Regional Science and Urban Economics 43 (2013) 250–271

Contents lists available at SciVerse ScienceDirect

Regional Science and Urban Economics

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

Model selection using J-test for the spatial autoregressive model vs. the matrix exponential spatial model ☆

Xiaoyi Han *, Lung-fei Lee

Department of Economics, Ohio State University, Columbus, OH 43210, USA

ARTICLE INFO

Article history:
Received 19 May 2012
Received in revised form 6 July 2012
Accepted 12 July 2012
Available online 21 July 2012

JEL classification:
C12
C21

Keywords:
Spatial autoregressive model
Matrix exponential spatial model
J-test
Pseudo true value
GMM

ABSTRACT

We consider using the J-test procedure for the non-nested model selection problem between the spatial autoregressive (SAR) model and the matrix exponential spatial specification (MESS) model. The 2SLS and GMM methods are used to implement the J-test procedure and derive several test statistics under the GMM framework. We investigate the behavior of those J-test statistics in terms of pseudo true values. We extend the J-test procedure into the setting when error terms in the model are with unknown heteroskedasticity. Monte Carlo results suggest with strong spatial dependence the J-test statistics can have good power to distinguish the SAR and MESS models.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Spatial econometric models applied in regional science and geography have been receiving more attention in various areas of economics. The most popular spatial econometric model is the spatial autoregressive (SAR) model. The SAR model implies a geometrical decline pattern of spillover effects or externalities from levels of neighbors in its reduced form.[1] There are other models which display different patterns of spillover effects or spatial externalities. Recently, LeSage and Pace (2007) introduce the matrix exponential spatial specification (MESS) model, which exhibits an exponential decline pattern of spatial externalities. The MESS model can produce estimates and inferences similar to those from the SAR model and it is computationally simpler. LeSage and Pace regard it as a substitute for the SAR model. However, with different features in their reduced forms, the two models cannot be perfect substitutes for each other. In practice, there is usually no formal theoretical guidance for which pattern of spatial externalities we should select to use. We are facing a non-nested model selection or testing problem among competitive models. Hence, it is of interest to construct a model discrimination procedure for them.[2]

For model selection among non-nested models, both classical approach and Bayesian approach are available in the literature. Bayesian model comparison procedure involves calculating and comparing the posterior probabilities of competitive models (Zellner, 1971) and is feasible for competitive non-nested models.[3] LeSage and Pace (2007) derive expressions for the log marginal likelihood of the MESS model, which could be used to produce Bayesian model comparison procedures for the SAR model and the MESS model.[4] For the classical approach, the J-test is a well-known test procedure for

☆ We are grateful to the editor and an anonymous referee for their helpful comments.
* Corresponding author. Tel.: +1 614 607 1414; fax: +1 614 292 3906.
E-mail addresses: han.293@osu.edu (X. Han), lee.1777@osu.edu (L. Lee).
[1] More discussions about spatial externalities and spatial econometric models can be found at Anselin (2003).

[2] As pointed out by a referee, in the face of no spatial dependence, both the SAR and MESS models collapse to independent linear models. Therefore it would be difficult to distinguish between spatial model specifications in the absence of dependence for any model comparison procedure. This is also confirmed by the Monte Carlo results for the classical J-test in this paper. With strong spatial dependence, the J-test statistics can have good power to distinguish between the two models. However, the powers of the test statistics decrease when we only have moderate spatial dependence.

[3] For more discussions regarding Bayesian model comparison procedures for spatial models, see, for example, Hepple (1995a,b), LeSage and Parent (2007) and LeSage and Pace (2009).

[4] We thank a referee for pointing out this.

0166-0462/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.regsciurbeco.2012.07.005

testing of non-nested models in a non-spatial content.[5] Davidson and MacKinnon (1981) propose a J-test procedure based on the comprehensive approach advocated by Atkinson (1970) for model selection among non-nested univariate linear and non-linear regression models. They also consider a linearized version of the J-test (the so-called P-test) for non-linear models if the computations are difficult. Since then, various extensions of the J-test, discussions of their finite sample properties and the corresponding bootstrap tests have appeared in the literature.[6] Furthermore, the J-test and its extensions can be derived as linear approximations to the Cox test statistic[7] (Pesaran and Weeks, 2001). Compared with other non-nested tests, it is both conceptually and computationally simpler (Davidson and MacKinnon, 1982). Therefore, it is relatively easy to implement in practice. Recently, Kelejian (2008) extends the J-test procedure into the spatial setting. His concern is to test competitive SAR models with different spatial weight matrices. The J-test in Kelejian (2008) is based on a Wald test statistic constructed from the 2SLS estimation of an augmented model. Kelejian and Piras (2011) modify the J-test in Kelejian (2008) by using available information in a more efficient way. Burridge (2012) improves the J-test in Kelejian (2008) by using the quasi-maximum likelihood estimation. Liu et al. (2011) extend the J-test in Kelejian (2008) to differentiate between two different social network models. Piras and Lozano-Gracia (2012), Burridge (2012), and Liu et al. (2011) evaluate the finite sample performance of their J-tests in Monte Carlo studies. Burridge and Fingleton (2010), and Burridge (2012) also conduct bootstrap J-tests to investigate finite sample properties of their J-test statistics.

In this paper we consider a J-test procedure in model selection between the SAR model and the competing MESS model. Our work is distinct from these studies in several ways. Firstly, our focus is to select an appropriate pattern of spatial externalities, rather than select a spatial weight matrix. Secondly, we consider the GMM method in Lee (2007) in addition to the 2SLS method in Kelejian and Prucha (1998) to estimate the augmented model and to set up test statistics. Thirdly, we construct the gradient (G) test statistic and the distance difference (DD) test statistic developed by Newey and West (1987), in addition to the Wald test statistic. Finally, we extend the spatial J-test procedure into the setting when error terms in the model are independent but with unknown heteroskedasticity. We provide rigorous statistical analysis for our test statistics in terms of pseudo true values of misspecified models under each of the null hypotheses.

The paper is organized as follows: Section 2 specifies the SAR and MESS models and considers the corresponding model selection problem. Section 3 discusses J-test procedures. We consider both the 2SLS and GMM estimation of the augmented model. Test statistics are constructed and their asymptotic distributions are analyzed. Section 4 extends J-test procedures into the setting when error terms in the model are independent but with unknown heteroskedasticity. Section 5 summarizes Monte Carlo results to illustrate some finite sample properties of the J-test statistics. Conclusions are drawn in Section 6. Technical details and tables are given in the Appendix.

[5] The J-test is not the only non-nested test that has been proposed in the literature. Several tests have been proposed based on Cox's two classic papers (Cox, 1961, 1962). There are also other non-nested tests based on the encompassing approach developed by Deaton (1982), Dastoor (1983) and Mizon and Richard (1986). For more discussions, see Pesaran and Weeks (2001).

[6] See, for example, Fisher and McAleer (1981), Davidson and MacKinnon (1982, 1983), Godfrey (1983, 1998), Davidson and MacKinnon (2002a, 2002b), Pesaran and Weeks (2001), Gourieroux and Monfort (1994) and the review in Davidson and MacKinnon (2004), pp. 665–675.

[7] The Cox test is based upon the pioneering work of Cox (1961, 1962). Cox extends the idea of a likelihood ratio test for non-nested models. For more discussions, see, for example, Pesaran (1974), Godfrey and Pesaran (1983), Pesaran and Weeks (2001) and Pesaran and Dupleich Ulloa (2008).

2. The models

The spatial autoregressive (SAR) model under consideration is

$$Y_n = \lambda W_n Y_n + X_n \beta + V_n, \tag{2.1}$$

where X_n is a $n \times k$ dimensional matrix of nonstochastic exogenous variables including the intercept term. W_n is a spatial weight matrix with a zero diagonal consisting of known constants. We impose the following basic assumptions about the SAR model:

Assumption 2.1. The v_{ni}'s in $V_n = (v_{n1}, v_{n2}, \ldots, v_{nn})'$ are i.i.d with zero mean, variance σ^2 and that a moment of order higher than the fourth exists.

Assumption 2.2. The elements of X_n are uniformly bounded constants. X_n has the full rank k and $\lim_{n\to\infty}\frac{1}{n}X_n'X_n$ exists and is nonsingular.

Assumption 2.3. The spatial weights matrices $\{W_n\}$ are uniformly bounded in both row and column sums in absolute value.

Assumption 2.4. The matrix $I_n - \lambda W_n$ is nonsingular for all λ in a compact parameter space Λ. In addition, $(I_n - \lambda W_n)^{-1}$ is uniformly bounded in both row and column sums in absolute value for all λ uniformly in Λ.

Assumption 2.1–2.3 are conventional regularity conditions for the SAR model. In particular, it will ensure finite variances for quadratic forms of V_n used in the GMM estimation. The higher than fourth moment condition is needed in order to apply a central limit theorem in Kelejian and Prucha (2001). The strong Assumption 2.4 is needed for the SAR model to ensure in particular that the variances of Y_n's remain bounded for large n. Furthermore, under Assumption 2.4, the reduced form of the SAR model reviews its implication in spatial externalities:

$$Y_n = X_n\beta + \sum_{m=1}^{\infty} W_n^m X_n \lambda^m \beta + (I_n - \lambda W_n)^{-1} V_n. \tag{2.2}$$

Here the nonzero elements of rows of W_n^m with $m \geq 1$ represent mth order contiguous neighbors.[8] Then the specification (2.2) has spillover effects or externalities generated by the regressors x's from one's different level of neighbors being geometrically declining.

As an alternative to the SAR specification, LeSage and Pace (2007) introduce the MESS model with the specification $S_n^{ex}(\mu)Y_n = X_n\beta^{ex} + V_n$, of which the reduced form is

$$Y_n = S_n^{ex}(\mu)^{-1}X_n\beta^{ex} + S_n^{ex}(\mu)^{-1}V_n, \tag{2.3}$$

where $S_n^{ex}(\mu) = e^{\mu W_n} = I_n + \sum_{t=1}^{\infty}\frac{1}{t!}(\mu W_n)^t$.[9] The model introduces an exponential decay pattern of spatial externalities. As emphasized by LeSage and Pace (2007), this model has computational advantage when it comes to estimation. With a zero diagonal W_n, the determinant of $e^{\mu W_n}$ is one, so the likelihood function of the MESS model is relatively simpler than that of the SAR model where the determinant of $(I_n - \lambda W_n)$ depends on λ.

[8] See page 14 of LeSage and Pace (2009).

[9] In the MESS model we set W_n to be a conventional spatial weight matrix consisting of known constants. As pointed out by a referee, LeSage and Pace (2009) has considered an extension of the MESS model, in which $W_n = \sum_{i=1}^{p}\left(\frac{\phi^i N_i}{\sum_{i=1}^{p}\phi^i}\right)$. Here p is the (unknown) number of nearest neighbors and $0<\phi<1$ represents an unknown decay factor applied to each of the nearest neighbor weight matrices N_i. In this paper we focus on the setting where W_n is a conventional spatial weight matrix for both the SAR model and the MESS model.

3. The J-test procedure

The basic idea of a J-test is to check whether predictors from the alternative model can add significantly to the explanatory power in the null model. Kelejian (2008) and Kelejian and Piras (2011) extend the J-test framework into the spatial setting. The focus of their J-test is to compare different specifications of the spatial weight matrix W_n in a SAR model. Here, the J-test procedure is to compare the SAR model vs the MESS model. Since a non-nested test works interchangeably between models, we conduct two groups of J-tests, where one has the null model being the SAR model and the other has the MESS model as the null.

3.1. The J-test using the SAR model as the null

The specified null model and the alternative model are:

$$\begin{aligned} &H_0: Y_n = \lambda W_n Y_n + X_n\beta + V_n, \\ &H_1: S_n^{ex}(\mu) Y_n = X_n \beta^{ex} + V_n. \end{aligned} \tag{3.1}$$

Let $\theta^{sar} = (\lambda, \beta', \sigma^2)'$ be the parameter vector of the SAR model. Similarly, let $\theta^{ex} = (\mu, \beta^{ex\prime}, \sigma^{ex2})'$ be the parameter vector of the MESS model. For a J-test procedure, we need to obtain predictors from the alternative model. Here we use the quasi-maximum likelihood (QML) method to estimate the MESS model. Let $\hat{\theta}_n^{ex} = (\hat{\mu}_n, \hat{\beta}_n^{ex\prime}, \hat{\sigma}_n^{ex2})'$ be the QMLE of the MESS model. According to Eq. (3.1), a predictor of Y_n can be from the reduced form of the MESS model, which is $\hat{Y}_{n|1} = S_n^{ex}(\hat{\mu}_n)^{-1} X_n \hat{\beta}_n^{ex}$. Alternatively, if we denote $U_n(\mu) = I_n - S_n^{ex}(\mu)$, then we can construct a predictor from the structural form of the MESS model: $Y_n = U_n(\mu) Y_n + X_n \beta^{ex} + V_n$, as $\hat{Y}_{n|2} = U_n(\hat{\mu}_n) Y_n + X_n \hat{\beta}_n^{ex}$. These predictors are motivated by Kelejian and Piras (2011) for the SAR model with different spatial weight matrices.

With a predictor, the null SAR model can be augmented into the following equation:

$$Y_n = \lambda W_n Y_n + X_n \beta + \hat{Y}_{n|r_1} \delta_{r_1} + V_n, \tag{3.2}$$

where the index r_1 is either 1 or 2, which provides the basic equation for a J-test. Obviously, the augmented model is just the SAR model plus an additional regressor, which is one of the two predictors from the MESS model. For estimation, the ML method might not be feasible as the augmented equation would not have a simple likelihood function. This is so, in particular, when the predictor is $\hat{Y}_{n|2}$, which contains the dependent variable Y_n. So instead, we consider the 2SLS method suggested by Kelejian and Prucha (1998), or the GMM procedure with both linear and quadratic moments proposed by Lee (2007) for estimating Eq. (3.2). The 2SLS method is simpler from a computational point of view as it has a closed form solution. The J-test in Kelejian and Piras (2011) is based on the 2SLS method. However, the GMM method in Lee (2007) uses quadratic moments in addition to the linear moments used in 2SLS and is relatively more efficient than the 2SLS method. To analyze asymptotic properties of the J-test procedure, it would be helpful to have an idea on how a predictor from the alternative model (the MESS model) would behave under the null SAR model. As the MESS model is a misspecified one under the null SAR model, the estimated parameters $\hat{\theta}_n^{ex}$ in the predictors would not converge to structural parameters but might converge to some limiting values. The detailed analysis of their limiting values, or the so-called pseudo true values of $\hat{\theta}_n^{ex}$ based upon the QML method is in Appendix B.

Denote $\eta_{r_1} = (\lambda, \beta', \delta_{r_1})'$. Let $\eta_{0r_1} = (\lambda_0, \beta_0', 0)'$ be the true value of η_{r_1} for $r_1 = 1, 2$, under the null SAR model. We first impose the following assumption on η_{0r_1}.

Assumption 3.1. η_{0r_1} is in the interior of the parameter space $\mathcal{H}_{r_1}$, which is a bounded subset of R^{k+2}.[10]

Since the 2SLS method can be viewed as a special case of GMM, we begin with J-test procedure based upon the GMM method. Let $V_n(\eta_{r_1}) = (I_n - \lambda W_n) Y_n - X_n \beta - \hat{Y}_{n|r_1} \delta_{r_1}$. The GMM method is based on an instrumental variable (IV) matrix Q_n and the IV functions $P_{jn} V_n(\eta_{r_1})$ where P_{jn} is a $n \times n$ square (constant) matrix with $tr(P_{jn}) = 0$ for $j = 1, 2, \ldots, q$ for some finite q. The GMM method uses the moment function vector

$$g_n(\eta_{r_1}) = \left(P_{1n} V_n(\eta_{r_1}), \ldots, P_{qn} V_n(\eta_{r_1}), Q_n\right)' V_n(\eta_{r_1}),$$

where $Q'_n V_n(\eta_{r_1})$ is the linear moment function and $V_n'(\eta_{r_1}) P_{jn} V_n(\eta_{r_1})$'s are the quadratic moment functions.

Consider first the linear moment $Q'_n V_n(\eta_{r_1})$. As suggested by Kelejian and Prucha (1998), one could specify the IV matrix Q_n as $Q_n = (X_n, W_n X_n, \ldots, W_n^d X_n)_{LI}$ where LI refers to the linearly independent columns, i.e., Q_n consists of all linearly independent columns of X_n, $W_n X_n$, ⋯, $W_n^d X_n$. In the augmented model, we have one additional predictor from the MESS model. Obviously we do not need IVs for $\hat{Y}_{n|1}$ since $\hat{Y}_{n|1} = S_n^{ex}(\hat{\mu}_n)^{-1} X_n \hat{\beta}_n^{ex}$ is essentially exogenous. However, we might need more IVs in order to accommodate $\hat{Y}_{n|2}$ because $\hat{Y}_{n|2}$ involves endogenous variables. Let $S_n(\lambda) = I_n - \lambda W_n$, $S_n = S_n(\lambda_0)$ and $\hat{U}_n = U_n(\hat{\mu}_n)$. Note that under the null SAR model

$$\hat{Y}_{n|2} = \hat{U}_n Y_n + X_n \hat{\beta}_n^{ex} = \hat{U}_n S_n^{-1} X_n \beta_0 + \hat{U}_n S_n^{-1} V_n + X_n \hat{\beta}_n^{ex}.$$

[10] We don't need this assumption for the 2SLS method. For nonlinear extremum estimation methods, compactness on the parameter space is usually needed to demonstrate consistency of the estimates (Amemiya, 1985). However, for the GMM method here, η_{r_1} appears nonlinearly in the linear and quadratic moments in terms of polynomials. So the boundness of $\mathcal{H}_{r_1}$ will be sufficient.

So we might still use Q_n as the IV matrix for $\hat{Y}_{n|2}$ since $S_n^{-1}X_n$ are correlated with $\hat{Y}_{n|2}$ as long as Q_n contains enough IVs.[11] We add some relevant rank conditions for Q_n:

Assumption 3.2. Assume the elements of Q_n are uniformly bounded in absolute value. Furthermore, $\lim_{n\to\infty}\frac{1}{n}Q_n'Q_n$ have finite full column rank.

Next, consider the quadratic moment functions $V_n'(\eta_{r_1})P_{jn}V_n(\eta_{r_1})$'s. Following Lee (2007), let $\mathcal{P}_{1n}$ be the class of constant $n\times n$ matrices which have a zero trace.[12] We impose the following assumption on $\mathcal{P}_{1n}$:

Assumption 3.3. The matrices P_{jn}'s from $\mathcal{P}_{1n}$ are uniformly bounded in both row and column sums in absolute value.

Denote $\gamma=(\lambda,\beta')'$ and $\eta_{r_1}=(\lambda,\beta',\delta_{r_1})'=(\gamma',\delta_{r_1})'$. Let $\mu^*_{n|sar}$ be the sequence of pseudo true values of $\hat{\mu}_n$ for the MESS model under the null SAR model and $\beta^{ex*}_{n|sar}$ be the sequence of pseudo true values of $\hat{\beta}^{ex}_n$. By Lemmas A.5 and B.1

$$\begin{aligned}
&\text{p}\lim\hat{Y}_{n|1}=S^{ex*-1}_{n|sar}X_n\beta^{ex*}_{n|sar}\\
&\text{p}\lim\hat{Y}_{n|2}=U^*_{n|sar}S_n^{-1}X_n\beta_0+X_n\beta^{ex*}_{n|sar}\\
&\text{p}\lim\frac{1}{n}Q_n'\hat{Y}_{n|1}-\text{p}\lim\frac{1}{n}Q_n'\left(S^{ex*-1}_{n|sar}X_n\beta^{ex*}_{n|sar}\right)=o_p(1)\\
&\text{p}\lim\frac{1}{n}Q_n'\hat{Y}_{n|2}-\text{p}\lim\frac{1}{n}Q_n'\left(U^*_{n|sar}S_n^{-1}X_n\beta_0+X_n\beta^{ex*}_{n|sar}\right)=o_p(1),
\end{aligned}$$

where $S^{ex*}_{n|sar}=S^{ex}_n(\mu^*_{n|sar})$ and $U^*_{n|sar}=U_n(\mu^*_{n|sar})$. Let $Y^*_{n|1}=S^{ex*-1}_{n|sar}X_n\beta^{ex*}_{n|sar}$ and $Y^*_{n|2}=U^*_{n|sar}S_n^{-1}X_n\beta_0+X_n\beta^{ex*}_{n|sar}$. With $Y^*_{n|r_1}$, we can derive the expression of $E(g_n(\eta_{r_1}))$. Denote $G_n(\lambda)=W_n(I_n-\lambda W_n)^{-1}$ and $G_n=G_n(\lambda_0)$. For any possible value η_{r_1}

$$E(g_n(\eta_{r_1}))=E\begin{pmatrix}\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n\right]'P_{1n}\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n\right]\\ \vdots\\ \left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n\right]'P_{qn}\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n\right]\\ Q_n'\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n\right]\end{pmatrix},$$

where $h_n(\gamma)=X_n(\beta_0-\beta)+(\lambda_0-\lambda)G_nX_n\beta_0$. According to Hansen (1982), in the GMM framework, the identification condition for η_{r_1} requires the unique solution of the limiting equations, $\lim_{n\to\infty}\frac{1}{n}Eg_n(\eta_{r_1})=0$ at η_{0r_1}. For the linear moment function, by the dominated convergence theorem

$$\begin{aligned}
\lim_{n\to\infty}\frac{1}{n}EQ_n'\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n\right]&=\lim_{n\to\infty}\frac{1}{n}Q_n'\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})\right]\\
&=\lim_{n\to\infty}\frac{1}{n}Q_n'\left[X_n,Y^*_{n|r_1},G_nX_n\beta_0\right]\begin{pmatrix}\beta_0-\beta\\ \delta_0-\delta_{r_1}\\ \lambda_0-\lambda\end{pmatrix}.
\end{aligned}$$

Therefore, η_{r_1} is identified if $\lim_{n\to\infty}\frac{1}{n}Q_n'\left[X_n,Y^*_{n|r_1},G_nX_n\beta_0\right]$ has full rank $k+2$. This sufficient rank condition implies the necessary rank condition that $[X_n,Y^*_{n|r_1},G_nX_n\beta_0]$ has full column rank for a large enough value of n. However, there could be some situations in which this necessary rank condition would not hold (Lee, 2007). A possible example is $\beta_0=0$. Under this circumstance, $[X_n,Y^*_{n|r_1},G_nX_n\beta_0]$'s rank will be $k+1$ if we assume $[X_n,Y^*_{n|r_1}]$ has rank $k+1$. In this case, β_0 and δ_0 can be identified only if λ_0 is identified (Lee, 2007). As suggested by Lee (2007), we can identify λ_0 by the quadratic moment function.

Consider $E\{[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n]'P_{jn}[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n]\}$. The corresponding limiting equation is

$$\begin{aligned}
&\lim_{n\to\infty}\frac{1}{n}E\left\{\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n\right]'P_{jn}\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n\right]\right\}\\
&=\lim_{n\to\infty}\frac{1}{n}\left\{\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})\right]'P_{jn}\left[h_n(\gamma)+Y^*_{n|r_1}(\delta_0-\delta_{r_1})\right]+\sigma_0^2tr\left(S_n^{-1'}S_n(\lambda)'P_{jn}S_n(\lambda)S_n^{-1}\right)\right\}.
\end{aligned}$$

for $r_1=1, 2$. Note that the first component of the above limiting equation would drop out when $\beta_0=0$, but λ can be identified by $\sigma_0^2tr(S_n^{-1'}S_n(\lambda)'P_{jn}S_n(\lambda)S_n^{-1})=0$. Let A^S denote the sum $(A+A')$ for any square matrix A. We can impose identification assumptions similar to Lee (2007).

Assumption 3.4. Either (i) $\lim_{n\to\infty}\frac{1}{n}Q_n'\left[X_n,Y^*_{n|r_1},G_nX_n\beta_0\right]$ has full rank $k+2$ for $r_1=1, 2$ or (ii) $\lim_{n\to\infty}\frac{1}{n}Q_n'\left[X_n,Y^*_{n|r_1}\right]$ has full rank $k+1$ for $r_1=1, 2$, $\lim_{n\to\infty}\frac{1}{n}tr\left(P^S_{jn}G_n\right)\neq0$ for some j, and $\lim_{n\to\infty}\frac{1}{n}\left[tr\left(P^S_{1n}G_n\right),\ldots,tr\left(P^S_{qn}G_n\right)\right]$ is linearly independent of $\lim_{n\to\infty}\frac{1}{n}\left[tr\left(G_n'P_{1n}G_n\right),\ldots,tr\left(G_n'P_{qn}G_n\right)\right]$.

[11] $\hat{U}_nS_n^{-1}X_n$ can be expressed as linear combination of X_n, W_nX_n, $W_n^2X_n$,…, $W_n^dX_n$,…. One could also use $Q_n=\left(\hat{U}_nX_n,\hat{U}_nW_nX_n,\ldots,\hat{U}_nW_n^dX_n\right)$ as the IV matrix.

[12] We could also consider a subclass $\mathcal{P}_{2n}$ of $\mathcal{P}_{1n}$, which consists of matrices with zero diagonal. $\mathcal{P}_{2n}$ would be useful in the GMM estimation when the model has unknown heteroskedastic disturbances, as discussed in Lin and Lee (2010).

Under the null, the augmented model evaluated at true parameters is just the SAR model. As in Lee (2007), the variance matrix of the moment functions of the SAR model involves variances and covariances of linear and quadratic forms of V_n. Denote $\Omega_n = \text{var}(g_n(\eta_{0r_1}))$. Also, let $vec_D(A) = (a_{11}, \ldots, a_{nn})'$ denote the column vector formed with the diagonal elements of a square $n \times n$ matrix A. By Lemma A.3

$$\Omega_n = \begin{pmatrix} \left(\mu_4 - 3\sigma_0^4\right)\omega_{qn}'\omega_{qn} & \mu_3\omega_{qn}'Q_n \\ \mu_3 Q_n'\omega_{qn} & 0 \end{pmatrix} + B_n, \tag{3.3}$$

where $\mu_3 = E(v_{ni}^3)$ and $\mu_4 = E(v_{ni}^4)$, $\omega_{qn} = [vec_D(P_{1n}), \ldots, vec_D(P_{qn})]$ and

$$B_n = \sigma_0^4 \begin{pmatrix} tr\left(P_{1n}P_{1n}^S\right) & \ldots & tr\left(P_{1n}P_{qn}^S\right) & 0 \\ \vdots & \vdots & \vdots & \vdots \\ tr\left(P_{qn}P_{1n}^S\right) & \ldots & tr\left(P_{qn}P_{qn}^S\right) & 0 \\ 0 & \ldots & 0 & \frac{1}{\sigma_0^2}Q_n'Q_n \end{pmatrix}. \tag{3.4}$$

Then following Lee (2007), we impose the following regularity condition on the limit of $\frac{1}{n}\Omega_n$.

Assumption 3.5. The limit of $\frac{1}{n}\Omega_n$ exists and is a nonsingular matrix.

As in Hansen (1982), with a linear transformation of the moment functions, $a_n g_n(\eta_{r_1})$,[13] we have the following proposition:

Proposition 1. *Under the null SAR model, given Assumptions 2.1–2.4, 3.1–3.5 and B.1, suppose that P_{jn} for $j = 1, \ldots, q$ are from $\mathcal{P}_{1n}$ and $a_0 \lim_{n\to\infty}\frac{1}{n}Eg_n(\eta_{r_1}) = 0$ has a unique root at $\eta_{0r_1} = (\gamma_0', 0)'$ in the parameter space for $r_1 = 1, 2$. Then, the GMME $\hat{\eta}_{n|r_1}$ derived from $\min_{\eta_{r_1}} g_n(\eta_{r_1})'a_n'a_n g_n(\eta_{r_1})$ is a consistent estimator of η_{0r_1}, and $\sqrt{n}(\hat{\eta}_{n|r_1} - \eta_{0r_1}) \xrightarrow{D} N(0, \Sigma)$, where*

$$\Sigma = \lim_{n\to\infty}\left[\left(\frac{1}{n}D_{n|r_1}'\right)a_n'a_n\left(\frac{1}{n}D_{n|r_1}\right)\right]^{-1}\left(\frac{1}{n}D_{n|r_1}'\right)a_n'a_n\left(\frac{1}{n}\Omega_n\right)a_n'a_n\left(\frac{1}{n}D_{n|r_1}\right) \times\left[\left(\frac{1}{n}D_{n|r_1}'\right)a_n'a_n\left(\frac{1}{n}D_{n|r_1}\right)\right]^{-1};$$

for $r_1 = 1$

$$D_{n|1} = \begin{pmatrix} \sigma_0^2 tr\left(P_{1n}^S G_n\right) & 0 & 0 \\ \vdots & \vdots & \vdots \\ \sigma_0^2 tr\left(P_{qn}^S G_n\right) & 0 & 0 \\ Q_n'G_nX_n\beta_0 & Q_n'X_n & Q_n'S_{n|sar}^{ex*-1}X_n\beta_{n|sar}^{ex*} \end{pmatrix};$$

and for $r_1 = 2$

$$D_{n|2} = \begin{pmatrix} \sigma_0^2 tr\left(P_{1n}^S G_n\right) & 0 & \sigma_0^2 tr\left(P_{1n}^S U_{n|sar}^* S_n^{-1}\right) \\ \vdots & \vdots & \vdots \\ \sigma_0^2 tr\left(P_{qn}^S G_n\right) & 0 & \sigma_0^2 tr\left(P_{qn}^S U_{n|sar}^* S_n^{-1}\right) \\ Q_n'G_nX_n\beta_0 & Q_n'X_n & Q_n'\left[U_{n|sar}^* S_n^{-1}X_n\beta_0 + X_n\beta_{n|sar}^{ex*}\right] \end{pmatrix}.$$

From Proposition 1, the optimal choice of a weighting matrix $a_n'a_n$ is $\left(\frac{1}{n}\Omega_n\right)^{-1}$ by the generalized Schwartz inequality. We have the following proposition:

Proposition 2. *Under the null SAR model, given Assumptions 2.1–2.4, 3.1–3.5 and B.1, suppose that $\left(\hat{\Omega}_n\right)^{-1} - \left(\Omega_n\right)^{-1} = o_p(1)$, then the feasible optimal GMME $\hat{\eta}_{on|r_1}$ derived from $\min_{\eta_{r_1}} g_n'(\eta_{r_1})\left(\hat{\Omega}_n\right)^{-1} g_n(\eta_{r_1})$ with P_{jn}'s from $\mathcal{P}_{1n}$ has the asymptotic distribution*

$$\sqrt{n}\left(\hat{\eta}_{on|r_1} - \eta_{0r_1}\right) \xrightarrow{D} N\left(0, \left(\lim_{n\to\infty} D_{n|r_1}'(\Omega_n)^{-1}D_{n|r_1}\right)^{-1}\right),$$

for $r_1 = 1, 2$.

Therefore, we could construct test statistics based on Proposition 1 to test whether δ_{r_1} is significantly different from zero or not. Also we could consider using the feasible optimal GMM (FOGMM) approach to construct the test statistics. Our J-test procedure based on the FOGMM approach can be summarized as follows:

Step 1: Estimate the parameters in the MESS model by the ML method in LeSage and Pace (2007) and calculate predictors $\hat{Y}_{n|r_1}$ for $r_1 = 1, 2$.
Step 2: Estimate the SAR model by the ML method or the GMM method, obtain estimates $\hat{\lambda}_n$ and $\hat{\beta}_n$. Then calculate the initial estimates of the variance of the residuals $\hat{\sigma}_n^2$ by $\hat{\sigma}_n^2 = \frac{1}{n}\hat{V}_n'\hat{V}_n$, where $\hat{V}_n = Y_n - \hat{\lambda}_nW_nY_n - X_n\hat{\beta}_n$.

[13] As in the usual GMM estimation framework, a_n is a matrix with a full rank. Also a_n is assumed to converge to a constant full rank matrix a_0.

Step 3: Use the results in the previous two steps to compute the weighting matrix $\left(\hat{\Omega}_n\right)^{-1}$.
Step 4: Use the FOGMM method to estimate the augmented model. In particular, $\hat{\eta}_{n|r_1}$ can be derived from $\min_{\eta_{r_1}} g_n'(\eta_{r_1})\left(\hat{\Omega}_n\right)^{-1} g_n(\eta_{r_1})$.

Let $R=(0_{1\times(k+1)}, 1)$. Then, the J statistic as the Wald test statistic is

$$\mathcal{W}_{ogmm|r_1} = \left(R\hat{\eta}_{n|r_1}\right)'\left(R\left(\hat{D}_{n|r_1}'\left(\hat{\Omega}_n\right)^{-1}\hat{D}_{n|r_1}\right)^{-1}R'\right)^{-1}\left(R\hat{\eta}_{n|r_1}\right). \tag{3.5}$$

Moreover, we could also construct a DD test statistic and a G test statistic in the GMM framework. The DD test statistic is:

$$\mathcal{DD}_{ogmm|r_1} = \min_{\eta_{r_1}|\delta_{r_1}=0} g_n'(\eta_{r_1})\hat{\Omega}_n^{-1}g_n(\eta_{r_1}) - \min_{\eta_{r_1}} g_n'(\eta_{r_1})\hat{\Omega}_n^{-1}g_n(\eta_{r_1}). \tag{3.6}$$

Lastly, denote $\hat{D}_{n|ogmmr_1}$ as the first derivative matrix of $g_n(\eta_{r_1})$, with respect to η_{r_1}, evaluated at the FOGMM estimates for the restricted parameters, $\hat{\eta}_{n|ogmm} = \left(\hat{\lambda}_{n|ogmm}, \hat{\beta}_{n|ogmm}'\right)'$. Also denote $\hat{S}_{n|ogmm}^{-1} = \left(I_n - \hat{\lambda}_{n|ogmm}W_n\right)^{-1}$, $\hat{G}_{n|ogmm} = W_n\left(I_n - \hat{\lambda}_{n|ogmm}W_n\right)^{-1}$ and $\hat{S}_n^{ex} = S_n^{ex}(\hat{\mu}_n)$. Let $\hat{\sigma}_{n|ogmm}^2$ be the FOGMM estimate of the variance of residuals of the restricted model (SAR). Explicitly,

$$\hat{D}_{n|ogmm1} = \begin{pmatrix} \hat{\sigma}_{n|ogmm}^2 tr\left(P_{1n}^S\hat{G}_{n|ogmm}\right) & 0 & 0 \\ \vdots & \vdots & \vdots \\ \hat{\sigma}_{n|ogmm}^2 tr\left(P_{qn}^S\hat{G}_{n|ogmm}\right) & 0 & 0 \\ Q_n'\hat{G}_{n|ogmm}X_n\hat{\beta}_{n|ogmm} & Q_n'X_n & Q_n'\hat{S}_n^{ex-1}X_n\hat{\beta}_n^{ex} \end{pmatrix}$$

and

$$\hat{D}_{n|ogmm2} = \begin{pmatrix} \hat{\sigma}_{n|ogmm}^2 tr\left(P_{1n}^S\hat{G}_{n|ogmm}\right) & 0 & \hat{\sigma}_{n|ogmm}^2 tr\left(P_{1n}^S\hat{U}_n\hat{S}_{n|ogmm}^{-1}\right) \\ \vdots & \vdots & \vdots \\ \hat{\sigma}_{n|ogmm}^2 tr\left(P_{qn}^S\hat{G}_{n|ogmm}\right) & 0 & \hat{\sigma}_{n|ogmm}^2 tr\left(P_{qn}^S\hat{U}_n\hat{S}_{n|ogmm}^{-1}\right) \\ Q_n'\hat{G}_{n|ogmm}X_n\hat{\beta}_{n|ogmm} & Q_n'X_n & Q_n'\left[\hat{U}_n\hat{S}_{n|ogmm}^{-1}X_n\hat{\beta}_{n|ogmm} + X_n\hat{\beta}_n^{ex}\right] \end{pmatrix}.$$

The G test statistic is:

$$\mathcal{G}_{ogmm|r_1} = g_n'\left(\hat{\eta}_{n|ogmm}\right)\hat{\Omega}_n^{-1}\hat{D}_{n|ogmmr_1}\left(\hat{D}_{n|ogmmr_1}'\hat{\Omega}_n^{-1}\hat{D}_{n|ogmmr_1}\right)^{-1}\hat{D}_{n|ogmmr_1}'\hat{\Omega}_n^{-1}g_n\left(\hat{\eta}_{n|ogmm}\right). \tag{3.7}$$

At the 5% level, H_0 would be rejected if $\mathcal{W}_{ogmm|r_1} > \chi_{0.95}^2(1)$, or $\mathcal{DD}_{ogmm|r_1} > \chi_{0.95}^2(1)$, or $\mathcal{G}_{ogmm|r_1} > \chi_{0.95}^2(1)$.

We could also use the 2SLS method to implement the J-test. The test procedure is a special case of the GMM using only linear moments, and in step 2, we will apply the 2SLS method to estimate the augmented model in Eq. (3.2). Let $F_{n|r_1} = \left(W_nY_n, X_n, \hat{Y}_{n|r_1}\right)$ and $P_n = Q_n(Q'_nQ_n)^{-1}Q'_n$. The 2SLS estimator of η_{r_1} is $\hat{\eta}_{n|r_1} = (F_{n|r_1}{}'P_nF_{n|r_1})^{-1}F_{n|r_1}{}'P_nY_n$. The Wald test statistic based on the 2SLS method is

$$\mathcal{W}_{sls|r_1} = \left(R\hat{\eta}_{n|r_1}\right)'\left(R\hat{\sigma}_n^2\left(F_{n|r_1}'P_nF_{n|r_1}\right)^{-1}R'\right)^{-1}\left(R\hat{\eta}_{n|r_1}\right). \tag{3.8}$$

As the 2SLS estimator $\hat{\eta}_{n|r_1}$ is derived from $\min_{\eta_{r_1}} V_n'(\eta_{r_1})Q_n\left(\hat{\sigma}_n^2Q_n'Q_n\right)^{-1}Q_n'V_n(\eta_{r_1})$, the DD test statistic is:

$$\mathcal{DD}_{sls|r_1} = \min_{\eta_{r_1}|\delta_{r_1}=0} V_n'(\eta_{r_1})Q_n\left(\hat{\sigma}_n^2Q_n'Q_n\right)^{-1}Q_n'V_n(\eta_{r_1}) - \min_{\eta_{r_1}} V_n'(\eta_{r_1})Q_n\left(\hat{\sigma}_n^2Q_n'Q_n\right)^{-1}Q_n'V_n(\eta_{r_1}). \tag{3.9}$$

Finally, let $\hat{D}_{n|slsr_1}$ stand for the first derivative matrix of $Q_n'V_n(\eta_{r_1})$ with respect to η_{r_1}, evaluated at the restricted parameters $\hat{\eta}_{n|sls} = \left(\hat{\lambda}_{n|sls}, \hat{\beta}_{n|sls}'\right)'$ from the 2SLS method. Let $\hat{S}_{n|sls} = S_n(\hat{\lambda}_{n|sls})$. Then, we have

$$\hat{D}_{n|sls1} = Q_n'\left[W_nY_n, X_n, \hat{S}_n^{ex-1}X_n\hat{\beta}_n^{ex}\right],$$
$$\hat{D}_{n|sls2} = Q_n'\left[W_nY_n, X_n, \hat{U}_nY_n + X_n\hat{\beta}_n^{ex}\right].$$

Note that $V_n(\hat{\eta}_{n|sls}) = \hat{S}_{n|sls}Y_n - X_n\hat{\beta}_{n|sls}$. The G test statistic is

$$\mathcal{G}_{sls|r_1} = V_n'(\hat{\eta}_{n|sls})Q_n\left(\hat{\sigma}_n^2Q_n'Q_n\right)^{-1}\hat{D}_{n|slsr_1}\left[\hat{D}_{n|slsr_1}'\left(\hat{\sigma}_n^2Q_n'Q_n\right)^{-1}\hat{D}_{n|slsr_1}\right]^{-1} \times \hat{D}_{n|slsr_1}'\left(\hat{\sigma}_n^2Q_n'Q_n\right)^{-1}Q_n'V_n(\hat{\eta}_{n|sls}).$$

3.2. The J-test using the MESS model as the null

Consider the J-test procedure using the MESS model as the null and the SAR model being the alternative:

$$H_0 : S_n^{ex}(\mu)Y_n = X_n\beta^{ex} + V_n, \quad H_1 : Y_n = \lambda W_n Y_n + X_n\beta + V_n. \tag{3.10}$$

Let $\hat{\theta}_n^{sar} = \left(\hat{\lambda}_n, \hat{\beta}_n', \hat{\sigma}_n^2\right)'$ denote the QMLE of the SAR model. The predictors are $\hat{Y}_{n|1} = \left(I_n - \hat{\lambda}_n W_n\right)^{-1} X_n\hat{\beta}_n$ and $\hat{Y}_{n|2} = \hat{\lambda}_n W_n Y_n + X_n\hat{\beta}_n$. The augmented MESS model is:

$$Y_n(\mu) = X_n\beta^{ex} + \hat{Y}_{n|r_2}\delta_{r_2} + V_n, \tag{3.11}$$

where $Y_n(\mu) = S_n^{ex}(\mu)Y_n$ for $r_2 = 1, 2$. Here we will use the nonlinear 2SLS (N2SLS) approach to estimate the augmented model.

Denote $\phi = (\mu, \beta^{ex\prime})'$, $\psi_{r_2} = (\mu, \beta^{ex\prime}, \delta_{r_2})'$ and $\psi_{0r_2} = (\mu, \beta^{ex\prime}, 0)'$. We impose the following regularity condition:

Assumption 3.6. ψ_{0r_2} is in the interior of the parameter space Ψ, which is a bounded subset of R^{k+2}.

In Appendix C, we investigate the limiting values, or the pseudo true values of $\hat{\theta}_n^{sar}$ based upon the QML method, under the null MESS model. Note that Assumption 2.3 implies that $S_n^{ex}(\mu)$ is uniformly bounded in both row sum and column sums in absolute value for all μ in the parameter space.[14]

Let $S_{n|ex}^* = S_n(\lambda_{n|ex}^*)$ where $\lambda_{n|ex}^*$ is the sequence of pseudo true values of $\hat{\lambda}_n$ under the null MESS model. Also denote $\beta_{n|ex}^*$ as the sequence of pseudo true values of $\hat{\beta}_n$. Let $S_n^{ex} = S_n^{ex}(\mu_0)$. Consider $Y_{n|r_2}^*$, the probability limit of $\hat{Y}_{n|r_2}$, which is

$$Y_{n|1}^* = S_{n|ex}^{*-1} X_n\beta_{n|ex}^*, \quad Y_{n|2}^* = \lambda_{n|ex}^* W_n S_n^{ex-1} X_n\beta_0^{ex} + X_n\beta_{n|ex}^*. \tag{3.12}$$

The J-test procedure for Eq. (3.10) is as follows:

Step 1: Estimate λ and β in the SAR model by the ML method and calculate the predictors $\hat{Y}_{n|r_2}$ for $r_2 = 1, 2$.
Step 2: Use the nonlinear 2SLS (N2SLS) method with the IV matrix $Q_n = (X_n, W_nX_n, \ldots, W_n^dX_n)$ to estimate the augmented Eq. (3.11).

Denote $g_n(\psi_{r_2}) = Q_n'V_n(\psi_{r_2})$. The N2SLS estimator can be derived from

$$\min_{\psi_{r_2}} V_n'(\psi_{r_2})Q_n\left(Q_n'Q_n\right)^{-1}Q_n'V_n(\psi_{r_2}), \tag{3.13}$$

where $V_n(\psi_{r_2}) = Y_n(\mu) - X_n\beta^{ex} - \hat{Y}_{n|r_2}\delta_{r_2}$. As the N2SLS estimation is just a special case of GMM estimation, the identification of ψ_{r_2} requires the unique solution of the limiting equations, $\lim_{n\to\infty}\frac{1}{n}E_{|ex}g_n(\psi_{r_2}) = 0$. Note that

$$\lim_{n\to\infty}\frac{1}{n}E_{|ex}g_n(\psi_{r_2}) = \lim_{n\to\infty}\frac{1}{n}Q_n'\left(S_n^{ex}(\mu)S_n^{ex-1}X_n\beta_0^{ex} - X_n\beta^{ex} - Y_{n|r_2}^*\delta_{r_2}\right) + o(1).$$

Thus, we impose the following identification condition:

Assumption 3.7. The limiting equations $\lim_{n\to\infty}\frac{1}{n}Q_n'\left(S_n^{ex}(\mu)S_n^{ex-1}X_n\beta_0^{ex} - X_n\beta^{ex} - Y_{n|r_2}^*\delta_{r_2}\right) = 0$ has a unique root in the parameter space in the augmented MESS model.

The asymptotic normality of the N2SLS estimator $\hat{\psi}_{n|r_2}$ follows from the next proposition:

Proposition 3. *Under the null MESS model, given Assumptions 2.1–2.4, 3.6–3.7 and C.1, the N2SLS estimator $\hat{\psi}_{n|r_2}$ derived from $\min_{\psi_{r_2}}V_n'(\psi_{r_2})Q_n(Q_n'Q_n)^{-1}Q_n'V_n(\psi_{r_2})$ is a consistent estimator of ψ_{0r_2}, and*

$$\sqrt{n}\left(\hat{\psi}_{n|r_2} - \psi_{0r_2}\right) \xrightarrow{D} N\left(0, \sigma_0^{ex2}\left(\operatorname{plim}\frac{1}{n}D_{n|r_2}'\left(Q_n'Q_n\right)^{-1}D_{n|r_2}\right)^{-1}\right),$$

where

$$D_{n|1} = Q_n'\left[W_nX_n\beta_0^{ex}, X_n, S_{n|ex}^{*-1}X_n\beta_{n|ex}^*\right], \quad D_{n|2} = Q_n'\left[W_nX_n\beta_0^{ex}, X_n, \lambda_{n|ex}^*W_nS_n^{ex-1}X_n\beta_0^{ex} + X_n\beta_{n|ex}^*\right].$$

Denote $\hat{\psi}_{n|r_2} = \left(\hat{\mu}_{n|r_2}, \hat{\beta}_{n|r_2}', \hat{\delta}_{n|r_2}\right)'$, $\hat{S}_n = S_n(\hat{\lambda}_n)$, $\hat{S}_{n|r_2}^{ex} = S_n^{ex}(\hat{\mu}_{n|r_2})$ *and*

$$\hat{D}_{n|1} = Q_n'\left[W_nX_n\hat{\beta}_{n|1}^{ex}, X_n, \hat{S}_n^{-1}X_n\hat{\beta}_n\right], \quad \hat{D}_{n|2} = Q_n'\left[W_nX_n\hat{\beta}_{n|2}^{ex}, X_n, \hat{\lambda}_nW_n\hat{S}_{n|2}^{ex-1}X_n\hat{\beta}_{n|2}^{ex} + X_n\hat{\beta}_n\right].$$

[14] This is so, because $\|e^{\mu W_n}\| \le \|I_n\| + |\mu|\|W_n\| + |\mu|^2\|W_n\|^2/2! + \ldots + |\mu|^t\|W_n\|^t/t! + \ldots = e^{|\mu|\|W_n\|}$. It follows that $\sup_n\|e^{\mu W_n}\| \le e^{|\mu|\sup\|W_n\|} < \infty$ under the assumption that $\sup\|W_n\| < \infty$.

The Wald-test statistic is:

$$\mathcal{W}_{sls|r_2} = \left(R\hat{\psi}_{n|r_2}\right)'\left(R\hat{\sigma}_n^{ex2}\left(\hat{D}'_{n|r_2}\left(Q'_nQ_n\right)^{-1}\hat{D}_{n|r_2}\right)^{-1}R'\right)^{-1}\left(R\hat{\psi}_{n|r_2}\right). \quad (3.14)$$

where $R=(0_{1\times(k+1)}, 1)$. The DD test statistic is:

$$\mathcal{DD}_{sls|r_2} = \min_{\psi_{r_2}|\delta_{r_2}=0} V'_n(\psi_{r_2})Q_n\left(\hat{\sigma}_n^{ex2}Q'_nQ_n\right)^{-1}Q'_nV_n(\psi_{r_2}) - \min_{\psi_{r_2}} V'_n(\psi_{r_2})Q_n\left(\hat{\sigma}_n^{ex2}Q'_nQ_n\right)^{-1}Q'_nV_n(\psi_{r_2}). \quad (3.15)$$

Finally, let $\hat{D}_{n|slsr_2}$ denote the first derivative of $Q'_nV_n(\psi_{r_2})$ with respect to ψ_{r_2}, evaluated at the N2SLS restricted estimate $\hat{\psi}_{n|sls} = \left(\hat{\mu}_{n|sls}, \hat{\beta}^{ex'}_{n|sls}\right)$. Note that $V_n(\hat{\psi}_{n|sls}) = S_n^{ex}(\hat{\mu}_{n|sls})Y_n - X_n\hat{\beta}^{ex}_{n|sls}$ for $r_2=1, 2$. The G test statistic is

$$\mathcal{G}_{sls|r_2} = V'_n(\hat{\psi}_{n|sls})Q_n\left(\hat{\sigma}_n^{ex2}Q'_nQ_n\right)^{-1}\hat{D}_{n|slsr_2}\left(\hat{D}'_{n|slsr_2}\left(\hat{\sigma}_n^{ex2}Q'_nQ_n\right)^{-1}\hat{D}_{n|slsr_2}\right)^{-1} \times \hat{D}'_{n|slsr_2}\left(\hat{\sigma}_n^{ex2}Q'_nQ_n\right)^{-1}Q'_nV_n(\hat{\psi}_{n|sls}). \quad (3.16)$$

4. The J-test for models with unknown heteroskedasticity

In the previous sections, the error terms of each of the SAR and the MESS models are i.i.d with mean zero and variance σ^2. However, this homoskedastic assumption may be restrictive. Therefore, it might be of interest to extend our J-test procedure to the setting where the error terms are independent but with unknown heteroskedasticity.

Assumption 4.1. The v_{ni}'s are independent $(0, \sigma_{ni}^2)$ with finite moments larger than the fourth order such that $E|v_{ni}|^{4+\varsigma}$ for some $\varsigma>0$ are uniformly bounded for all n and i.

Let $\Sigma_n = Diag(\sigma_{n1}^2, \ldots, \sigma_{nn}^2)$ represent the variance matrix of the error terms, where $\sigma_{ni}^2=E(v_{ni}^2)$ for $i=1, \ldots, n$. Recently, Lin and Lee (2010) and Kelejian and Prucha (2010) propose the Robust GMM (RGMM) estimation method of the SAR model. We follow their RGMM approach[15] to derive a J-test procedure for the two models in the presence of unknown heteroskedasticity.

4.1. The J-test under the SAR model with unknown heteroskedasticity as the null

Recall that our null model and alternative models are:

$$H_0: Y_n = \lambda W_nY_n + X_n\beta + V_n,$$
$$H_1: S_n^{ex}(\mu)Y_n = X_n\beta^{ex} + V_n,$$

but with unknown heteroskedastic variances. Note that $\phi=(\mu, \beta^{ex'})'$ is the vector of parameters of the MESS model without σ^{ex2} and $\hat{\phi}_n$ is the estimated parameter ϕ. The predictors from the MESS model are the same as in Section 3.1, namely $\hat{Y}_{n|1} = S_n^{ex}(\hat{\mu}_n)^{-1}X_n\hat{\beta}_n$ and $\hat{Y}_{n|2} = U_n(\hat{\mu}_n)Y_n + X_n\hat{\beta}_n^{ex}$. Here we will apply the N2SLS method to estimate the MESS model in order to obtain the predictors $\hat{Y}_{n|r_1}$ for $r_1=1, 2.$[16] Specifically, $\hat{\phi}_n$ is obtained from $\min_\phi V'_n(\phi)Q_n(Q'_nQ_n)^{-1}Q'_nV_n(\phi)$ where Q_n is the same IV matrix in Section 3 and $V_n(\phi) = S_n^{ex}(\mu)Y_n - X_n\beta^{ex}$. We discuss the pseudo true values of $\hat{\phi}_n$ based on the N2SLS method in Appendix D.

The augmented SAR equation is:

$$Y_n = \lambda W_nY_n + X_n\beta + \hat{Y}_{n|r_1}\delta_{r_1} + V_n.$$

Recall that $\eta_{r_1}=(\lambda, \beta', \delta_{r_1})'$ and $V_n(\eta_{r_1}) = S_n(\lambda)Y_n - X_n\beta - \hat{Y}_{n|r_1}\delta_{r_1}$. We will construct our RGMM estimation for this augmented equation through the linear and the quadratic moments. The IV matrix Q_n used in the linear moment function will be the same as in Section 3. For quadratic moments, we consider matrix P_n in $\mathcal{P}_{2n}$ with $Diag(P_n)=0$. As in Lin and Lee (2010), by taking P_n from $\mathcal{P}_{2n}$, we maintain the uncorrelatedness between V_n and P_nV_n because $E(V'_nP_nV_n) = tr[P_nE(V_nV'_n)] = tr[Diag(P_n)E(V_nV'_n)] = 0$. We impose the following conditions on $\mathcal{P}_{2n}$:

Assumption 4.2. The matrices P_{jn}'s from $\mathcal{P}_{2n}$ are uniformly bounded in both row and column sum norms.

The set of moment functions form the vector $g_n(\eta_{r_1}) = (P_{1n}V_n(\eta_{r_1}), \ldots, P_{qn}V_n(\eta_{r_1}), Q_n)'V_n(\eta_{r_1})$. For identification of the parameters, the first part of the identification condition of Assumption 3.4 will be maintained but the second part needs to be modified.

Assumption 4.3. Either (i) $\lim_{n\to\infty}\frac{1}{n}Q'_n\left[X_n, Y^*_{n|r_1}, G_nX_n\beta_0\right]$ has full rank $k+2$ for $r_1=1, 2$ or (ii) $\lim_{n\to\infty}\frac{1}{n}Q'_n\left[X_n, Y^*_{n|r_1}\right]$ has full rank $k+1$ for $r_1=1, 2$, $\lim_{n\to\infty}\frac{1}{n}tr\left(\Sigma_nP^S_{jn}G_n\right)\neq 0$ for some $j\in 1, \cdots, q$, and $\lim_{n\to\infty}\frac{1}{n}\left[tr\left(\Sigma_nP^S_{1n}G_n\right), \ldots, tr\left(\Sigma_nP^S_{qn}G_n\right)\right]$ and $\lim_{n\to\infty}\frac{1}{n}\left[tr\left(\Sigma_nG'_nP_{1n}G_n\right), \ldots, tr\left(\Sigma_nG'_nP_{qn}G_n\right)\right]$ are linearly independent.

Recall that $\mu^*_{n|sar}$ is the sequence of pseudo true values of $\hat{\mu}_n$ under the null SAR model and $\beta^{ex*}_{n|sar}$ is the sequence of pseudo true values of $\hat{\beta}^{ex}_n$. Let $S^{ex*}_{n|sar} = S^{ex}_n(\mu^*_{n|sar})$, $U^*_{n|sar} = U_n(\mu^*_{n|sar})$. Similar to Lin and Lee (2010), the consistency and asymptotic normality of the RGMM estimator can be derived as follows:

Proposition 4. *Under the null SAR model, given Assumptions 2.2–2.4, 3.1–3.2, 4.1–4.3 and D.1, suppose that P_{jn} are from $\mathcal{P}_{2n}$, a_0 $\lim_{n\to\infty}\frac{1}{n}Eg_n(\eta_{r_1}) = 0$ has a unique root at $\eta_{0r_1} = (\gamma'_0, 0)'$ in its parameter space for $r_1=1, 2$. Then, the RGMME $\hat{\eta}_{n|r_1}$ derived from $\min_{\eta_{r_1}} g_n(\eta_{r_1})'a'_na_ng_n(\eta_{r_1})$ is a consistent estimator of η_{0r_1}, and $\sqrt{n}\left(\hat{\eta}_{n|r_1} - \eta_{0r_1}\right) \xrightarrow{D} N(0, \Gamma)$, where*

$$\Gamma = \lim_{n\to\infty}\frac{1}{n}\left(D'_{nh|r_1}a'_na_nD_{nh|r_1}\right)^{-1}D'_{nh|r_1}a'_na_n\Omega_{nh}a'_na_nD_{nh|r_1}\left(D'_{nh|r_1}a'_na_nD_{nh|r_1}\right)^{-1}$$

$$\Omega_{nh} = Var(g_n(\eta_{0r_1})) = \begin{pmatrix} tr\left[\Sigma_nP_{1n}\Sigma_nP^S_{1n}\right] & tr\left[\Sigma_nP_{1n}\Sigma_nP^S_{2n}\right] & \ldots & 0 \\ tr\left[\Sigma_nP_{2n}\Sigma_nP^S_{1n}\right] & tr\left[\Sigma_nP_{2n}\Sigma_nP^S_{2n}\right] & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \ldots & Q'_n\Sigma_nQ_n \end{pmatrix}$$

and when $r_1=1$

$$D_{nh|1} = \begin{pmatrix} tr\left(\Sigma_nP^S_{1n}G_n\right) & 0 & 0 \\ \vdots & \vdots & \vdots \\ tr\left(\Sigma_nP^S_{qn}G_n\right) & 0 & 0 \\ Q'_nG_nX_n\beta_0 & Q'_nX_n & Q'_nS^{ex*-1}_{n|sar}X_n\beta^{ex*}_{n|sar} \end{pmatrix};$$

[15] The Robust 2SLS method is just a special case of the RGMM method.

[16] There may be other methods useful for the estimation of the MESS model with unknown heteroskedasticity. Here we use the N2SLS as the model equation has a form well suited for that estimation method.

when $r_1=2$

$$D_{nh|2}=\begin{pmatrix} tr\left(\Sigma_n P_{1n}^S G_n\right) & 0 & tr\left(\Sigma_n P_{1n}^S U_{n|sar}^* S_n^{-1}\right) \\ \vdots & \vdots & \vdots \\ tr\left(\Sigma_n P_{qn}^S G_n\right) & 0 & tr\left(\Sigma_n P_{qn}^S U_{n|sar}^* S_n^{-1}\right) \\ Q_n' G_n X_n \beta_0 & Q_n' X_n & Q_n'\left[U_{n|sar}^* S_n^{-1} X_n \beta_0 + X_n \beta_{n|sar}^{ex*}\right] \end{pmatrix}.$$

As expected, the Ω_{nh} and $D_{nh|r_1}$ can be consistently estimated via a similar procedure to the robust variance construction in White (1980):

Proposition 5. *Under the null SAR model, given Assumption 2.2–2.4, 3.1–3.2, 4.1–4.3 and D.1,* $\frac{1}{n}\left(\hat{D}_{nh|r_1}-D_{nh|r_1}\right)=o_p(1)$ *for* $r_1=1,\ 2$ *and* $\frac{1}{n}\left(\hat{\Omega}_{nh}-\Omega_{nh}\right)=o_p(1)$*, where* $\frac{1}{n}\hat{D}_{nh|r_1}$ *and* $\frac{1}{n}\hat{\Omega}_{nh}$ *are, respectively, estimators of* $\frac{1}{n}D_{nh|r_1}$ *and* $\frac{1}{n}\Omega_{nh}$*, with all the parameters replaced by their consistent estimators, and* Σ_n *by* $\hat{\Sigma}_n$*, where* $\hat{\Sigma}_n=Diag\left(\hat{v}_{n1}^2,\ldots,\hat{v}_{nn}^2\right)$ *and* $\hat{v}_{ni}$*'s are the residuals obtained from the initial estimates of the SAR model.*

With a consistently estimated Ω_{nh}, a feasible optimal RGMM (FORGMM) estimation for the augmented model can be derived.

Assumption 4.4. $\lim_{n\to\infty}\frac{1}{n}\Omega_{nh}$ exists and is nonsingular.

Proposition 6. *Suppose that* $\left(\frac{1}{n}\hat{\Omega}_{nh}\right)^{-1}-\left(\frac{1}{n}\Omega_{nh}\right)^{-1}=o_p(1)$*, under the null SAR model, given Assumption 2.2–2.4, 3.1–3.2, 4.1–4.4 and D.1, then the FORGMME* $\hat{\eta}_{on|r_1}$ *derived from* $\min_{\eta_{r_1}} g_n'(\eta_{r_1})\hat{\Omega}_{nh}^{-1}g_n(\eta_{r_1})$ *has the asymptotic distribution*

$$\sqrt{n}\left(\hat{\eta}_{on|r_1}-\eta_{0r_1}\right)\xrightarrow{D}N\left(0,\left(\lim_{n\to\infty}\frac{1}{n}D_{nh|r_1}'\Omega_{nh}^{-1}D_{nh|r_1}\right)^{-1}\right)$$

for $r_1=1,\ 2.$

Our J-test procedure based on the FORGMM method can be summarized as follows:

Step 1: Estimate the MESS model by the N2SLS method and obtain estimates of the relevant predictors.
Step 2: Estimate the SAR model by the RGMM method to obtain initial consistent estimates $\hat{\lambda}_n$ and $\hat{\beta}_n$. Then use the estimated residuals to compute $\hat{\Sigma}_n$.
Step 3: Use the results in the previous two steps to compute the weighting matrix $\hat{\Omega}_{nh}^{-1}$.
Step 4: Use the FORGMM method to estimate the augmented model.

The Wald test statistic is

$$\mathcal{W}_{orgmm|r_1}=\left(R\hat{\eta}_{on|r_1}\right)'\left(R\left(\hat{D}_{nh|r_1}'\left(\hat{\Omega}_{nh}\right)^{-1}\hat{D}_{nh|r_1}\right)^{-1}R'\right)^{-1}\left(R\hat{\eta}_{on|r_1}\right). \tag{4.1}$$

We can also construct a DD test statistic and a G test statistic based on the FORGMM method. The DD test statistic is:

$$\mathcal{DD}_{orgmm|r_1}=\min_{\eta_{r_1}|\delta_{r_1}=0} g_n'(\eta_{r_1})\hat{\Omega}_{nh}^{-1}g_n(\eta_{r_1})-\min_{\eta_{r_1}} g_n'(\eta_{r_1})\hat{\Omega}_{nh}^{-1}g_n(\eta_{r_1}). \tag{4.2}$$

Lastly, denote $\hat{D}_{nh|r_1}$ as the first derivative of $g_n(\eta_{r_1})$, with respect to η_{r_1}, evaluated at the restricted FORGMM estimate $\hat{\eta}_{n|rgmm}=\left(\hat{\lambda}_{n|rgmm},\hat{\beta}_{n|rgmm}'\right)'$, which is:

$$\hat{D}_{nh|1}=\begin{pmatrix} tr\left(\hat{\Sigma}_n P_{1n}^S \hat{G}_{n|rgmm}\right) & 0 & 0 \\ \vdots & \vdots & \vdots \\ tr\left(\hat{\Sigma}_n P_{qn}^S \hat{G}_{n|rgmm}\right) & 0 & 0 \\ Q_n'\hat{G}_{n|rgmm}X_n\hat{\beta}_{n|rgmm} & Q_n'X_n & Q_n'\hat{S}_n^{ex-1}X_n\hat{\beta}_n^{ex} \end{pmatrix}$$

and

$$\hat{D}_{nh|2}=\begin{pmatrix} tr\left(\hat{\Sigma}_n P_{1n}^S \hat{G}_{n|rgmm}\right) & 0 & tr\left(\hat{\Sigma}_n P_{1n}^S \hat{U}_n \hat{S}_{n|rgmm}^{-1}\right) \\ \vdots & \vdots & \vdots \\ tr\left(\hat{\Sigma}_n P_{qn}^S \hat{G}_{n|rgmm}\right) & 0 & tr\left(\hat{\Sigma}_n P_{qn}^S \hat{U}_n \hat{S}_{n|rgmm}^{-1}\right) \\ Q_n'\hat{G}_{n|rgmm}X_n\hat{\beta}_{n|rgmm} & Q_n'X_n & Q_n'\left[\hat{U}_n\hat{S}_{n|rgmm}^{-1}X_n\hat{\beta}_{n|rgmm}+X_n\hat{\beta}_n^{ex}\right] \end{pmatrix}.$$

where $\hat{G}_{n|rgmm}=W_n(I_n-\hat{\lambda}_{n|rgmm})^{-1}$ and $\hat{S}_{n|rgmm}=\left(I_n-\hat{\lambda}_{n|rgmm}W_n\right)^{-1}$. As a result, our G test statistic is:

$$\mathcal{G}_{orgmm|r_1}=g_n'\left(\hat{\eta}_{n|rgmm}\right)\hat{\Omega}_{nh}^{-1}\hat{D}_{nh|r_1}\left(\hat{D}_{nh|r_1}'\hat{\Omega}_{nh}^{-1}\hat{D}_{nh|r_1}\right)^{-1}\hat{D}_{nh|r_1}'\hat{\Omega}_{nh}^{-1}g_n(\hat{\eta}_{n|rgmm}).$$

4.2. The J-test under the MESS model with unknown heteroskedasticity as the null

Here, our null and alternative models are:

$$H_0: S_n^{ex}(\mu)Y_n=X_n\beta^{ex}+V_n,$$
$$H_1: Y_n=\lambda W_nY_n+X_n\beta+V_n.$$

And the augmented model is

$$Y_n(\mu)=X_n\beta^{ex}+\hat{Y}_{n|r_2}\delta_{r_2}+V_n,$$

where $Y_n(\mu)=S_n^{ex}(\mu)Y_n$. Recall that $\gamma=(\lambda,\ \beta')'$. Let $\hat{\gamma}_n=\left(\hat{\lambda}_n,\hat{\beta}_n'\right)'$ represent the 2SLS or RGMM estimate of γ of the SAR model. The predictors from the SAR model are $\hat{Y}_{n|1}=(I_n-\hat{\lambda}_nW_n)^{-1}X_n\hat{\beta}_n$ and $\hat{Y}_{n|2}=\hat{\lambda}_nW_nY_n+X_n\hat{\beta}_n$. The detailed analysis of the pseudo true values of $\hat{\gamma}_n$ is given in Appendix E.

We estimate the MESS model by the N2SLS method and use the estimated residuals to obtain consistent estimates of the variance matrix Σ_n as in White (1980). Finally we will use a generalized N2SLS (GN2SLS) method to estimate the augmented MESS equation. Recall that $\psi_{r_2}=(\mu,\beta^{ex\prime},\delta_{r_2})'$ and $V_n(\psi_{r_2})=Y_n(\mu)-X_n\beta^{ex}-\hat{Y}_{n|r_2}\delta_{r_2}$. Let $S_{n|ex}^*=S_n(\lambda_{n|ex}^*)$ where $\lambda_{n|ex}^*$ is the sequence of pseudo true values of $\hat{\lambda}_n$. $\beta_{n|ex}^*$ is the sequence of pseudo true values of $\hat{\beta}_n$. We have the following proposition:

Proposition 7. *Under the null MESS model, given Assumptions 2.2–2.4, 3.6–3.7, 4.1–4.3 and E.1, the GN2SLS estimator* $\hat{\psi}_{n|r_2}$ *derived from* $\min_{\psi_{r_2}} V_n'(\psi_{r_2})Q_n\left(Q_n'\hat{\Sigma}_nQ_n\right)^{-1}Q_n'V_n(\psi_{r_2})$ *is a consistent estimator of* ψ_{0r_2}*, and*

$$\sqrt{n}\left(\hat{\psi}_{n|r_2}-\psi_{0r_2}\right)\xrightarrow{D}N\left(0,\left(\text{p}\lim\frac{1}{n}D_{n|r_2}'\left(Q_n'\Sigma_nQ_n\right)^{-1}D_{n|r_2}\right)^{-1}\right),$$

where $D_{n|r_2}$ is

$$D_{n|1}=Q_n'\left(W_nX_n\beta_0^{ex},X_n,S_{n|ex}^{*-1}X_n\beta_{n|ex}^*\right),$$
$$D_{n|2}=Q_n'\left(W_nX_n\beta_0^{ex},X_n,\left[\lambda_{n|ex}^*W_nS_n^{ex-1}X_n\beta_0^{ex}+X_n\beta_{n|ex}^*\right]\right).$$

Our J-test procedure can be summarized as follows:

Step 1: Estimate the SAR model by the 2SLS or the RGMM method and calculate the relevant predictors.
Step 2: Estimate the MESS model by the N2SLS method and use the estimated residuals to compute the variance matrix of the error terms Σ_n.
Step 3: Use the GN2SLS method to estimate the augmented MESS equation based on the results in the previous two steps.
Step 4: Construct the corresponding Wald, DD and G test statistics.

Recall that $R=(0_{1\times(k+1)},1)$, the Wald test statistic is

$$\mathcal{W}_{sls|r_2}=\left(R\hat{\psi}_{n|r_2}\right)'\left(R\left(\hat{D}'_{n|r_2}\left(Q'_n\hat{\sum}_nQ_n\right)^{-1}\hat{D}_{n|r_2}\right)^{-1}R'\right)^{-1}\left(R\hat{\psi}_{n|r_2}\right).$$

The DD test statistic is:

$$\mathcal{DD}_{sls|r_2}=\min_{\psi_{r_2}|\delta_{r_2}=0}V'_n(\psi_{r_2})Q_n\left(Q'_n\hat{\sum}_nQ_n\right)^{-1}Q'_nV_n(\psi_{r_2})-\min_{\psi_{r_2}}V'_n(\psi_{r_2})Q_n\left(Q'_n\hat{\sum}_nQ_n\right)^{-1}Q'_nV_n(\psi_{r_2}).$$

Lastly, let $\hat{D}_{n|slsr_2}$ denote the first derivative of $Q'_nV_n(\psi_{r_2})$ with respect to ψ_{r_2}, evaluated at the restricted GN2SLS estimate $\hat{\psi}_{n|sls}=\left(\hat{\mu}_{n|sls},\hat{\beta}^{ex'}_{n|sls}\right)'$. The G test statistic is

$$\mathcal{G}_{sls|r_2}=V'_n\left(\hat{\psi}_{n|sls}\right)Q_n\left(Q'_n\hat{\sum}_nQ_n\right)^{-1}\hat{D}_{n|slsr_2}\left(\hat{D}'_{n|slsr_2}\left(Q'_n\hat{\sum}_nQ_n\right)^{-1}\hat{D}_{n|slsr_2}\right)^{-1}\times\hat{D}'_{n|slsr_2}\left(Q'_n\hat{\sum}_nQ_n\right)^{-1}Q'_nV_n(\hat{\psi}_{n|sls}).$$

5. Monte Carlo experiment

5.1. Experiment design

We consider the following two pairs of experiments:

$$\begin{aligned}H_0&:Y_n=\lambda W_nY_n+l_n\beta_1+X_{2n}\beta_2+V_n\\H_1&:S^{ex}_n(\mu)Y_n=l_n\beta^{ex}_1+X_{2n}\beta^{ex}_2+V_n;\end{aligned}\tag{5.1}$$

$$\begin{aligned}H_0&:S^{ex}_n(\mu)Y_n=l_n\beta^{ex}_1+X_{2n}\beta^{ex}_2+V_n\\H_1&:Y_n=\lambda W_nY_n+l_n\beta_1+X_{2n}\beta_2+V_n\end{aligned}\tag{5.2}$$

where $l_n=(1,\ldots,1)'$ is the $n\times1$ column vector of ones for the intercept, X_{2n} is a column vector of exogenous regressors, and β_1, β^{ex}_1 are, respectively, the coefficients of the intercept term of the SAR model and the MESS model. The IV matrix used in the experiment is

$$Q_n=\left[l_n,X_{2n},W_nX_{2n},W^2_nX_{2n},W^3_nX_{2n}\right].$$

The J-test statistics considered are:

$\mathcal{W}_{sls}$: the Wald test statistic based on the 2SLS or the N2SLS methods, using Q_n as the IV matrix.
$\mathcal{G}_{sls}$: the G test statistic based on the 2SLS or the N2SLS methods.
$\mathcal{DD}_{sls}$: the DD test statistic based on the 2SLS or the N2SLS methods.
$\mathcal{W}_{ogmm}$: the Wald test statistic based on the FOGMM method, using Q_n as the IV matrix for the linear moments, and W_n, $W^2_n-\frac{1}{n}tr\left(W^2_n\right)I_n$ for the quadratic moments.
$\mathcal{G}_{ogmm}$: the G test statistic based on the FOGMM method.
$\mathcal{DD}_{ogmm}$: the DD test statistic based on the FOGMM method.

The spatial weight matrix W_n is constructed by the function "makeneighborsw",[17] which generates a row-normalized spatial weight matrix based on m nearest neighbors. Specifically, the function first computes a distance measure $d(i,j)$ between any two points, i and j, that have coordinates (x_i,y_i) and (x_j,y_j). Then for each i, the function selects the m nearest neighbors based on $d(i,j)$, $j\neq i$. If $d(i,j)$ is among the m closest distances, then $W^*_{ij}=1$ and $W_{ij}=\frac{W^*_{ij}}{\sum_{j=1}^nW^*_{ij}}$. In all sets of experiment, m is set to be 5.

[17] This function is taken from LeSage's matlab code for spatial econometrics, which can be found at http://www.spatial-econometrics.com/.

Following Kelejian and Piras (2011), we consider two distributions for the exogenous regressor X_{2n}. The first distribution is $\chi^2(3)$, a chi-squared distribution with three degrees of freedom. The second is the uniform distribution $U(0,10)$ over (0, 10). The V_{ni}'s are randomly generated from a normal distribution with zero mean and a finite variance. With homoskedastic disturbances, the estimation procedure in the first step of the J-test for constructing predictors is the ML method. For the first pair of experiments (Eq. (5.1)), we consider several sets of parameter values for the two models. Specifically, if the data generating process (dgp) is the SAR model, parameter value 1 (P-V1) has $(\lambda_0,\beta_{10},\beta_{20},\sigma_0)=(0.6,2,1,1)$ and value 2 (P-V2) has $(\lambda_0,\beta_{10},\beta_{20},\sigma_0)=(0.6,2,0.5,\sqrt{2})$. If the dgp is the MESS model, parameter value 1 (P-V1) has $(\mu_0,\beta^{ex}_{10},\beta^{ex}_{20},\sigma^{ex}_0)=(-1.6094,2,1,1)$ and value 2 (P-V2) has $(\mu_0,\beta^{ex}_{10},\beta^{ex}_{20},\sigma^{ex}_0)=(-1.6094,2,0.5,\sqrt{2})$. The variation in the error terms with P-V2 is relatively more dominant than that of P-V1 since the coefficient of the exogenous regressor in P-V2 becomes smaller and the standard deviation of the error terms become larger. In addition to $\lambda_0=0.6$ and $\mu_0=-1.6094$, we also consider a moderate spatial interaction effect model with $\lambda_0=0.4$ and $\mu_0=-0.5108$.[18] Thus if the dgp is the SAR model, parameter value 3 (P-V3) has $(\lambda_0,\beta_{10},\beta_{20},\sigma_0)=(0.4,2,1,1)$ and value 4 (P-V4) has $(\lambda_0,\beta_{10},\beta_{20},\sigma_0)=(0.4,2,0.5,\sqrt{2})$. If the dgp is the MESS model, parameter value 3 (P-V3) has $(\mu_0,\beta^{ex}_{10},\beta^{ex}_{20},\sigma^{ex}_0)=(-0.5108,2,1,1)$ and value 4 (P-V4) has $(\mu_0,\beta^{ex}_{10},\beta^{ex}_{20},\sigma^{ex}_0)=(-0.5108,2,0.5,\sqrt{2})$.

Lastly, for the second pair of experiments (Eq. (5.2)), we still consider two values for the spatial parameters λ and μ. If the dgp is the MESS model, parameter value 5 (P-V5) has $(\mu_0,\beta^{ex}_{10},\beta^{ex}_{20},\sigma^{ex}_0)=(-1.6094,2,1,1)$ and value 6 (P-V6) has $(\mu_0,\beta^{ex}_{10},\beta^{ex}_{20},\sigma^{ex}_0)=(-0.5108,2,1,1)$. If the dgp is the SAR model, parameter value 5 (P-V5) has $(\lambda_0,\beta_{10},\beta_{20},\sigma_0)=(0.6,2,1,1)$ and value 6 (P-V6) has $(\lambda_0,\beta_{10},\beta_{20},\sigma_0)=(0.4,2,1,1)$.

We use 1000 repetitions for each case in the Monte Carlo experiment. The regressors are randomly redrawn for each repetition. We consider 4 sample sizes here: 100, 300, 500 and 700. We calculate all the test statistics and compute the relevant empirical sizes and powers. These test statistics are evaluated at 5% critical values of the chi-squared distribution with one degree of freedom.

For the J-test procedure for models with unknown heteroskedasticity, we follow the variance design in Arraiz et al. (2010). Explicitly, we take the ith element of V_n as

$$\begin{aligned}v_{n,i}&=\sigma_{n,i}\epsilon_{n,i},\\\sigma_{n,i}&=c\frac{Ne_{n,i}}{\sum_{j=1}^nNe_{n,j}/n}\end{aligned}\tag{5.3}$$

where $\epsilon_{n,i}$ is generated from i.i.d $N(0,1)$ for all sample sizes considered and $Ne_{n,i}$ is the number of neighbors that the ith unit has. The c is set to be 2 in all experiments. As we need variation in the number of neighbors for each unit, we construct the spatial weight matrix following the specifications given by Arraiz et al. (2010). That is, we consider W_n in terms of a square grid. Let x_i and y_i, which only take values $1,\ 1.5,\ 2,\ 2.5,\ldots,\ \bar{L}$, denote the coordinates for unit i. For the units in the northeastern quadrant, both coordinates take discrete values $L,\ L+0.5,\ L+1,\ L+1.5,\ L+2,\ldots,\bar{L}$. The coordinates of the remaining units only take integer values $1,2,\ldots,L-1$. Then, we can define a distance measure between any two units i and j, whose coordinates are (x_i,y_i) and (x_j,y_j), as follows:

$$d(i,j)=\sqrt{\left(x_i-x_j\right)^2+\left(y_i-y_j\right)^2}$$

and the row normalized spatial weight matrix W_n is defined as $W_{i,j}=W^*_{i,j}/\sum_{j=1}^nW^*_{i,j}$, where $W^*_{i,j}=1$ if $d(i,j)$ is between 0 and 1, and 0 otherwise. In the experiment, we consider two cases of this matrix, namely

[18] LeSage and Pace (2007) suggest an approximate mean of relating the magnitude of λ and μ by letting $\lambda=1-e^{\mu}$. So $\mu=-1.6094$ is approximately equivalent to a value of 0.8 for λ in the SAR model while $\mu=-0.5108$ corresponds to $\lambda=0.4$.

$(L=5, \bar{L}=15)$ and $(L=14, \bar{L}=20)$. These values of L and $\bar{L}$ are selected because they have different implications for the proportion of units located in the northeastern part of the grid. $(L=5, \bar{L}=15)$ refers to a case where about 80% of the units are located in the northeastern quadrant while $(L=14, \bar{L}=20)$ implies that about 32% of the units are located in the northeastern quadrant. According to Arraiz et al. (2010), these two cases of W_n correspond to a "space" where units located in the northeastern portion of that space are closer to each other and have more neighbors than units located in other portions of that space.[19] The value of other parameters in the two models are the same as in the homoskedastic case. The test statistics considered here are still the Wald test statistic, the DD test statistic and the G test statistic. However, the differences are: first we use, respectively, the RGMM method[20] or the N2SLS method to estimate the SAR or the MESS model in order to obtain their predictors; second, the estimation method of the augmented model is, respectively, the FORGMM method and the GNSLS method for the SAR model and the MESS model.[21]

Finally, we conduct bootstrap J-tests to investigate the finite sample properties of the test statistics. Burridge and Fingleton (2010) suggest the bootstrap method for the J-tests for the SAR model with various W_n's in order to correct the size-inflation problem for the test statistics. We also utilize the bootstrap method for comparison purpose. The bootstrap method applied here is the residual bootstrap.[22] Consider the homoskedastic case first. If the null model is the SAR model,[23] then the resampling scheme is:

Step 1: Compute the J-test statistics as in Section 3.1

Step 2: Use $\hat{V}_n$ from the ML estimation of the SAR model as the building block, draw a random sample using sampling with replacement; call this resampled residuals $\hat{V}_n^b$.

Step 3: Use $\hat{\lambda}_n$, $\hat{\beta}_{1n}$ and $\hat{\beta}_{2n}$ from the ML estimation, generate

$$Y_n^b = \left(I_n - \hat{\lambda}_n W_n\right)^{-1}\left(l_n\hat{\beta}_{1n} + X_{2n}\hat{\beta}_{2n} + \hat{V}_n^b\right).$$

Step 4: Calculate the J-test statistics using the Y_n^b sample.

Step 5: Repeat steps 2–4 for 99 times to create a bootstrap sample for the J-test statistics. If the proportion of the 99 bootstrap repetitions that exceed the observed J-test statistics is less than 5%, then reject the null hypothesis.

For the models with unknown heteroskedasticity, we use the wild bootstrap approach suggested by MacKinnon (2009).[24] Denote $\tilde{X}_n = (l_n, X_{2n})$. We use the diagonals of $\tilde{X}_n(\tilde{X}_n'\tilde{X}_n)^{-1}\tilde{X}_n'$ to rescale the residuals. If the null model is the SAR model, the resampling scheme is:

Step 1: Compute the J-test statistics as in Section 4.1

Step 2: Rescale the estimated residuals $\hat{V}_n$ derived from the RGMM method by

$$f\left(\hat{V}_{ni}\right) = \frac{\hat{V}_{ni}}{(1-B_i)^{\frac{1}{2}}}$$

where B_i is the ith diagonal of $\tilde{X}_n(\tilde{X}_n'\tilde{X}_n)^{-1}\tilde{X}_n'$;

Step 3: generate n random numbers τ_i^b for $i=1, 2, \ldots, n$, from the Rademacher distribution, where $\tau_i^b = 1$ with probability $\frac{1}{2}$ and $\tau_i^b = -1$ with probability $\frac{1}{2}$.

Step 4: Denote $\xi_i = f\left(\hat{V}_{ni}\right) \times \tau_i^b$ for $i=1, \ldots, n$ and $\xi = (\xi_1', \ldots, \xi_n')'$. Using $\hat{\lambda}_n$, $\hat{\beta}_{1n}$ and $\hat{\beta}_{2n}$ derived from the RGMM method, generate

$$Y_n^b = \left(I_n - \hat{\lambda}_n W_n\right)^{-1}\left(l_n\hat{\beta}_{1n} + X_{2n}\hat{\beta}_{2n} + \xi\right).$$

Step 5: Calculate the J-test statistics using the Y_n^b sample.

Step 6: Repeat steps 2–5 for 99 times to create a bootstrap sample for the J-test statistics. If the proportion of the 99 bootstrap repetitions that exceed the observed J-test statistics is less than 5%, then reject the null hypothesis.

Table 1
Size and power of the J-test statistics under H_0: $Y_n = \lambda W_n Y_n + l_n\beta_1 + X_{2n}\beta_2 + V_n$.

	n	$x=\chi^2(3)$				$x=U(0, 10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{sls}$	100	0.043	0.053	0.939	0.987	0.054	0.055	0.987	1
	300	0.047	0.05	1	1	0.041	0.044	1	1
	500	0.061	0.064	1	1	0.06	0.053	1	1
	700	0.049	0.048	1	1	0.058	0.066	1	1
$\mathcal{W}_{ogmm}$	100	0.144	0.217	0.482	0.873	0.134	0.185	0.574	0.94
	300	0.082	0.104	0.559	1	0.082	0.117	0.676	1
	500	0.072	0.087	0.627	1	0.066	0.09	0.782	1
	700	0.057	0.071	0.672	1	0.068	0.079	0.882	1
$\mathcal{G}_{sls}$	100	0.043	0.053	0.939	0.987	0.054	0.055	1	1
	300	0.047	0.05	1	1	0.041	0.044	1	1
	500	0.061	0.064	1	1	0.06	0.053	1	1
	700	0.049	0.048	1	1	0.058	0.066	1	1
$\mathcal{G}_{ogmm}$	100	0.05	0.045	0.587	0.998	0.041	0.056	0.739	1
	300	0.046	0.056	0.77	1	0.05	0.044	0.897	1
	500	0.053	0.059	0.803	1	0.047	0.066	0.954	1
	700	0.04	0.048	0.841	1	0.045	0.062	0.983	1
$\mathcal{DD}_{sls}$	100	0.043	0.053	0.939	0.987	0.054	0.055	0.987	1
	300	0.047	0.05	1	1	0.041	0.044	1	1
	500	0.061	0.064	1	1	0.06	0.053	1	1
	700	0.049	0.048	1	1	0.058	0.066	1	1
$\mathcal{DD}_{ogmm}$	100	0.048	0.051	0.432	1	0.045	0.059	0.521	1
	300	0.047	0.053	0.581	1	0.053	0.045	0.677	1
	500	0.055	0.061	0.626	1	0.044	0.064	0.763	1
	700	0.038	0.046	0.65	1	0.047	0.059	0.806	1

The SAR model: $\lambda_0 = 0.6$, $\beta_{10} = 2$, $\beta_{20} = 1$, and $\sigma_0 = 1$.
The MESS model: $\mu_0 = -1.6094$, $\beta_{10}^{ex} = 2$, $\beta_{20}^{ex} = 1$, and $\sigma_0^{ex} = 1$.

5.2. Monte Carlo results

Tables 1 and 2 summarize the sizes and powers of the J-test statistics when the null model is the SAR model with parameter values P-V1 and P-V2. The empirical sizes of the Wald test statistics based upon the 2SLS method are reasonable for all the sample sizes. However, there are some size distortions for the Wald test statistics based upon the FOGMM method. For instance, in Table 2, when the sample size is 100, we observe a size of 0.238 for the Wald statistic based upon the FOGMM method, using the second predictor. The size distortions decrease as sample sizes increase. For the G test statistics and the DD test statistics, the empirical sizes seem reasonable although the DD test statistics using the first predictor based on the FOGMM method do not have enough power when the sample size is small. All three test statistics from the second predictor tend to be more powerful than test statistics from the first predictor, suggesting that calculating our predictor based on the structural form of the MESS model can help us to reject the wrong null model specification more frequently.[25] Lastly, compare

[19] Arraiz et al. (2010) argue that one can think of the locations of the states in the US. The states in the northeastern part of the US are closer to each other and have more neighbors, compared to the western states.

[20] In the experiment we use the RGMM method with identity matrix as the weight matrix to estimate the SAR model.

[21] The IV matrix used is still $Q_n = [l_n, X_{2n}, W_nX_{2n}, W_n^2X_{2n}, W_n^3X_{2n}]$. For the FORGMM method, we use Q_n as the IV matrix for the linear moments, and W_n, $W_n^2 - \frac{1}{n}tr\left(W_n^2\right)I_n$ for the quadratic moments.

[22] For more details, see MacKinnon (2009).

[23] If the null model is the MESS model, then the resampling scheme is similar.

[24] We cannot use the residual bootstrap when the error terms are independent but with unknown heteroskedasticity. To simulate the wild bootstrap error terms, we multiply the rescaled residuals by some random variable with mean 0 and variance 1. So the wild bootstrap error terms will have about the same variance as the true error terms. And the wild bootstrap dgp should capture the essential features of the true dgp. For more discussion, see MacKinnon (2009).

[25] For the J-test for various W_n's in Kelejian and Piras (2011), they have a similar conclusion.

Table 2
Size and power of the J-test statistics under H_0: $Y_n = \lambda W_n Y_n + l_n\beta_1 + X_{2n}\beta_2 + V_n$.

	n	$x=\chi^2(3)$				$x=U(0, 10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{sls}$	100	0.057	0.053	0.279	0.382	0.054	0.052	0.376	0.507
	300	0.05	0.047	0.742	0.866	0.054	0.045	0.855	0.951
	500	0.066	0.066	0.898	0.975	0.054	0.052	0.96	0.998
	700	0.042	0.048	0.962	0.996	0.053	0.065	0.987	1
$\mathcal{W}_{ogmm}$	100	0.093	0.238	0.405	0.664	0.094	0.234	0.432	0.691
	300	0.062	0.149	0.577	0.996	0.059	0.181	0.507	1
	500	0.058	0.114	0.662	1	0.048	0.13	0.51	1
	700	0.052	0.093	0.727	1	0.059	0.113	0.588	1
$\mathcal{G}_{sls}$	100	0.057	0.053	0.279	0.382	0.054	0.052	0.376	0.507
	300	0.05	0.047	0.742	0.866	0.054	0.045	0.855	0.955
	500	0.066	0.066	0.898	0.975	0.054	0.052	0.96	0.998
	700	0.042	0.048	0.962	0.996	0.053	0.065	0.987	1
$\mathcal{G}_{ogmm}$	100	0.046	0.044	0.15	0.686	0.046	0.056	0.162	0.803
	300	0.045	0.047	0.281	0.997	0.041	0.042	0.238	1
	500	0.048	0.054	0.322	1	0.042	0.053	0.294	1
	700	0.056	0.04	0.408	1	0.046	0.059	0.332	1
$\mathcal{DD}_{sls}$	100	0.057	0.053	0.279	0.382	0.054	0.052	0.376	0.507
	300	0.05	0.047	0.742	0.866	0.054	0.045	0.855	0.951
	500	0.066	0.066	0.898	0.975	0.054	0.052	0.96	0.998
	700	0.042	0.048	0.962	0.996	0.053	0.065	0.987	1
$\mathcal{DD}_{ogmm}$	100	0.056	0.055	0.299	0.801	0.051	0.061	0.278	0.876
	300	0.049	0.047	0.523	0.998	0.045	0.041	0.426	1
	500	0.048	0.056	0.63	1	0.041	0.053	0.459	1
	700	0.055	0.043	0.708	1	0.047	0.059	0.513	1

The SAR model: $\lambda_0=0.6$, $\beta_{10}=2$, $\beta_{20}=0.5$, and $\sigma_0=\sqrt{2}$.
The MESS model: $\mu_0=-1.6094$, $\beta_{10}^{ex}=2$, $\beta_{20}^{ex}=0.5$, and $\sigma_0^{ex}=\sqrt{2}$.

Table 1 with Table 2, all of the test statistics are more powerful when the variation in the exogenous regressor X_{2n} is dominant.

Tables 3 and 4 provide the sizes and powers of the J-test statistics when the null model is the SAR model with parameter values P-V3 and P-V4. The empirical sizes of all test statistics are reasonable except that there are still some over-rejections for the Wald test statistics based upon the FOGMM method. More importantly, compared with Tables 1 and 2, the empirical powers of all statistics are smaller

Table 3
Size and power of the J-test statistics under H_0: $Y_n = \lambda W_n Y_n + l_n\beta_1 + X_{2n}\beta_2 + V_n$.

	n	$x=\chi^2(3)$				$x=U(0, 10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{sls}$	100	0.058	0.051	0.063	0.078	0.051	0.055	0.061	0.083
	300	0.04	0.046	0.137	0.197	0.042	0.047	0.167	0.21
	500	0.055	0.06	0.235	0.295	0.045	0.049	0.323	0.371
	700	0.049	0.051	0.331	0.381	0.064	0.062	0.457	0.494
$\mathcal{W}_{ogmm}$	100	0.165	0.341	0.234	0.223	0.162	0.299	0.211	0.205
	300	0.098	0.185	0.147	0.171	0.094	0.189	0.161	0.118
	500	0.086	0.14	0.131	0.131	0.084	0.115	0.163	0.104
	700	0.066	0.107	0.112	0.107	0.073	0.116	0.143	0.117
$\mathcal{G}_{sls}$	100	0.058	0.051	0.063	0.078	0.051	0.055	0.061	0.083
	300	0.04	0.046	0.137	0.197	0.042	0.047	0.167	0.21
	500	0.055	0.06	0.235	0.295	0.045	0.049	0.323	0.371
	700	0.049	0.051	0.331	0.381	0.064	0.062	0.457	0.494
$\mathcal{G}_{ogmm}$	100	0.051	0.041	0.06	0.071	0.043	0.048	0.043	0.079
	300	0.045	0.044	0.052	0.218	0.05	0.047	0.062	0.239
	500	0.052	0.065	0.066	0.325	0.046	0.058	0.067	0.407
	700	0.042	0.042	0.054	0.465	0.048	0.065	0.067	0.536
$\mathcal{DD}_{sls}$	100	0.058	0.051	0.063	0.078	0.051	0.055	0.061	0.083
	300	0.04	0.046	0.137	0.197	0.042	0.047	0.167	0.21
	500	0.055	0.06	0.235	0.295	0.045	0.049	0.323	0.371
	700	0.049	0.051	0.331	0.381	0.064	0.062	0.457	0.494
$\mathcal{DD}_{ogmm}$	100	0.05	0.043	0.049	0.065	0.044	0.049	0.045	0.078
	300	0.046	0.042	0.044	0.219	0.052	0.05	0.06	0.235
	500	0.053	0.061	0.067	0.323	0.046	0.057	0.067	0.409
	700	0.041	0.042	0.052	0.467	0.05	0.063	0.067	0.537

The SAR model: $\lambda_0=0.4$, $\beta_{10}=2$, $\beta_{20}=1$, and $\sigma_0=1$.
The MESS model: $\mu_0=-0.5108$, $\beta_{10}^{ex}=2$, $\beta_{20}^{ex}=1$, and $\sigma_0^{ex}=1$.

Table 4
Size and power of the J-test statistics under H_0: $Y_n = \lambda W_n Y_n + l_n\beta_1 + X_{2n}\beta_2 + V_n$.

	n	$x=\chi^2(3)$				$x=U(0, 10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{sls}$	100	0.061	0.051	0.06	0.041	0.06	0.051	0.054	0.052
	300	0.059	0.044	0.069	0.063	0.058	0.045	0.061	0.059
	500	0.068	0.053	0.077	0.08	0.054	0.046	0.064	0.082
	700	0.055	0.05	0.068	0.08	0.056	0.066	0.071	0.106
$\mathcal{W}_{ogmm}$	100	0.071	0.182	0.077	0.298	0.076	0.174	0.081	0.282
	300	0.047	0.1	0.121	0.216	0.056	0.101	0.131	0.205
	500	0.051	0.091	0.18	0.186	0.048	0.079	0.157	0.175
	700	0.043	0.058	0.156	0.55	0.055	0.058	0.167	0.171
$\mathcal{G}_{sls}$	100	0.061	0.051	0.06	0.041	0.06	0.051	0.054	0.052
	300	0.059	0.044	0.069	0.063	0.058	0.045	0.061	0.059
	500	0.068	0.053	0.077	0.08	0.054	0.046	0.064	0.082
	700	0.055	0.05	0.068	0.08	0.056	0.066	0.071	0.106
$\mathcal{G}_{ogmm}$	100	0.052	0.042	0.045	0.03	0.049	0.049	0.042	0.04
	300	0.045	0.047	0.039	0.094	0.046	0.041	0.035	0.077
	500	0.047	0.056	0.044	0.127	0.042	0.053	0.036	0.118
	700	0.055	0.04	0.043	0.159	0.045	0.063	0.043	0.197
$\mathcal{DD}_{sls}$	100	0.061	0.051	0.06	0.041	0.06	0.051	0.054	0.052
	300	0.059	0.044	0.069	0.063	0.058	0.045	0.061	0.059
	500	0.068	0.053	0.077	0.08	0.054	0.046	0.064	0.082
	700	0.055	0.05	0.068	0.08	0.056	0.066	0.071	0.106
$\mathcal{DD}_{ogmm}$	100	0.051	0.035	0.046	0.038	0.048	0.036	0.044	0.045
	300	0.047	0.032	0.042	0.09	0.047	0.021	0.036	0.076
	500	0.046	0.038	0.044	0.122	0.04	0.029	0.036	0.121
	700	0.054	0.022	0.045	0.157	0.045	0.029	0.043	0.192

The SAR model: $\lambda_0=0.4$, $\beta_{10}=2$, $\beta_{20}=0.5$, and $\sigma_0=\sqrt{2}$.
The MESS model: $\mu_0=-0.5108$, $\beta_{10}^{ex}=2$, $\beta_{20}^{ex}=0.5$, and $\sigma_0^{ex}=\sqrt{2}$.

when we only have moderate spatial dependence for λ and μ. The situation becomes obvious in Table 4, with larger $\sigma=\sigma^{ex}=\sqrt{2}$. All three statistics do not have enough power, which suggests that it is more difficult to distinguish between the SAR and the MESS models with moderate spatial dependence and large variance in disturbances.

Tables 5 and 6 provide the sizes and powers of the J-test statistics when the null model is the MESS model with parameter values P-V5 and P-V6. The empirical sizes and powers of the G test statistics and the DD test statistics are reasonable for the two sets of parameter values. However, there are serious over-rejections for the Wald test statistics with P-V6, in which we only have moderate spatial dependence.

Tables 7 and 8 summarize the sizes and powers of the J-test statistics using the SAR model with unknown heteroskedasticity as the null. With $\lambda_0=0.6$ and $\mu_0=-1.6094$, all of the empirical sizes and powers seem reasonable except that the test statistics based upon the FORGMM method using the first predictor do not have enough power. With $\lambda_0=0.4$ and $\mu_0=-0.5108$, the empirical powers of all test statistics decrease. Thus it is again harder to distinguish between

Table 5
Size and power of the J-test statistics under H_0: $S_n^{ex}(\mu)Y_n = l_n\beta_1^{ex} + X_{2n}\beta_2^{ex} + V_n$.

	n	$x=\chi^2(3)$				$x=U(0, 10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{sls}$	100	0.073	0.089	0.571	0.267	0.056	0.075	0.712	0.198
	300	0.052	0.066	0.887	0.238	0.056	0.046	0.954	0.232
	500	0.073	0.062	0.954	0.22	0.054	0.054	0.985	0.228
	700	0.047	0.047	0.987	0.182	0.052	0.055	0.995	0.198
$\mathcal{G}_{sls}$	100	0.054	0.052	0.336	0.331	0.05	0.049	0.336	0.331
	300	0.039	0.036	0.674	0.666	0.054	0.055	0.674	0.666
	500	0.071	0.06	0.868	0.867	0.054	0.047	0.868	0.867
	700	0.044	0.048	0.961	0.963	0.065	0.065	0.988	0.986
$\mathcal{DD}_{sls}$	100	0.053	0.048	0.199	0.257	0.052	0.049	0.348	0.428
	300	0.037	0.036	0.613	0.648	0.05	0.052	0.786	0.825
	500	0.07	0.057	0.817	0.845	0.055	0.045	0.932	0.945
	700	0.044	0.049	0.92	0.944	0.047	0.048	0.971	0.975

The SAR model: $\lambda_0=0.6$, $\beta_{10}=2$, $\beta_{20}=1$, and $\sigma_0=1$.
The MESS model: $\mu_0=-1.6094$, $\beta_{10}^{ex}=2$, $\beta_{20}^{ex}=1$, and $\sigma_0^{ex}=1$.

Table 6
Size and power of the J-test statistics under H_0: $S_n^{ex}(\mu)Y_n = l_n\beta_1^{ex} + X_{2n}\beta_2^{ex} + V_n$.

	n	$x=\chi^2(3)$				$x=U(0, 10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{sls}$	100	0.044	0.373	0.082	0.389	0.058	0.326	0.087	0.398
	300	0.098	0.266	0.317	0.243	0.11	0.199	0.482	0.255
	500	0.1	0.196	0.643	0.187	0.1	0.138	0.766	0.228
	700	0.085	0.131	0.786	0.148	0.092	0.129	0.849	0.183
$\mathcal{G}_{sls}$	100	0.051	0.052	0.112	0.116	0.049	0.049	0.269	0.207
	300	0.037	0.041	0.153	0.146	0.052	0.055	0.264	0.263
	500	0.06	0.053	0.219	0.218	0.045	0.05	0.331	0.326
	700	0.045	0.046	0.252	0.256	0.064	0.065	0.387	0.384
$\mathcal{DD}_{sls}$	100	0.026	0.023	0.059	0.055	0.028	0.021	0.273	0.273
	300	0.028	0.036	0.044	0.057	0.037	0.042	0.284	0.312
	500	0.043	0.043	0.083	0.144	0.042	0.048	0.339	0.388
	700	0.037	0.043	0.162	0.215	0.06	0.063	0.439	0.48

The SAR model: $\lambda_0 = 0.4$, $\beta_{10} = 2$, $\beta_{20} = 1$, and $\sigma_0 = 1$.
The MESS model: $\mu_0 = -0.5108$, $\beta_{10}^{ex} = 2$, $\beta_{20}^{ex} = 1$, and $\sigma_0^{ex} = 1$.

Table 7
Size and power of J-test statistics with unknown heteroskedasticity under H_0: $Y_n = \lambda W_n Y_n + l_n\beta_1 + X_{2n}\beta_2 + V_n$.

	$L, \overline{L}$	$x=\chi^2(3)$				$x=U(0, 10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{rsls}$	$L=5, \overline{L}=15$	0.049	0.045	0.998	1	0.055	0.048	1	1
	$L=14, \overline{L}=20$	0.047	0.048	0.999	1	0.047	0.044	1	1
$\mathcal{W}_{orgmm}$	$L=5, \overline{L}=15$	0.053	0.054	0.228	1	0.058	0.072	0.159	1
	$L=14, \overline{L}=20$	0.063	0.056	0.173	1	0.076	0.037	0.182	1
$\mathcal{G}_{rsls}$	$L=5, \overline{L}=15$	0.049	0.045	0.998	1	0.055	0.048	1	1
	$L=14, \overline{L}=20$	0.047	0.048	0.999	1	0.047	0.044	1	1
$\mathcal{G}_{orgmm}$	$L=5, \overline{L}=15$	0.044	0.051	0.247	1	0.042	0.051	0.836	1
	$L=14, \overline{L}=20$	0.041	0.051	0.243	1	0.056	0.046	0.818	1
$\mathcal{DD}_{rsls}$	$L=5, \overline{L}=15$	0.049	0.045	0.998	1	0.055	0.048	1	1
	$L=14, \overline{L}=20$	0.047	0.048	0.999	1	0.047	0.044	1	1
$\mathcal{DD}_{orgmm}$	$L=5, \overline{L}=15$	0.046	0.049	0.214	1	0.043	0.053	0.489	1
	$L=14, \overline{L}=20$	0.044	0.062	0.17	1	0.059	0.049	0.514	1

The SAR model: $\lambda_0 = 0.6$, $\beta_{10} = 2$, and $\beta_{20} = 1$.
The MESS model: $\mu_0 = -1.6094$, $\beta_{10}^{ex} = 2$, and $\beta_{20}^{ex} = 1$.
Sample size is 545 for $L = 5, \bar{L} = 15$. Sample size is 520 for $L = 14, \bar{L} = 20$.
$\mathcal{W}_{rsls}$: the Wald test statistic based on the Robust 2SLS method.
$\mathcal{G}_{rsls}$: the G test statistic based on the Robust 2SLS method.
$\mathcal{DD}_{rsls}$: the DD test statistic based on the Robust 2SLS method.
$\mathcal{W}_{orgmm}$: the Wald test statistic based on the FORGMM method.
$\mathcal{G}_{orgmm}$: the G test statistic based on the FORGMM method.
$\mathcal{DD}_{orgmm}$: the DD test statistic based on the FORGMM method.

the SAR model and the MESS model unless we have strong spatial interaction effect. Also there are some size distortions for the Wald test statistics based on the FORGMM method.

Tables 9 and 10 summarize the sizes and powers of the J-test statistics using the MESS model with unknown heteroskedasticity as the null. Otherwise, the sizes and powers of all test statistics are reasonable in Table 9. But the empirical power decreases when we only have moderate spatial dependence in Table 10. Specifically, the DD test statistics do not have power. There are over-rejections for the Wald test statistics using the second predictor in Table 10.

Tables 11–14 provide the sizes and powers of the bootstrapped Wald test statistics based upon the FOGMM method, using the SAR model as the null. All of the bootstrapped test statistics have empirical sizes much closer to the nominal 5% level than the asymptotic test statistics in the previous tables and they seem to have a better control over size for different parameter values.[26] However, they do not have enough empirical power with moderate spatial dependence.

Table 15 provides the sizes and powers of the bootstrapped Wald test statistics based on the N2SLS method, using the MESS model as the null, with parameter values P-V6. All the empirical sizes of test statistics are closer to the nominal 5% level than that of the asymptotic test statistics. For instance, when the sample size is 100, the empirical size of the Wald test statistic from the second predictor is 0.05, compared with a size of 0.373 of the asymptotic Wald test statistic in Table 6. Again those bootstrapped Wald test statistics do not have enough power.

Table 16 summarizes the sizes and powers of the bootstrapped Wald test statistics based on the FORGMM method, using the SAR model with unknown heteroskedasticity as the null. With $\lambda_0 = 0.4$ and $\mu_0 = -0.5108$, the sizes of test statistics are closer to the nominal 5% level than the asymptotic test statistics in Table 8. But the empirical powers are not strong, especially for the Wald test statistics from the first predictor.

Finally, Table 17 summarizes the sizes and powers of the bootstrapped Wald test statistics based on the GN2SLS method, using the MESS model with unknown heteroskedasticity as the null. With $\lambda_0 = 0.4$ and $\mu_0 = -0.5108$, the sizes of test statistics are closer to the nominal 5% level than the asymptotic test statistics in Table 10. But the empirical powers are not strong.

6. Conclusion

Empirical researchers in spatial studies frequently utilize the SAR model, which implies a geometrical decay pattern of spatial externalities. On the other hand, the MESS model has an exponential decay pattern. In this paper, we consider using the J-test procedure for the selection between the SAR model and the MESS model. We construct J-test statistics by using both the 2SLS method as well as the extended GMM method in Lee (2007). We derive several test statistics under the GMM framework. In addition to the testing procedures, we investigate the behavior of those J-test statistics in terms of pseudo true values, which provide a clearer view of the augmented variables used in testing. We also extend the J-test procedure into the setting when error terms in the model are independent but with unknown heteroskedasticity. We have also used bootstrapped procedures to compare with those based on conventional asymptotic distributions of the test statistics. From our Monte Carlo results, we can conclude that when the spatial dependence is strong and the sample size is not small, the J-test statistics can have good power to distinguish the SAR and MESS models.

One limitation of this paper is that we rely on setting W_n in the MESS model to be a conventional spatial weight matrix without any unknown parameter. LeSage and Pace (2009) has considered a more flexible extension of the MESS model, in which $W_n = \sum_{i=1}^{p}\left(\frac{\phi^i N_i}{\sum_{i=1}^{p}\phi^i}\right)$. Here p is the number of nearest neighbors and $0<\phi<1$ represents a decay factor applied to each of the nearest neighbor weight matrices N_i, which is an $n \times n$ weight matrix consisting non-zero elements for the ith closest neighbor. Both p and ϕ are unknown parameters that must be estimated. As suggested by LeSage and Pace (2009), this weight matrix would be flexible enough to approximate more conventional spatial weight matrices.[27] Therefore it would be desirable to consider the model selection problem between the SAR model and that extension of the MESS model. LeSage and Pace (2009) consider Bayesian MCMC estimation to produce estimates of parameters of the extension of the MESS model. Thus we could follow Bayesian model comparison procedures in Zellner (1971) in principle to calculate and compare the posterior probabilities of the SAR model and the extended MESS model. It would also be a promising research to consider classical estimation methods for that extension of the MESS model and derive J-test procedures for this model selection problem.

[26] For the bootstrap J-test for various W_n's in Burridge and Fingleton (2010), they have similar results.

[27] As pointed out by a referee, the more flexible specification would make it possible for the MESS exponential decay specification to more closely replicate the SAR geometric pattern of decay.

Table 8
Size and power of J-test statistics with unknown heteroskedasticity under H_0: $Y_n=\lambda W_nY_n+l_n\beta_1+X_{2n}\beta_2+V_n$.

	$L, \overline{L}$	$x=\chi^2(3)$				$x=U(0,10)$			
		Size		Power		Size		Power	
		$Y_{n\vert 1}$	$Y_{n\vert 2}$	$Y_{n\vert 1}$	$Y_{n\vert 2}$	$Y_{n\vert 1}$	$Y_{n\vert 2}$	$Y_{n\vert 1}$	$Y_{n\vert 2}$
$\mathcal{W}_{rsls}$	$L=5, \overline{L}=15$	0.04	0.044	0.124	0.158	0.049	0.053	0.182	0.219
	$L=14, \overline{L}=20$	0.049	0.049	0.129	0.17	0.045	0.05	0.177	0.205
$\mathcal{W}_{orgmm}$	$L=5, \overline{L}=15$	0.047	0.093	0.108	0.33	0.06	0.134	0.119	0.339
	$L=14, \overline{L}=20$	0.059	0.147	0.09	0.402	0.076	0.111	0.122	0.347
$\mathcal{G}_{rsls}$	$L=5, \overline{L}=15$	0.04	0.044	0.124	0.158	0.049	0.053	0.182	0.219
	$L=14, \overline{L}=20$	0.049	0.049	0.129	0.17	0.045	0.05	0.177	0.205
$\mathcal{G}_{orgmm}$	$L=5, \overline{L}=15$	0.045	0.047	0.268	0.547	0.043	0.052	0.139	0.514
	$L=14, \overline{L}=20$	0.037	0.058	0.233	0.564	0.048	0.043	0.154	0.518
$\mathcal{DD}_{rsls}$	$L=5, \overline{L}=15$	0.04	0.044	0.124	0.158	0.049	0.053	0.182	0.219
	$L=14, \overline{L}=20$	0.049	0.049	0.129	0.17	0.045	0.05	0.177	0.205
$\mathcal{DD}_{orgmm}$	$L=5, \overline{L}=15$	0.049	0.041	0.27	0.551	0.044	0.053	0.15	0.528
	$L=14, \overline{L}=20$	0.04	0.068	0.236	0.582	0.056	0.049	0.167	0.529

The SAR model: $\lambda_0=0.4$, $\beta_{10}=2$, and $\beta_{20}=1$.
The MESS model: $\mu_0=-0.5108$, $\beta_{10}^{ex}=2$, and $\beta_{20}^{ex}=1$.
Sample size is 545 for $L=5, \bar{L}=15$. Sample size is 520 for $L=14, \bar{L}=20$.
$\mathcal{W}_{rsls}$: the Wald test statistic based on the Robust 2SLS method.
$\mathcal{G}_{rsls}$: the G test statistic based on the Robust 2SLS method.
$\mathcal{DD}_{rsls}$: the DD test statistic based on the Robust 2SLS method.
$\mathcal{W}_{orgmm}$: the Wald test statistic based on the FORGMM method.
$\mathcal{G}_{orgmm}$: the G test statistic based on the FORGMM method.
$\mathcal{DD}_{orgmm}$: the DD test statistic based on the FORGMM method.

Appendix A. Some useful lemma

Lemmas A.1–A.6 are some basic results on quadratic form, law of large numbers, and central limit theorems, which are useful for our analysis. They can be found, e.g., in Kelejian and Prucha (1998) and/or Lee (2007).

Lemma A.1. *Suppose elements of the sequences of n-dimensional column vectors $\{Z_{1n}\}$ and $\{Z_{2n}\}$ are uniformly bounded. If $\{A_n\}$ are uniformly bounded in either row or column sums in absolute value, then $|Z'_{1n}A_nZ_{2n}|=O(n)$.*

Lemma A.2. *Suppose that $\{A_n\}$ are uniformly bounded in both row and column sums in absolute value. The $v_{n1}, \cdots, v_{nn}$ are i.i.d with zero mean and its fourth moment exists. Then, $E(V_n'A_nV_n)=O(n)$, $var(V_n'A_nV_n)=O(n)$, $V_n'A_nV_n=O_p(n)$, and $\frac{1}{n}V'_nA_nV_n-\frac{1}{n}EV'_nA_nV_n=o_p(1)$.*

Lemma A.3. *Suppose that $v_{n1}, \ldots, v_{nn}$ are i.i.d random variables with zero mean, finite variance σ^2 and finite fourth moment μ_4. Then, for any two square $n\times n$ matrices A and B,*

$$E(V_n'AV_nV_n'BV_n)=(\mu_4-3\sigma_0^4)vec_D'(A)vec_D(B)+\sigma_0^4\left[tr(A)tr(B)+tr\left(A(B+B')\right)\right].$$

Lemma A.4. *Suppose that the elements of the $n\times k$ matrices X_n are uniformly bounded for all n; and $\lim_{n\to\infty}\left(\frac{1}{n}X'_nX_n\right)$ exists and is nonsingular, then the projectors, $X_n(X_n'X_n)^{-1}X_n'$ and $I_n-X_n(X_n'X_n)^{-1}X_n'$, are uniformly bounded in both row and column sum norms.*

Lemma A.5. *Suppose that $\{A_n\}$ is a sequence of $n\times n$ matrices with its column sums being uniformly bounded in absolute value, elements of the $n\times k$ matrix C_n are uniformly bounded, and $v_{n1}, \cdots, v_{nn}$ are i.i.d with zero mean and finite variance σ^2. Then, $\frac{1}{\sqrt{n}}C_n'A_nV_n=O_p(1)$ and $\frac{1}{n}C_n'A_nV_n=o_p(1)$. Furthermore, if the limit of $\frac{1}{n}C_n'A_nA'_nC_n$ exists and is positive definite, then $\frac{1}{\sqrt{n}}C_n'A_nV_n\xrightarrow{D}N\left(0,\sigma_0^2\lim_{n\to\infty}\frac{1}{n}C_n'A_nA_n'C_n\right)$.*

Lemma A.6. *Suppose that $\{A_n\}$ is a sequence of symmetric $n\times n$ matrices with row and column sums uniformly bounded in absolute value and $b_n=(b_{n1}, \ldots, b_{nn})'$ is a n-dimensional vector such that $\sup_n\frac{1}{n}\sum_{i=1}^n|b_{ni}|^{2+\eta_1}<\infty$ for some $\eta_1>0$. The $V_{n1}, \ldots, V_{nn}$ are i.i.d random variables with zeros mean and finite variance σ^2, and its moment $E(|V|^{4+2\delta})$ for some $\delta>0$ exists. Let $\sigma_{Q_n}^2$ be the variance of Q_n where $Q_n=V_n'A_nV_n+b_n'V_n-\sigma^2tr(A_n)$. Assume that the variance $\sigma_{Q_n}^2$ is bounded away from zero at the rate n. Then, $\left(\frac{Q_n}{\sigma_{Q_n}}\right)\xrightarrow{D}N(0,1)$.*

Lemmas A.7–A.11 are from, for example, Lin and Lee (2010) and also Kelejian and Prucha (2010).

Lemma A.7. *Assume that v_{ni}'s have zero mean and finite variances, and are mutually independent. Let $A_n=(a_{n,ij})$ and $B_n=(b_{n,ij})$ be two square matrices of dimension n. Then, $E(A_nV_n(B_nV_n)')=A_n\Sigma_nB_n'$. If the diagonal of B_n is zero, $E(A_nV_nV_n'B_nV_n)=0$. Furthermore, if both A_n and B_n have zero diagonals,*

$$E\left(V_n'A_nV_nV_n'B_nV_n\right)=\sum_{i=1}^n\sum_{j=1}^n a_{n,ij}\left(b_{n,ij}+b_{n,ji}\right)\sigma_{ni}^2\sigma_{nj}^2=tr\left[\sum_nA_n\sum_n(B_n'+B_n)\right],$$

where $\Sigma_n=Diag(\sigma_{n1}^2, \ldots, \sigma_{nn}^2)$ with $\sigma_{ni}^2=E(v_{ni}^2)$ and $V_n=(v_{n1}, \ldots, v_{nn})'$.

Lemma A.8. *For any square matrices $A_n=[a_{n,ij}]$ of dimension n, assume that v_{ni}'s have a zero mean and are mutually independent. Then*

$$E(V_n'A_nV_n)=\sum_{i=1}^n a_{n,ii}\sigma_{ni}^2=tr(\sum_nA_n),$$

$$Var(V_n'A_nV_n)=\sum_{i=1}^n a_{n,ii}^2\left[E(v_{ni}^4)-3\sigma_{ni}^4\right]+\sum_{i=1}^n\sum_{j=1}^n a_{n,ij}\left(a_{n,ij}+a_{n,ji}\right)\sigma_{ni}^2\sigma_{nj}^2=\sum_{i=1}^n a_{n,ii}^2\left[E(v_{ni}^4)-3\sigma_{ni}^4\right]+tr\left[\sum_nA_n\sum_n\left(A_n'+A_n\right)\right];$$

where $\Sigma_n=Diag(\sigma_{n1}^2, \ldots, \sigma_{nn}^2)$ with $\sigma_{ni}^2=E(v_{ni}^2)$ and $V_n=(v_{n1}, \ldots, v_{nn})'$.

Lemma A.9. *Suppose $\{A_n\}$ are uniformly bounded in both row and column sums in absolute value and v_{ni}'s have a zero mean and are mutually independent where its sequence of variance $\{\sigma_{ni}^2\}$ is bounded, and, in addition, if $a_{n,ii}\neq 0$ for some i, the sequence of fourth moments $\{\mu_{ni,4}\}$ is bounded. Then, $E(V_n'A_nV_n)=O(n)$, $Var(V_n'A_nV_n)=O(n)$, $V_n'A_nV_n=O_p(n)$ and $\frac{1}{n}V_n'A_nV_n-\frac{1}{n}E(V_n'A_nV_n)=o_p(1)$.*

Table 9
Size and power of J-test statistics with unknown heteroskedasticity under H_0: $S_n^{ex}(\mu)Y_n = l_n\beta_1^{ex} + X_{2n}\beta_2^{ex} + V_n$.

	$L, \bar{L}$	$x=\chi^2(3)$				$x=U(0,10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{rsls}$	$L=5, \bar{L}=15$	0.063	0.073	0.77	0.391	0.052	0.064	0.829	0.392
	$L=14, \bar{L}=20$	0.057	0.07	0.741	0.352	0.059	0.061	0.796	0.378
$\mathcal{G}_{rsls}$	$L=5, \bar{L}=15$	0.064	0.053	0.482	0.484	0.052	0.051	0.606	0.619
	$L=14, \bar{L}=20$	0.047	0.056	0.439	0.449	0.056	0.05	0.589	0.599
$\mathcal{DD}_{rsls}$	$L=5, \bar{L}=15$	0.065	0.047	0.452	0.49	0.051	0.052	0.603	0.618
	$L=14, \bar{L}=20$	0.047	0.052	0.407	0.446	0.058	0.047	0.577	0.611

The SAR model: $\lambda_0=0.6$, $\beta_{10}=2$, and $\beta_{20}=1$.
The MESS model: $\mu_0=-1.6094$, $\beta_{10}^{ex}=2$, and $\beta_{20}^{ex}=1$.
Sample size is 545 for $L=5, \bar{L}=15$. Sample size is 520 for $L=14, \bar{L}=20$.
$\mathcal{W}_{rsls}$: the Wald test statistic based on the GN2SLS method.
$\mathcal{G}_{rsls}$: the G test statistic based on the GN2SLS method.
$\mathcal{DD}_{rsls}$: the DD test statistic based on the GN2SLS method.

Table 10
Size and power of J-test statistics with unknown heteroskedasticity under H_0: $S_n^{ex}(\mu)Y_n = l_n\beta_1^{ex} + X_{2n}\beta_2^{ex} + V_n$.

	$L, \bar{L}$	$x=\chi^2(3)$				$x=U(0,10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{rsls}$	$L=5, \bar{L}=15$	0.118	0.171	0.268	0.386	0.087	0.153	0.346	0.387
	$L=14, \bar{L}=20$	0.107	0.183	0.267	0.323	0.131	0.149	0.13	0.141
$\mathcal{G}_{rsls}$	$L=5, \bar{L}=15$	0.052	0.051	0.164	0.174	0.052	0.048	0.221	0.245
	$L=14, \bar{L}=20$	0.054	0.054	0.13	0.141	0.053	0.053	0.236	0.256
$\mathcal{DD}_{rsls}$	$L=5, \bar{L}=15$	0.041	0.034	0.08	0.084	0.033	0.03	0.125	0.146
	$L=14, \bar{L}=20$	0.038	0.037	0.062	0.069	0.044	0.043	0.147	0.17

The SAR model: $\lambda_0=0.4$, $\beta_{10}=2$, and $\beta_{20}=1$.
The MESS model: $\mu_0=-0.5108$, $\beta_{10}^{ex}=2$, and $\beta_{20}^{ex}=1$.
Sample size is 545 for $L=5, \bar{L}=15$. Sample size is 520 for $L=14, \bar{L}=20$.
$\mathcal{W}_{rsls}$: the Wald test statistic based on the GN2SLS method.
$\mathcal{G}_{rsls}$: the G test statistic based on the GN2SLS method.
$\mathcal{DD}_{rsls}$: the DD test statistic based on the GN2SLS method.

Lemma A.10. *Suppose that A_n is an $n\times n$ matrix with its column sums being uniformly bounded in absolute value, elements of the $n\times k$ matrix C_n are uniformly bounded, and elements v_{ni} of $V_n=(v_{n1}, \ldots, v_{nn})'$ are mutually independent with zero mean and finite third absolute moments, which are uniformly bounded for all n and i.*

Then, $\frac{1}{\sqrt{n}}C_n'A_nV_n = O_p(1)$ and $\frac{1}{n}C_n'A_nV_n = o_p(1)$. Furthermore, if the limit of $\frac{1}{n}C_n'A_n\sum_nA_n'C_n$ exists and is positive definite, then $\frac{1}{\sqrt{n}}C_n'A_nV_n \xrightarrow{D} N\left(0, \lim_{n\to\infty}\frac{1}{n}C_n'A_n\sum_nA_n'C_n\right)$.

Lemma A.11. *Suppose that $\{A_n\}$ is a sequence of symmetric $n\times n$ matrices with row and column sums uniformly bounded in absolute value and $b_n=[b_{ni}]$ is a n-dimensional column vector such that $\sup_n\frac{1}{n}\sum_{i=1}^n|b_{ni}|^{2+\varsigma_1}<\infty$ for some $\varsigma_1>0$. Furthermore, $v_{n1}, \ldots, v_{nn}$ are mutually independent, with zero mean and moments higher than four exist such that $E(|v_{ni}|^{4+\varsigma_2})$ for some $\varsigma_2>0$ are uniformly bounded for all n and i.*

Let $\sigma_{Q_n}^2$ be the variance of Q_n where $Q_n=V_n'A_nV_n+b_n'V_n-tr(A_n\Sigma_n)$. Assume that $\frac{1}{n}\sigma_{Q_n}^2$ is bounded away from zero. Then, $\frac{Q_n}{\sigma_{Q_n}}\xrightarrow{D}N(0,1)$.

Appendix B. Pseudo true values of $\hat{\theta}_n^{ex}$ based upon the QML method

Let $\hat{\theta}_n^{ex}$ be the quasi-maximum likelihood (QML) estimate of θ^{ex} for the MESS model. Based on the normal distribution, the log quasi-likelihood function of the MESS model[28] is:

$$L_n(\theta^{ex}) = -\frac{n}{2}\ln 2\pi\sigma^{ex2} - \frac{1}{2\sigma^{ex2}}\left(S_n^{ex}(\mu)Y_n - X_n\beta^{ex}\right)'\left(S_n^{ex}(\mu)Y_n - X_n\beta^{ex}\right). \tag{B.1}$$

For simplicity, denote $E_{|sar}(L_n(\theta^{ex}))\equiv E(L_n(\theta^{ex})|H_0)$, the expectation of the above equation under the null SAR model, which is

$$E_{|sar}(L_n(\theta^{ex})) = -\frac{n}{2}\ln 2\pi\sigma^{ex2} - \frac{1}{2\sigma^{ex2}}E_{|sar}\left[\left(S_n^{ex}(\mu)Y_n - X_n\beta^{ex}\right)'\left(S_n^{ex}(\mu)Y_n - X_n\beta^{ex}\right)\right]. \tag{B.2}$$

With a sample of size n, the pseudo-true value $\theta_{n|sar}^{ex*}$ is defined as

$$\theta_{n|sar}^{ex*} = \arg\max_{\theta^{ex}} E_{|sar}\left(L_n(\theta^{ex})\right). \tag{B.3}$$

For the MESS model, some components of $\theta_{n|sar}^{ex*}$ have simple expressions that can be reviewed from the concentrated expected function from Eq. (B.2). Note that the concentrated likelihood function of Eq. (B.1) in terms of μ is

$$L_n(\mu) = -\frac{n}{2}(\ln 2\pi + 1) - \frac{n}{2}\ln\sigma_n^{ex2}(\mu), \tag{B.4}$$

with

$$\beta_n^{ex}(\mu) = (X_n'X_n)^{-1}X_n'S_n^{ex}(\mu)Y_n,$$
$$\sigma_n^{ex2}(\mu) = \frac{1}{n}Y_n'S_n^{ex}(\mu)'M_nS_n^{ex}(\mu)Y_n,$$

where $M_n=I_n-X_n(X_n'X_n)^{-1}X_n'$. Correspondingly, the concentrated expected function for μ from Eq. (B.2) is $H_{n|sar}(\mu)=\max_{\beta^{ex},\sigma^{ex2}}E_{|sar}(L_n(\theta^{ex}))$. As

$$\begin{aligned}
&E_{|sar}\left[\left(S_n^{ex}(\mu)Y_n - X_n\beta^{ex}\right)'\left(S_n^{ex}(\mu)Y_n - X_n\beta^{ex}\right)\right] \\
&= \left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0 - X_n\beta^{ex}\right)'\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0 - X_n\beta^{ex}\right) \\
&\quad + \sigma_0^2 tr\left(S_n^{-1\prime}S_n^{ex}(\mu)'S_n^{ex}(\mu)S_n^{-1}\right).
\end{aligned}$$

One has $\beta_{n|sar}^{ex}(\mu)=(X_n'X_n)^{-1}X_n'S_n^{ex}(\mu)S_n^{-1}X_n\beta_0$ and

$$\begin{aligned}
\sigma_{n|sar}^{ex2}(\mu) &= \frac{1}{n}E_{|sar}\left[\left(S_n^{ex}(\mu)Y_n - X_n\beta_{n|sar}^{ex}(\mu)\right)'\left(S_n^{ex}(\mu)Y_n - X_n\beta_{n|sar}^{ex}(\mu)\right)\right] \\
&= \frac{1}{n}\left[\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right)'M_n\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right) + \sigma_0^2 tr\left(S_n^{-1\prime}S_n^{ex}(\mu)'S_n^{ex}(\mu)S_n^{-1}\right)\right].
\end{aligned} \tag{B.5}$$

Thus

$$H_{n|sar}(\mu) = \max_{\beta^{ex},\sigma^{ex2}}E_{|sar}\left(L_n(\theta^{ex})\right) = -\frac{n}{2}(\ln 2\pi + 1) - \frac{n}{2}\ln\sigma_{n|sar}^{ex2}(\mu). \tag{B.6}$$

Based on Eqs. (B.5) and (B.6), the pseudo-true values are $\mu_{n|sar}^* = argmax_\mu H_{n|sar}(\mu)$, $\beta_{n|sar}^{ex*}=\beta_{n|sar}^{ex}(\mu_{n|sar}^*)$, and $\sigma_{n|sar}^{ex2*}=\sigma_{n|sar}^{ex2}(\mu_{n|sar}^*)$.

Following White (1994), we can discuss the corresponding identification uniqueness condition of the parameters in the likelihood function (B.1) of the MESS model under the null SAR process in terms of $\mu_{n|sar}^*$. Let Θ_μ, a compact subset of R, be the parameter space of μ, and let $\Theta_{n\mu|sar}$ be a non-empty compact subset of Θ_μ for $n=1, 2, \ldots$. Suppose $H_{n|sar}(\mu)$ is maximized in $\Theta_{n\mu|sar}$, at $\mu_{n|sar}^*$ for $n=1, 2, \ldots$. Furthermore, let $S_{n\mu|sar}(\epsilon)$ be an open ball in R centered at $\mu_{n|sar}^*$ with a radius $\epsilon>0$. Define the neighborhood $N_{n\mu|sar}(\epsilon)=S_{n\mu|sar}(\epsilon)\cap\Theta_{n\mu|sar}$ and its complement $N_{n\mu|sar}^c(\epsilon)$. The

[28] Here $|S^{ex}(\mu)|=|\exp(\mu W_n)|=\exp(\text{trace}(\mu W_n))=1$ as W_n has a zero diagonal. See LeSage and Pace (2007) for more details.

Table 11
Bootstrap size and power of the J-test statistics under H_0: $Y_n=\lambda W_nY_n+l_n\beta_1+X_{2n}\beta_2+V_n$.

	n	$x=\chi^2(3)$				$x=U(0,10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{ogmm}$	100	0.045	0.051	0.333	0.517	0.045	0.039	0.376	0.499
	300	0.048	0.045	0.516	1	0.045	0.051	0.643	1
	500	0.058	0.04	0.593	1	0.053	0.036	0.761	1
	700	0.054	0.034	0.692	1	0.057	0.03	0.853	1

The SAR model: $\lambda_0=0.6$, $\beta_{10}=2$, $\beta_{20}=1$, and $\sigma_0=1$.
The MESS model: $\mu_0=-1.6094$, $\beta_{10}^{ex}=2$, $\beta_{20}^{ex}=1$, and $\sigma_0^{ex}=1$.

Table 12
Bootstrap size and power of the J-test statistics under H_0: $Y_n=\lambda W_nY_n+l_n\beta_1+X_{2n}\beta_2+V_n$.

	n	$x=\chi^2(3)$				$x=U(0,10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{ogmm}$	100	0.055	0.054	0.324	0.568	0.057	0.036	0.348	0.563
	300	0.05	0.048	0.592	0.998	0.052	0.062	0.478	0.998
	500	0.042	0.057	0.665	1	0.05	0.048	0.546	1
	700	0.047	0.05	0.672	1	0.062	0.046	0.541	1

The SAR model: $\lambda_0=0.6$, $\beta_{10}=2$, $\beta_{20}=0.5$, and $\sigma_0=\sqrt{2}$.
The MESS model: $\mu_0=-1.6094$, $\beta_{10}^{ex}=2$, $\beta_{20}^{ex}=0.5$, and $\sigma_0^{ex}=\sqrt{2}$.

sequence of pseudo-true value $\mu_{n|sar}^*$ is said to be identifiably unique on $\Theta_{n\mu|sar}$ if either for all $\epsilon>0$ and all n, $N_{n\mu|sar}^c(\epsilon)$ is empty, or

$$\limsup_{n\to\infty}\left[\max_{\mu\in N_{n\mu|sar}^c(\epsilon)}\frac{1}{n}H_{n|sar}(\mu)-\frac{1}{n}H_{n|sar}\left(\mu_{n|sar}^*\right)\right]<0.$$

The following assumption will ensure that $\mu_{n|sar}^*$ is uniquely identified:

Assumption B.1. For any $\mu\neq\mu_{n|sar}^*$

$$\lim_{n\to\infty}\left[\ln\sigma_{n|sar}^{ex2}(\mu)-\ln\sigma_{n|sar}^{ex2}\left(\mu_{n|sar}^*\right)\right]\neq 0.$$

Based on the above assumption, we have the stochastic convergence of the estimator $\hat\theta_{n|sar}^{ex*}$.

Lemma B.1. *Under the null SAR model, given regularity Assumptions 2.1–2.4 and B.1, $\hat\theta_n^{ex}$ is a consistent estimator of the pseudo-true value $\theta_{n|sar}^{ex*}$, in the sense that $\hat\theta_n^{ex}-\theta_{n|sar}^{ex*}=o_p(1)$.*

Appendix C. Pseudo true values of $\hat\theta_n^{sar}$ based upon the QML method

The QML of the SAR model is:

$$L_n\left(\theta^{sar}\right)=-\frac{n}{2}\ln 2\pi\sigma^2+\ln|S_n(\lambda)|-\frac{1}{2\sigma^2}(S_n(\lambda)Y_n-X_n\beta)'(S_n(\lambda)Y_n-X_n\beta).$$

Its concentrated likelihood function of the SAR model at λ is:

$$L_n(\lambda)=-\frac{n}{2}(\ln(2\pi)+1)+\ln|S_n(\lambda)|-\frac{n}{2}\ln\sigma_n^2(\lambda),\tag{C.1}$$

where $\beta_n(\lambda)=(X_n'X_n)^{-1}X_n'S_n(\lambda)Y_n$ and $\sigma_n^2(\lambda)=\frac{1}{n}Y_n'S_n(\lambda)'M_nS_n(\lambda)\times Y_n$. Denote $E_{|ex}(L_n(\theta^{sar}))=E(L_n(\theta^{sar})|H_0)$. Consequently, the expectation of the likelihood function under the null MESS model is

$$E_{|ex}\left(L_n\left(\theta^{sar}\right)\right)=-\frac{n}{2}\ln 2\pi\sigma^2+\ln|S_n(\lambda)|-\frac{1}{2\sigma^2}E_{|ex}\left[(S_n(\lambda)Y_n-X_n\beta)'(S_n(\lambda)Y_n-X_n\beta)\right].$$

Let $\theta_{n|ex}^{sar*}$ be the pseudo true value of $\hat\theta_n^{sar}$, which is defined as

$$\theta_{n|ex}^{sar*}=\arg\max_{\theta^{sar}}E_{|ex}\left(L\left(\theta^{sar}\right)\right).$$

Let $H_{n|ex}(\lambda)=\max_{\beta,\sigma^2}E_{ex}(L_n(\theta^{sar}))$. To derive the exact expression of $H_{n|ex}(\lambda)$, we simplify the term $E_{|ex}[(S_n(\lambda)Y_n-X_n\beta)'(S_n(\lambda)Y_n-X_n\beta)]$, which is

$$\left(S_n(\lambda)S_n^{ex-1}X_n\beta_0^{ex}-X_n\beta\right)'\left(S_n(\lambda)S_n^{ex-1}X_n\beta_0^{ex}-X_n\beta\right)+\sigma_0^2tr\left(S_n^{ex-1\prime}S_n(\lambda)'S_n(\lambda)S_n^{ex-1}\right),$$

where $S_n^{ex}=S_n^{ex}(\mu_0)$. With $\beta_{n|ex}(\lambda)=(X_n'X_n)^{-1}X_n'S_n(\lambda)S_n^{ex-1}X_n\beta_0^{ex}$ and

$$\sigma_{n|ex}^2(\lambda)=\frac{1}{n}\left[\left(S_n(\lambda)S_n^{ex-1}X_n\beta_0^{ex}\right)'M_n\left(S_n(\lambda)S_n^{ex-1}X_n\beta_0^{ex}\right)+\sigma_0^2tr\left(S_n^{ex-1\prime}S_n(\lambda)'S_n(\lambda)S_n^{ex-1}\right)\right].$$

$H_{n|ex}(\lambda)$ can be written as

$$H_{n|ex}(\lambda)=-\frac{n}{2}(\ln 2\pi+1)+\ln|S_n(\lambda)|-\frac{n}{2}\ln\sigma_{n|ex}^2(\lambda).\tag{C.2}$$

The pseudo-true value of λ is defined as $\lambda_{n|ex}^*=\arg\max_\lambda H_{n|ex}(\lambda)$. Then, correspondingly, $\beta_{n|ex}^*=\beta_{n|ex}(\lambda_{n|ex}^*)$ and $\sigma_{n|ex}^{2*}=\sigma_{n|ex}^2(\lambda_{n|ex}^*)$.

Let Θ_λ, a compact subset of R, represent the parameter space of λ, and let $\Theta_{n\lambda|ex}$ be a sequence of non-empty compact subsets of Θ_λ for $n=1,2,\ldots$ such that $H_{n|ex}(\lambda)$ is maximized on $\Theta_{n\lambda|ex}$ at $\lambda_{n|ex}^*$. Furthermore, let $S_{n\lambda|ex}(\epsilon)$ be an open ball in R centered at $\lambda_{n|ex}^*$ with a radius $\epsilon>0$. Define the neighborhood $N_{n\lambda|ex}(\epsilon)=S_{n\lambda|ex}(\epsilon)\cap\Theta_{n\lambda|ex}$ with its compact complement $N_{n\lambda|ex}^c(\epsilon)$. The sequence of pseudo-true value $\lambda_{n|ex}^*$ is identifiably unique on $\Theta_{n\lambda|ex}$ if either for all $\epsilon>0$ and all n, $N_{n\lambda|ex}^c(\epsilon)$ is empty, or

$$\limsup_{n\to\infty}\left[\max_{\lambda\in N_{n\lambda|ex}^c(\epsilon)}\frac{1}{n}H_{n|ex}(\lambda)-\frac{1}{n}H_{n|ex}(\lambda_{n|ex}^*)\right]<0.$$

The following assumption will ensure that $\lambda_{n|ex}^*$ is uniquely identified:

Assumption C.1. For any $\lambda\neq\lambda_{n|ex}^*$

$$\lim_{n\to\infty}\left(\frac{1}{n}\left[\ln|S_n(\lambda)|-\ln|S_n(\lambda_{n|ex}^*)|\right]-\frac{1}{2}\left[\ln\sigma_{n|ex}^2(\lambda)-\ln\sigma_{n|ex}^2(\lambda_{n|ex}^*)\right]\right)\neq 0.$$

The following lemma shows that $\hat\theta_n^{sar}-\theta_{n|ex}^{sar*}=o_p(1)$.

Lemma C.1. *Under the null MESS model and given Assumptions 2.1–2.4, 3.6 and C.1, $\hat\theta_n^{sar}$ is a consistent estimator of the pseudo-true values $\theta_{n|ex}^{sar*}$.*

Appendix D. Pseudo true values of $\hat\theta_n^{ex}$ based upon the N2SLS method

Let $g_n(\phi)=Q_n'V_n(\phi)$ represent the moment equation where $V_n(\phi)=S_n^{ex}(\mu)Y_n-X_n\beta^{ex}$. The pseudo true value $\phi_{n|sar}^*$ based on the N2SLS approach can be defined as

$$\phi_{n|sar}^*=\arg\min E_{|sar}g_n(\phi)'\left(Q_n'Q_n\right)^{-1}E_{|sar}g_n(\phi).$$

We impose the following assumption on $\phi_{n|sar}^*$.

Table 13
Bootstrap size and power of the J-test statistics under H_0: $Y_n = \lambda W_n Y_n + l_n\beta_1 + X_{2n}\beta_2 + V_n$.

	n	$x=\chi^2(3)$				$x=U(0,10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{ogmm}$	100	0.058	0.038	0.086	0.019	0.051	0.033	0.071	0.022
	300	0.051	0.03	0.083	0.015	0.051	0.05	0.101	0.014
	500	0.034	0.047	0.099	0.03	0.05	0.04	0.112	0.027
	700	0.051	0.046	0.092	0.042	0.056	0.03	0.108	0.036

The SAR model: $\lambda_0=0.4$, $\beta_{10}=2$, $\beta_{20}=1$, and $\sigma_0=1$.
The MESS model: $\mu_0=-0.5108$, $\beta_{10}^{ex}=2$, $\beta_{20}^{ex}=1$, and $\sigma_0^{ex}=1$.

Table 14
Bootstrap size and power of the J-test statistics under H_0: $Y_n = \lambda W_n Y_n + l_n\beta_1 + X_{2n}\beta_2 + V_n$.

	n	$x=\chi^2(3)$				$x=U(0,10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{ogmm}$	100	0.061	0.049	0.061	0.04	0.059	0.028	0.054	0.031
	300	0.049	0.042	0.087	0.039	0.055	0.034	0.101	0.034
	500	0.045	0.046	0.156	0.048	0.049	0.031	0.145	0.03
	700	0.059	0.054	0.165	0.037	0.048	0.023	0.142	0.025

The SAR model: $\lambda_0=0.4$, $\beta_{10}=2$, $\beta_{20}=0.5$, and $\sigma_0=\sqrt{2}$.
The MESS model: $\mu_0=-0.5108$, $\beta_{10}^{ex}=2$, $\beta_{20}^{ex}=0.5$, and $\sigma_0^{ex}=\sqrt{2}$.

Assumption D.1. $\phi^*_{n|sar}$ is the unique solution of the moment equations $E_{|sar}g_n(\phi)=0$.

With Assumption D.1 and regularity conditions given in previous sections, the following lemma shows that $\hat{\phi}_n - \phi^*_{n|sar} = o_p(1)$:

Lemma D.1. *Under the null SAR model and given Assumption 2.2–2.4, 3.6, 4.1 and D.1, $\hat{\phi}_n$ is a consistent estimator of $\phi^*_{n|sar}$.*

Appendix E. Pseudo true values of $\hat{\theta}_n^{sar}$ based upon the RGMM method

We investigate the pseudo true value of $\hat{\gamma}_n$ based upon the RGMM method under the null MESS model. Note that $V_n(\gamma)=(I_n-\lambda W_n)Y_n - X_n\beta$. The moment vector is

$$g_n(\gamma) = \left(P_{1n}V_n(\gamma), \ldots, P_{qn}V_n(\gamma), Q_n\right)' V_n(\gamma)$$

where P_{jn}, $j=1,\cdots,q$, have zero diagonals. Let $a_n g_n(\gamma)$ represent a linear combination of $g_n(\gamma)$. $\hat{\gamma}_n$ is obtained from $\min_\gamma g_n'(\gamma)a_n'a_n g_n(\gamma)$. Hence, the pseudo-true values of $\hat{\gamma}_n$ are defined as

$$\gamma^*_{n|ex} = \arg\min_\gamma E_{|ex}g_n'(\gamma)a_n'a_n E_{|ex}g_n(\gamma).$$

We assume that $\gamma^*_{n|ex}$ is unique:

Assumption E.1. $\gamma^*_{n|ex}$ is the unique solution of the moment equations $E_{|ex}g_n(\gamma)=0$.

With Assumption E.1 and regularity conditions given in previous sections, we have the following lemma:

Lemma E.1. *Under the null MESS model and given Assumptions 2.2–2.4, 3.6, 4.1–4.3 and E.1, $\hat{\gamma}_n$ is a consistent estimator of $\gamma^*_{n|ex}$ in the sense that $\hat{\gamma}_n - \gamma^*_{n|ex} = o_p(1)$.*

Appendix F. Proof of propositions and lemmas

Proof of Lemma B.1. The proof basically follows the proof of Theorems 3.1 and 4.1 in Lee (2004). By definition, $H_{n|sar}(\mu) \le H_{n|sar}(\mu^*_{n|sar})$. According to White (1994, Theorem 3.4), we shall show the uniform convergence of $\frac{1}{n}(L_n(\mu) - H_{n|sar}(\mu))$ to zero in the parameter space of μ, check the uniform equicontinuity of $H_{n|sar}(\mu)$ and show the identification uniqueness condition.

Denote $\sigma_n^2(\mu,\lambda_0,\sigma_0^2) = \frac{1}{n}\sigma_0^2 tr\left(S_n^{-1'}S_n^{ex}(\mu)'S_n^{ex}(\mu)S_n^{-1}\right)$. We shall show $\sigma_n^2(\mu, \lambda_0, \sigma_0^2)$ is uniformly bounded away from zero in μ. To begin with, consider the log likelihood function of a pure SAR process $Y_n = S_n(\lambda)^{-1}V_n$. That is

$$L_{p,n}(\lambda,\sigma^2) = -\frac{n}{2}\ln 2\pi\sigma^2 - \frac{1}{2\sigma^2}Y_n'S_n(\lambda)'S_n(\lambda)Y_n + \ln|S_n(\lambda)|;$$

and $E_{|psar}(L_{p,n}(\lambda_0,\sigma_0^2)) = -\frac{n}{2}(\ln 2\pi + 1) - \frac{n}{2}\ln\sigma_0^2 + \ln|S_n|$. Similarly, the log likelihood function of a pure MESS process $Y_n = S_n^{ex}(\mu)^{-1}V_n$ is

$$L_{p,n}(\mu,\sigma^{ex2}) = -\frac{n}{2}\ln 2\pi\sigma^{ex2} - \frac{1}{2\sigma^{ex2}}Y_n'S_n^{ex}(\mu)'S_n^{ex}(\mu)Y_n.$$

Denote $H_{psar,n}(\mu) = \max_{\sigma^{ex2}} E_{|psar}(L_{p,n}(\mu,\sigma^{ex2}))$ as the concentrated likelihood function for the conditional expectation of $L_{p,n}(\mu,\sigma^{ex2})$, given the true pure SAR process. $H_{psar,n}(\mu)$ can be written as

$$H_{psar,n}(\mu) = -\frac{n}{2}(\ln 2\pi + 1) - \frac{n}{2}\ln\sigma_n^2(\mu,\lambda_0,\sigma_0^2).$$

By the information inequality, $H_{psar,n}(\mu) \le E_{|psar}(L_{p,n}(\lambda_0, \sigma_0^2))$, which means that for all μ in its parameter space, we have

$$-\frac{1}{2}\ln\sigma_n^2(\mu,\lambda_0,\sigma_0^2) \le -\frac{1}{2}\ln\sigma_0^2 + \frac{1}{n}\ln|S_n|. \tag{F.1}$$

Based on Eq. (F.1), we can argue that $\sigma_n^2(\mu, \lambda_0, \sigma_0^2)$ is uniformly bounded away from zero. Suppose not, then there would exist a sequence μ_n in its parameter space such that $\lim_{n\to\infty}\sigma_n^2(\mu_n, \lambda_0, \sigma_0^2)=0$. However, in Eq. (F.1), $-\frac{1}{2}\ln\sigma_n^2(\mu_n,\lambda_0,\sigma_0^2)\to\infty$ while $-\frac{1}{2}\ln\sigma_0^2 + \frac{1}{n}\ln|S_n|$ is bounded, a contradiction. As a result, $\sigma_n^2(\mu, \lambda_0, \sigma_0^2)$ must be uniformly bounded away from zero.

Next, we show the uniform convergence of $\frac{1}{n}(L_n(\mu) - H_{n|sar}(\mu))$ to zero. Note that

$$\frac{1}{n}\left(L_n(\mu) - H_{n|sar}(\mu)\right) = -\frac{1}{2}\left(\ln\sigma_n^{ex2}(\mu) - \ln\sigma_{n|sar}^{ex2}(\mu)\right). \tag{F.2}$$

Recall that under the null SAR model

$$\begin{aligned}\sigma_n^{ex2}(\mu) &= \frac{1}{n}\left(S_n^{-1}X_n\beta_0 + S_n^{-1}V_n\right)'S_n^{ex}(\mu)'M_nS_n^{ex}(\mu)\left(S_n^{-1}X_n\beta_0 + S_n^{-1}V_n\right)\\ &= \frac{1}{n}\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right)'M_n\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right)\\ &\quad + \frac{2}{n}\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right)'M_nS_n^{ex}(\mu)S_n^{-1}V_n\\ &\quad + \frac{1}{n}V_n'S_n^{-1'}S_n^{ex}(\mu)'M_nS_n^{ex}(\mu)S_n^{-1}V_n\end{aligned}$$

and

$$\begin{aligned}\sigma_{n|sar}^{ex2}(\mu) = \frac{1}{n}\Big[&\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right)'M_n\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right)\\ &+\sigma_0^2 tr\left(S_n^{-1'}S_n^{ex}(\mu)'S_n^{ex}(\mu)S_n^{-1}\right)\Big].\end{aligned}$$

Table 15
Bootstrap size and power of the J-test statistics under H_0: $S_n^{ex}(\mu)Y_n = l_n\beta_1^{ex} + X_{2n}\beta_2^{ex} + V_n$.

	n	$x=\chi^2(3)$				$x=U(0,10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{sls}$	100	0.056	0.05	0.099	0.05	0.043	0.034	0.094	0.03
	300	0.057	0.045	0.173	0.059	0.047	0.049	0.252	0.069
	500	0.039	0.048	0.335	0.059	0.045	0.037	0.432	0.07
	700	0.036	0.047	0.497	0.062	0.062	0.044	0.668	0.12

The SAR model: $\lambda_0=0.4$, $\beta_{10}=2$, $\beta_{20}=1$, and $\sigma_0=1$.
The MESS model: $\mu_0=-0.5108$, $\beta_{10}^{ex}=2$, $\beta_{20}^{ex}=1$, and $\sigma_0^{ex}=1$.

Table 16
Bootstrap size and power of J-test statistics with unknown heteroskedasticity under H_0: $Y_n=\lambda W_nY_n+l_n\beta_1+X_{2n}\beta_2+V_n$.

	$L, \overline{L}$	$x=\chi^2(3)$				$x=U(0,10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{orgmm}$	$L=5,\overline{L}=15$	0.041	0.053	0.077	0.227	0.047	0.044	0.086	0.253
	$L=14,\overline{L}=20$	0.051	0.058	0.078	0.297	0.056	0.027	0.082	0.291

The SAR model: $\lambda_0=0.4$, $\beta_{10}=2$, and $\beta_{20}=1$.
The MESS model: $\mu_0=-0.5108$, $\beta_{10}^{ex}=2$, and $\beta_{20}^{ex}=1$.
Sample size is 545 for $L=5, \bar{L}=15$. Sample size is 520 for $L=14, \bar{L}=20$.
$\mathcal{W}_{orgmm}$: the Wald test statistics based on the FORGMM method.

Therefore

$$\begin{aligned}\sigma_n^{ex2}(\mu)-\sigma_{n|sar}^{ex2}(\mu) &= \frac{2}{n}\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right)'M_nS_n^{ex}(\mu)S_n^{-1}V_n \\ &+\frac{1}{n}V'_nS_n^{-1'}S_n^{ex}(\mu)'M_nS_n^{ex}(\mu)S_n^{-1}V_n \\ &-\frac{\sigma_0^2}{n}tr\left(S_n^{-1'}S_n^{ex}(\mu)'S_n^{ex}(\mu)S_n^{-1}\right) \\ &= \frac{2}{n}\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right)'M_nS_n^{ex}(\mu)S_n^{-1}V_n \\ &+\frac{1}{n}V'_nS_n^{-1'}S_n^{ex}(\mu)'S_n^{ex}(\mu)S_n^{-1}V_n \\ &-\frac{\sigma_0^2}{n}tr\left(S_n^{-1'}S_n^{ex}(\mu)'S_n^{ex}(\mu)S_n^{-1}\right)-C_n(\mu)\end{aligned}$$

where $C_n(\mu)=\frac{1}{n}V'_nS_n^{-1'}S_n^{ex}(\mu)'X_n(X'_nX_n)^{-1}X'_nS_n^{ex}(\mu)S_n^{-1}V_n$.

By Lemma A.5 and the series expansion form of the matrix exponential, $\frac{1}{n}X'_nS_n^{ex}(\mu)S_n^{-1}V_n=o_p(1)$ uniformly in μ. Thus,

$$C_n(\mu)=\left(\frac{1}{n}V'_nS_n^{-1'}S_n^{ex}(\mu)'X_n\right)\left(\frac{X'_nX_n}{n}\right)^{-1}\left(\frac{1}{n}X'_nS_n^{ex}(\mu)S_n^{-1}V_n\right)=o_p(1)$$

uniformly in μ. Furthermore, by Lemmas A.4 and A.5,

$$\frac{2}{n}\left(S_n^{ex}(\mu)S^{-1}X_n\beta_0\right)'M_nS_n^{ex}(\mu)S_n^{-1}V_n=o_p(1)$$

uniformly in μ. Also by Lemma A.2,

$$\frac{1}{n}\left[V'_nS_n^{-1'}S_n^{ex}(\mu)'S_n^{ex}(\mu)S_n^{-1}V_n-\sigma_0^2tr\left(S_n^{-1'}S_n^{ex}(\mu)'S_n^{ex}(\mu)S_n^{-1}\right)\right]=o_p(1)$$

uniformly in μ. Therefore, $\sigma_n^{ex2}(\mu)-\sigma_{n|sar}^{ex2}(\mu)=o_p(1)$ uniformly in μ.

Finally, by the mean-value theorem, $\left|\ln\sigma_n^{ex2}(\mu)-\ln\sigma_{n|sar}^{ex2}(\mu)\right|=\left|\sigma_n^{ex2}(\mu)-\sigma_{n|sar}^{ex2}(\mu)\right|/\tilde{\sigma}_n^{ex2}(\mu)$, where $\tilde{\sigma}_n^{ex2}(\mu)$ lies between $\sigma_n^{ex2}(\mu)$ and $\sigma_{n|sar}^{ex2}(\mu)$. Notice that $\sigma_{n|sar}^{ex2}(\mu)\geq\sigma_n^2(\mu,\lambda_0,\sigma_0)$ since

$$\sigma_{n|sar}^{ex2}(\mu)=\frac{1}{n}\left(S_n^{ex}(\mu)S_n^{-1}X_n\beta_0\right)'M_nS_n^{ex}(\mu)S_n^{-1}X_n\beta_0+\sigma_n^2\left(\mu,\lambda_0,\sigma_0^2\right).$$

As $\sigma_n^2(\mu,\lambda_0,\sigma_0^2)$ is uniformly bounded away from zero in μ, $\sigma_{n|sar}^{ex2}(\mu)$ will be so too. As a result, $\sigma_n^{ex2}(\mu)$ will be uniformly bounded away from zero in μ in probability. Therefore, $|\ln\sigma_n^{ex2}(\mu)-\ln\sigma_{n|sar}^{ex2}(\mu)|=o_p(1)$ uniformly in μ. Hence $\sup_\mu\left|\frac{1}{n}(L_n(\mu)-H_{n|sar}(\mu))\right|=o_p(1)$.

Next we show the uniform equicontinuity of $\frac{1}{n}H_{n|sar}(\mu)$. Note that $\frac{1}{n}H_{n|sar}(\mu)=-\frac{1}{2}(\ln2\pi+1)-\frac{1}{2}\ln\sigma_{n|sar}^{ex2}(\mu)$. $\sigma_{n|sar}^{ex2}(\mu)$ is uniformly continuous in μ since it is essentially a polynomial of μ. The uniform continuity of $\ln\sigma_{n|sar}^{ex2}(\mu)$ follows because $\frac{1}{\sigma_{n|sar}^{ex2}(\mu)}$ is uniformly bounded in μ. Hence $\frac{1}{n}H_{n|sar}(\mu)$ is uniformly equicontinuous in μ.

For the identification uniqueness condition, note that

$$\frac{1}{n}H_{n|sar}(\mu)-\frac{1}{n}H_{n|sar}(\mu_{n|sar}^*)=-\frac{1}{2}\left[\ln\sigma_{n|sar}^{ex2}(\mu)-\ln\sigma_{n|sar}^{ex2}(\mu_{n|sar}^*)\right]\leq0.$$

Then, by Assumption B.1, the identification uniqueness condition is satisfied.

In conclusion, $\text{plim}\hat{\mu}_{n|sar}=\mu_{n|sar}^*$ and thus $\text{plim}\hat{\theta}_{n|sar}^{ex}=\theta_{n|sar}^{ex*}$ follows from the identification uniqueness and uniform convergence (White, 1994, Theorem 3.4). □

Proof of Lemma C.1. To show that $\hat{\theta}_n^{sar}-\theta_{n|ex}^{sar*}=o_p(1)$, we follow similar steps in the proof of Lemma B.1. □

Proof of Lemma D.1. The proof is similar to the proof of Proposition 1 in Lee (2007). Specifically, the parameter space is bounded, $\frac{1}{n}a_ng_n(\phi)$ is continuous in ϕ with $a_n=\left(\frac{Q'_nQ_n}{n}\right)^{-\frac{1}{2}}$, the identification uniqueness condition is satisfied by Assumption D.1, and $\frac{1}{n}a_ng_n(\phi)-\frac{1}{n}a_nE_{|sar}g_n(\phi)$ converges in probability to zero in ϕ uniformly in its parameter space. □

Proof of Lemma E.1. The proof is similar to the proof of Proposition 1 in Lee (2007) and Lemma D.1. □

Proof of Proposition 1. The proof follows similarly the proof of Proposition 1 in Lee (2007). For consistency, we shall first show that $\frac{1}{n}a_ng_n(\eta_{r_1})-\frac{1}{n}a_nEg_n(\eta_{r_1})$ will converge in probability uniformly in η_{r_1} to zero. Following Lee (2007), let $a_n=(a_{n1},\ldots,a_{nq},a_{nx})$ where a_{nx} is a row subvector so that $a_ng_n(\eta_{r_1})=V'_n(\eta_{r_1})(\Sigma_{j=1}^q a_{nj}P_{jn})V_n(\eta_{r_1})+a_{nx}Q'_nV_n(\eta_{r_1})$. Note that

$$\begin{aligned}V_n(\eta_{r_1})&=d_n(\eta_{r_1})+S_n(\lambda)S_n^{-1}V_n=h_n(\gamma)+\hat{Y}_{n|r_1}(\delta_0-\delta_{r_1})+S_n(\lambda)S_n^{-1}V_n \\ &=h_n(\gamma)+\hat{Y}_{n|r_1}(\delta_0-\delta_{r_1})+V_n+(\lambda_0-\lambda)G_nV_n\end{aligned} \quad \text{(F.3)}$$

where $h_n(\gamma)=X_n(\beta_0-\beta)+(\lambda_0-\lambda)G_nX_n\beta_0$. By Lemma A.5,

$$\frac{1}{n}a_{nx}Q'_nV_n(\eta_{r_1})=\frac{1}{n}a_{nx}Q'_nh_n(\gamma)+\frac{1}{n}a_{nx}Q'_nY_{n|r_1}^*(\delta_0-\delta_{r_1})+o_p(1) \quad \text{(F.4)}$$

for $r_1=1,2$ where $Y_{n|r_1}^*$ is defined in Section 3. The quadratic moment function can be decomposed into three terms:

$$\begin{aligned}V'_n(\eta_{r_1})\left(\sum_{j=1}^q a_{nj}P_{jn}\right)V_n(\eta_{r_1})&=d'_n(\eta_{r_1})\left(\sum_{j=1}^q a_{nj}P_{jn}\right)d_n(\eta_{r_1}) \\ &+l_n(\eta_{r_1})+t_n(\eta_{r_1})\end{aligned}$$

Table 17
Bootstrap size and power of J-test statistics with unknown heteroskedasticity under H_0: $S_n^{ex}(\mu)Y_n = l_n\beta_1^{ex} + X_{2n}\beta_2^{ex} + V_n$.

	$L, \overline{L}$	$x=\chi^2(3)$				$x=U(0, 10)$			
		Size		Power		Size		Power	
		$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$	$Y_{n\|1}$	$Y_{n\|2}$
$\mathcal{W}_{rsls}$	$L=5, \overline{L}=15$	0.053	0.047	0.094	0.074	0.051	0.056	0.084	0.079
	$L=14, \overline{L}=20$	0.049	0.045	0.096	0.059	0.071	0.048	0.083	0.091

The SAR model: $\lambda_0=0.4$, $\beta_{10}=2$, and $\beta_{20}=1$.
The MESS model: $\mu_0=-0.5108$, $\beta_{10}^{ex}=2$, and $\beta_{20}^{ex}=1$.
Sample size is 545 for $L=5, \bar{L}=15$. Sample size is 520 for $L=14, \bar{L}=20$.
$\mathcal{W}_{rsls}$: the Wald test statistics based on the GN2SLS method.

where

$$l_n(\eta_{r_1}) = d_n'(\eta_{r_1})\left(\sum_{j=1}^q a_{nj}P_{jn}^S\right)(V_n + (\lambda_0-\lambda)G_nV_n), \text{ and}$$

$$t_n(\eta_{r_1}) = (V_n + (\lambda_0-\lambda)G_nV_n)'\left(\sum_{j=1}^q a_{nj}P_{jn}\right)(V_n + (\lambda_0-\lambda)G_nV_n).$$

By Lemma A.2,

$$\begin{aligned}\frac{1}{n}t_n(\eta_{r_1}) &= \frac{1}{n}V_n'\left(\sum_{j=1}^q a_{nj}P_{jn}\right)V_n + (\lambda_0-\lambda)\frac{1}{n}V_n'G_n'\left(\sum_{j=1}^q a_{nj}P_{jn}^S\right)V_n \\ &\quad +(\lambda_0-\lambda)^2\frac{1}{n}V_n'G_n'\left(\sum_{j=1}^q a_{nj}P_{jn}\right)G_nV_n \\ &= (\lambda_0-\lambda)\frac{\sigma_0^2}{n}\sum_{j=1}^q a_{nj}tr\left(G_n'P_{jn}^S\right) + (\lambda_0-\lambda)^2\frac{\sigma_0^2}{n}\sum_{j=1}^q a_{nj}tr\left(G_n'P_{jn}G_n\right) + o_p(1).\end{aligned} \tag{F.5}$$

For $d_n'(\eta_{r_1})\left(\sum_{j=1}^q a_{nj}P_{jn}\right)d_n(\eta_{r_1})$ and $l_n(\eta_{r_1})$, we need to consider two cases $r_1=1$ and $r_1=2$ separately. For $r_1=1$, by Lemmas A.1 and B.1,

$$\begin{aligned}\frac{1}{n}d_n'(\eta_1)\left(\sum_{j=1}^q a_{nj}P_{jn}\right)d_n(\eta_1) &= \frac{1}{n}h_n'(\gamma)\left(\sum_{j=1}^q a_{nj}P_{jn}\right)h_n(\gamma) \\ &\quad +\frac{1}{n}h_n'(\gamma)\left(\sum_{j=1}^q a_{nj}P_{jn}^S\right)Y_{n|1}^*(\delta_0-\delta_1) \\ &\quad +\frac{1}{n}\left(Y_{n|1}^*(\delta_0-\delta_1)\right)'\left(\sum_{j=1}^q a_{nj}P_{jn}\right) \\ &\quad \times\left(Y_{n|1}^*(\delta_0-\delta_1)\right) + o_p(1)\end{aligned} \tag{F.6}$$

uniformly in η_1. For $\frac{1}{n}l_n(\eta_1)$, by Lemma A.5, it is

$$\frac{1}{n}l_n(\eta_1) = \frac{1}{n}\left(h_n(\gamma) + Y_{n|1}^*(\delta_0-\delta)\right)'\left(\sum_{j=1} a_{nj}P_{jn}^S\right)(V_n + (\lambda_0-\lambda)G_nV_n) + o_p(1) = o_p(1).$$

For $r_1=2$, by Lemmas A.1, A.2 and B.1

$$\begin{aligned}\frac{1}{n}d_n'(\eta_2)\left(\sum_{j=1}^q a_{nj}P_{jn}\right)d_n(\eta_2) &= \frac{1}{n}\left(h_n(\gamma) + Y_{n|2}^*(\delta_0-\delta_2)\right)'\left(\sum_{j=1}^q a_{nj}P_{jn}\right) \\ &\quad \times\left(h_n(\gamma) + Y_{n|2}^*(\delta_0-\delta_2)\right) \\ &\quad +(\delta_0-\delta_2)^2\frac{\sigma_0^2}{n}\sum_{j=1}^q a_{nj}tr\left(S_n^{-1'}U_{n|sar}^{*'}P_{jn}U_{n|sar}^*S_n^{-1}\right) + o_p(1).\end{aligned} \tag{F.7}$$

For $\frac{1}{n}l_n(\eta_2)$, $\hat{Y}_{n|2} = \hat{U}_nY_n + X_n\hat{\beta}_n^{ex} = \hat{U}_n(S_n^{-1}X_n\beta_0 + S_n^{-1}V_n) + X_n\hat{\beta}_n^{ex}$, therefore

$$\begin{aligned}\frac{1}{n}l_n(\eta_2) &= \frac{1}{n}(h_n(\gamma) + \hat{Y}_{n|2}(\delta_0-\delta_2))'\left(\sum_{j=1}^q a_{nj}P_{jn}^S\right)(V_n + (\lambda_0-\lambda)G_nV_n) \\ &= (\delta_0-\delta_2)\frac{\sigma_0^2}{n}\sum_{j=1}^q a_{nj}tr\left(S_n^{-1'}U_{n|sar}^*P_{jn}^S(I_n + (\lambda_0-\lambda)G_n)\right) \\ &\quad + o_p(1)\end{aligned} \tag{F.8}$$

under the null SAR model.

With Eqs. (F.5)–(F.8) together, we have

$$\begin{aligned}\frac{1}{n}V_n'(\eta_1)\left(\sum_{j=1}^q a_{nj}P_{jn}\right)V_n(\eta_1) &= \frac{1}{n}h_n'(\gamma)\left(\sum_{j=1}^q a_{nj}P_{jn}\right)h_n'(\gamma) \\ &\quad +\frac{1}{n}h_n'(\gamma)\left(\sum_{j=1}^q a_{nj}P_{jn}^S\right)Y_{n|1}^*(\delta_0-\delta_1) \\ &\quad +\frac{1}{n}\left(Y_{n|1}^*(\delta_0-\delta_1)\right)'\left(\sum_{j=1}^q a_{nj}P_{jn}\right)(Y_{n|1}^*(\delta_0-\delta_1)) \\ &\quad +(\lambda_0-\lambda)\frac{\sigma_0^2}{n}\sum_{j=1}^q a_{nj}tr\left(G_n'P_{jn}^S\right) \\ &\quad +(\lambda_0-\lambda)^2\frac{\sigma_0^2}{n}\sum_{j=1}^q a_{nj}tr\left(G_n'P_{jn}G_n\right) + o_p(1); \\ \frac{1}{n}V_n'(\eta_2)\left(\sum_{j=1}^q a_{nj}P_{jn}\right)V_n(\eta_2) &= \frac{1}{n}(h_n(\gamma)+Y_{n|2}^*(\delta_0-\delta_2))'\left(\sum_{j=1}^q a_{nj}P_{jn}\right)(h_n(\gamma)+Y_{n|2}^*(\delta_0-\delta_2)) \\ &\quad +(\delta_0-\delta_2)^2\frac{\sigma_0^2}{n}\sum_{j=1}^q a_{nj}tr\left(S_n^{-1'}U_{n|sar}^{*'}P_{jn}U_{n|sar}^*S_n^{-1}\right) \\ &\quad +(\delta_0-\delta_2)\frac{\sigma_0^2}{n}\sum_{n}^q a_{nj}tr\left(S_n^{-1'}U_{n|sar}^*P_{jn}^S(I_n + (\lambda_0-\lambda)G_n)\right) \\ &\quad +(\lambda_0-\lambda)\frac{\sigma_0^2}{n}\sum_{j=1}^q a_{nj}tr\left(G_n'P_{jn}^S\right) \\ &\quad +(\lambda_0-\lambda)^2\frac{\sigma_0^2}{n}\sum_{j=1}^q a_{nj}tr\left(G_n'P_{jn}G_n\right) + o_p(1)\end{aligned} \tag{F.9}$$

uniformly in η_1 and η_2, respectively.

With Eq. (F.9) $\frac{1}{n}a_ng_n(\eta_{r_1}) - \frac{1}{n}a_nEg_n(\eta_{r_1})$ converges in probability uniformly in η_{r_1} to zero. Since $g_n(\eta_{r_1})$ is a quadratic function of η_{r_1} and the parameter space is bounded, $\frac{1}{n}a_nEg_n(\eta_{r_1})$ is uniformly equicontinuous in η_{r_1}. Thus, by a similar argument in Lee (2007), the identification uniqueness condition for $\left(\frac{1}{n^2}\right)E\left(g_n'(\eta_{r_1})\right)\times a_n'a_nE\left(g_n(\eta_{r_1})\right)$ must be satisfied. The consistency of the GMME $\hat{\eta}_{n|r_1}$ follows from the uniform convergence and the identification uniqueness condition (White, 1994).

For the asymptotic distribution of $\hat{\eta}_{n|r_1}$, by the Taylor expansion,

$$\begin{aligned}&\sqrt{n}(\hat{\eta}_{n|r_1} - \eta_{0r_1}) \\ &= -\left[\frac{1}{n}\frac{\partial g_n'(\hat{\eta}_{n|r_1})}{\partial\eta_{r_1}}a_n'a_n\frac{1}{n}\frac{\partial g_n(\bar{\eta}_{n|r_1})}{\partial\eta_{r_1}}\right]^{-1}\frac{1}{n}\frac{\partial g_n'(\hat{\eta}_{n|r_1})}{\partial\eta_{r_1}}a_n'\frac{1}{\sqrt{n}}a_ng_n(\eta_{0r_1}).\end{aligned}$$

Note that $\frac{\partial g_n'(\eta_{r_1})}{\partial\eta_{r_1}} = -(P_{1n}^SV_n(\eta_{r_1}), \ldots, P_{qn}^SV_n(\eta_{r_1}), Q_n)'(-W_nY_n, -X_n, -\hat{Y}_{n|r_1})$. Then, for any j,

$$\frac{1}{n}V_n'(\eta_{r_1})P_{jn}^SW_nY_n = \frac{1}{n}V_n'(\eta_{r_1})P_{jn}^SG_nX_n\beta_0 + \frac{1}{n}V_n'(\eta_{r_1})P_{jn}^SG_nV_n.$$

By Lemmas A.2 and A.5,

$$\frac{1}{n}V'_n(\eta_{r_1})P^S_{jn}G_nX_n\beta_0 = \frac{1}{n}d'_n(\eta_{r_1})P^S_{jn}G_nX_n\beta_0 + \frac{1}{n}V'_nP^S_{jn}G_nX_n\beta_0 + (\lambda_0-\lambda)\frac{1}{n}V'_nG'_nP^S_{jn}G_nX_n\beta_0 = \frac{1}{n}\left(h_n(\gamma)+Y^*_{n|r_1}(\delta_{0r_1}-\delta_{r_1})\right)'P^S_{jn}G_nX_n\beta_0 + o_p(1) \quad \text{(F.10)}$$

for $r_1=1, 2$. Moreover, for $r_1=1$

$$\frac{1}{n}V'_n(\eta_1)P^S_{jn}G_nV_n = \frac{\sigma_0^2}{n}tr\left(P^S_{jn}G_n\right) + (\lambda_0-\lambda)\frac{\sigma_0^2}{n}tr\left(G'_nP^S_{jn}G_n\right) + o_p(1);$$

for $r_1=2$, $\hat{Y}_{n|2} = \hat{U}_n\left(S_n^{-1}X_n\beta_0 + S_n^{-1}V_n\right) + X_n\hat{\beta}_n^{ex}$, thus

$$\frac{1}{n}V'_n(\eta_2)P^S_{jn}G_nV_n = \frac{\sigma_0^2}{n}(\delta_0-\delta_2)tr\left(S_n^{-1'}U^{*'}_{n|sar}P^S_{jn}G_n\right) + \frac{\sigma_0^2}{n}tr\left(P^S_{jn}G_n\right) + (\lambda_0-\lambda)\frac{\sigma_0^2}{n}tr\left(G'_nP^S_{jn}G_n\right) + o_p(1)$$

under the null SAR model. Hence,

$$\frac{1}{n}V'_n(\eta_1)P^S_{jn}W_nY_n = \frac{1}{n}\left(h_n(\gamma)+Y^*_{n|1}(\delta_{01}-\delta_1)\right)'P^S_{jn}G_nX_n\beta_0 + \frac{\sigma_0^2}{n}tr\left(P^S_{jn}G_n\right) + (\lambda_0-\lambda)\frac{\sigma_0^2}{n}tr\left(G'_nP^S_{jn}G_n\right) + o_p(1); \text{ and}$$

$$\frac{1}{n}V'_n(\eta_2)P^S_{jn}W_nY_n = \frac{1}{n}\left(h_n(\gamma)+Y^*_{n|2}(\delta_{02}-\delta_2)\right)'P^S_{jn}G_nX_n\beta_0 + \frac{\sigma_0^2}{n}tr\left(P^S_{jn}G_n\right) + (\lambda_0-\lambda)\frac{\sigma_0^2}{n}tr\left(G'_nP^S_{jn}G_n\right) + \frac{\sigma_0^2}{n}(\delta_{02}-\delta_2)tr\left(S_n^{-1'}U^{*'}_{n|sar}P^S_{jn}G_n\right) + o_p(1).$$

At η_{0r_1}, $d_n(\eta_{0r_1})=0$. Thus for $r_1=1, 2$, $\frac{1}{n}V'_n(\eta_{0r_1})P^S_{jn}W_nY_n = \frac{\sigma_0^2}{n}\times tr\left(P^S_{jn}G_n\right)+o_p(1)$. Furthermore, for any j, $\frac{1}{n}V'_n(\eta_{r_1})P^S_{jn}X_n = \frac{1}{n}(d_n(\eta_{r_1}) + S_n(\lambda)S_n^{-1}V_n)'P^S_{jn}X_n$. So at η_{0r_1}, $\frac{1}{n}V'_n(\eta_{0r_1})P^S_{jn}X_n = o_p(1)$. Next, consider $\frac{1}{n}V'_n(\eta_{r_1})P^S_{jn}\hat{Y}_{n|r_1}$. When $r_1=1$,

$$\frac{1}{n}V'_n(\eta_1)P^S_{jn}\hat{Y}_{n|1} = \frac{1}{n}\left(d_n(\eta_1)+S_n(\lambda)S_n^{-1}V_n\right)'P^S_{jn}Y^*_{n|1} + o_p(1).$$

Thus at η_{01}, $\frac{1}{n}V'_n(\eta_{01})P^S_{jn}\hat{Y}_{n|1} = o_p(1)$. Moreover, for $r_1=2$,

$$\frac{1}{n}V'_n(\eta_2)P^S_{jn}\hat{Y}_{n|2} = \frac{1}{n}d'_n(\eta_2)P^S_{jn}\hat{Y}_{n|2} + \frac{1}{n}(V_n+(\lambda_0-\lambda)G_nV_n)'P^S_{jn}\hat{Y}_{n|2}.$$

Under the null, by Lemma A.2, $\frac{1}{n}V'_nP^S_{jn}\hat{Y}_{n|2} = \frac{\sigma_0^2}{n}tr\left(P^S_{jn}U^*_{n|sar}S_n^{-1}\right)+o_p(1)$. Hence,

$$\frac{1}{n}V'_n(\eta_2)P^S_{jn}\hat{Y}_{n|2} = \frac{1}{n}d'_n(\eta_2)P^S_{jn}Y^*_{n|2} + \frac{\sigma_0^2}{n}tr\left(P^S_{jn}U^*_{n|sar}S_n^{-1}\right) + (\lambda_0-\lambda)\frac{\sigma_0^2}{n}tr\left(G'_nP^S_{jn}U^*_{n|sar}S_n^{-1}\right) + o_p(1).$$

Thus, at η_{02}, $\frac{1}{n}V'_n(\eta_{02})P^S_{jn}\hat{Y}_{n|2} = \frac{\sigma_0^2}{n}tr\left(P^S_{jn}U^*_{n|sar}S_n^{-1}\right)+o_p(1)$. Lastly, $\frac{1}{n}Q'_nW_nY_n = \frac{1}{n}Q'_nG_nX_n\beta_0 + \frac{1}{n}Q'_nG_nV_n = \frac{1}{n}Q'_nG_nX_n\beta_0 + o_p(1)$ and $\frac{1}{n}Q'_n\hat{Y}_{n|r_1} = \frac{1}{n}Q'_nY^*_{n|r_1}+o_p(1)$ for $r_1=1, 2$.

In conclusion, because $\bar{\eta}_{n|r_1}-\eta_{0r_1} = o_p(1)$, $\frac{1}{n}\frac{\partial g_n(\bar{\eta}_{n|r_1})}{\partial \eta_{r_1}} = -\frac{1}{n}D_{n|r_1}+o_p(1)$ for $r_1=1, 2$, with $D_{n|1}$ and $D_{n|2}$ defined in Proposition 1. On the other hand, Lemma A.6 implies

$$\frac{1}{\sqrt{n}}a_ng_n(\eta_{0r_1}) = \frac{1}{\sqrt{n}}\left[V'_n\left(\sum_{j=1}^q a_{nj}P_{jn}\right)V_n + a_{nx}Q'_nV_n\right] \xrightarrow{D} N\left(0, \lim_{n\to\infty}\frac{1}{n}a_n\Omega_na'_n\right).$$

The asymptotic distribution of $\sqrt{n}\left(\hat{\eta}_{n|r_1}-\eta_{0r_1}\right)$ follows.

Proof of Proposition 2. The proof is basically similar to the proof of Proposition 2 in Lee (2007). As usual, by the generalized Schwartz inequality, that the optimal weighting matrix for $a_n'a_n$ in the GMM estimation of Eq. (3.2) will be $\left(\frac{1}{n}\Omega_n\right)^{-1}$. Consider

$$\frac{1}{n}g'_n(\eta_{r_1})\hat{\Omega}_n^{-1}g_n(\eta_{r_1}) = \frac{1}{n}g'_n(\eta_{r_1})\Omega_n^{-1}g_n(\eta_{r_1}) + \frac{1}{n}g'_n(\eta_{r_1})\left(\hat{\Omega}_n^{-1}-\Omega_n^{-1}\right)g_n(\eta_{r_1}).$$

Under Assumptions 3.1–3.5 and B.1, the uniform convergence in probability of $\frac{1}{n}g'_n(\eta_{r_1})\Omega_n^{-1}g_n(\eta_{r_1})$ to a well defined limit uniformly in η_{r_1} can be established as in the proof of Proposition 1. It remains to show $\frac{1}{n}g'_n(\eta_{r_1})\left(\hat{\Omega}_n^{-1}-\Omega_n^{-1}\right)g_n(\eta_{r_1}) = o_p(1)$ uniformly in η_{r_1} for $r_1=1, 2$. Let ||.|| be the Euclidean norm for vector and matrix. Thus

$$\left\|\frac{1}{n}g'_n(\eta_{r_1})\left(\hat{\Omega}_n^{-1}-\Omega_n^{-1}\right)g_n(\eta_{r_1})\right\| \le \left(\left\|\frac{1}{n}g_n(\eta_{r_1})\right\|\right)^2\left\|\left(\frac{\hat{\Omega}_n}{n}\right)^{-1}-\left(\frac{\Omega_n}{n}\right)^{-1}\right\|.$$

Since $\left(\frac{\hat{\Omega}_n}{n}\right)^{-1}-\left(\frac{\Omega_n}{n}\right)^{-1} = o_p(1)$, it is sufficient to show that $\frac{1}{n}g_n(\eta_{r_1}) = O_p(1)$ uniformly in η_{r_1}. From the proof of Proposition 1, $\frac{1}{n}\left[g_n(\eta_{r_1})-Eg_n(\eta_{r_1})\right]\xrightarrow{p}0$ uniformly in η_{r_1}. Hence we may check the order of $\frac{1}{n}Eg_n(\eta_{r_1})$. Note that $g_n(\eta_{r_1})=(P_{1n}V_n(\eta_{r_1}),\ldots,P_{qn}V_n(\eta_{r_1}),Q_n)'V_n(\eta_{r_1})$. For the linear moment functions, by Lemma A.1,

$$\frac{1}{n}E\left(Q'_nV_n(\eta_{r_1})\right) = \frac{1}{n}Q'_nh_n(\gamma) + \frac{1}{n}Q'_nY^*_{n|r_1}\left(\delta_{0r_1}-\delta_{r_1}\right) = (\lambda_0-\lambda)\frac{1}{n}Q'_nG_nX_n\beta_0 + \frac{1}{n}Q'_nX_n(\beta_0-\beta) + \frac{1}{n}Q'_nY^*_{n|r_1}\left(\delta_{0r_1}-\delta_{r_1}\right) = O(1).$$

uniformly in η_{r_1} for $r_1=1, 2$. For the quadratic moments, following the proof of Proposition 1, for any j

$$\frac{1}{n}E\left[V'_n(\eta_{r_1})P_{jn}V_n(\eta_{r_1})\right] = \frac{1}{n}E\left[d'_n(\eta_{r_1})P_{jn}d_n(\eta_{r_1}) + l_n(\eta_{r_1}) + t_n(\eta_{r_1})\right].$$

First,

$$\frac{1}{n}Et_n(\eta_{r_1}) = (\lambda_0-\lambda)\frac{\sigma_0^2}{n}tr\left(G'_nP^S_{jn}\right) + (\lambda_0-\lambda)^2\frac{\sigma_0^2}{n}tr\left(G'_nP_{jn}G_n\right) = O(1)$$

for $r_1=1, 2$. Next we need to check $\frac{1}{n}E\left[d'_n(\eta_{r_1})P_{jn}d_n(\eta_{r_1})\right]$ and $\frac{1}{n}El_n(\eta_{r_1})$. When $r_1=1$, $\frac{1}{n}El_n(\eta_1) = \lim_{n\to\infty}\left[\frac{1}{n}\left(h_n(\gamma)+Y^*_{n|1}\right)'P^S_{jn}(V_n+(\lambda_0-\lambda)G_nV_n)\right] = 0$ and also by Lemma A.1,

$$\lim_{n\to\infty}\frac{1}{n}E\left[d'_n(\eta_1)P_{jn}d_n(\eta_1)\right] = E\frac{1}{n}\left[\left(h_n(\gamma)+Y^*_{n|1}(\delta_0-\delta_1)\right)'P_{jn}\left(h_n(\gamma)+Y^*_{n|1}(\delta_0-\delta_1)\right)\right] = O(1).$$

When $r_1 = 2$, by Lemmas A.1 and A.2,

$$\frac{1}{n}El_n(\eta_2) = (\delta_0-\delta_2)\frac{\sigma_0^2}{n}tr\left(S_n^{-1'}U_{n|sar}^{*'}P_{jn}^S(I_n+(\lambda_0-\lambda)G_n)\right) = O(1)$$

and

$$\frac{1}{n}Ed_n'(\eta_2)P_{jn}d_n(\eta_2) = \frac{1}{n}\left(h_n(\gamma)+Y_{n|2}^*(\delta_0-\delta)\right)'P_{jn}\left(h_n(\gamma)+Y_{n|2}^*(\delta_2-\delta_0)\right) + (\delta_0-\delta_2)^2\frac{\sigma_0^2}{n}tr\left(S_n^{-1'}U_{n|sar}^*P_{jn}U_{n|sar}^*S_n^{-1}\right) = O(1)$$

uniformly in η_{r_1}. These together imply that $\frac{1}{n}Eg_n(\eta_{r_1}) = O(1)$ uniformly in η_{r_1}. Consequently, $\frac{1}{n}\left\|g_n(\eta_{r_1})\right\| = O_p(1)$ uniformly in η_{r_1}. Thus, $\left\|\frac{1}{n}g_n'(\eta_{r_1})\left(\hat{\Omega}_n^{-1}-\Omega_n^{-1}\right)g_n(\eta_{r_1})\right\| = O_p(1)$, uniformly in η_{r_1}. The consistency of FOGMME follows. For the asymptotic distribution, as $\frac{1}{n}\frac{\partial g_n(\bar{\eta}_{on|r_1})}{\partial\eta_{r_1}} = -\frac{D_{n|r_1}}{n}+O_p(1)$ from Proposition 1,

$$\sqrt{n}\left(\hat{\eta}_{on|r_1}-\eta_{0r_1}\right) = -\left[\frac{1}{n}\frac{\partial g_n'(\hat{\eta}_{on|r_1})}{\partial\eta_{r_1}}\left(\frac{\hat{\Omega}_n}{n}\right)^{-1}\frac{1}{n}\frac{\partial g_n'(\bar{\eta}_{n|r_1})}{\partial\eta_{r_1}}\right]^{-1} \times\frac{1}{n}\frac{\partial g_n'(\hat{\eta}_{on|r_1})}{\partial\eta_{r_1}}\left(\frac{\hat{\Omega}_n}{n}\right)^{-1}\frac{1}{\sqrt{n}}g_n(\eta_{0r_1}) = \left[\frac{D_{n|r_1}'}{n}\left(\frac{\Omega_n}{n}\right)^{-1}\frac{D_{n|r_1}}{n}\right]^{-1}\frac{D_{n|r_1}'}{n}\left(\frac{\Omega_n}{n}\right)^{-1}\frac{1}{\sqrt{n}}g_n(\eta_{0r_1}) + o_p(1).$$

The asymptotic distribution of $\sqrt{n}\left(\hat{\eta}_{on|r_1}-\eta_{0r_1}\right)$ follows. □

Proof of Proposition 3. The proof is similar to the proof of Proposition 1 but is simpler as we only use the linear moment function $Q_n'V_n(\psi_{r_2})$. Therefore, the N2SLS is a special case of GMM estimation with $a_n = \left(\frac{Q_n'Q_n}{n}\right)^{-\frac{1}{2}}$ and $\frac{1}{n}a_ng_n(\psi_{r_2}) = \left(\frac{Q_n'Q_n}{n}\right)^{-\frac{1}{2}}\frac{1}{n}Q_n'V_n(\psi_{r_2})$. □

Proof of Proposition 4. The proof is similar to those proofs of Proposition 1 and Proposition 1 in Lin and Lee (2010). □

Proof of Proposition 5. The proof of the consistency of $\frac{1}{n}\hat{\Omega}_{nh}$ will be similar to that in Lin and Lee (2010). Here we shall show that $\frac{1}{n}\left(\hat{D}_{nh|r_1}-D_{nh|r_1}\right) = o_p(1)$ for $r_1 = 1, 2$.

Note that two generic forms of the elements in $\frac{1}{n}D_{nh|r_1}$ are $\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SG_n\right)_{ii}\sigma_{ni}^2$ and $\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\times\sigma_{ni}^2$. Since P_n's, G_n, $U_{n|sar}^*$ and S_n^{-1} are all uniformly bounded in both row and column sum norms, so are the matrices $P_{jn}^SG_n$'s and $P_{jn}^SU_{n|sar}^*S_n^{-1'}$s. To prove $\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\hat{v}_{ni}^2-\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\sigma_{ni}^2 = o_p(1)$, we note that, by Lemma A.10,

$$\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}v_{ni}^2-\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\sigma_{ni}^2 = \frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\left(v_{ni}^2-\sigma_{ni}^2\right) = o_p(1).$$

It remains to show $\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\left(\hat{v}_{ni}^2-v_{ni}^2\right) = o_p(1)$. As $\hat{V}_n = S_n(\hat{\lambda}_n)Y_n-X_n\hat{\beta}_n = V_n+\left(\lambda_0-\hat{\lambda}_n\right)G_nV_n+X_n\left(\beta_0-\hat{\beta}_n\right)+(\lambda_0-\hat{\lambda}_n)G_nX_n\beta_0$, $\hat{v}_{ni}$ can be decomposed into three terms:

$$\hat{v}_{ni} = v_{ni}+b_{ni}+d_{ni}$$
$$b_{ni} = \left(\lambda_0-\hat{\lambda}_n\right)e_{i,n}G_nV_n$$
$$d_{ni} = e_{i,n}X_n\left(\beta_0-\hat{\beta}_n\right)+\left(\lambda_0-\hat{\lambda}_n\right)e_{i,n}G_nX_n\beta_0$$

where $e_{i,n}$ refers to the ith row in the $n\times n$ identity matrix. Thus $\hat{v}_{ni}^2 = v_{ni}^2+b_{ni}^2+d_{ni}^2+2v_{ni}b_{ni}+2v_{ni}d_{ni}+2b_{ni}d_{ni}$. Then,

$$\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\left(\hat{v}_{ni}^2-v_{ni}^2\right) = \frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\left(b_{ni}^2+d_{ni}^2+2v_{ni}b_{ni}+2v_{ni}d_{ni}+2b_{ni}d_{ni}\right),$$

which is $o_p(1)$ as follows. For illustration, we shall check the higher order terms of v_{ni}'s. For example,

$$\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}v_{ni}b_{ni} = \left(\lambda_0-\hat{\lambda}_n\right)\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}v_{ni}\left(e_{i,n}G_nV_n\right) = \left(\lambda_0-\hat{\lambda}_n\right)\frac{1}{n}\sum_{i=1}^n\sum_{l=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}G_{n,i,l}v_{ni}v_{nl}.$$

By Cauchy's inequality, $E|v_{ni}v_{nl}|\le(Ev_{ni}^2)^{\frac{1}{2}}(Ev_{nl}^2)^{\frac{1}{2}} = \sigma_{ni}\sigma_{nl}\le C$ for some constant C, uniformly in i, l and n since σ_{ni} is uniformly bounded, it follows that

$$E\left|\frac{1}{n}\sum_{i=1}^n\sum_{l=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}G_{n,i,l}v_{ni}v_{nl}\right|\le C\frac{1}{n}\sum_{i=1}^n\left|\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\right|\left(\sum_{l=1}^n|G_{n,i,l}|\right) = O(1).$$

Therefore, as $\lambda_0-\hat{\lambda}_n$ is $o_p(1)$, $\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}v_{ni}b_{ni} = o_p(1)$. Another higher order term of v_{ni} is $\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}b_{ni}^2$, which is $o_p(1)$ because

$$\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}b_{ni}^2 = \left(\lambda_0-\hat{\lambda}_n\right)^2\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}G_{n,i,j}G_{n,i,l}v_{nj}v_{nl}$$

and

$$E\left|\frac{1}{n}\sum_{i=1}^n\sum_{j=1}^n\sum_{l=1}^n\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}G_{n,i,j}G_{n,i,l}v_{nj}v_{nl}\right| \le C\frac{1}{n}\left(\sum_{i=1}^n\left|\left(P_{jn}^SU_{n|sar}^*S_n^{-1}\right)_{ii}\right|\right)\left(\sum_{j=1}^n\left|G_{n,ij}\right|\right)\left(\sum_{l=1}^n|G_{n,i,l}|\right) = O(1).$$

The term $\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SG_n\right)_{ii}\hat{v}_{ni}^2-\frac{1}{n}\sum_{i=1}^n\left(P_{jn}^SG_n\right)_{ii}\sigma_{ni}^2 = o_p(1)$ can be proved with similar arguments. Following all the above arguments, $\frac{1}{n}\left(\hat{D}_{nh|r_1}-D_{nh|r_1}\right) = o_p(1)$. Together, these prove Proposition 5. □

Proof of Proposition 6. The proof is similar to those proofs of Proposition 2 and Proposition 3 in Lin and Lee (2010). □

Proof of Proposition 7. The proof is similar to those proofs of Proposition 1 and Proposition 1 in Lin and Lee (2010). The GN2SLS is a special case of RGMM estimation with $a_n = \left(\frac{Q'_n \hat{\Sigma}_n Q_n}{n}\right)^{-\frac{1}{2}}$ and $\frac{1}{n} a_n g_n(\psi_{r_2}) = \left(\frac{Q'_n \hat{\Sigma}_n Q_n}{n}\right)^{-\frac{1}{2}} \frac{1}{n} Q'_n V_n(\psi_{r_2})$. □

References

Amemiya, T., 1985. Advanced Econometrics. Basil Blackwell, Oxford.

Anselin, L., 2003. Spatial externalities, spatial multipliers, and spatial econometrics. International Regional Science Review 26 (2), 153–166.

Arraiz, I., Drukker, D.M., Kelejian, H.H., Prucha, I.R., 2010. A spatial Cliff–Ord-type model with heteroskedastic innovations: small and large sample results. Journal of Regional Science 50, 592–614.

Atkinson, A., 1970. A method for discriminating between models (with discussion). Journal of the Royal Statistical Society, Series B 32, 323–353.

Burridge, P., 2012. Improving the *J* test in the SARAR model by likelihood-based estimation. Spatial Economic Analysis 7, 75–107.

Burridge, P., Fingleton, B., 2010. Bootstrap inference in spatial econometrics: the *J*-test. Spatial Economic Analysis 5, 93–119.

Cox, D., 1961. Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, 1, pp. 105–123.

Cox, D., 1962. Further results on tests of separate families of hypotheses. Journal of the Royal Statistical Society, Series B 24, 406–424.

Dastoor, N.K., 1983. Some aspects of testing non-nested hypotheses. Journal of Econometrics 21, 213–228.

Davidson, R., MacKinnon, J., 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 47, 781–793.

Davidson, R., MacKinnon, J., 1982. Some non-nested hypothesis tests and the relations among them. Review of Economic Studies 49, 551–565.

Davidson, R., MacKinnon, J., 1983. Testing the specification of multivariate models in the presence of alternative hypotheses. Journal of Econometrics 23, 301–313.

Davidson, R., MacKinnon, J., 2002a. Bootstrap J tests of nonnested linear regression models. Journal of Econometrics 109, 167–193.

Davidson, R., MacKinnon, J., 2002b. Fast double bootstrap tests of nonnested linear regression models. Econometric Reviews 21, 417–427.

Davidson, R., MacKinnon, J., 2004. Econometric Theory and Methods. Oxford University Press, New York.

Deaton, A.S., 1982. Model selection procedures, or, does the consumption function exist? In: Chow, G.C., Corsi, P. (Eds.), Evaluating the Reliability of Macroeconomic Models. Wiley, New York.

Fisher, G., McAleer, M., 1981. Alternative procedures and associated tests of significance for non-nested hypotheses. Journal of Econometrics 16, 103–119.

Godfrey, L.G., 1983. Testing non-nested models after estimation by instrumental variables or least squares. Econometrica 51, 355–365.

Godfrey, L.G., 1998. Tests of non-nested regression models: some results on small sample behaviour and the bootstrap. Journal of Econometrics 84, 59–74.

Godfrey, L.G., Pesaran, M.H., 1983. Tests of non-nested regression models: small sample adjustments and Monte Carlo evidence. Journal of Econometrics 21, 133–154.

Gourieroux, C., Monfort, A., 1994. Testing non-nested hypotheses. In: Engle, R.F., McFadden, D.L. (Eds.), Handbook of Econometrics. North-Holland Elsevier, Amsterdam.

Hansen, L., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50, 1029–1054.

Hepple, L.W., 1995a. Bayesian techniques in spatial and network econometrics: 1. Model comparison and posterior odds. Environment and Planning A 27, 447–469.

Hepple, L.W., 1995b. Bayesian techniques in spatial and network econometrics: 2. Computational methods and algorithms. Environment and Planning A 27, 615–644.

Kelejian, H.H., 2008. A spatial *J*-test for model specification against a single or a set of non-nested alternatives. Letters in Spatial and Resource Sciences 1, 3–11.

Kelejian, H.H., Piras, G., 2011. An extension of Kelejian's *J*-test for non-nested spatial models. Regional Science and Urban Economics 41, 281–292.

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

Kelejian, H.H., Prucha, I.R., 2001. On the asymptotic distribution of the Moran *I* test statistic with applications. Journal of Econometrics 104, 219–257.

Kelejian, H.H., Prucha, I.R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. Journal of Econometrics 157, 53–67.

Lee, L.F., 2004. Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Econometrica 72, 1899–1925.

Lee, L.F., 2007. GMM and 2SLS estimation of mixed regressive, spatial autoregressive models. Journal of Econometrics 137, 489–514.

LeSage, J., Pace, R.K., 2007. A matrix exponential spatial specification. Journal of Econometrics 140, 190–214.

LeSage, J., Pace, R.K., 2009. Introduction to Spatial Econometrics. CRC Press, Boca Raton.

LeSage, J., Parent, O., 2007. Bayesian model averaging for spatial econometric models. Geographical Analysis 39, 241–267.

Lin, X., Lee, L.F., 2010. GMM estimation of spatial autoregressive models with unknown heteroskedasticity. Journal of Econometrics 157, 34–52.

Liu, X., Patacchini, E., Zenou, Y., 2011. Peer effects in education, sport, and screen activities: local aggregate or local average? Working paper.

MacKinnon, J., 2009. Bootstrap hypothesis testing. In: Belsley, D.A., Kontoghiorghes, E.J. (Eds.), Handbook of Computational Econometrics. John Wiley & Sons, Ltd., Chichester, UK, pp. 183–213.

Mizon, G.E., Richard, J.F., 1986. The encompassing principle and its application to non-nested hypotheses. Econometrica 54, 654–678.

Newey, W., West, K., 1987. Hypothesis testing with efficient method of moments testing. International Economic Review 28, 777–787.

Pesaran, M.H., 1974. On the general problem of model selection. Review of Economic Studies 41, 153–171.

Pesaran, M.H., Dupleich Ulloa, M.R., 2008. Non-nested hypotheses, In: Durlauf, S.N., Blume, L.E. (Eds.), The New Palgrave Dictionary of Economics, Second edition. Palgrave Macmillan.

Pesaran, M.H., Weeks, M., 2001. Nonnested hypothesis testing: an overview. In: Baltagi, B.H. (Ed.), A Companion to Theoretical Econometrics. Basil Blackwell, Oxford.

Piras, G., Lozano-Gracia, N., 2012. Spatial *J*-test: some Monte Carlo evidence. Statistics and Computing 22, 169–183.

White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica 48, 817–838.

White, H., 1994. Estimation, Inference and Specification Analysis. Cambridge University Press, New York.

Zellner, A., 1971. An Introduction to Bayesian Inference in Econometrics. J. Wiley and Sons, Inc., New York.