Regional Science and Urban Economics 52 (2015) 50–70

Contents lists available at ScienceDirect

Regional Science and Urban Economics

journal homepage: www.elsevier.com/locate/regec

Modified QML estimation of spatial autoregressive models with unknown heteroskedasticity and nonnormality☆

Shew Fan Liu, Zhenlin Yang *

School of Economics, Singapore Management University, 90 Stamford Road, 178903 Singapore

A R T I C L E I N F O

Article history:
Received 23 July 2014
Received in revised form 16 January 2015
Accepted 4 February 2015
Available online 11 February 2015

JEL classification:
C10
C13
C15
C21

Keywords:
Spatial dependence
Unknown heteroskedasticity
Nonnormality
Modified QML estimator
Robust standard error
SARAR models

A B S T R A C T

In the presence of heteroskedasticity, Lin and Lee (2010) show that the quasi-maximum likelihood (QML) estimator of the spatial autoregressive (SAR) model can be inconsistent as a 'necessary' condition for consistency can be violated, and thus propose robust GMM estimators for the model. In this paper, we first show that this condition may hold in certain situations and when it does the regular QML estimator can still be consistent. In cases where this condition is violated, we propose a simple modified QML estimation method robust against unknown heteroskedasticity. In both cases, asymptotic distributions of the estimators are derived, and methods for estimating robust variances are given, leading to robust inferences for the model. Extensive Monte Carlo results show that the modified QML estimator outperforms the GMM and QML estimators even when the latter is consistent. The proposed methods are then extended to the more general SARAR models.

1. Introduction

Spatial dependence is increasingly becoming an integral part of empirical works in economics as a means of modelling the effects of 'neighbours' (see, e.g., Cliff and Ord (1972, 1973, 1981), Ord (1975), Anselin (1988, 2003), Anselin and Bera (1998), LeSage and Pace (2009) for some early and comprehensive works). Spatial interaction in general can occur in many forms. For instance peer interaction can cause stratified behaviour in the sample such as herd behaviour in stock markets, innovation spillover effects, localized purchase decisions, etc., while spatial relationships can also occur more naturally due to structural differences in space/cross-section such as geographic proximity, trade agreements, demographic characteristics, etc. See Case (1991), Pinkse and Slade (1998), Pinkse et al. (2002), Hanushek et al. (2003), Baltagi et al. (2007) to name a few. Among the various spatial econometrics models that have been extensively treated, the most popular one may be the spatial autoregressive (SAR) model.

While heteroskedasticity is common in regular cross-section studies, it may be more so for a spatial econometrics model due to aggregation, clustering, etc. Anselin (1988) identifies that heteroskedasticity can broadly occur due to "idiosyncrasies in model specification and affect the statistical validity of the estimated model". This may be due to the misspecification of the model that feeds to the disturbance term or may occur more naturally in the presence of peer interactions. Data related heteroskedasticity may also occur for example if the model deals with a mix of aggregate and nonaggregate data, the aggregation may cause errors to be heteroskedastic. See, e.g., Glaeser et al. (1996), LeSage and Pace (2009), Lin and Lee (2010), Kelejian and Prucha (2010), for more discussions. As such, the assumption of homoskedastic disturbances is likely to be invalid in a spatial context in general. However, much of the present spatial econometrics literature has focused on estimators developed under the assumption that the errors are homoskedastic. This is in a clear contrast to the standard cross-section econometrics literature where the use of heteroskedasticity robust estimators is a standard practice.

Although Anselin raised the issue of heteroskedasticity in spatial models as early as in 1988, and made an attempt to provide tests of

☆ The authors are grateful to the School of Economics, Singapore Management University, for research support, and to the editor, Daniel McMillen, a referee, and the participants of the VIII World Conference of the Spatial Econometrics Association, Zürich 2014, for their helpful comments.
* Corresponding author at: 90 Stamford Road, 178903 Singapore. Tel.: +65 6828 0852; fax: +65 6828 0833.
E-mail addresses: shewfan.liu.2011@smu.edu.sg (S.F. Liu), zlyang@smu.edu.sg (Z. Yang).

http://dx.doi.org/10.1016/j.regsciurbeco.2015.02.003
0166-0462/

spatial effects robust to unknown heteroskedasticity, comprehensive treatments of estimation related issues were not considered until recent years by, e.g., Kelejian and Prucha (2007, 2010), LeSage (1997), Lin and Lee (2010), Arraiz et al. (2010), Badinger and Egger (2011), Jin and Lee (2012), Baltagi and Yang (2013b), and Do an and Taşpinar (2014). Lin and Lee (2010) formally illustrate that the traditional quasi-maximum likelihood (QML) and generalized method of moments (GMM) estimators are inconsistent in general when the SAR model suffers from heteroskedasticity, and provide heteroskedasticity robust GMM estimators by modifying the usual quadratic moment conditions.

Inspired by Lin and Lee (2010), we introduce a modified QML estimator (QMLE) for the SAR model by modifying the concentrated score function for the spatial parameter to make it robust against unknown heteroskedasticity. It turns out that the method is very simple and more importantly, it can be easily generalized to suit more general models.[1] For heteroskedasticity robust inferences, we propose an outer-product-of-gradient (OPG) method for estimating the variance of the modified QMLE. We provide formal theories for the consistency and asymptotic normality of the proposed estimator, and the consistency of the robust standard error estimate. Extensive Monte Carlo results show that the modified QML estimator generally outperforms its GMM counter parts in terms of efficiency and sensitivity to the magnitude of model parameters in particular the regression coefficients. The Monte Carlo results also show that the proposed robust standard error estimate performs well. We also study the cases under which the regular QMLE is robust against unknown heteroskedasticity and provide a set of robust inference methods. It is interesting to note that the modified QMLE is computationally as simple as the regular QMLE, and it also outperforms the regular QMLE when the latter is heteroskedasticity robust. This is because the modified QMLE captures the extra variability inherent from the estimation of the regression coefficients and the average of error variances.

To demonstrate their flexibility and generality, the proposed methods are then extended to the popular spatial autoregressive model with spatial autoregressive disturbances (SARAR(1, 1)) with heteroskedastic innovations. Kelejian and Prucha (2010) formally treat this model with a three-step estimation procedure. Monte Carlo results show that the modified QMLE performs better in finite sample than the three-step estimator. Further possible extensions of the proposed methods are discussed. In summary, the proposed set of QML-based robust inference methods are simple and reliable, and can be easily adopted by applied researchers.

The rest of the paper is organized as follows. Section 2 examines the cases where the regular QML estimator of the SAR model is consistent under unknown heteroskedasticity, and provides methods for robust inferences. Section 3 introduces the modified QML estimator that is generally robust against unknown heteroskedasticity, and presents methods for robust inferences. Section 4 presents the Monte Carlo results for the SAR model. Section 5 extends the proposed methods to the popular SARAR(1, 1) model and discusses further possible extensions. Section 6 concludes the paper. All technical details are given in Appendix B.

2. QML estimation of spatial autoregressive models

In this section, we first outline the QML estimation of the SAR model under the assumptions that the errors are independent and identically distributed (iid). Then, we examine the properties of the QMLE of the SAR model when the errors are independent but not identically distributed (inid). We provide conditions under which the regular QMLE is robust against heteroskedasticity of unknown form, derive its asymptotic distribution, and provide heteroskedasticity robust estimator of its asymptotic variance.

Some general notation will be followed in this paper: $|\cdot|$ and $\mathrm{tr}(\cdot)$ denote, respectively, the determinant and trace of a square matrix; A' denotes the transpose of a matrix A; $\mathrm{diag}(\cdot)$ denotes the diagonal matrix formed by a vector or the diagonal elements of a square matrix; $\mathrm{diagv}(\cdot)$ denotes the column vector formed by the diagonal elements of a square matrix; and a vector raised to a certain power is operated elementwise.

2.1. The model and the QML estimation

Consider the spatial autoregressive or SAR model of the form:

$$Y_n = \lambda_0 W_n Y_n + X_n\beta_0 + \epsilon_n, \tag{1}$$

where X_n is an $n \times k$ matrix of exogenous variables, W_n is a known $n \times n$ spatial weights matrix, ϵ_n is an $n \times 1$ vector of disturbances of independent and identically distributed (iid) elements with mean zero and variance σ^2, β is a $k \times 1$ vector of regression coefficients and λ is the spatial parameter. The Gaussian loglikelihood of $\theta = (\beta', \sigma^2, \lambda)$ is,

$$\ell_n(\theta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\left(\sigma^2\right) + \ln|A_n(\lambda)| - \frac{1}{2\sigma^2}\epsilon_n'(\beta,\lambda)\epsilon_n(\beta,\lambda), \tag{2}$$

where $A_n(\lambda) = I_n - \lambda W_n$, I_n is an $n \times n$ identity matrix, and $\epsilon_n(\beta, \lambda) = A_n(\lambda)Y_n - X_n\beta$. Given λ, $\ell_n(\theta)$ is maximized at $\hat\beta_n(\lambda) = (X_n'X_n)^{-1}X_n'A_n(\lambda)Y_n$ and $\hat\sigma_n^2(\lambda) = \frac{1}{n}Y_n'A_n'(\lambda)M_nA_n(\lambda)Y_n$, where $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. By substituting $\hat\beta_n(\lambda)$ and $\hat\sigma_n^2(\lambda)$ into $\ell_n(\theta)$, we arrive at the concentrated Gaussian loglikelihood function for λ as,

$$\ell_n^c(\lambda) = -\frac{n}{2}[\ln(2\pi) + 1] - \frac{n}{2}\ln\left(\hat\sigma_n^2(\lambda)\right) + \ln|A_n(\lambda)|. \tag{3}$$

Maximizing $\ell_n^c(\lambda)$ gives the unconstrained QMLE $\hat\lambda_n$ of λ, and thus the QMLEs of β and σ^2 as $\hat\beta_n \equiv \hat\beta(\hat\lambda_n)$ and $\hat\sigma_n^2 \equiv \hat\sigma_n^2(\hat\lambda_n)$. Denote $\hat\sigma_n = (\hat\beta_n', \hat\sigma_n^2, \hat\lambda_n)'$, the QMLE of θ.

Under regularity conditions, Lee (2004) establishes the consistency and asymptotic normality of the QMLE $\hat\theta_n$. In particular, he shows that $\hat\lambda_n$ and $\hat\beta_n$ may have a slower than $\sqrt{n}$-rate of convergence if the degree of spatial dependence (or the number of neighbours each spatial unit has) grows with the sample size n. The QMLE and its asymptotic distribution developed by Lee are robust against nonnormality of the error distribution. However, some important issues need to be further considered: (*i*) conditions under which the regular QMLE $\hat\theta_n$ remains consistent when errors are heteroskedastic, (*ii*) methods to modify the regular QMLE $\hat\theta_n$ so that it becomes generally consistent under unknown heteroskedasticity, and (*iii*) methods for estimating the variance of the (modified) QMLE robust against unknown heteroskedasticity.

2.2. Robustness of QMLE against unknown heteroskedasticity

It is accepted that the regular QMLE of the usual linear regression model without spatial dependence, developed under homoskedastic errors, is still consistent when the errors are in fact heteroskedastic. However, for correct inferences the standard error of the estimator has to be adjusted to account for this unknown heteroskedasticity (White, 1980). Suppose now we have a linear regression model with

[1] The efficiency of an MLE may be the driving force for exploiting a likelihood-based estimator for achieving robustness against various model misspecifications such as heteroskedasticity and nonnormality. The computational complexity may be the key factor that hinders the application of the ML-type estimation method. However, with the modern computing technologies this is no longer an issue of major concern, unless n is very large.