the indirect effects estimates show a slight advantage for the ML estimator. This arises from slightly less bias and increased precision indicated by the lower empirical standard deviations for the indirect effects presented in the table. Given the controlled nature of our experiment, this arose from additional accuracy provided by use of the analytical variance–covariance matrix to produce simulated draws that were in turn used to calculate empirical measures of dispersion for the scalar summary indirect effects estimates.

As a validity check, we report results for an experiment with $n = 3107$ and moderate SNR producing a median R^2 over the 1000 trials equal to 0.6729. We would expect ML and GMM methods to produce similar performance for this larger sample given our theoretical knowledge about consistency of the GMM estimator and variance–covariance matrix. Results presented in Table 2 show this to be the case. These results also make the point that any bias or imprecision in β, ρ estimates tend to become amplified for the direct and indirect effects estimates, presumably due to the nonlinear relationships involved. In addition, indirect effects will exhibit more amplification because of cumulation over all off-diagonal elements of the $n \times n$ matrices, than direct effects based only on the main diagonal elements.

Table 3 shows the impact of lowering the SNR to 0.3528, consistent with a median R-squared of 0.35 over the 1000 trials. This set of results indicates similarly equivalent MSE across the three methods for the parameters β, ρ. In contrast, the MSE for indirect effects provides a distinct advantage for the ML procedure. The source of this advantage can be seen in the standard deviations for the indirect effects reported in the Table.

Table 4 shows results for a further lowering of SNR to produce a model with R^2 around 0.19. These results show a dramatic increase in bias and dispersion of

12 A random-walk Metropolis–Hastings approach was used to sample the parameter ρ with a tuning procedure described in LeSage and Pace (2009, Chapter 5).

13 See Dogan and Taspinar (2014) for details regarding this.

Table 2 Results for $n = 3107$, SNR = 0.6737, $R^2 = 0.6729$

	Bias			MSE		
	ML	RBayes	RGMM	ML	RBayes	RGMM
β_1	0.0001	0.0001	−0.0004	0.0003	0.0003	0.0003
β_2	−0.0008	−0.0009	−0.0008	0.0003	0.0003	0.0003
β_3	−0.0004	−0.0007	0.0002	0.0003	0.0003	0.0003
ρ	−0.0025	−0.0040	−0.0007	0.0001	0.0002	0.0001
Direct						
x_1	−0.0423	−0.0026	−0.0421	0.0022	0.0005	0.0022
x_2	−0.0009	−0.0010	−0.0009	0.0004	0.0004	0.0004
x_3	0.0419	0.0019	0.0420	0.0021	0.0005	0.0021
Indirect						
x_1	0.0194	−0.0358	0.0381	0.0171	0.0180	0.0185
x_2	−0.0017	−0.0020	−0.0019	0.0016	0.0015	0.0016
x_3	−0.0199	0.0346	−0.0385	0.0169	0.0179	0.0182

Table 3 Results for $n = 506$, SNR = 0.3528, $R^2 = 0.3534$

	Bias			MSE		
	ML	RBayes	RGMM	ML	RBayes	RGMM
β_1	0.0181	0.0131*	0.0150	0.0079	0.0078	0.0077*
β_2	0.0081*	0.0155	0.0090	0.0081	0.0084	0.0080*
β_3	0.0177*	0.0206	0.0215	0.0078	0.0079	0.0077*
ρ	−0.0044	−0.0050	0.0030*	0.0003*	0.0014	0.0013
Direct						
x_1	−0.0220	0.0149*	−0.0225	0.0096*	0.0108	0.0096*
x_2	−0.0094*	−0.0174	0.0097	0.0098*	0.0109	0.0098*
x_3	0.0608	−0.0228*	0.0620	0.0124*	0.0107	0.0124*
Indirect						
x_1	0.0710	−0.0154*	0.1426	0.1028*	0.2060	0.2466
x_2	0.0212*	0.0343	0.0226	0.0425*	0.0446	0.0483
x_3	−0.0012*	0.0507	−0.0695	0.0942*	0.1975	0.2143
	Standard deviation					
Indirect						
x_1	0.3128*	0.4536	0.4757			
x_2	0.2051*	0.2083	0.2185			
x_3	0.3069*	0.4416	0.4577			

the RGMM indirect effects estimates. Examining only MSE for the parameters β, ρ would provide no indication of the magnitude of this problem. The problem seems to stem from large bias in estimation of the parameter ρ, which of course is magnified

Table 4 Result for n = 506, SNR = 0.1951, R^2 = 0.1907

	Bias			MSE		
	ML	RBayes	RGMM	ML	RBayes	RGMM
β_1	0.0269	0.0196	−0.0251	0.0178	0.0176*	0.0179
β_2	0.0126	0.0234	0.0484	0.0183*	0.0189	0.0206
β_3	0.0270	0.0309	0.0750	0.0171*	0.0176	0.0223
ρ	−0.0033	−0.0054	0.2376	0.0004*	0.0016	0.0570
Direct						
x_1	−0.0120	0.0224	0.0335	0.0207*	0.0236	0.0267
x_2	0.0145	−0.0264	0.0590	0.0221*	0.0245	0.0305
x_3	0.0703	−0.0340	0.0253	0.0246	0.0234*	0.0250
Indirect						
x_1	0.1059	0.0245*	21.5002	0.1825*	0.2856	478.9915
x_2	0.0324*	0.0519	1.6604	0.0982*	0.1018	16.0252
x_3	−0.0018*	0.0729	−20.4693	0.1639*	0.2709	434.9305
	Standard deviation					
Indirect						
x_1	0.4139*	0.5338	4.0907			
x_2	0.3116*	0.3148	3.6426			
x_3	0.4048*	0.5154	3.9923			

when calculating indirect effects as well as the empirical measure of dispersion for these. This is because of the nonlinear transformation: $(I_n - \rho W)^{-1} = I_n + \rho W + \rho^2 W^2 + \rho^3 W^3 \ldots$, which plays an important role in calculating empirical measures of dispersion for the indirect effects.

The divergence between conclusions drawn from examining only outcomes for the parameters β, ρ while ignoring direct and indirect effect outcomes can be seen more clearly for a larger sample of $n = 3109$ observations. Table 5 reports results from this experiment, where low SNR has been used. In this experiment, the MSE for the parameters β, ρ from all three methods is identical. The conclusion that would be drawn is that all three estimators provide equally good performance. However, an examination of the MSE for these methods based on indirect effects estimates would lead to a very different conclusion. Here we see a five- to sixfold increase in MSE for the RGMM estimator relative to ML and RBayes. This increased MSE arises in large part from a fifty-fold increase in bias of the indirect effects estimates.

4.2 A comparison of ML, RBayes, RGMM when disturbances are heteroscedastic

In this section, we turn to a second experiment that follows the Monte Carlo study by Dogan and Taspinar (2014). They rely on a DGP that produces heteroscedasticity that is related to the number of neighboring observations. We use the same consistent estimate

Table 5 Result for $n = 3, 109$, SNR = 0.0925, $R^2 = 0.0905$

	Bias			MSE		
	ML	RBayes	RGMM	ML	RBayes	RGMM
β_1	−0.0008*	−0.0016	−0.0034	0.0065*	0.0065*	0.0065*
β_2	−0.0037	−0.0043	−0.0035*	0.0065*	0.0066*	0.0065*
β_3	−0.0002*	−0.0005	0.0026	0.0065*	0.0065*	0.0065*
ρ	−0.0031*	−0.0058	0.0070	0.0003*	0.0003*	0.0003*
Direct						
x_1	−0.0435	−0.0054	−0.0433	0.0097	0.0085*	0.0097
x_2	−0.0040	−0.0049	−0.0038	0.0079*	0.0085	0.0079*
x_3	0.0421	0.0029	0.0424	0.0095	0.0084*	0.0096
Indirect						
x_1	0.0141*	−0.0560	0.5091	0.0622*	0.0624	0.3492
x_2	−0.0087*	−0.0094	0.0242	0.0328	0.0308*	0.0544
x_3	−0.0173*	0.0506	−0.5614	0.0616*	0.0621	0.4042
	Standard deviation					
Indirect						
x_1	0.2490	0.2436*	0.3001			
x_2	0.1808	0.1754*	0.2319			
x_3	0.2476	0.2439*	0.2984			

of the asymptotic covariance matrix to construct empirical measures of dispersion for the effects estimates as in Dogan and Taspinar (2014). This experiment should produce systematic bias in ML estimates. A Monte Carlo study of this situation can be used to examine the classic trade-off between bias and precision, which we examine using mean squared error.

An experiment similar to those reported in Dogan and Taspinar (2014), for $n = 945$ was carried out. The nature of heteroscedasticity in their experiment is such that:

$$\varepsilon_{ni} = \sigma_{ni} u_{ni}$$
$$\sigma_{ni} = c \frac{h_i}{\sum_{j=1}^{n} h_j / n}$$

where h_i is the number of neighbors to observation i, and $u_{ni} \sim N(0, 1)$, a standard normal distribution. Neighbors are defined as in Arraiz et al. (2010), which is intended to reflect a scenario based on the fact that for the U.S. states there are relatively more states located in the northeast.

Table 6 shows results for a case where SNR is moderate. The results mirror those reported in Dogan and Taspinar (2014), who indicate that RGMM produces lower bias, but RBayes estimates are more precise, with the MSE trade-off leading to RBayes being the preferred estimator in terms of β, ρ, direct and indirect effects in our results. It should be noted that results reported in Table 6 follow the proposed Monte Carlo

Table 6 Results for $n = 945$, SNR = 0.5804, $R^2 = 0.6045$

	Bias			MSE		
	ML	RBayes	RGMM	ML	RBayes	RGMM
β_1	0.0055	0.0034	−0.0004*	0.0014	0.0009*	0.0012
β_2	0.0013	0.0012	−0.0005*	0.0014	0.0009*	0.0015
β_3	−0.0043	−0.0023*	0.0045	0.0014	0.0010*	0.0015
ρ	−0.0428	−0.0205	−0.0012*	0.0019	0.0009	0.0007*
	Standard deviation					
β_1	0.0369	0.0296*	0.0350			
β_2	0.0375	0.0307*	0.0383			
β_3	0.0378	0.0309*	0.0386			
ρ	0.0047	0.0213*	0.0264			
Direct						
x_1	−0.0627	−0.0234*	−0.0529	0.0056	00019*	0.0043
x_2	0.0014	0.0014	−0.0005*	0.0017	00012*	0.0019
x_3	0.0641	0.0247*	0.0575	0.0059	00020*	0.0051
Indirect						
x_1	−0.3128	−0.1630	0.0632*	0.1055	0.0699*	0.0900
x_2	0.0021	0.0025	−0.0012*	0.0048	0.0038*	0.0077
x_3	0.3162	0.1656	−0.0535*	0.1077	0.0709*	0.0858
	Standard deviation					
Indirect						
x_1	0.0873*	0.2082	0.2932			
x_2	0.0691	0.0619*	0.0876			
x_3	0.0880*	0.2084	0.2879			

study approach that does not produce variances of the estimates β, ρ using a variance calculated over the 1000 Monte Carlo trial outcomes. Rather, we calculate the mean of standard deviations that would be reported by the estimator using either MCMC draws, an analytical Hessian, or asymptotic variance–covariance matrix. These are what practitioners would use to draw inferences in applied practice. As noted earlier, there are a number of reasons why these would diverge from a variance determined using M Monte Carlo point estimates of the effects. The nature of the experiments conducted here is such that we attempt to eliminate all reasons for divergence between these measures of dispersion other than the nature of the approach taken to determining estimated standard deviations by the different estimators (e.g., MCMC draws, an analytical Hessian, or asymptotic variance–covariance matrix).

A second experiment where the SNR is lowered to 0.2569, producing a median R^2 over the 1000 trials equal to 0.2718 was carried out, with results reported in Table 7. These results also mirror those reported in Dogan and Taspinar (2014) for the parameters β, ρ. The RGMM estimator produces lower bias, but RBayes better precision, leading to superior MSE for RBayes. The improved precision of RBayes

Table 7 Results for $n = 945$, SNR = 0.2569, $R^2 = 0.2718$

	Bias			MSE		
	ML	RBayes	RGMM	ML	RBayes	RGMM
β_1	0.0084	0.0057	−0.0023*	0.0056	0.0035*	0.0049
β_2	0.0018	0.0022	−0.0009*	0.0057	0.0038*	0.0059
β_3	−0.0046	−0.0028*	0.0114	0.0058	0.0038*	0.0060
ρ	−0.0653	−0.0330	0.0086*	0.0043	0.0019	0.0014*
	Standard deviation					
β_1	0.0740	0.0592*	0.0697			
β_2	0.0753	0.0615*	0.0766			
β_3	0.0757	0.0617*	0.0764			
ρ	0.0046	0.0277*	0.0365			
Direct						
x_1	−0.0671	−0.0275*	−0.0473	0.0111	0.0056*	0.0085
x_2	0.0019	0.0025	−0.0012*	0.0069	0.0049*	0.0075
x_3	0.0716	0.0309*	0.0580	0.0120	0.0062*	0.0107
Indirect						
x_1	−0.4646	−0.2565	1.3281	0.2347	0.1387*	2.1271
x_2	0.0024	0.0041	−0.0580	0.0161	0.0142*	0.0856
x_3	0.4735	0.2625	−1.2256	0.2437	0.1424*	1.8636
	Standard deviation					
Indirect						
x_1	0.1375*	0.2700	0.6027			
x_2	0.1269	0.1192*	0.2868			
x_3	0.1394*	0.2711	0.6012			

is shown by reported standard deviations, where again these were calculated on the basis of MCMC draws used in *each* of the 1000 Monte Carlo trials. An examination of the indirect effects estimates shows that conclusions based on the parameters β, ρ do not carry over to these estimates.

We see bias across all three estimators, with the bias for indirect effects estimates calculated by RGMM being the largest. We would theoretically expect ML estimates to be biased, but the RGMM method proposed by Dogan and Taspinar (2014) to correct for this results in larger bias than that from an ML estimation procedure. This reversal in conclusions comes about because we have considered all facets of the three different estimators, which has not been traditionally done in applied Monte Carlo practice. The three estimators behave as theoretically expected based on examining only parameters β, ρ, leading to a conclusion that RGMM helps solve the bias problem introduced for ML estimators in the face of this type of heteroscedasticity. However, in applied practice practitioners would not experience this type of improved performance when they attempt to draw inferences regarding the direct and indirect effects estimates.

5 Conclusion

We argue that we have not payed enough attention to direct and indirect effects estimates and inference in past Monte Carlo studies. If we want to provide Monte Carlo results that will be meaningful with regard to the experience that applied practitioners will have using alternative estimation methods, we need to start doing this.

We cannot ignore the role played by estimated variance–covariance matrices used to simulate draws for calculating empirical measures of dispersion for the direct and indirect effects estimates. These are frequently the major focus of inference by practitioners who are using spatial regression models that include a spatial lag of the dependent variable.

Differences in the quality of estimated dispersion arising from use of an analytical or numerical ML Hessian, Bayesian MCMC draws or asymptotic covariance matrices may have an important impact on practitioner experience when using algorithms constructed to implement different estimation methodologies. Monte Carlo practices should evolve to consider this very important aspect of spatial econometric work on different approaches to estimating our models.

References

Arraiz I, Drukker DM, Kelejian HH, Prucha IR (2010) A spatial Clifford-type model with heteroskedastic innovations: small and large sample results. J Reg Sci 50(2):592–614
Debarsy N, Ertur C, LeSage JP (2012) Interpreting dynamic space-time panel models. Stat Methodol 9(1–2):158–171
Dogan O, Taspinar S (2014) Spatial autoregressive models with unknown heteroskedasticity: a comparison of Bayesian and Robust GMM approach. Reg Sci Urb Econ 45:1–21
Drukker DM, Prucha IR, Raciborski R (2013a) A command for estimating spatial-autoregressive models with spatial-autoregressive disturbances and additional endogenous variables. Stata J 13:287–301
Drukker DM, Prucha IR, Raciborski R (2013b) Maximum likelihood and generalized spatial two-stage least-squares estimators for a spatial-autoregressive model with spatial-autoregressive disturbances. Stata J 13:221–241
Elhorst JP (2010) Applied spatial econometrics: raising the bar. Spat Econ Anal 5(1):9–28
Elhorst JP (2014) Spatial econometrics: from cross-sectional data to spatial panels. Springer, Berlin
Gelfand AE, Hills SE, Racine-Poon A, Smith AFM (1990) Illustration of Bayesian inference in normal data models using Gibbs sampling. J Am Stat Assoc 85:972–985
Gelfand AE, Smith AFM (1990) Sampling-based approaches to calculating marginal densities. J Am Stat Assoc 85:398–409
Gilley OW, Pace RK (1996) On the Harrison and Rubinfeld data. J Environ Econ Manag 31(3):403–405
Griffiths WE (2007) Heteroskedasticity. In: Baltagi BH (ed) A companion to theoretical econometrics. Blackwell Publishing Ltd, pp 82–100
Harrison D, Rubinfeld DL (1978) Hedonic prices and the demand for clean air. J Environ Econ Manag 5(1):81–102
Kelejian HH, Prucha IR (2010) Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. J Econ 157(1):53–67
Lee LF (2007) GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. J Econ 137(2):489–514
Lee M (2010) Microeconometrics: methods of moments and limited dependent variables. Springer, New York
Lin X, Lee LF (2010) GMM estimation of spatial autoregressive models with unknown heteroskedasticity. J Econ 157(1):34–52
LeSage JP, Pace RK (2009) Introduction to spatial econometrics. CRC Press Taylor & Francis Group, Boca Raton

LeSage JP, Pace RK, Lam N, Campanella R, Liu X (2011) New Orleans business recovery in the aftermath of Hurricane Katrina. J R Stat Soc Ser A 174(4):1007–1027
LeSage JP, Sheng Y (2014) A spatial econometric panel data examination of endogenous versus exogenous interaction in Chinese province-level patenting. J Geogr Syst 16(3):233–262
LeSage JP, Thomas-Agnan C (2015) Interpreting spatial autoregressive interaction models. J Reg Sci 85(2):188–208
Pace RK, Barry RP (1997) Quick computation of spatial autoregressive estimators. Geogr Anal 29(3):232–246
Pace RK, LeSage JP, Zhu S (2012) Spatial dependence in regressors. In: Terrell D, Millimet D (eds) Advances in econometrics, vol 30. Elsevier, Amsterdam, pp 257–295