A Simple and Robust Method of Inference for Spatial Lag Dependence*

Zhenlin YANG

School of Economics, Singapore Management University[†]

Yan SHEN

Department of Statistics, School of Economics, Xiamen University, P. R. China

emails: zlyang@smu.edu.sg, sheny@xmu.edu.cn

August 20, 2014

Abstract

A simple and reliable method of inference for the spatial parameter in spatial autoregressive models is introduced, based on a statistic obtained by centering and rescaling the numerator of the concentrated Gaussian score function. The resulted tests and confidence intervals are robust against the distributional misspecifications and are insensitive to the spatial layouts and the error standard deviation. In contrast, the standard methods based on Gaussian score and information matrix may lead to inconsistent inference when errors are nonnormal, and can be quite sensitive to the spatial layouts and the error standard deviation even when errors are normally distributed. Extensive Monte Carlo results are reported and an empirical illustration is given.

Key Words: Spatial dependence; Confidence interval; LM Tests; Centering; Rescaling; Finite sample performance; Robustness.

JEL Classification: C12, C13, C21

1 Introduction.

Consider the mixed regressive, spatial autoregressive (SAR) model:

$$Y_n = \lambda W_n Y_n + X_n \beta + u_n, \tag{1}$$

where n is the total number of spatial units, Y_n is an $n \times 1$ vector of observations on these spatial units, X_n is an $n \times k$ matrix containing the values of the exogenous regressors, W_n is a specified $n \times n$ spatial weights matrix, and u_n is an n-dimensional vector of independent and identically distributed (iid) disturbances of zero mean and finite variance σ^2, λ is the scalar spatial parameter, and β is a $k \times 1$ vector of regression coefficients. When there are no regressors X_n in the model, the SAR model becomes a pure SAR process.

*We benefited from the Vth World Conference of the Spatial Econometrics Association, Toulouse, 2011. We are grateful to three anonymous referees and the Editor Daniel A. Griffith for their helpful comments. Zhenlin Yang gratefully acknowledges the support from a research grant (Grant number: C244/MSS9E005) from Singapore Management University.

[†]Corresponding Author: 90 Stamford Road, Singapore 178903. Tel. No.: 65 68280852; Fax: 65 6828-0833

Due to its popularity in modelling cross-sectional dependence induced by neighborhood effects, spillover effects, copy-catting, peer-group effects, etc., the SAR model of Cliff and Ord (1973, 1981) has been extensively studied and applied in recent years.[1] One popular method for estimating the SAR model is the maximum likelihood (ML) or quasi-maximum likelihood (QML) (Ord, 1975; Smirnov and Anselin, 2001; Lee, 2004a,b). Let $\theta = (\beta', \sigma^2, \lambda)$. If the disturbances are exactly normal, we have the true loglikelihood function,

$$\ell_n(\theta) = -\frac{n}{2}\log(2\pi\sigma^2) + \log|A_n(\lambda)| - \frac{1}{2\sigma^2}\left[A_n(\lambda)Y_n - X_n\beta\right]'\left[A_n(\lambda)Y_n - X_n\beta\right], \quad (2)$$

where $A_n(\lambda) = I_n - \lambda W_n$ and I_n is an $n \times n$ identity matrix. Maximizing $\ell_n(\theta)$ gives the ML estimator (MLE) of θ. If the errors are not exactly normal, as assumed in this paper, $\ell_n(\theta)$ can still be used as a working log-likelihood called the *quasi-loglikelihood* and maximizing it would still produce a consistent estimator of θ provided that certain regularity conditions are satisfied (Lee, 2004a). The resulted estimator is called the quasi-maximum likelihood estimator (QMLE). Now, given λ, $\ell_n(\theta)$ can be partially maximized, giving the constrained QMLEs of β and σ^2, respectively, as

$$\hat{\beta}_n(\lambda) = (X_n'X_n)^{-1}X_n'A_n(\lambda)Y_n, \text{ and} \quad (3)$$

$$\hat{\sigma}_n^2(\lambda) = \frac{1}{n}Y_n'A_n'(\lambda)M_nA_n(\lambda)Y_n, \quad (4)$$

where $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$. The concentrated loglikelihood of λ is thus

$$\ell_n^c(\lambda) = -\frac{n}{2}[\log(2\pi) + 1] - \frac{n}{2}\log\hat{\sigma}_n^2(\lambda) + \log|A_n(\lambda)|. \quad (5)$$

Maximizing $\ell_n^c(\lambda)$ gives the unconstrained QMLE $\hat{\lambda}_n$ of λ, and substituting $\hat{\lambda}_n$ into $\hat{\beta}_n(\lambda)$ and $\hat{\sigma}_n^2(\lambda)$ gives the unconstrained QMLEs of β and σ^2 as $\hat{\beta}_n \equiv \hat{\beta}_n(\hat{\lambda}_n)$ and $\hat{\sigma}_n^2 \equiv \hat{\sigma}_n^2(\hat{\lambda}_n)$, respectively. Denote $\hat{\theta}_n = (\hat{\beta}_n', \hat{\sigma}_n^2, \hat{\lambda}_n)'$, the unconstrained QMLE of θ.

Lee (2004a) gives a detailed study on the asymptotic properties of QML estimation of Model (1). In particular, he showed that the QMLEs of β and λ are $\sqrt{n}$-consistent if each spatial unit depends on a fixed number of neighbors, otherwise they are $\sqrt{n/h_n}$-consistent if the number of neighbors is of order h_n such that as $n \to \infty$, $h_n \to \infty$ and $h_n/n \to 0$. The QMLE of σ^2 is always $\sqrt{n}$-consistent. Lee's results lay the theoretical bases for the likelihood-based inferences, under the likelihood ratio, Wald, or the LM principle, for testing and confidence interval (CI) construction for the SAR model.

Clearly, inference for spatial parameter λ is central to the SAR model. The likelihood ratio and Wald methods require the estimation of the full model, which needs the numerical maximization of the concentrated loglikelihood function $\ell_n^c(\lambda)$ to obtain the QMLE of λ.

[1]The representative theoretical works include Kelejian and Prucha (1999, 2001), Lee (2002, 2003, 2004a, 2007a,b), Bao and Ullah (2007), Robinson (2010), Born and Breitung (2011), and Yang (2010b). The representative empirical applications include Case (1991), Case, et al. (1993), Besley and Case (1995), Brueckner (1998), Bell and Bockstael (2000), Bertrand, et al. (2000), and Topa (2001).

In contrast, the LM method requires only the estimation of the model for a given value of λ, which can be done by the ordinary least squares (OLS) method. While computationally the estimation of the SAR model is no longer a burden (see LeSage and Pace, 2009), the LM method is still preferred due to its easy implementation, and more importantly the fact that it can be easily 'corrected' for enhancing its robustness and finite sample performance. This is because the standard LM tests and the test-based CIs (i.e., the CIs obtained by inverting the tests) are derived under the assumption that the errors are normal, and thus they may not be robustness against the distributional misspecification; and even when the error distribution is known (e.g., normal) or the test is asymptotically robust against the distributional misspecification (e.g., testing for lack of spatial effect in SAR model considered in this paper), the standard LM tests may still suffer from finite sample size distortion due to the facts that the concentrated score is not centered and its variance estimator is biased (see the end of Sec. 3 for a discussion on this). A simple and reliable method for testing and CI construction for λ is thus desirable.

Section 2 introduces the standard LM tests, and the test-based CIs for λ. Section 3 introduces a robust version of the LM test, through which a robust CI is given. Section 4 presents Monte Carlo results for comparing the finite sample behaviors of the standard and the robust LM tests as well as the corresponding CIs. Section 5 presents an empirical application. Section 6 concludes the paper.

2 LM Tests and Confidence Intervals for Spatial Parameter

We are interested in testing and confidence interval (CI) construction for the spatial parameter λ in the SAR model. In particular, we are interested in the score-based inferences as they do not require the estimation of the spatial parameter, and thus avoid the numerical optimization which can be computationally demanding for large sample sizes and general spatial weight matrices. The classical inferences of this type under normality assumption are readily available based on the results of Anselin (1988a,b), Anselin and Rey (1991) and Lee (2004a). In particular, the score-based or LM test of the hypothesis of no SAR effect in the regression model, i.e., $H_0 : \lambda = 0$ vs $H_a : \lambda \neq 0$, is given in Anselin (1988a):

$$\text{LM}_A = \frac{\hat{u}'_{n0} W_n Y_n}{\hat{\sigma}_{n0}\sqrt{T_{0n}\hat{\sigma}^2_{n0} + \hat{\eta}'_{n0} M_n \hat{\eta}_{n0}}}, \tag{6}$$

where $T_{0n} = \text{tr}(W_n^2 + W_n' W_n)$, $\text{tr}(\cdot)$ denotes the trace of a square matrix, $\hat{\eta}_{n0} = W_n X_n \hat{\beta}_{n0}$, $\hat{u}_{n0} = Y_n - X_n \hat{\beta}_{n0}$, $\hat{\beta}_{n0} = \hat{\beta}_n(0)$, and $\hat{\sigma}^2_{n0} = \hat{\sigma}^2_n(0)$. Alternatively, LM_A can be written as

$$\text{LM}_A = \frac{Y_n' M_n W_n Y_n}{\hat{\sigma}_{n0}\sqrt{Y_n'(M_n T_{0n}/n + P_n' W_n' M_n W_n P_n) Y_n}},$$

where $P_n = I_n - M_n$ and M_n is defined below (4). When the errors are iid normal, LM_A is asymptotically $N(0, 1)$ under the null hypothesis of no spatial lag effect. However, it is not clear whether this asymptotic normality holds when the errors are nonnormal.

A more general test of spatial effect in the SAR model is the test of the null hypothesis $H_0 : \lambda = \lambda_0$ versus the alternative hypothesis $H_a : \lambda \neq \lambda_0$ where λ_0 is the hypothesized value for the spatial parameter, not necessarily zero. This general test is more interesting in the sense that (i) it can be used to decide a plausible (non-rejection) value for λ, e.g., .5, so that model estimation proceeds by simply running an OLS of $A_n(.5)Y_n$ on X_n and inferences are carried out conditionally on the chosen $\lambda = .5$,[2] and (ii) it can be inverted to give a confidence interval for λ without having to estimate λ, and more importantly it can be easily 'corrected', resulting in a robust confidence interval for λ with an improved finite sample performance (see Section 3 for details).

Let $S_n^c(\lambda) = \frac{d}{d\lambda}\ell_n^c(\lambda)$ be the concentrated score function. We have,

$$S_n^c(\lambda) = -\text{tr}(G_n(\lambda)) + \hat{\sigma}_n^{-2}(\lambda)Y_n'A_n'(\lambda)M_nW_nY_n = \hat{\sigma}_n^{-2}(\lambda)\hat{u}_n(\lambda)'G_n^\circ(\lambda)A_n(\lambda)Y_n, \tag{7}$$

where $\hat{u}_n(\lambda) = A_n(\lambda)Y_n - X_n\hat{\beta}_n(\lambda) = M_nA_n(\lambda)Y_n$, $G_n(\lambda) = W_nA_n^{-1}(\lambda)$, and $G_n^\circ(\lambda) = G_n(\lambda) - \frac{1}{n}\text{tr}(G_n(\lambda))I_n$.

The variance of $S_n^c(\lambda)$ can be estimated in at least two different ways in the context of the SAR model. One is based on the expected information matrix and the other is based on the observed information matrix, resulting in two versions of LM tests of the general hypothesis. The expected information matrix, $I_n(\theta) = -\text{E}\left(\frac{\partial^2}{\partial\theta\partial\theta'}\ell_n(\theta)\right)$, is given as

$$I_n(\theta) = \frac{1}{\sigma^2}\begin{pmatrix} X_n'X_n, & 0, & X_n'\eta_n(\lambda) \\ 0, & \frac{n}{2\sigma^2}, & \text{tr}G_n(\lambda) \\ \eta_n(\lambda)'X_n, & \text{tr}G_n(\lambda), & \eta_n(\lambda)'\eta_n(\lambda) + \sigma^2\text{tr}(G_n^2(\lambda) + G_n'(\lambda)G_n(\lambda)) \end{pmatrix},$$

where $\eta_n(\lambda) = G_n(\lambda)X_n\beta$. Partition $I_n(\theta)$ according to (β, σ^2) and λ, and denote the submatrices by $I_{n,11}, I_{n,12}, I_{n,21}$ and $I_{n,22}$. Then, the asymptotic variance of $S_n^c(\lambda)$ is

$$\begin{aligned} \text{AVar}[S_n^c(\lambda)] &= I_{n,22} - I_{n,21}I_{n,11}^{-1}I_{n,12} \\ &= \sigma^{-2}\eta_n(\lambda)'M_n\eta_n(\lambda) + \text{tr}[G_n^2(\lambda) + G_n'(\lambda)G_n(\lambda)] - 2[\text{tr}(G_n(\lambda))]^2. \end{aligned} \tag{8}$$

Combining (7) and (8), evaluating at the constrained MLEs and simplifying, we obtain an LM statistic for inference for λ,

$$\text{LM}_E(\lambda) = \frac{\hat{u}_n(\lambda)'G_n^\circ(\lambda)A_n(\lambda)Y_n}{\hat{\sigma}_n(\lambda)\sqrt{\hat{\eta}_n(\lambda)'M_n\hat{\eta}_n(\lambda) + \hat{\sigma}_n^2(\lambda)T_{1n}(\lambda)}}, \tag{9}$$

where $\hat{\eta}_n(\lambda) = G_n(\lambda)X_n\hat{\beta}_n(\lambda)$ and $T_{1n}(\lambda) = \text{tr}[G_n^\circ(\lambda)^2 + G_n^\circ(\lambda)'G_n^\circ(\lambda)]$. When $\lambda = 0$, we have $A_n(0) = I_n$, $G_n^\circ(0) = G_n(0) = W_n$, $\eta_{n0} = W_nX_n\beta$, and $T_{1n}(0) = \text{tr}(W_n^2 + W_n'W_n)$. Thus, $\text{LM}_E(0)$ simplifies to LM_A given in (6).

[2]This is in principle similar to the case of testing $H_0 : \lambda = 0$: when H_0 is not rejected one may simply run an OLS of Y_n on X_n and make inference conditional on $\lambda = 0$. See the empirical illustration given in Section 5. These ideas are also similar to those in the Box-Cox transformation model (Box and Cox, 1964).

An alternative way to estimate the variance of $S_n^c(\lambda)$ is to replace the expected information submatrices by the corresponding observed information submatrices evaluated at the constrained MLEs, resulting in an expression that is identical to

$$H_n(\lambda) = -\frac{d^2}{d\lambda^2}\ell_n^c(\lambda) = \text{tr}(G_n^2(\lambda)) + R_{2n}(\lambda) - \frac{2}{n}R_{1n}^2(\lambda),$$

where $R_{1n}(\lambda) = \hat{\sigma}_n^{-2}(\lambda)Y_n'A_n'(\lambda)M_nW_nY_n$ and $R_{2n}(\lambda) = \hat{\sigma}_n^{-2}(\lambda)Y_n'W_n'M_nW_nY_n$. This leads to the Hessian-based LM statistic of the form,

$$\text{LM}_H(\lambda) = \frac{\hat{u}_n'(\lambda)G_n^\circ(\lambda)A_n(\lambda)Y_n}{\sigma_n^2(\lambda)\sqrt{\text{tr}(G_n^2(\lambda)) + R_{2n}(\lambda) - \frac{2}{n}R_{1n}^2(\lambda)}}. \tag{10}$$

Under the assumptions that the model disturbances are iid normal, $\text{LM}_H(\lambda) \xrightarrow{D} N(0,1)$ (see Lee, 2004a, p. 1911), and a similar result holds for $\text{LM}_E(\lambda)$. The two inferential statistics are asymptotically equivalent and they lead immediately to two asymptotically equivalent tests and two asymptotically equivalent CIs for λ.

Thus, for testing $H_0 : \lambda = \lambda_0$ versus $H_a : \lambda \neq \lambda_0$, one rejects H_0 at α level of significance if $|\text{LM}_E(\lambda_0)| > Z_{\alpha/2}$, or if $|\text{LM}_H(\lambda_0)| > Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the upper $\alpha/2$-quantile of the standard normal distribution. Both tests are very simple to implement and the most interesting case is to test $H_0 : \lambda = 0$. However, if such a test is rejected, one would be interested in making a more precise statement about the true value of λ. Thus, a confidence interval statement for λ is desirable, which can simply be obtained by inverting the tests. A $100(1-\alpha)\%$ large sample CI for λ obtained by inverting $\text{LM}_E(\lambda)$ is defined as

$$\text{CI}_E(\lambda) = \Big(\min\{\lambda : \text{LM}_E(\lambda) \geq -Z_{\alpha/2}\},\ \ \max\{\lambda : \text{LM}_E(\lambda) \leq Z_{\alpha/2}\}\Big), \tag{11}$$

and similarly, a $100(1-\alpha)\%$ large sample CI for λ based on $\text{LM}_H(\lambda)$ is defined as

$$\text{CI}_H(\lambda) = \Big(\min\{\lambda : \text{LM}_H(\lambda) \geq -Z_{\alpha/2}\},\ \ \max\{\lambda : \text{LM}_H(\lambda) \leq Z_{\alpha/2}\}\Big). \tag{12}$$

Lee (2004a, p. 1911) commented that even when $\{u_i\}$ are not normally distributed the $\text{LM}_H(\lambda)$ test can still be asymptotically valid as long as $\lim_{n\to\infty} h_n = \infty$ and $\gamma = 0$, where γ is the third central moment of u_i. Thus, one would expect a similar conclusion holds for $\text{LM}_E(\lambda)$. This conclusion implies that when the error distribution is skewed, the tests $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$ can be asymptotically invalid. However, he did not proceed to provide results that correct the non-robustness of the LM tests against the skewness. Furthermore, when h_n is bounded and the disturbances are nonnormal, the asymptotic behaviors of these tests are not clear. Also, to the best of our knowledge, there is no result available in the literature about the finite sample performance of these tests and the corresponding test-based CIs even when the disturbances are iid normal.

In this paper, we show that the two LM tests discussed above are in general not robust against nonnormality. We introduce a robust LM test statistic by centering and then rescaling the numerators of $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$, which captures the effects of both skewness and excess kurtosis and thus is robust against the nonnormality of the error distribution whether h_n is bounded or unbounded.[3] We show that such corrections are also effective in improving the finite sample performance of the LM tests even when the disturbances are iid normal. This robust test can be inverted to give a more reliable CI for λ. We further show that $\text{LM}_A = \text{LM}_E(0)$ and $\text{LM}_H(0)$ are asymptotically robust against skewness and excess kurtosis, but Monte Carlo results show that their finite sample behavior can be quite dependent on the spatial layout and the magnitude of error standard deviation. Monte Carlo results also show that the robust LM test and the corresponding confidence interval perform well in finite sample, and they clearly outperform the non-robust counterparts.

3 Robust LM Tests and CIs for Spatial Parameter

From the discussion above, we see that it is highly desirable to derive a test that is not only asymptotically robust against the distributional misspecification, but also insensitive to the spatial layouts and error standard deviation in finite sample. Motivated by Yang (2010a), we first note that the key quantity, $\hat{u}_n'(\lambda)G_n^\circ(\lambda)A_n(\lambda)Y_n$, in the concentrated score function $S_n^c(\lambda)$ given in (7) can be written as

$$\hat{u}_n'(\lambda)G_n^\circ(\lambda)A_n(\lambda)Y_n = u_n'M_nG_n^\circ(\lambda)u_n + u_n'M_nG_n^\circ(\lambda)X_n\beta,$$

because $\hat{u}_n(\lambda) = M_nA_n(\lambda)Y_n$, $A_n(\lambda)Y_n = X_n\beta + u_n$, and $M_nX_n = 0$. It follows that

$$\text{E}\left[\hat{u}_n'(\lambda)G_n^\circ(\lambda_0)A_n(\lambda)Y_n\right] = \sigma^2\text{tr}[M_nG_n^\circ(\lambda)], \tag{13}$$

which is clearly not zero in general, although it approaches to zero when $n \to \infty$. This indicates that the standard LM statistics may not be centered properly for a finite n, which suggests that one should work with the centered quantity

$$\hat{u}_n'(\lambda)G_n^\circ(\lambda)A_n(\lambda)Y_n - \sigma^2\text{tr}[M_nG_n^\circ(\lambda)]$$

or its feasible version, obtained by replacing σ^2 by its unbiased (constrained) estimator,

$$\hat{u}_n'(\lambda)G_n^\circ(\lambda)A_n(\lambda)Y_n - \frac{n}{n-k}\hat{\sigma}_n^2(\lambda)\text{tr}[M_nG_n^\circ(\lambda)] = \hat{u}_n'(\lambda)D_n(\lambda)A_n(\lambda)Y_n, \tag{14}$$

where $D_n(\lambda) = G_n^\circ(\lambda) - \frac{1}{n-k}\text{tr}(M_nG_n^\circ(\lambda))I_n$. Clearly, the quantity in (14) has a zero mean.

Second, we note that the estimators of the variance of the score function are obtained under the assumption that the errors of the model are normally distributed. These variance estimators may not be consistent when the errors are not normally distributed. As a result, the distributions of $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$ may not converge to $N(0,1)$. Thus, a correction on the variance is also necessary after the mean correction. It is easy to see that

[3] Clearly, the nonnormality of the error distribution includes the case where there exist gross errors. See the case of the mixture normal in the Monte Carlo simulations.

$\hat{u}_n'(\lambda)D_n(\lambda)A_n(\lambda)Y_n = u_n'M_nD_n(\lambda)u_n + u_n'M_nD_n(\lambda)X_n\beta = u_n'M_nD_n(\lambda)u_n + u_n'M_n\eta_n(\lambda)$. By Lemma A.4 (ii) in the appendix, we have

$$\begin{aligned}&\text{Var}\left(\hat{u}_n'(\lambda)D_n(\lambda)A_n(\lambda)Y_n\right)\\ =\;& \sigma^4 T_{2n}(\lambda) + \sigma^2\eta_n'(\lambda)M_n\eta_n(\lambda) + \sigma^4\kappa d_n'(\lambda)d_n(\lambda) + 2\sigma^3\gamma\eta_n'(\lambda)M_nd_n(\lambda), \qquad (15)\end{aligned}$$

where $T_{2n}(\lambda) = \text{tr}[M_n(D_n(\lambda) + D_n'(\lambda))M_nD_n(\lambda)]$, $d_n(\lambda) = \text{diagv}(M_nD_n(\lambda))$, the vector of diagonal elements of $M_nD_n(\lambda)$, and γ and κ are, respectively, the measures of skewness and excess kurtosis of $u_{n,i}$. This variance formula captures the effects of skewness and excess kurtosis of the errors, and is thus robust against nonnormality in these senses. Using (14) and (15), one obtains a modified LM-type statistic that is properly centered and rescaled, and thus would be robust against distributional misspecifications and spatial layouts.[4]

Some regularity conditions are necessary before we introduce the new robust test and confidence interval for λ.

Assumption 1: *The innovations $\{u_i\}$ are iid with mean zero, variance σ^2, skewness γ and excess kurtosis κ. Also, the moment $E|u_i|^{4+\epsilon}$ exists for some $\epsilon > 0$.*

Assumption 2: *The elements of the $n \times k$ matrix X_n are uniformly bounded for all n, and $\lim_{n\to\infty} \frac{1}{n}X_n'X_n$ exists and is nonsingular.*

Assumption 3: *The elements $\{w_{n,ij}\}$ of W_n are at most of order h_n^{-1} uniformly for all i, j, with the rate sequence $\{h_n\}$, bounded or divergent, satisfying $h_n^{1+\delta}/n \to 0$ as $n \to \infty$ for some $\delta > 0$.*

Assumption 4: *The sequences of matrices $\{W_n\}$ and $\{A_n^{-1}(\lambda)\}$ are uniformly bounded in both row and column sums.*[5] *As a normalization, $w_{n,ii} = 0$, for all i.*

Assumption 5: *$\{A_n^{-1}(\lambda^*)\}$ is uniformly bounded in either row or column sums uniformly in λ^* in a compact set containing in its interior the true value λ.*

Assumption 6: *The elements of $M_n\eta_n(\lambda)$ are of uniform order $O(1/\sqrt{h_n})$, and for $0 \le c < \infty$, $\lim_{n\to\infty}(h_n/n)\eta_n'(\lambda)M_n\eta_n(\lambda) = c$.*

These assumptions are essentially adapted from Lee (2004a). Assumption 1 is required for the application of the central limit theorem for linear-quadratic forms of Kelejian and Prucha (2001) for the cases when h_n is bounded, and its extended version by Lee (2004a,

[4] While the idea of centering and rescaling are not new (see, e.g., Koenker, 1981; Moulton and Randolph, 1989; and Robinson, 2008), there is an issue of how to implement them. Our method is clearly the simplest. The same idea is following by Baltagi and Yang (2013a) for standardizing the LM tests for the lack of spatial error dependence in linear or panel regressions. Anselin et al. (1996) presents LM tests robust against local misspecifications, i.e., LM test for spatial lag dependence in the possible presence of a missing spatial error dependence, and vise versa, which are labeled as `RLMlag` and `RLMerr`, respectively, in the `sedep R` package. See for example Bivand et al. (2013) for more details on these.

[5] That is, e.g., for W_n, $\sup_i \sum_{j=1}^n |w_{n,ij}| < \infty$ and $\sup_j \sum_{i=1}^n |w_{n,ij}| < \infty$.

Appendix A) for the cases when h_n is unbounded. Assumption 3 identifies the different types of spatial dependence considered. Typically, one type of spatial dependence corresponds to the case where each unit has a fixed number of neighbors, which in turn means that h_n is bounded. The other type of spatial dependence corresponds to the case where the number of neighbors of each spatial unit grows as n goes to infinity, and in this case h_n is divergent. See Case (1991) and the discussions in Lee (2004a, p. 1903) for the practical situations when this might occur. However, h_n can only increase at a slower rate than n (i.e., one needs to limit the spatial dependence to a manageable degree) to ensure the proper $\sqrt{n/h_n}$-consistency of $\hat{\lambda}_n$.[6] Assumptions 3 and 4 provide conditions for this. Assumptions 5 and 6 are, respectively, Assumptions 7 and 10 of Lee (2004a).

Now, recall the quantities defined earlier: $T_{2n}(\lambda) = \text{tr}[M_n(D_n(\lambda) + D_n'(\lambda))M_n D_n(\lambda)]$, $d_n(\lambda) = \text{diagv}(M_n D_n(\lambda))$. Let $\hat{\eta}_{n0} \equiv \hat{\eta}_n(\lambda) = G_n(\lambda)X_n\hat{\beta}_n(\lambda)$, $\hat{\sigma}_{n0}^2 \equiv \hat{\sigma}_n^2(\lambda)$, and $d_{n0} \equiv d_n(\lambda)$. Let $\hat{\gamma}_{n0}$ and $\hat{\kappa}_{n0}$ are, respectively, the sample skewness and excess kurtosis of $\hat{u}_n(\lambda)$. The following theorem presents a robustified version of the LM test statistics given above.

Theorem 1. *Under the Assumptions 1-6, a robustified LM-type inferential statistic for λ takes the following form*

$$LM_R(\lambda) = \frac{\hat{u}_n'(\lambda)D_n(\lambda)A_n(\lambda)Y_n}{\hat{\sigma}_{n0}\sqrt{\hat{\eta}_{n0}'M_n\hat{\eta}_{n0} + \hat{\sigma}_{n0}^2 T_{2n}(\lambda) + \hat{\sigma}_{n0}^2\hat{\kappa}_{n0}d_{n0}'d_{n0} + 2\hat{\sigma}_{n0}\hat{\gamma}_{n0}\hat{\eta}_{n0}'M_n d_{n0}}}, \tag{16}$$

such that (i) $LM_R(\lambda) \xrightarrow{D} N(0,1)$; (ii) $LM_E(\lambda)$ and $LM_H(\lambda)$ are in general not asymptotically equivalent to $LM_R(\lambda)$ when h_n is bounded, but they are when h_n is divergent, and (iii) when $\lambda = 0$, $LM_E(0)$, $LM_H(0)$ and $LM_R(0)$ are asymptotically equivalent.

The proof of Theorem 1 is given in the appendix. Theorem 1 leads immediately to a robust test for testing $H_0 : \lambda = \lambda_0$ against $H_a : \lambda \neq \lambda_0$, which rejects H_0 in favor of H_a if $\text{LM}_R(\lambda_0) > Z_{\alpha/2}$, and a robust CI for λ as

$$\text{CI}_R(\lambda) = \Big(\min\{\lambda : \text{LM}_R(\lambda) \geq -Z_{\alpha/2}\},\ \max\{\lambda : \text{LM}_R(\lambda) \leq Z_{\alpha/2}\}\Big). \tag{17}$$

The results of Theorem 1 imply that if one knows that h_n is bounded, one should use $\text{LM}_R(\lambda)$ as $\text{LM}_E(\lambda)$ or $\text{LM}_H(\lambda)$ may not lead to correct inference statements for the spatial effect λ even when n is large unless one knows for sure the error distribution is normal; if one knows that $\lim_{n\to\infty} h_n = \infty$, one can choose any of the three LM statistics as they are asymptotically equivalent and are robust against the distributional misspecification.[7]

[6] Lee (2004a, Footnote 8) commented that whether h_n is bounded or divergent has interesting implications on the least square estimation of β and λ, i.e., the least square estimators are inconsistent when h_n is bounded, but can be consistent when h_n is divergent. The results presented in this paper show that the behavior of h_n has interesting implications on the robustness of the standard LM statistics as well.

[7] Lee (2004a, p.1911) stated that the classical inference methods are valid as long as $\lim_{n\to\infty} h_n = \infty$ and $\gamma = 0$. However, our results show that $\gamma = 0$ is not required for the asymptotic validity of the classical inference methods. See the proof of Theorem 1 given in the appendix and the Monte Carlo results provided in the next section.

However, simple but tedious derivations, following the proof of Theorem 1, show that

$$\mathrm{E}[\mathrm{LM}_E(\lambda)] = O((h_n/n)^{1/2}) \text{ and } \mathrm{E}[\mathrm{LM}_R(\lambda)] = O((h_n/n)^{3/2}),$$

which implies that the mean of $\mathrm{LM}_E(\lambda)$, and hence the mean of $\mathrm{LM}_H(\lambda)$, can differ from zero much more significantly than the mean of $\mathrm{LM}_R(\lambda)$ if each spatial unit has many neighbors. Furthermore, $\mathrm{LM}_R(\lambda)$ is constructed by using the variance of its numerator directly, but $\mathrm{LM}_E(\lambda)$ is not though an asymptotically equivalent quantity is used. Thus, it is expected that the variance of $\mathrm{LM}_E(\lambda)$ may differ from its nominal value, 1, more than that of $\mathrm{LM}_R(\lambda)$. Monte Carlo results given in the following section provide a strong support to these statements. Thus, it is suggested that one should use the robust statistic $\mathrm{LM}_R(\lambda)$ to conduct statistical inference for λ. When $\lambda = 0$, the three statistics are asymptotically equivalent, meaning that any of the three can be used for testing $H_0 : \lambda = 0$. However, Monte Carlo results show that $\mathrm{LM}_R(0)$ has a much better finite sample performance and is much less sensitive to the spatial layouts and the error standard deviation than $\mathrm{LM}_E(0)$ and $\mathrm{LM}_H(0)$. Thus, $\mathrm{LM}_R(0)$ is deemed more reliable than $\mathrm{LM}_E(0)$ and $\mathrm{LM}_H(0)$.

4 Monte Carlo Study

The finite sample performance of the inference methods for the spatial parameter in the spatial autoregressive model introduced in this paper is evaluated based on a series of Monte Carlo experiments. These experiments involve a number of different error distributions and a number of different spatial layouts. Comparisons are made between the usual LM tests and the corresponding CIs and their robust counterparts to see the effects of the error distributions and the spatial layouts.

4.1 Spatial layouts and error distributions

Two general spatial layouts are considered in the Monte Carlo experiments and they are applied to different test statistics involved in the experiments. The first is based on the Queen contiguity, and the second is based on the notion of group or social interactions (Case, 1991; Lee, 2004a) with the number of groups $G = n^\delta$ where $0 < \delta < 1$. In the case of Queen contiguity, the number of neighbors is between 3 and 8, which does not change with the sample size n, whereas in the case of group interaction, the number of neighbors for each spatial unit increases with the increase of n but at a slower rate. Also, the number of neighbors is allowed to change from group to group.

The details for generating the W_n matrix under Queen contiguity are as follows: (i) index the n spatial units by $\{1, 2, \ldots, n\}$, randomly permute these indices and then allocate them into a lattice of $r \times m (\geq n)$ squares, (ii) let $W_{n,ij} = 1$ if the index j is in a square which shares either a common side or a vertex with the square containing the index i, otherwise

$W_{n,ij} = 0$, and (iii) divide each element of W_n by its row sum. Other weight matrices based on spatial contiguity can be constructed in a similar manner. See, e.g., Anselin (1988b).

To generate the W_n matrix according to the group interaction scheme, (i) calculate the number of groups according to $G = \texttt{Round}(n^\delta)$, and the approximate average group size $m = n/G$, (ii) generate the group sizes $(n_1, n_2, \ldots, n_G)$ according to a discrete uniform distribution from $m/2$ to $3m/2$, (iii) adjust the group sizes so that $\sum_{g=1}^{G} n_g = n$, and (iv) define $W_n = \text{diag}\{W_g/(n_g - 1), g = 1, \ldots, G\}$, a matrix formed by placing the submatrices W_g along the diagonal direction, where W_g is an $n_g \times n_g$ matrix with ones on the off-diagonal positions and zeros on the diagonal positions. In our Monte Carlo experiments, we choose $\delta = 0.3$, 0.5, and 0.7, representing respectively the situations where (i) there are few groups with many spatial units in each group, (ii) the number of groups and the sizes of the groups are of the same magnitude, and (iii) there are many groups with few elements in each. Clearly, under Queen contiguity, h_n defined in the theorems is bounded, whereas under group interaction, h_n is divergent with the rate $n^{1-\delta}$. Note that the latter spatial layout contains that of Case (1991) as a special case.

Figure 1 gives a graphical representation illustrating the neighboring structure of the spatial layouts used in the Monte Carlo experiments. See the book by Bivand et al. (2013, p.272) for the method and the book's website, `http://www.asdar-book.org`, for the `R` code in producing this type of plots based on a given spatial weight matrix.

The reported Monte Carlo results correspond to the following three error distributions: (i) standard normal, (ii) mixture normal, standardized to have mean zero and variance 1, and (iii) log-normal, also standardized to have mean zero and variance one. The standardized normal-mixture variates are generated according to

$$u_i = ((1 - \xi_i)Z_i + \xi_i \tau Z_i)/(1 - p + p * \tau^2)^{0.5},$$

where ξ is a Bernoulli random variable with probability of success p and Z_i is standard normal independent of ξ. The parameter p in this case also represents the proportion of mixing the two normal populations. In our experiments, we choose $p = 0.1$, meaning that 90% of the random variates are from standard normal and the remaining 10% are from another normal population with standard deviation τ. We choose $\tau = 4$ to simulate the situation where there are gross errors in the data. The standardized lognormal random variates are generated according to

$$u_i = [\exp(Z_i) - \exp(0.5)]/[\exp(2) - \exp(1)]^{0.5}.$$

This gives an error distribution that is both skewed and leptokurtic. The normal mixture gives an error distribution that is still symmetric like normal but leptokurtic. All the Monte Carlo experiments are based on 10,000 replications.

4.2 Performance of the tests

The performance of the robustified LM statistic, $\text{LM}_R(\lambda)$, introduced in Section 3 is compared with that of the usual LM statistics $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$. The Monte Carlo experiments are carried out based on the following data generating process:

$$Y_i = \lambda w'_{n,i} Y_n + \beta_0 + X_{1i}\beta_1 + X_{2i}\beta_2 + u_i.$$

When the Queen-contiguity spatial layout is used, X_{1i}'s are drawn from $\sqrt{12}U(0,1)$ and X_{2i}'s are drawn from $N(0,1)$. When the group-interaction spatial layout is used, the regressors are generated as in Lee (2004a) to allow the values within a group to be correlated. Specifically, the regressors X_{1ig} and X_{2ig} of the ith member in the gth group are generated as $X_{1ig} = (2z_{1g} + z_{1ig})/\sqrt{5}$ and $X_{2ig} = (2z_{2g} + z_{2ig})/\sqrt{5}$, where all the random variates z_{1g}, z_{1ig}, z_{2g} and z_{2ig} are iid $N(0,1)$. Furthermore, the parameters $\beta = \{5,1,1\}'$ and $\sigma = 2$. Four different sample sizes are considered, i.e., $n = 50, 100, 200$, and 500. The Monte Carlo results are reported in the form of plots for conciseness, and the detailed tabular information is available from `http://www.mysmu.edu/faculty/zlyang/`.

Size of Tests and Coverage Probability of CIs. The empirical mean, standard deviation (SD), and the 5% equi-tailed probability of the three statistics, $\text{LM}_E(\lambda)$, $\text{LM}_H(\lambda)$, and $\text{LM}_R(\lambda)$, are reported in Figures 2-5 where the horizontal index $\{1, 2, \ldots, 28\}$ corresponds to $\lambda = \{.75, .5, .25, .0, -.25, -.5, -.75\}$ for each of $n = \{50, 100, 200, 500\}$. Figures 2-4 corresponds to group interaction spatial layout with, respectively, $G = n^{0.3}$, $G = n^{0.5}$ and $G = n^{0.7}$, and Figures 5 corresponds to Queen contiguity. The results generally show that both $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$ can perform poorly in the sense that their empirical means, SDs and tail probabilities can be far from their nominal levels which are 0, 1 and 0.05, respectively. The true value of λ also affects the performance of these two tests. In contrast, $\text{LM}_R(\lambda)$ performs well in general, irrespective of the error distributions, spatial layouts, the magnitude of the error standard deviation, and the true value of the spatial parameter. In particular, the empirical mean of $\text{LM}_R(\lambda)$ is always very close to 0, showing that our mean correction procedure works very well. The empirical SD of $\text{LM}_R(\lambda)$ is also fairly close to its nominal level 1, which shows that our rescaling procedure also works well. These two adjustments lead to a simple and reliable inference procedure for λ. More details on the finite sample performance of $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$ are as follows.

The empirical mean, SD, and tail probability of $\text{LM}_E(\lambda)$ can be far below their nominal levels (0, 1, 0.05). As a result, the inference based on $\text{LM}_E(\lambda)$ can be quite misleading. For example, when $n = 50$ and 100 with large group interactions (i.e., few large groups as in the case where $G = n^{0.3}$, Figure 2), the empirical mean can be as low as -0.6566 (corresponding to $\lambda = 0.25$ and $n = 100$), the empirical SD can be as low as 0.6737 (corresponding to $\lambda = -0.5$ and $n = 50$), and the empirical tail probability can be as low as 0.0069 (corresponding to $\lambda = -0.5$ and $n = 50$). Similar to $\text{LM}_E(\lambda)$, the $\text{LM}_H(\lambda)$ can also perform quite poorly. It performs worse than $\text{LM}_E(\lambda)$ in terms of empirical mean, but better in terms of empirical SD. Unlike $\text{LM}_E(\lambda)$ whose tail probability is almost always below and sometimes far below its nominal level, the tail probability of $\text{LM}_H(\lambda)$ tends to be above its nominal level and can often be far above its nominal level, in particular when sample size is small and spatial dependence is strong, e.g., in Figure 2 with $\lambda = -0.75$ and $n = 50$, the empirical tail probability is 0.1238 compared with nominal level 0.05.[8]

The results in the tables show that one of the major factors affecting the distribution of the two standard LM statistics is the spatial layout, or rather the degree of spatial dependence. In contrast, the new test is much more robust to the spatial layout. In situations of a large group interaction, e.g., $G = \texttt{Round}(n^{0.3})$ as in Figure 2, the number of groups ranges from 3 to 6 for n ranging from 50 to 500. Thus, there are only a few groups, each containing many spatial units which are all neighbors of each other. This heavy spatial dependence distorts severely the distributions of $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$. In comparison, in situations of small group interaction, e.g., $G = \texttt{Round}(n^{0.7})$ as in Figure 4, the number of groups ranges from 15 to 77 for n ranging from 50 to 500. In this case, there are many groups each having only 3 to 8 units, giving a spatial layout with a very weak spatial dependence. As a result, the distributions of $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$ are much closer to N(0,1). The results (not reported for brevity) further show that the error standard deviation also heavily affects the performance of the two standard statistics $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$, but has little effect on the robust LM statistic $\text{LM}_R(\lambda)$.

Power of the tests. Empirical frequencies of rejection of the three tests for testing $H_0 : \lambda = 0$ are plotted in Figures 6 & 7 against the values of λ from -0.75 to 0.75 (horizontal line). In our power comparison, simulated critical values for each test are used, which means that the reported powers of the tests are size-adjusted. Figure 6 corresponds to group interaction spatial layout with $G = n^{0.5}$ while Figure 7 is for Queen contiguity. Each figure contains nine plots, corresponding to the combinations of three error distributions and three sample sizes.

The two figures reveal that the spatial layout and the sample size are the two important factors affecting the power of these tests. With less neighbors or with a larger sample, the tests become more powerful. It is interesting to note that when there is spatial dependence, it is harder to detect the spatial dependence when the spatial parameter is negative than when it is positive (see Figure 6). The error distribution does not seem to affect the power of the tests much, as the three plots in the same line look very similar.

The two figures also show that the power of $\text{LM}_E(\lambda)$ and $\text{LM}_R(\lambda)$ is very close to each

[8] We note that both $\text{LM}_E(\lambda)$ and $\text{LM}_H(\lambda)$ can perform worse when $n = 100$ than when $n = 50$. The reason is that from $n = 50$ to $n = 100$, the number of groups increase only from 3 to 4, and the average group size increases from 16.7 to 25. This means that although a large sample contains more information, under this particular spatial layout, increasing sample size from 50 to 100 is not enough to compensate the increase in the degree of spatial dependence.

other, as their curves almost overlap; but surprisingly, the power of $\mathrm{LM}_H(\lambda)$ behaves in an odd way. As shown in Figure 6, for negative λ, $\mathrm{LM}_H(\lambda)$ seems to have a slightly better performance than the other two tests. But this advantage fades away when λ becomes positive, and the three tests performs very similar for λ from 0 to 0.5. When λ exceeds 0.5, the power of $\mathrm{LM}_H(\lambda)$ starts to drop sharply. This phenomenon can also be observed in Figure 7, though milder. The reason for this abnormal behavior of $\mathrm{LM}_H(\lambda)$ may be due to the fact that observed information matrix does not guarantee a positive variance estimate.

5 An Empirical Illustration

To illustrate the applications of the three tests and compare their performances, we adopt a well known data set here: the cigarettes demand for United States. The data contains a panel of 46 states over 30 years (1963-1992) and is listed as `CIGAR.TXT` on the Wiley web site related to Baltagi (2001). In the data set, the independent variable is `cigarette sales` (in packs per capita). The covariates are `price` (per pack of cigarettes); `population`; `population16` (above the age of 16); `consumer price index` (with 1983=100); per capita `disposable income`; and `minimum price` (in adjoining states per pack of cigarettes). In our study, only cross-sectional data are needed, thus without loss of generality, we focus on the three specified years: 1970, 1980 and 1990. Note that the covariate `consumer price index` is omitted in our SAR model, as for a given year, the consumer price index is fixed and is no longer a useful variable.

We consider two SAR models with Rook contiguity spatial weights: (I) both response and covariates are original; (II) both response and covariates are log transformed. The null hypothesis $H_0 : \lambda = \lambda_0$ is tested with λ_0 values ranging from -0.75 to 0.75. Also the CIs for λ are computed. The test results and CIs are summarized in Tables 1 and 2. Based on the data of 1970 and 1980, the three statistics lead to the same conclusion: the spatial effect is not significant. However, based on the 1990 data, the three statistics lead to different conclusions with $\mathrm{LM}_R(\lambda)$ showing a positive λ which is significant at 5% level based on both models, but the other two tests showing a non-significant λ based on the model with log scale, and a barely significant result based on the model with original scale. Thus the new statistic shows a stronger evidence for the existence of the spatial dependence among the cigarette sales in 1990 at the different states. This result is reasonable considering the fast developments in transformation and communications over the period 1970-1990.

6 Conclusion

This paper introduces a robust statistic $\mathrm{LM}_R(\lambda)$ for making inferences for the spatial lag dependence parameter λ in a spatial autoregressive model. The new statistic is constructed by first centering the numerator of the concentrated (quasi-) score function of λ, and then finding the variance of the feasible version of the centered quantity, allowing the errors to be nonnormal. This corrects both the mean and the variance of the standard LM statistics. The mean adjustment is, however, often neglected in the literature, which happens to be more important in spatial models as the degree of spatial dependence can increase with the sample size (Lee, 2004a), making the concentrated score function more biased.

Compared with the inferences based on the two standard LM statistics, the inference based on the robust LM statistic is much more reliable. The robust statistic is seen to be very simple as well, thus it is recommended for the practical applications. The same idea can potentially be applied to many other spatial econometrics models, e.g., the spatial regression models with both spatial lag and spatial error dependence (Anselin, 1988a, b; Kelejian and Prucha, 2001), spatial panel data models with fixed effects, etc., where in particular the issue of constructing the robust confidence intervals/regions for the spatial parameter(s) has not been studied. The key is that the concentrated score function or in general the concentrated estimating equation can be written as linear-quadratic forms of a random vector of iid elements. However, each model has its own unique feature, we plan to pursue these issues in future research.

Throughout the paper, we have emphasized on producing inference methods for spatial effects that are not only robust against distributional misspecification but also possess good finite sample performance. There are other forms of model misspecifications popular in spatial econometrics models such as local misspecification (Anselin et al., 1996; Fang et al., 2014) and heteroskedasticity (Born and Breitung, 2011; Baltagi and Yang, 2013b). It would be of great interest to extend the ideas and methods followed in this paper to allow for these types of model misspecifications to give robust tests and confidence intervals for the spatial effects with an improved finite sample performance. As these issues are highly non-trivial, it will be pursued in a separate paper.

Appendix: Lemmas and Proof of the Theorem

For the proofs of the theorem and its corollary, we need the following lemmas.

Lemma A.1 (Lee, 2004a, p.1918): *Suppose that the elements of the* $n \times k$ *matrix* X_n *are uniformly bounded; and* $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ *exists and is nonsingular. Then the projectors* $P_n = X_n(X_n'X_n)^{-1}X_n'$ *and* $M_n = I_n - X_n(X_n'X_n)^{-1}X_n'$ *are uniformly bounded in both row and column sums.*

Lemma A.2 (Lemma A.9, Lee, 2004b): *Let* $\{A_n\}$ *be a sequence of* $n \times n$ *matrices that are uniformly bounded in both row and column sums. For* M_n *defined in Lemma A.1,*

(i) $\mathrm{tr}(M_n A_n) = \mathrm{tr}(A_n) + O(1)$,
(ii) $\mathrm{tr}(A_n' M_n A_n) = \mathrm{tr}(A_n' A_n) + O(1)$,
(iii) $\mathrm{tr}[(M_n A_n)^2] = \mathrm{tr}(A_n^2) + O(1)$, *and*
(iv) $\mathrm{tr}[(A_n' M_n A_n)^2] = \mathrm{tr}[(M_n A_n' A_n)^2] = \mathrm{tr}[(A_n' A_n)^2] + O(1)$.

Furthermore, if the elements $a_{n,ij}$ *of* A_n *are* $O(h_n^{-1})$ *uniformly in all* i *and* j, *then,*

(v) $\mathrm{tr}^2(M_n A_n) = \mathrm{tr}^2(A_n) + O(\frac{n}{h_n})$, *and*
(vi) $\sum_{i=1}^n ((M_n A_n)_{ii})^2 = \sum_{i=1}^n a_{n,ii}^2 + O(h_n^{-1})$,

where $(M_n A_n)_{ii}$ *is the* i*th diagonal element of* $M_n A_n$.

Lemma A.3 (Kelejian and Prucha, 1999; Lee, 2002): *Let* $\{A_n\}$ *and* $\{B_n\}$ *be two sequences of* $n \times n$ *matrices that are uniformly bounded in both row and column sums. Let* C_n *be a sequence of conformable matrices whose elements are uniformly* $O(h_n^{-1})$. *Then*

(i) the sequence $\{A_n B_n\}$ *are uniformly bounded in both row and column sums,*
(ii) the elements of A_n *are uniformly bounded and tr*$(A_n) = O(n)$, *and*
(iii) the elements of $A_n C_n$ *and* $C_n A_n$ *are uniformly* $O(h_n^{-1})$.

Lemma A.4 (Kelejian and Prucha, 2001, p.227, extended): *Let* $\{A_n\}$ *be an* $n\times n$ *matrix of elements* $\{a_{n,ij}\}$, b_n *be an* $n \times 1$ *vector of elements* $\{b_{n,i}\}$, *and* u_n *be an* $n \times 1$ *random vector of iid elements, having mean zero, variance* σ^2, *skewness* γ, *and excess kurtosis* κ. *Let* $Q_n = u_n' A_n u_n + b_n' u_n$. *Let* $a_n = \mathrm{diagv}(A_n)$, *the column vector formed by* $\{a_{n,ii}\}$. *Then,*

(i) $E(Q_n) = \sigma^2 \mathrm{tr}(A_n)$,
(ii) $Var(Q_n) = \sigma^4 \mathrm{tr}(A_n A_n' + A_n^2) + \sigma^4 \kappa a_n' a_n + \sigma^2 b_n' b_n + 2\sigma^3 \gamma a_n' b_n$.

Furthermore, if $\{a_{n,ij}\}$ *are of uniform order* $O_p(h_n^{-1})$, $\{b_{n,i}\}$ *are of uniform order* $O_p(h_n^{-\frac{1}{2}})$, *and* $\{A_n\}$ *are uniformly bounded in either row or column sums, then*

(iii) $E(Q_n) = O(\frac{n}{h_n})$, *and*
(iv) $Var(Q_n) = O(\frac{n}{h_n})$.

Subsequently, if h_n *is bounded, then* $E(Q_n) = O(n)$ *and* $Var(Q_n) = O(n)$.

Proof of Theorem 1: For (i), the derivation of $\text{LM}_R(\lambda)$ is already given before the appearance of Theorem 1. Now, the numerator of $\text{LM}_R(\lambda)$ can be written as

$$\hat{u}_n'(\lambda)D_n(\lambda)A_n(\lambda)Y_n = u_n'M_nD_n(\lambda)u_n + u_n'M_n\eta_n(\lambda),$$

which is a linear-quadratic form in u_n of iid elements. Recall $G_n(\lambda) = W_nA_n^{-1}(\lambda)$, $A_n(\lambda) = I_n - \lambda W_n$, and $D_n(\lambda) = G_n(\lambda) - \frac{1}{n}\text{tr}(M_nG_n(\lambda))I_n$. Under Assumption 2, Lemma A.1 shows that M_n is uniformly bounded in both row and column sums. Under Assumptions 3 and 4, Lemma A.3 shows that $G_n(\lambda)$ is uniformly bounded in both row and column sums, and that the elements $G_n(\lambda)$ are uniformly $O(h_n^{-1})$. Lemma A.2 (i) shows that $\frac{1}{n}\text{tr}(M_nG_n(\lambda)) = O(h_n^{-1})$. It follows that $D_n(\lambda)$, and hence $M_nD_n(\lambda)$, are uniformly bounded in both row and column sums and that the elements of $D_n(\lambda)$, and hence the elements of $M_nD_n(\lambda)$, are uniformly $O(h_n^{-1})$. Thus, the central limit theorem for the linear-quadratic form of Lee (2004a) is applicable to $u_n'M_nD_n(\lambda)u_n + u_n'M_n\eta_n(\lambda)$, which shows that

$$\frac{\hat{u}_n'(\lambda)D_n(\lambda)A_n(\lambda)Y_n}{\sigma\sqrt{\sigma^2T_{2n}(\lambda) + \eta_n'(\lambda)M_n\eta_n(\lambda) + \sigma^2\kappa d_n'(\lambda)d_n(\lambda) + 2\sigma\gamma\eta_n'(\lambda)M_nd_n(\lambda)}} \xrightarrow{D} N(0,1).$$

Replacing $\sigma^2, \eta_n(\lambda), \gamma$, and κ by their consistent estimators defined in the theorem, the result (i) thus follows by Slutsky's theorem.

For (ii), it suffices to show that
(a) $\eta_n'M_n\eta_n = O(n/h_n)$,
(b) $T_{2n}(\lambda) = O(n/h_n)$,
(c) $d_n'(\lambda)d_n(\lambda) = O(n/h_n^2)$,
(d) $\eta_n'(\lambda)M_nd_n(\lambda) = O(n/h_n^{3/2})$, and
(e) $T_{1n}(\lambda) \sim T_{2n}(\lambda)$,

which are all quite straightforward. These results allow us to conclude that when h_n is bounded, the denominator of $\text{LM}_R(\lambda)$ differs from that of $\text{LM}_E(\lambda)$ essentially by a term $\kappa d_n'(\lambda)d_n(\lambda) + 2\sigma\gamma\eta_n'(\lambda)M_nd_n(\lambda)$, which can be of the same order as the leading terms in the denominator. Thus, asymptotically, $\text{LM}_E(\lambda)$ does not converge to $N(0,1)$ in distribution. It is well known that $\text{LM}_H(\lambda)$ is asymptotically equivalent to $\text{LM}_E(\lambda)$ and thus it does not converge to $N(0,1)$ in distribution either. When h_n is divergent, the difference term is of a smaller order, and thus the three statistics are asymptotically equivalent.

For (iii), we note that when $\lambda = 0$, $A_n(\lambda) = I_n$, $G_n(\lambda) = W_n$, and $D_n(\lambda) = W_n - \frac{1}{n-k}\text{tr}(M_nW_n)I_n$. It follows from Lemma A.2 (i) that $\frac{1}{n-k}\text{tr}(M_nW_n) = O(n^{-1})$. Thus, from Lemma A.2 (vi), we have $d_n'(\lambda)d_n(\lambda) = O(h_n^{-1})$. By Cauchy-Schwarz inequality, one sees that $\eta_n'(\lambda)M_nd_n(\lambda) \leq [d_n'(\lambda)d_n(\lambda)]^{\frac{1}{2}}[\eta_n'M_n\eta_n]^{\frac{1}{2}} = O(n^{\frac{1}{2}}/h_n)$. Thus, the term $\kappa d_n'(\lambda)d_n(\lambda) + 2\sigma\gamma\eta_n'(\lambda)M_nd_n(\lambda)$ is always of smaller order than $\sigma^2T_{2n}(\lambda) + \eta_n'(\lambda)M_n\eta_n(\lambda)$. Hence, the three statistics are asymptotically equivalent whether h_n is bounded or unbounded.

References

[1] Anselin, L. (1988a). Lagrange multiplier test diagnostics for spatial dependence and heterogeneity. *Geographical Analysis* 20, 1-17.

[2] Anselin, L. (1988b). *Spatial Econometrics: Methods and Models.* Kluwar Academic Publishing.

[3] Anselin, L., and S. Rey (1991). Properties of tests for spatial dependence in linear regression models. *Geographical Analysis* 23, 112-131.

[4] Anselin, L., A. K. Bera, R. Florax, and M. J. Yoon (1996). Simple diagnostic tests for spatial dependence. *Regional Science and Urban Economics* 26, 77-104.

[5] Anselin, L. (2001). Rao's score test in spatial econometrics. *Journal of Statistical Planning and Inferences* 97, 113-139.

[6] Baltagi, B. H. (2001). *Econometric Analysis of Panel Data, 2nd ed.* John Wiley & Sons, Ltd.

[7] Baltagi, B. H., and Z. L. Yang (2013a). Standardized LM tests for spatial error dependence in linear or panel regressions. *Econometrics Journal* 16, 103-134.

[8] Baltagi, B. H., and Z. L. Yang (2013b). Heteroskedasticity and non-normality robust LM tests of spatial dependence. *Regional Science and Urban Economics* 43, 725-739.

[9] Bao, Y., and A. Ullah (2007). Finite sample properties of maximum likelihood estimator in spatial models. *Journal of Econometrics* 137, 396-413.

[10] Bell, K. P., and N. E. Bockstael (2000). Applying the generalized-moment estimation approach to spatial problems involving microlevel data. *Review of Economics and Statistics* 82, 72-82.

[11] Besley, T., and A. C. Case (1995). Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition. *The American Economic Review* 85, 25-45.

[12] Bertrand, M., E. F. P. Luttmer, and S. Mullaninathan (2000). Network effects and welfare cultures. *Quarterly Journal of Economics* 115, 1019-1055. spatial problems involving microlevel data. *Review of Economics and Statistics* **82**, 72-82.

[13] Bivand, R. S., E. Pebesma, and V. Gómez-Rubio (2013). *Applied Spatial Data Analysis with R, 2nd ed.* Springer.

[14] Born, B., and J. Breitung (2011). Simple regression based tests for spatial dependence. *Econometrics Journal* 14, 330-342.

[15] Box, G. E. P., and D. R. Cox (1964). An analysis of transformation (with discussion). *Journal of the Royal Statistical Society, Ser B* 26, 211-252.

[16] Brueckner, J. K. (1998). Testing for Strategic Interaction Among Local Governments: The Case of Growth Controls. *Journal of Urban Economics* 44, 438-467.

[17] Case, A. C. (1991). Spatial Patterns in Household Demand. *Econometrica* 59, 953-965.

[18] Case, A. C., H. S. Rosen, and J. R. Hines (1993). Budget Spillovers and Fiscal Policy Interdependence: Evidence from the States. *Journal of Public Economics* 52, 285-307.

[19] Cliff, A. D., and J. K. Ord (1973). *Spatial Autocorrelation.* London: Pion.

[20] Cliff, A. D., and J. K. Ord (1981). *Spatial Process, Models and Applications.* London: Pion.

[21] Fang, Y., S. Y. Park, and J. F. Zhang (2014). A simple spatial dependence test robust to local and distributional misspecifications. *Economics Letters* 124, 203-206.

[22] Kelejian, H. H., and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* 40, 509-533.

[23] Kelejian H. H., and I. R. Prucha (2001). On the asymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* 104, 219-257.

[24] Koenker, R. (1981). A note on studentising a test for heteroscedasticity. *Journal of Econometrics* 17, 107-112.

[25] Lee, L. F. (2002). Consistency and efficiency of least squares estimation for mixed regressive spatial autoregressive models. *Econometric Theory* 18, 252-277.

[26] Lee, L. F. (2003). Best spatial two-stage least squares estimators for a spatial autoregressive models with autoregressive disturbances. *Econometric Reviews* 22, 307-335.

[27] Lee, L. F. (2004a). Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica* 72, 1899-1925.

[28] Lee, L. F. (2004b). A supplement to 'Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models'. Working paper, Department of Economics, Ohio State University.

[29] Lee, L. F. (2007a). GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. *Journal of Econometrics* 137, 489-514.

[30] Lee, L. F. (2007b). Identification and estimation of econometric models with group interaction, contextual factors and fixed effects. *Journal of Econometrics* 140, 333-374.

[31] LeSage, J. P., and R. K. Pace (2009). *Introduction to Spatial Econometrics.* CRC Press Taylor and Francis Group, London.

[32] Moulton, B. R. and W. C. Randolph (1989). Alternative tests of the error components model. *Econometrica* 57, 685-693.

[33] Ord, J. (1975). Estimation methods for models of spatial interaction. *Journal of the American Statistical Association* 70, 120-126.

[34] Pace, R. K., and J. P. LeSage (2002). Semiparametric maximum likelihood estimates of spatial dependence. *Geographical Analysis* 34, 76-90.

[35] Robinson, P. M. (2008). Correlation testing in time series, spatial and cross-sectional data. *Journal of Econometrics* 147, 5-16.

[36] Robinson, P. M. (2010). Efficient estimation of the semiparametric spatial autoregressive model. *Journal of Econometrics* 157, 6-17.

[37] Smirnov, O., and L. Anselin. (2001). Fast maximum likelihood estimation of very large spatial autoregressive models: a characteristic polynomial approach. *Computational Statistics and Data Analysis* 35, 301-319.

[38] Topa, G. (2001). Social interaction, local spillovers and unemployment. *Review of Economic Studies* 68, 261-295.

[39] Wang, S. F., S. C. Wang, and P. Smith (2013). Two spatial non-nested tests for weight structure in the spatial autoregressive model. *Geographical Analysis* 45, 345-358.

[40] Yang, Z. L. (2010a). A robust LM test for spatial error components. *Regional Science and Urban Economics* 40, 299-310.

[41] Yang, Z. L. (2010b). Bias-Corrected Estimation for Spatial Autocorrelation. *Working Paper, Singapore Management University.*

Table 1. Tests of Spatial Dependence Based on Cigarettes Sales Data

		Original			Log Transformed		
Year	λ	$\text{LM}_E(\lambda)$	$\text{LM}_H(\lambda)$	$\text{LM}_R(\lambda)$	$\text{LM}_E(\lambda)$	$\text{LM}_H(\lambda)$	$\text{LM}_R(\lambda)$
1970	0.75	-3.2923	-4.9678	-3.3882	-3.1523	-4.6773	-3.2230
	0.50	-3.4321	-4.0558	-3.4237	-3.2126	-3.8432	-3.1717
	0.25	-2.1948	-1.9151	-2.0025	-2.0657	-1.8950	-1.8339
	0	0.2004	0.1510	0.6071	0.0449	0.0359	0.4956
	-0.25	2.8019	2.2509	3.4107	2.3660	1.9803	3.0048
	-0.50	4.5944	4.6845	5.3270	4.0725	4.1505	4.8117
	-0.75	5.2592	7.1883	5.9724	4.8213	6.3388	5.5360
1980	0.75	-2.7093	-3.7047	-2.7680	-2.7235	-3.7691	-2.7809
	0.50	-2.4012	-2.6371	-2.3406	-2.5735	-2.9843	-2.5106
	0.25	-1.0990	-0.9940	-0.8367	-1.5538	-1.4966	-1.2951
	0	0.7884	0.6638	1.2729	0.0649	0.0566	0.5419
	-0.25	2.6420	2.3691	3.2985	1.8253	1.6186	2.4795
	-0.50	3.9563	4.1715	4.6799	3.2487	3.2368	3.9901
	-0.75	4.5396	5.7516	5.1976	4.0467	4.7545	4.7587
1990	0.75	-1.8229	-2.2717	-1.6732	-2.1401	-3.0326	-1.9965
	0.50	-0.8020	-0.8688	-0.3895	-1.4281	-1.6781	-1.1210
	0.25	0.6563	0.6735	1.2831	-0.0355	-0.0370	0.4464
	0	2.0887	2.2325	2.8523	1.5592	1.6209	2.1839
	-0.25	3.2107	3.8154	4.0292	2.9266	3.3646	3.6401
	-0.50	3.9094	5.2455	4.7114	3.8221	5.1242	4.5599
	-0.75	4.1720	6.0593	4.8954	4.1828	6.3617	4.8760

Table 2. 95% CIs for λ Based on Cigarettes Sales Data

Year	$\text{LM}_E(\lambda)$	$\text{LM}_H(\lambda)$	$\text{LM}_R(\lambda)$
1970	(-0.1642, 0.2205)	(-0.2170, 0.2552)	(-0.1159, 0.2450)
	(-0.2034, 0.2348)	(-0.2475, 0.2582)	(-0.1417, 0.2667)
1980	(-0.1522, 0.3953)	(-0.1914, 0.3949)	(-0.0796, 0.4200)
	(-0.2705, 0.3295)	(-0.3035, 0.3247)	(-0.1800, 0.3658)
1990	(0.0243, *)	(0.0433, 0.6864)	(0.1475, *)
	(-0.0666, 0.6473)	(-0.0499, 0.5442)	(0.0334, 0.7273)

Note: * means that rational solution is unavailable.
Two rows in each year, models based on original and logged data.

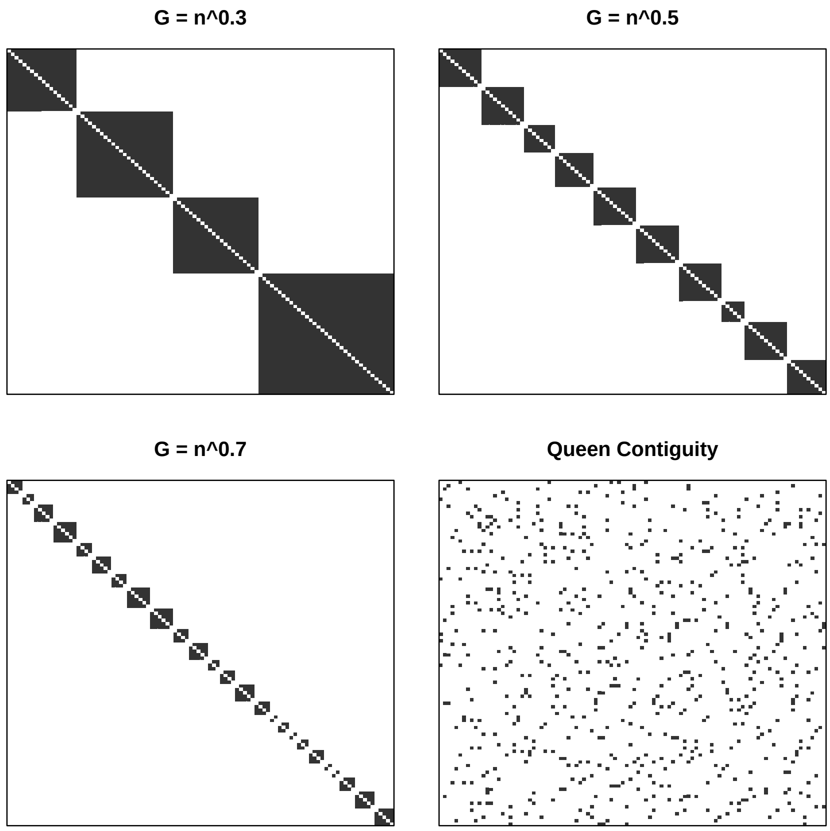

Figure 1: A Graphical Representation of the Spatial Weight Matrices, $n = 100$.

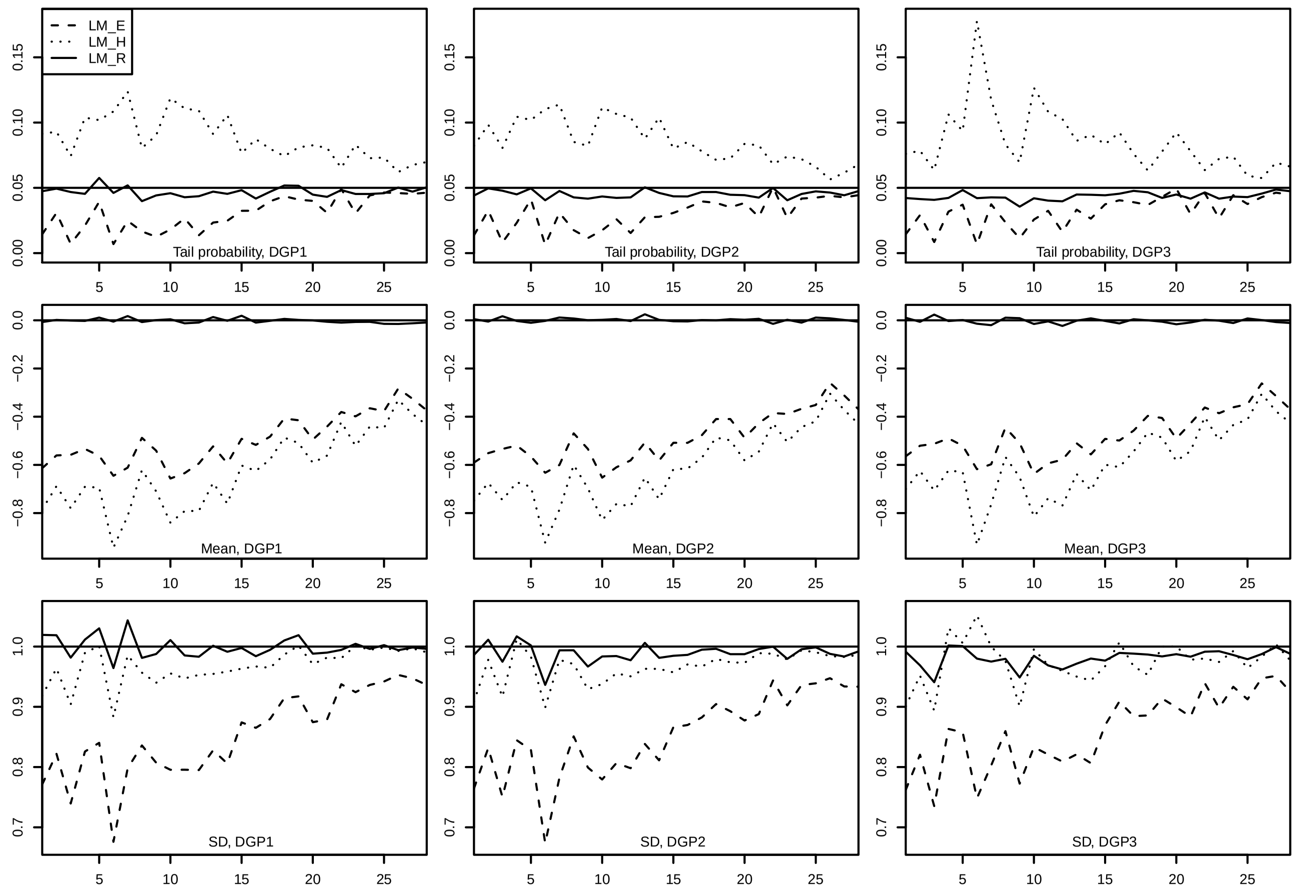

Figure 2: Empirical Tail Probabilities, Means and SDs of LM_E, LM_H and LM_R: $G = n^{0.3}$.

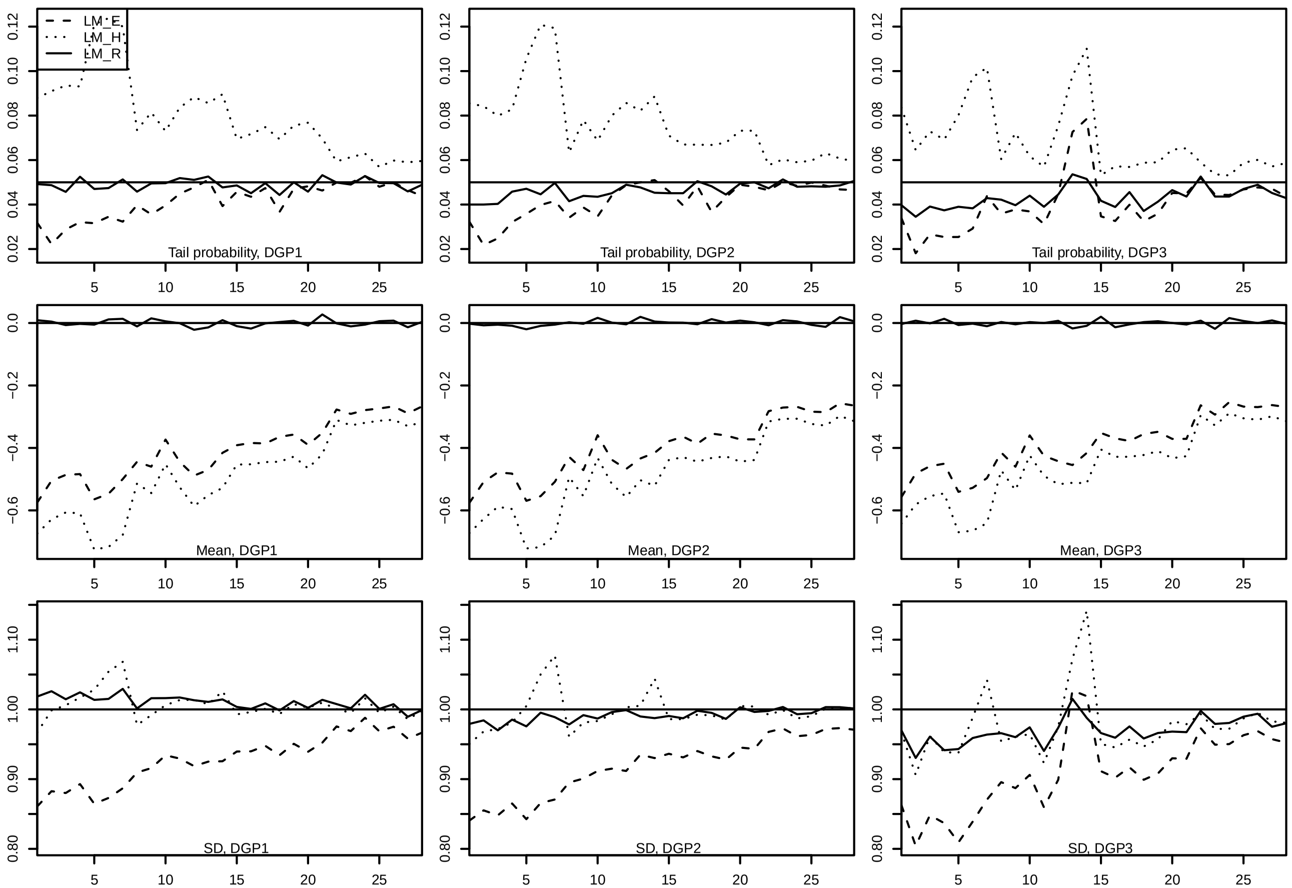

Figure 3: Empirical Tail Probabilities, Means and SDs of LM_E, LM_H and LM_R: $G = n^{0.5}$.

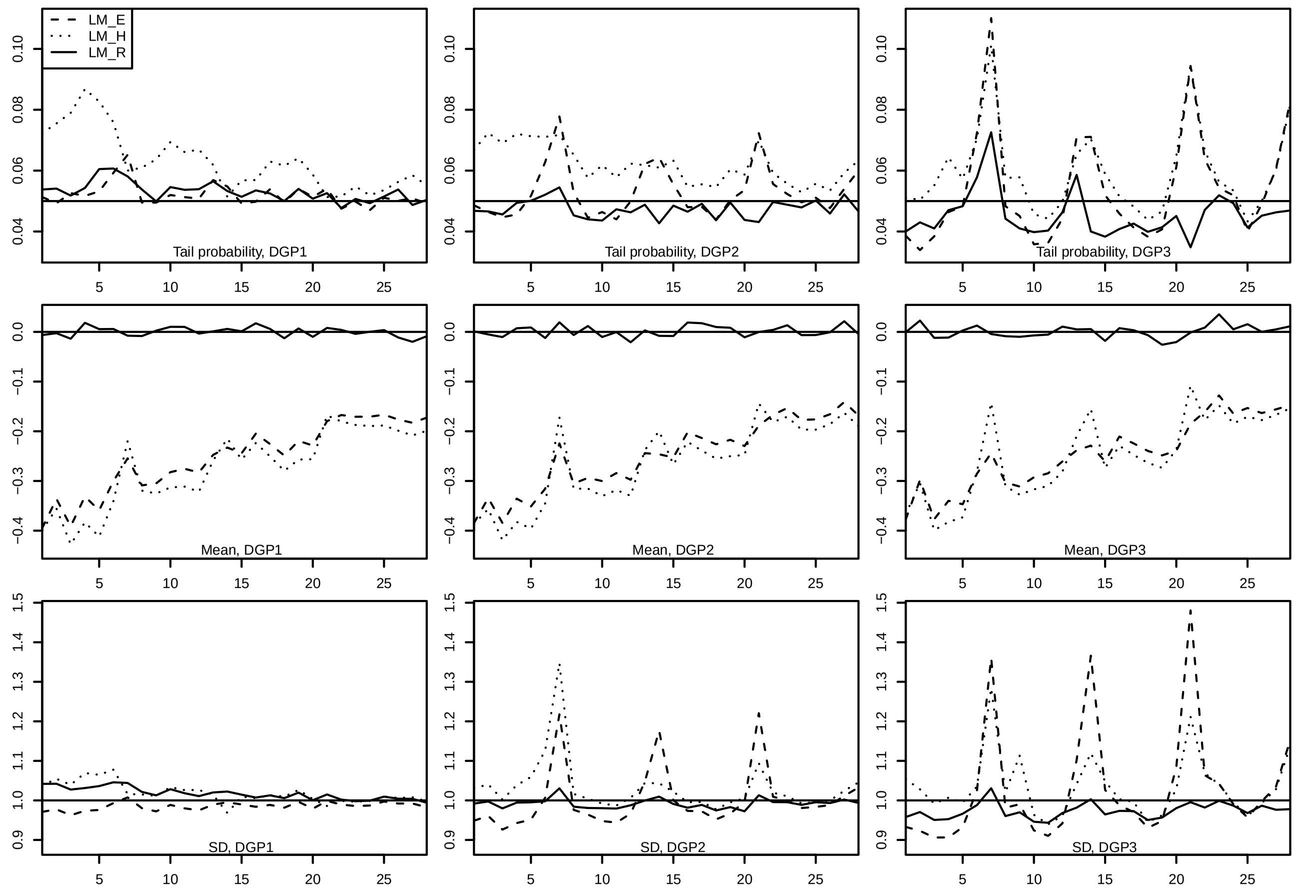

Figure 4: Empirical Tail Probabilities, Means and SDs of LM_E, LM_H and LM_R: $G = n^{0.7}$.

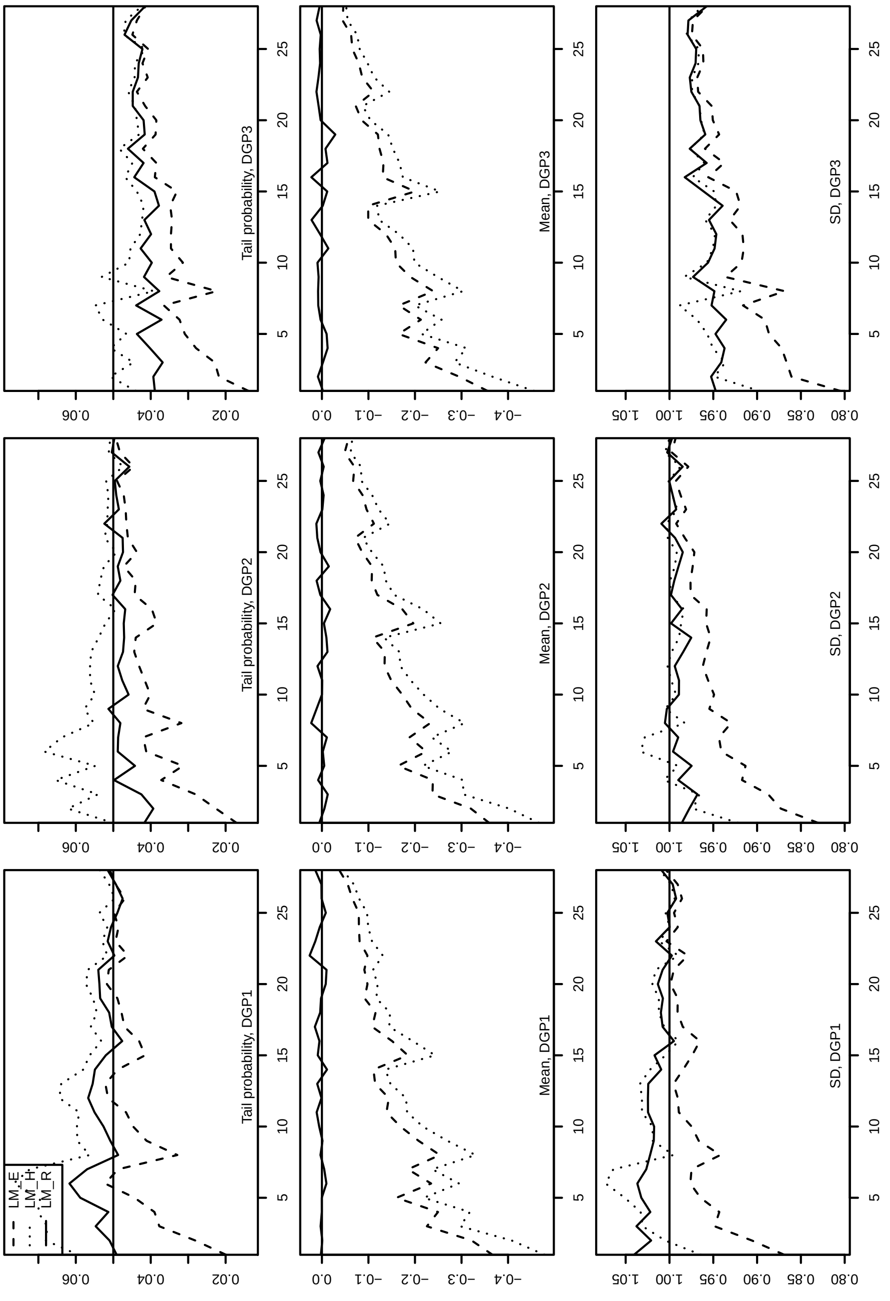

Figure 5: Empirical Tail Probabilities, Means and SDs of LM_E, LM_H and LM_R: Queen.

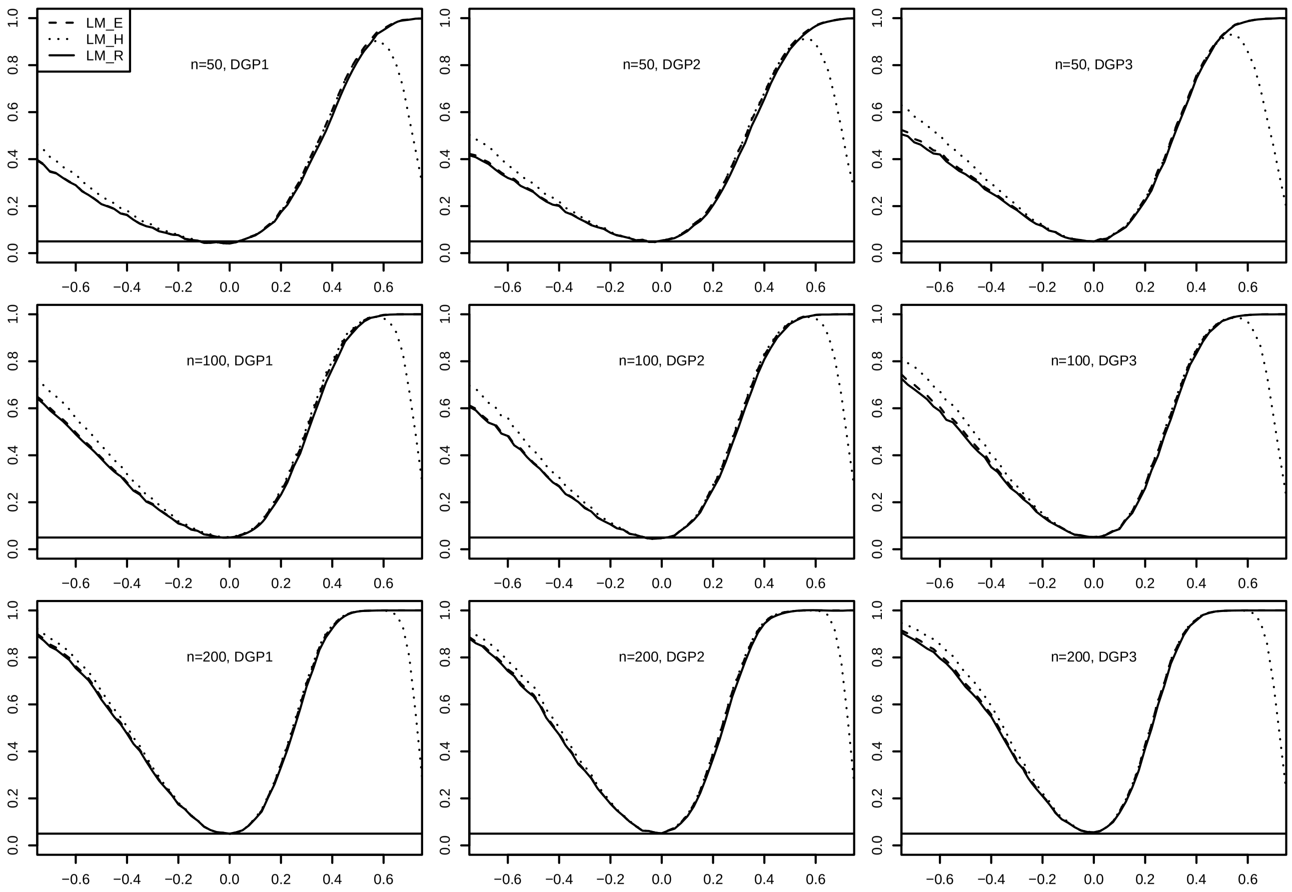

Figure 6: Size-Adjusted Empirical Power of LM_E, LM_H and LM_R: $G = N^{0.5}$.

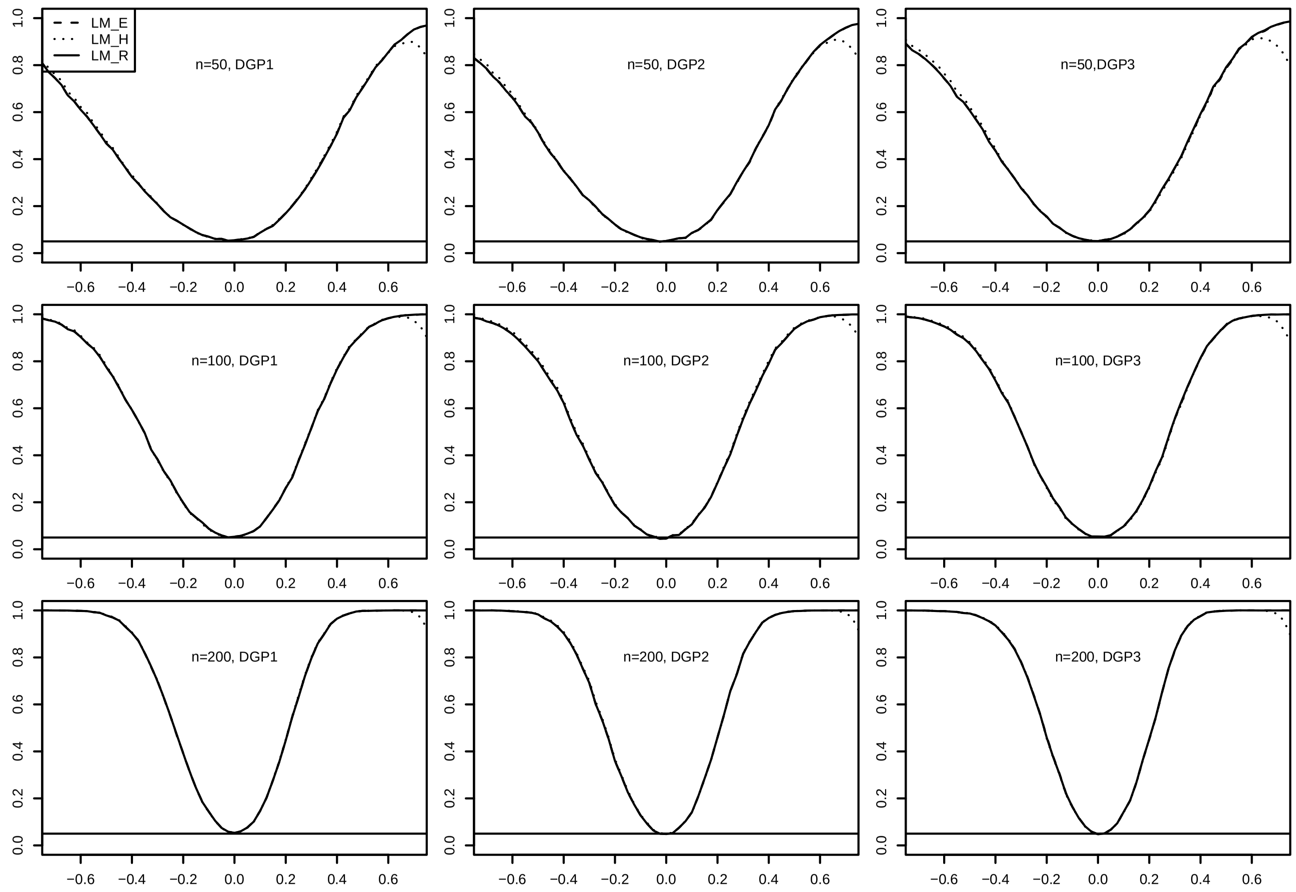

Figure 7: Size-Adjusted Empirical Power of LM_E, LM_H and LM_A: Queen.