Cox-type Tests for Competing Spatial Autoregressive Models with Spatial Autoregressive Disturbances[☆]

Fei Jin[a,*], Lung-fei Lee[a]

[a] *Department of Economics, The Ohio State University, Columbus, OH 43210 USA*

Abstract

In this paper, we consider the Cox-type tests of non-nested hypotheses for spatial autoregressive (SAR) models with SAR disturbances. We formally derive the asymptotic distributions of the test statistics. In contrast to regression models, we show that the Cox-type and J-type tests for non-nested hypotheses in the framework of SAR models are not asymptotically equivalent under the null hypothesis. The Cox test in non-spatial setting has been found often to have large size distortion, which can be removed by the bootstrap. Cox-type tests for SAR models with SAR disturbances may also have large size distortion. We show that the bootstrap is consistent for Cox-type tests in our framework. Performances of the Cox-type and J-type tests as well as their bootstrapped versions in finite samples are compared via a Monte Carlo study. These tests are of particular interest when there are competing models with different spatial weights matrices. Using bootstrapped p-values, the Cox tests have relatively high power in all experiments and can outperform J-type and several other related tests in some cases.

Keywords: Specification, Spatial autoregressive model, Non-nested, Cox test, J test, QMLE

JEL classification: C12, C21, C52, R15

1. Introduction

There are three general approaches in testing non-nested hypotheses: the centered log-likelihood ratio procedure, known as the Cox test (Cox, 1961, 1962); the comprehensive model approach, which involves constructing artificial general models including non-nested models as special cases (Cox, 1962; Atkinson, 1970); and the encompassing approach that tests directly the ability of one model to explain features of an alternative model (Deaton, 1982; Dastoor, 1983; Mizon and Richard, 1986; Gourieroux and Monfort, 1995).[1] In a contribution related to the encompassing approach, Gourieroux et al. (1983) extend the Wald

[☆] We are grateful to the editor and two anonymous referees for their helpful comments.

[*] Corresponding author. Tel.: +1 614 247 8481; fax: +1 614 292 3906.

Email addresses: jin.120@osu.edu (Fei Jin), lee.1777@osu.edu (Lung-fei Lee)

[1] For the definition and overviews of non-nested hypotheses, see McAleer and Pesaran (1986), Gourieroux and Monfort (1994), Pesaran and Weeks (2001), Pesaran and Dupleich Ulloa (2008), among others.

and score tests to non-nested hypotheses based on the difference between two estimators for the alternative model. The comprehensive model approach suffers from the Davies's problem (Davies, 1977), which can be circumvented in various ways. Davidson and MacKinnon (1981)'s J test can be seen as a way to deal with the problem. These well-established procedures may also be very useful for model specifications in spatial econometrics.

There are many spatial econometric models, e.g., spatial autoregressive models, spatial moving average models (Cliff and Ord, 1981) and spatial error components models (Kelejian and Robinson, 1993), that cannot nest other models as special cases. In addition, spatial econometric models usually involve spatial weights matrices which are assumed to be exogenous. As economic theories are often ambiguous about spatial weights, we may construct spatial weights matrices in different ways, which also lead to non-nested models. The J test, as the most widely used procedure for testing non-nested hypotheses due to its simplicity (McAleer, 1995), has been discussed in spatial econometrics by several authors, while other procedures have seldom been focused on.[2] Anselin (1984) illustrates the use of the J test for spatial autoregressive (SAR) models with an empirical example and Anselin (1986) presents Monte Carlo results of the J-type tests for SAR models where only an intercept term is included as the exogenous variable. Kelejian (2008) formally extends the J test to SAR models with SAR disturbances (SARAR models, for short). Piras and Lozano-Gracia (2012) present some Monte Carlo evidence in support of Kelejian's spatial J test. Burridge (2012) proposes to improve Kelejian's spatial J test by using parameter estimates constructed from the likelihood based moment conditions. Kelejian and Piras (2011) modify Kelejian (2008)'s spatial J test so that available information is used in a more effective way and thus may have higher power in finite samples. Liu et al. (2013) extend Kelejian (2008)'s spatial J test to differentiate between models with a non-row-normalized spatial weights matrix versus a row-normalized one in a social-interaction model. No formal results on other non-nested procedures, as far as we are aware of, have been derived for spatial econometric models.

In this paper, we derive asymptotic distributions of the Cox-type tests for SARAR models and compare them with spatial J test statistics. It is of interest to derive the Cox-type test statistics. For regression models, it has been established that the Cox and J statistics are asymptotically equivalent under the null hypothesis (Atkinson, 1970; Davidson and MacKinnon, 1981; Gourieroux and Monfort, 1994). For the SARAR models, we shall show that the Cox statistics and the proposed spatial J test statistics in Kelejian (2008) and Kelejian and Piras (2011) are, in general, not asymptotically equivalent under the null hypothesis. The different ways that the Cox-type tests use available information might lead to distinct size and power properties. For comparison purposes, we also present the extended Wald and extended score tests (Gourieroux et al., 1983) for the SARAR models as supplements (in Appendix B).

For the non-spatial setting, many Monte Carlo experiments (see, e.g., Godfrey and Pesaran 1983) have

[2]See Anselin (1984) for a general discussion of applying tests of non-nested hypotheses in spatial econometrics.

shown that the Cox and J tests can have large size distortion and typically reject a true null hypothesis too frequently. Horowitz (1994) considers the use of the bootstrap in econometric testing and finds that it can overcome the well-known problem of the excessive size of variants of the information matrix test. Fan and Li (1995) and Godfrey (1998) have suggested bootstrapping the J test and other non-nested hypothesis tests. Davidson and MacKinnon (2002) provides a theoretical analysis of why bootstrapping the J test often works well. Burridge and Fingleton (2010) numerically demonstrate that Kelejian (2008)'s spatial J test is excessively liberal in some leading cases and the bootstrap approach is superior to the asymptotic test. For spatial econometric models, Jin and Lee (2012) have shown that the bootstrap is in general consistent for statistics that may be approximated by a linear-quadratic form of disturbances.[3] Using the result, we show that the bootstrap is consistent for Cox-type tests in our framework. We compare the finite sample performances of various tests as well as their bootstrapped versions by a Monte Carlo study. Our Monte Carlo experiments show that although the Cox-type tests have larger size distortions than the J-type tests in some cases, the bootstrap can essentially remove size distortions of both types of tests. The bootstrapped Cox-type tests have relatively high power in all experiments and outperform the bootstrapped J-type and several other tests in some cases.

The rest of the paper is laid out as follows. Section 2 formally derives the asymptotical distributions of the Cox-type test statistics. Section 3 shows that the Cox-type and J-type tests for SARAR models are not asymptotically equivalent under the null hypothesis, and also briefly compares the two types of tests. Section 4 shows that the bootstrap is consistent for Cox-type tests. Section 5 compares the performances of various test statistics as well as their bootstrapped versions in finite samples by a Monte Carlo study. Section 6 illustrates the use of Cox-type tests with a housing data set. Finally, Section 7 concludes. Some assumptions, expressions, lemmas and proofs are collected in the appendices.

2. Cox-type Tests

We derive the Cox-type tests for SARAR models in this section. The setting of the non-nested testing problem is as follows. A SARAR model as the null hypothesis H_0 is tested against another SARAR model as the alternative hypothesis H_1:

$$H_0: \quad y_n = \lambda_1 W_{1n} y_n + X_{1n}\beta_1 + u_{1n}, \quad u_{1n} = \rho_1 M_{1n} u_{1n} + \epsilon_{1n}, \tag{1}$$

[3] Consistency of the bootstrap for a statistic means that the bootstrap can provide a consistent estimator for the asymptotic distribution of the statistic. On the question that whether the bootstrap can provide asymptotic refinements, i.e., whether the bootstrap can be more accurate than the first-order asymptotic theory, only preliminary results are available. Jin and Lee (2012) establish the Edgeworth expansion for a linear-quadratic form with normal disturbances, which can be used to show the asymptotic refinements of the bootstrap for a linear-quadratic form. Then for a statistic that can be approximated by a linear-quadratic form, with proper regularity conditions on the remainder term, the bootstrap can provide asymptotic refinements. For a linear-quadratic form with non-normal disturbances, the Edgeworth expansion has not be established.

$$H_1: \quad y_n = \lambda_2 W_{2n} y_n + X_{2n}\beta_2 + u_{2n}, \quad u_{2n} = \rho_2 M_{2n} u_{2n} + \epsilon_{2n}, \tag{2}$$

where n is the sample size, y_n is an n-dimensional vector of observations, W_{jn} and M_{jn} are $n \times n$ spatial weights matrices with zero diagonals, X_{jn} is an $n \times k_j$ matrix of exogenous variables, elements of an n-dimensional vector of disturbances ϵ_{jn} are i.i.d. with mean zero and finite variance σ_j^2, and $\theta_j = (\lambda_j, \rho_j, \beta_j', \sigma_j^2)'$ for $j = 1, 2$ are vectors of parameters to be estimated. Denote $S_{jn}(\lambda_j) = I_n - \lambda_j W_{jn}$ and $R_{jn}(\rho_j) = I_n - \rho_j M_{jn}$ with I_n being an $n \times n$ identity matrix. Let the true parameter vector of the model (1) be θ_{10}, $S_{1n} = S_{1n}(\lambda_{10})$ and $R_{1n} = R_{1n}(\rho_{10})$ for short. The X_{1n} and X_{2n} may have different dimensions. The W_{jn} and M_{jn} are in general different, but could be the same in empirical applications. A particularly interesting case in practice is the one in which we have different spatial weights matrices W_{1n} vs W_{2n} or M_{1n} vs M_{2n} in the two models. Let $L_{jn}(\theta_j)$ be the log likelihood function of the model (j), for $j = 1, 2$, as if the disturbances were normally distributed:

$$\begin{aligned} L_{jn}(\theta_j) = &-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_j^2 + \ln|S_{jn}(\lambda_j)| + \ln|R_{jn}(\rho_j)| \\ &- \frac{1}{2\sigma_j^2}[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j]' R_{jn}'(\rho_j) R_{jn}(\rho_j)[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j]. \end{aligned} \tag{3}$$

Let $\hat{\theta}_{jn}$ be the corresponding quasi-maximum likelihood estimator (QMLE) by maximizing $L_{jn}(\theta_j)$. The idea of the Cox-type tests is to modify the log-likelihood ratio $[L_{2n}(\hat{\theta}_{2n}) - L_{1n}(\hat{\theta}_{1n})]$ so that it is approximately centered at zero under the null hypothesis, and then test whether the modified statistic after being properly scaled is significantly different from zero.[4] As the test statistics involve the QMLEs $\hat{\theta}_{1n}$ and $\hat{\theta}_{2n}$, we first investigate their properties, and then derive the Cox-type test statistics with the QMLEs.

For a correctly specified first order SAR model without spatially correlated disturbances, Lee (2004a) has proved that the QMLE is consistent under suitable regularity conditions. We can extend the analysis to SARAR models. When we estimate the alternative model, generally it might have a different number of parameters and/or variables from that of the data generating process (DGP), let alone the consistency to the true values of the DGP. We use the so-called pseudo-true values to study the behavior of the QMLE for the alternative model.[5] For the model (2), we define the pseudo true value $\bar{\theta}_{2n,1}$ to be the vector that maximizes $\mathrm{E}\, L_{2n}(\theta_2)$, and we shall show that $n^{1/2}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1})$ is asymptotically normal. With the pseudo-true values, we can derive the asymptotic distribution of the Cox-type test statistics by using the central limit theorem for linear-quadratic forms $\epsilon_n' A_n \epsilon_n - \sigma_0^2 \operatorname{tr}(A_n) + b_n' \epsilon_n$ (Kelejian and Prucha, 2001), where ϵ_n is an n-dimensional vector of i.i.d. disturbances with mean zero and variance σ_0^2, and the elements of the

[4] Since the data generating process is not assumed to have normally distributed disturbances and we will construct the tests with the centered log quasi-maximum likelihood ratio, the tests correspond to Aguirre-Torres and Gallant (1983)'s generalized, distribution-free Cox tests.

[5] For the definition of pseudo-true values, see, e.g., Sawa (1978) and White (1982). The pseudo-true values are often used for non-nested hypothesis testing problems, see, among others, Gourieroux et al. (1983) and Gourieroux and Monfort (1994).

$n \times n$ matrix A_n and n-dimensional vector b_n are all non-stochastic.[6]

Similar to that in Lee (2004a), the consistency of $\hat{\theta}_{1n}$ can be established by investigating the concentrated log likelihood function $L_{1n}(\phi_1) = \max_{\beta_1, \sigma_1^2} L_{1n}(\theta_1)$ with $\phi_1 = (\lambda_1, \rho_1)'$. For $i, j = 1, 2$, let $\bar{L}_{jn}(\theta_j; \theta_i)$ be the expected value of $L_{jn}(\theta_j)$ when the model (i) with parameter θ_i generates the data. Thus, in particular, $\bar{L}_{1n}(\theta_1; \theta_{10}) = \mathrm{E}\, L_{1n}(\theta_1)$ and $\bar{L}_{2n}(\theta_2; \theta_{10}) = \mathrm{E}\, L_{2n}(\theta_2)$. Denote $\bar{L}_{jn}(\phi_j; \theta_{10}) = \max_{\beta_j, \sigma_j^2} \bar{L}_{jn}(\theta_j; \theta_{10})$ with $\phi_j = (\lambda_j, \rho_j)'$ for $j = 1, 2$. We make the following assumptions for the consistency of $\hat{\theta}_{1n}$.

Assumption 1. *$\{\epsilon_{1n,i}\}$'s in $\epsilon_{1n} = (\epsilon_{1n,1}, \ldots, \epsilon_{1n,n})'$, $i = 1, \ldots, n$, are i.i.d. with mean zero and variance σ_{10}^2. The moment $\mathrm{E}(\epsilon_{1n,i}^{4+\zeta})$ for some $\zeta > 0$ exists.*

Assumption 2. *The elements of X_{1n} are uniformly bounded constants, X_{1n} has full column rank k_1, and $\lim_{n\to\infty} \frac{1}{n} X_{1n}' X_{1n}$ exists and is nonsingular.*

Assumption 3. *Matrices S_{1n} and R_{1n} are nonsingular.*

Assumption 4. *$\{W_{1n}\}$ and $\{M_{1n}\}$ have zero diagonals. The sequences of matrices $\{W_{1n}\}$, $\{M_{1n}\}$, $\{R_{1n}^{-1}\}$ and $\{S_{1n}^{-1}\}$ are bounded in both row and column sum norms (for short, UB).*[7]

Assumption 5. *$\{S_{1n}^{-1}(\lambda_1)\}$ is bounded in either row or column sum norm uniformly in λ_1 in a compact parameter space Λ_1, and $\{R_{1n}^{-1}(\rho_1)\}$ is bounded in either row or column sum norm uniformly in ρ_1 in a compact parameter space ϱ_1. The true λ_{10} is in the interior of Λ_1 and the true ρ_{10} is in the interior of ϱ_1.*

Assumption 6. *The limit $\lim_{n\to\infty} \frac{1}{n} X_{1n}' R_{1n}'(\rho_1) R_{1n}(\rho_1) X_{1n}$ exists and is nonsingular for any $\rho_1 \in \varrho_1$, and the sequence of the smallest eigenvalues of $R_{1n}'(\rho_1) R_{1n}(\rho_1)$ is bounded away from zero uniformly in ρ_1.*[8]

Assumption 7. *Either (i) $\lim_{n\to\infty} \frac{1}{n}[\ln|\sigma_{10}^2 S_{1n}^{-1} R_{1n}^{-1} R_{1n}'^{-1} S_{1n}'^{-1}| - \ln|\bar{\sigma}_{1n,a}^2(\phi_1) S_{1n}^{-1}(\lambda_1) R_{1n}^{-1}(\rho_1) R_{1n}'^{-1}(\rho_1) S_{1n}'^{-1}(\lambda_1)|]$ exists and is nonzero for any $\phi_1 \neq \phi_{10}$, where $\bar{\sigma}_{1n,a}^2(\phi_1) = \frac{\sigma_{10}^2}{n} \operatorname{tr}[R_{1n}'^{-1} S_{1n}'^{-1} S_{1n}'(\lambda_1) R_{1n}'(\rho_1) R_{1n}(\rho_1) S_{1n}(\lambda_1) S_{1n}^{-1} R_{1n}^{-1}]$, or (ii) $\lim_{n\to\infty} \frac{1}{n}(Q_{1n} X_{1n} \beta_{10}, X_{1n})'(Q_{1n} X_{1n} \beta_{10}, X_{1n})$ exists and is nonsingular, and for any $\rho_1 \neq \rho_{10}$, $\lim_{n\to\infty} \frac{1}{n}[\ln|\sigma_{10}^2 S_{1n}^{-1} R_{1n}^{-1} R_{1n}'^{-1} S_{1n}'^{-1}| - \ln|\bar{\sigma}_{1n,a}^2(\lambda_{10}, \rho_1) S_{1n}^{-1} R_{1n}^{-1}(\rho_1) R_{1n}'^{-1}(\rho_1) S_{1n}'^{-1}|]$ exists and is nonzero, where $Q_{1n} = W_{1n} S_{1n}^{-1}$.*

---

[6]In Kelejian and Piras (2011), the pseudo-true values are not explicitly discussed for spatial J tests. This is because their tests are based on two-stage least squares (2SLS) estimators, which have closed forms. Thus, by assuming that some matrices involving the estimators for the alternative model converge to positive definite matrices in probability, there is no need to explicitly consider the pseudo-true values.

[7]A sequence of $n \times n$ matrices $\{A_n = [a_{n,ij}]\}$ is bounded in row sum norm if there is a constant c such that $\sup_{1\le i\le n} \sum_{j=1}^n |a_{n,ij}| < c$ for all n, and is bounded in column sum norm if there is a constant c such that $\sup_{1\le j\le n} \sum_{i=1}^n |a_{n,ij}| < c$. See Horn and Johnson (1985).

[8]Let μ_{n,ρ_1} be the smallest eigenvalue of $R_{1n}'(\rho_1) R_{1n}(\rho_1)$. Then the second part of the assumption means that there is some constant $c > 0$ such that $\inf_{\rho_1 \in \varrho_1} \mu_{n,\rho_1} > c$ for all n.

Assumptions 1–5 are similar to those in Lee (2004a), except for the additional conditions on $R_{1n}(\rho_1)$ which resemble those on $S_{1n}(\lambda_1)$. In practice, the λ and ρ are typically assumed to be in the interval $(-1, 1)$ such that $|S_{1n}(\lambda_1)|$ and $|R_{1n}(\rho_1)|$ are positive, while for the theoretical purpose, the parameter space can be taken to be the compact interval contained in $(-1, 1)$ so that the consistency of the estimator would still hold.[9] Note that $R_{1n}(\rho_1)$ is linear in ρ_1, a sufficient condition for the first part of Assumption 6 is that the limit of $n^{-1}X'_{1n}[X_{1n}, (M'_{1n} + M_{1n})X_{1n}, M'_{1n}M_{1n}X_{1n}]$ exists and has full column rank.[10] The second part of Assumption 6 is required to guarantee the uniform convergence of $\frac{1}{n}[L_{1n}(\phi_1) - \bar{L}_{1n}(\phi_1; \theta_{10})]$ to zero in probability. As $R'_{1n}(\rho_1)R_{1n}(\rho_1)$ is positive semi-definite, its eigenvalues are non-negative. The assumption further limits the eigenvalues to be strictly positive for all n. Assumption 7 provides sufficient conditions for global identification, where (i) is related to the uniqueness of the variance-covariance (VC) matrix of y_n and (ii) states that a part of the identification can be from the asymptotically non-multicollinearity of $Q_{1n}X_{1n}\beta_{10}$ and X_{1n}. The first part of (ii) does not hold if X_{1n} contains a vector of ones and W_{1n} is a matrix of equal weights.[11]

Proposition 1. *Under H_0 and Assumptions 1–7, $\hat{\theta}_{1n} - \theta_{10} = o_P(1)$.*

The asymptotic distribution of $\hat{\theta}_{1n}$ can be derived by applying the mean value theorem to the first order condition $\frac{\partial L_{1n}(\hat{\theta}_{1n})}{\partial \theta_1} = 0$ at the true θ_{10}:

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) = -\Big(\frac{1}{n}\frac{\partial^2 L_{1n}(\tilde{\theta}_{1n})}{\partial \theta_1 \partial \theta'_1}\Big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1}, \tag{4}$$

[9]To make $|S_{1n}(\lambda_1)|$ positive, the admissible interval for λ_1 is $(1/\mu_{n,\min}, 1/\mu_{n,\max})$, where $\mu_{n,\min}$ and $\mu_{n,\max}$ are, respectively, the minimum and maximum real eigenvalue of W_n. If W_n with non-negative elements is row normalized, then $\mu_{n,\max} = 1$ and $-1 \le \mu_{n,\min} < 0$. Thus the interval is $(1/\mu_{n,\min}, 1)$, where $1/\mu_{n,\min} \le -1$. The admissible interval for ρ_1 is similar, thus we only focus on the admissible interval for λ_1. The concentrated quasi log likelihood function over n is $\frac{1}{n}L_{1n}(\phi_1) = -\frac{1}{2}[\ln(2\pi) + 1] - \frac{1}{2}\ln\hat{\sigma}^2_{1n}(\phi_1) + \frac{1}{n}\ln|S_{1n}(\lambda_1)| + \frac{1}{n}\ln|R_{1n}(\rho_1)|$, where $\hat{\sigma}^2_{1n}(\phi_1) = n^{-1}y'_nS'_{1n}(\lambda_1)R'_{1n}(\rho_1)H_{1n}(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)y_n$ with $H_{1n}(\rho_1) = I_n - R_{1n}(\rho_1)X_{1n}[X'_{1n}R'_{1n}(\rho_1)R_{1n}(\rho_1)X_{1n}]^{-1}X'_{1n}R'_{1n}(\rho_1)$, from (A.1). By the proof of Proposition 3, $\hat{\sigma}^2_{1n}(\phi_1) - \bar{\sigma}^2_{1n}(\phi_1; \theta_{10}) = o_P(1)$, where $\bar{\sigma}^2_{1n}(\phi_1; \theta_{10}) = \frac{\sigma^2_{10}}{n}\,\mathrm{tr}[R'^{-1}_{1n}S'^{-1}_{1n}S'_{1n}(\lambda_1)R'_{1n}(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S^{-1}_{1n}R^{-1}_{1n}] + \frac{1}{n}(X_{1n}\beta_{10})'S'^{-1}_{1n}S'_{1n}(\lambda_1)R'_{1n}(\rho_1)H_{1n}(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S^{-1}_{1n}X_{1n}\beta_{10}$ is bounded away from zero. Then $\ln\hat{\sigma}^2_{1n}(\phi_1)$ is bounded in probability. In the case that $\mu_{n,\max} = 1$, when λ_1 approaches 1, $\frac{1}{n}\ln|S_{1n}(\lambda_1)|$ approaches minus infinity, thus $\frac{1}{n}L_{1n}(\phi_1)$ approaches minus infinity in probability, which implies that $\frac{1}{n}L_{1n}(\phi_1)$ at a λ_1 very close to 1 will be smaller than its value at some λ_1 in the interior of $(-1, 1)$ in probability one. Similarly, when $1/\mu_{n,\min} = -1$, $\frac{1}{n}L_{1n}(\phi_1)$ approaches minus infinity in probability as λ_1 approaches -1. When $1/\mu_{n,\min} < -1$, $|S_{1n}(\lambda_1)|$ at -1 is positive and finite. Thus the interval for λ_1 can be taken to be $(-1, 1)$ in practice, while it makes no harm to assume the parameter space to be compact. This view is in Amemiya (1985, p. 108). In this paper, the QMLE is proved to be consistent only for a compact parameter space.

[10]When X_{1n} contains a vector of ones and M_{1n} is a matrix of equal weights, $n^{-1}X'_{1n}[X_{1n}, (M'_{1n} + M_{1n})X_{1n}, M'_{1n}M_{1n}X_{1n}]$ doest not have full column rank, but the first part of Assumption 6 may still hold in this case.

[11]The condition is equivalent to that the limit $n^{-1}[Q_{1n}X_{1n}\beta_{10}]'M_{X_{1n}}Q_{1n}X_{1n}\beta_{10}$ exists and is non-zero when the limit of $n^{-1}X'_{1n}X_{1n}$ exists and is nonsingular, where $M_{X_{1n}} = I_n - X_{1n}(X'_{1n}X_{1n})^{-1}X'_{1n}$. Let $W_{1n} = (l_nl'_n - I_n)/(n-1)$, where l_n is an n-dimensional vector of ones. Then $M_{X_{1n}}W^k_{1n} = (1-n)^{-k}M_{X_{1n}}$. Thus $M_{X_{1n}}Q_{1n}X_{1n}\beta_{10} = 0$ and $n^{-1}[Q_{1n}X_{1n}\beta_{10}]'M_{X_{1n}}Q_{1n}X_{1n}\beta_{10} = 0$.

where $\tilde{\theta}_{1n}$ is between $\hat{\theta}_{1n}$ and θ_{10}.[12] In the above equation, every element of $\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1}$ is a linear-quadratic form of the disturbances ϵ_{1n}, thus the central limit theorem in Kelejian and Prucha (2001) is applicable.[13] The term $\frac{1}{n}\frac{\partial^2 L_{1n}(\tilde{\theta}_{1n})}{\partial \theta_1 \partial \theta_1'}$ can be shown (see the proof of Proposition 4) to be equal to $\frac{1}{n}\,\mathrm{E}\big(\frac{\partial^2 L_{1n}(\theta_{10})}{\partial \theta_1 \partial \theta_1'}\big)$ plus a term converging to zero in probability. The following assumption is needed for the limit of $\Sigma_{1n,1} = -\frac{1}{n}\,\mathrm{E}\big(\frac{\partial^2 L_{1n}(\theta_{10})}{\partial \theta_1 \partial \theta_1'}\big)$ to exist and be nonsingular.

Assumption 8. *The limit* $\lim_{n\to\infty}\frac{1}{n}\frac{\partial^2 \bar{L}_{1n}(\phi_{10};\theta_{10})}{\partial \phi_1 \partial \phi_1'}$ *exists and is nonsingular.*

Proposition 2. *Under* H_0 *and Assumptions 1–8,*

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) \xrightarrow{d} N\big(0, \lim_{n\to\infty}(\Sigma_{1n,1}^{-1}\Omega_{1n,1}\Sigma_{1n,1}^{-1})\big), \tag{5}$$

where $\Omega_{1n,1} = \frac{1}{n}\,\mathrm{E}\big(\frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1}\frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1'}\big)$ *and* $\Sigma_{1n,1} = -\frac{1}{n}\,\mathrm{E}\big(\frac{\partial^2 L_{1n}(\theta_{10})}{\partial \theta_1 \partial \theta_1'}\big)$. *In the case that* $\epsilon_{1n,i}$*'s are normally distributed,* $\sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) \xrightarrow{d} N(0, \lim_{n\to\infty}\Sigma_{1n,1}^{-1})$.

The $\Omega_{1n,1}$ generally involves the third and fourth moments of the disturbances if they are not normally distributed, thus it has a form more complicated than that of $\Sigma_{1n,1}$. When $\epsilon_{1n,i}$'s are normally distributed, the information matrix equality holds, i.e., $\Sigma_{1n,1} = \Omega_{1n,1}$, so the VC matrix has a simpler form.

For the alternative model (2), the following assumptions are made for the convergence of $\hat{\theta}_{2n} - \bar{\theta}_{2n,1}$ to zero in probability under the null hypothesis of the model (1). Denote $S_{2n} = S_{2n}(\bar{\lambda}_{2n,1})$ and $R_{2n} = R_{2n}(\bar{\rho}_{2n,1})$ for short.

Assumption 9. *The elements of* X_{2n} *are uniformly bounded constants,* X_{2n} *has full column rank* k_2*, and* $\lim_{n\to\infty}\frac{1}{n}X_{2n}'X_{2n}$ *exists and is nonsingular.*

Assumption 10. *Matrices* S_{2n} *and* R_{2n} *are nonsingular.*

Assumption 11. $\{W_{2n}\}$ *and* $\{M_{2n}\}$ *have zero diagonals. The sequences of matrices* $\{W_{2n}\}$, $\{M_{2n}\}$, $\{R_{2n}^{-1}\}$ *and* $\{S_{2n}^{-1}\}$ *are UB.*

Assumption 12. $\{S_{2n}^{-1}(\lambda_2)\}$ *is bounded in either row or column sum norm uniformly in* λ_2 *in a compact parameter space* Λ_2*, and* $\{R_{2n}^{-1}(\rho_2)\}$ *is bounded in either row or column sum norm uniformly in* ρ_2 *in a compact parameter space* ϱ_2.

Assumption 13. *The limit* $\lim_{n\to\infty}\frac{1}{n}X_{2n}'R_{2n}'(\rho_2)R_{2n}(\rho_2)X_{2n}$ *exists and is nonsingular for any* $\rho_2 \in \varrho_2$*, and the sequence of the smallest eigenvalues of* $R_{2n}'(\rho_2)R_{2n}(\rho_2)$ *is bounded away from zero uniformly in* ρ_2.

[12] The mean value theorem is applicable to a function but not a vector-valued mapping. So $\tilde{\theta}_{1n}$ can be different for each row of the Hessian matrix.

[13] The expressions for $\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1}$ and some other terms in the text are collected in Appendix A.

Assumption 14. *For $\eta > 0$, there exists $\kappa > 0$ such that, when $||\phi_2 - \bar{\phi}_{2n,1}|| > \eta$, $n^{-1}\big(\bar{L}_{2n}(\bar{\phi}_{2n,1};\theta_{10}) - \bar{L}_{2n,1}(\phi_2;\theta_{10})\big) > \kappa$ for any large enough n.*

Assumption 15. *The limit of $n^{-1}\operatorname{tr}[R_{1n}'^{-1}S_{1n}'^{-1}S_{2n}'R_{2n}'R_{2n}S_{2n}S_{1n}^{-1}R_{1n}^{-1}]$ or $n^{-1}(X_{1n}\beta_{10})'S_{1n}'^{-1}S_{2n}'R_{2n}'H_{2n}R_{2n}S_{2n}S_{1n}^{-1}X_{1n}\beta_{10}$ exists and is non-zero.*

Assumptions 9–13 are similar to those for the estimation of the model (1). With a misspecified model being estimated, it is not straightforward to find primitive identification conditions, so Assumption 14 is imposed. Assumption 15 implies that $\{\bar{\sigma}^2_{2n,1}\}$, the sequence of pseudo true values for σ_2^2, is bounded away from zero by (A.4), which is necessary to prove the uniform convergence of $n^{-1}\big(L_{2n}(\phi_2) - \bar{L}_{2n}(\phi_2;\theta_{10})\big)$ to zero in probability on $\Lambda_2 \times \varrho_2$. Without this assumption, $n^{-1}\bar{L}_{2n}(\phi_2;\theta_{10})$ can be arbitrarily large.

Proposition 3. *Under H_0 and Assumptions 1–4, and 9–15, $\hat{\theta}_{2n} - \bar{\theta}_{2n,1} = o_P(1)$.*

The asymptotic distribution for $\hat{\theta}_{2n} - \bar{\theta}_{2n,1}$ can be derived by an expansion of the first order condition that $\frac{\partial L_{2n}(\hat{\theta}_{2n})}{\partial\theta_2} = 0$ at $\bar{\theta}_{2n,1}$:

$$\sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1}) = -\Big(\frac{1}{n}\frac{\partial^2 L_{2n}(\tilde{\theta}_{2n})}{\partial\theta_2\partial\theta_2'}\Big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2}, \tag{6}$$

where $\tilde{\theta}_{2n}$ is between $\hat{\theta}_{2n}$ and $\bar{\theta}_{2n,1}$. Noting that $\frac{\partial \operatorname{E} L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2} = 0$ and $\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2} = \frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2} - \frac{\partial \operatorname{E} L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2}$, every element of $\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2}$ can be written as a linear-quadratic form of the vector of disturbances ϵ_{1n}. Since $\frac{1}{n}\frac{\partial^2 L_{2n}(\tilde{\theta}_{2n})}{\partial\theta_2\partial\theta_2'} = \frac{1}{n}\operatorname{E}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'} + o_P(1)$, we make the following assumption which guarantees that $\frac{1}{n}\operatorname{E}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'}$ is nonsingular in the limit.

Assumption 16. *The limit $\lim_{n\to\infty}\frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\bar{\phi}_{2n,1};\theta_{10})}{\partial\phi_2\partial\phi_2'}$ exists and is nonsingular.*

Proposition 4. *Under H_0 and Assumptions 1–4, 9–16,*

$$\sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1}) \xrightarrow{d} N\big(0, \lim_{n\to\infty}(\Sigma_{2n,1}^{-1}\Omega_{2n,1}\Sigma_{2n,1}^{-1})\big), \tag{7}$$

where $\Sigma_{2n,1} = -\frac{1}{n}\operatorname{E}\big(\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'}\big)$ and $\Omega_{2n,1} = \frac{1}{n}\operatorname{E}\big(\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2'}\big)$.

With asymptotic distributions of the estimators, we are now ready to derive the Cox-type test statistics. As mentioned earlier, the Cox-type tests are based on the recentered log likelihood ratio $L_{2n}(\hat{\theta}_{2n}) - L_{1n}(\hat{\theta}_{1n})$. Thus we need to find an expression for the asymptotic mean of the ratio. Because of the results in Propositions 2 and 4, we shall show that $n^{-1/2}[L_{2n}(\hat{\theta}_{2n}) - L_{1n}(\hat{\theta}_{1n})] = n^{-1/2}[L_{2n}(\bar{\theta}_{2n,1}) - L_{1n}(\theta_{10})] + o_P(1)$. The leading order term of $n^{-1/2}[L_{2n}(\bar{\theta}_{2n,1}) - L_{1n}(\theta_{10})]$ is the expected value $n^{-1/2}[\operatorname{E} L_{2n}(\bar{\theta}_{2n,1}) - \operatorname{E} L_{1n}(\theta_{10})]$, which can be shown by applying Chebyshev's inequality, as $L_{2n}(\bar{\theta}_{2n,1}) - \operatorname{E} L_{2n}(\bar{\theta}_{2n,1})$ and $L_{1n}(\theta_{10}) - \operatorname{E} L_{1n}(\theta_{10})$ are both linear-quadratic forms of ϵ_{1n}. The $\operatorname{E} L_{2n}(\bar{\theta}_{2n,1})$ involves the unknown parameters $\bar{\theta}_{2n,1}$ and θ_{10} because an expectation is taken, and $\operatorname{E} L_{1n}(\theta_{10})$ involves θ_{10}. Except for $\hat{\theta}_{2n}$, another estimate for $\bar{\theta}_{2n,1}$ can be the vector that maximizes $\bar{L}_{2n}(\theta_2;\hat{\theta}_{1n})$. Denote $\bar{\theta}_{2n}(\theta_1) = \max_{\theta_2} \bar{L}_{2n}(\theta_2;\theta_1)$. The difference between $\bar{\theta}_{2n}(\hat{\theta}_{1n})$ and $\hat{\theta}_{2n}$ is expected to be small under the null hypothesis, since they are maximizers of two functions whose difference is small in probability.[14] Hence, we investigate the asymptotic distribution of the statistic

$$\frac{1}{\sqrt{n}}\Big[[L_{2n}(\hat{\theta}_{2n}) - L_{1n}(\hat{\theta}_{1n})] - \big[\bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big) - \bar{L}_{1n}(\hat{\theta}_{1n};\hat{\theta}_{1n})\big]\Big],$$

or

$$\frac{1}{\sqrt{n}}\big[[L_{2n}(\hat{\theta}_{2n}) - L_{1n}(\hat{\theta}_{1n})] - [\bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n}) - \bar{L}_{1n}(\hat{\theta}_{1n};\hat{\theta}_{1n})]\big],$$

under H_0. But note that $L_{1n}(\hat{\theta}_{1n}) = \bar{L}_{1n}(\hat{\theta}_{1n};\hat{\theta}_{1n})$,[15] so essentially the tests are based on the statistics

$$\frac{1}{\sqrt{n}}\big[L_{2n}(\hat{\theta}_{2n}) - \bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big)\big], \tag{8}$$

or

$$\frac{1}{\sqrt{n}}\big[L_{2n}(\hat{\theta}_{2n}) - \bar{L}_{2n}\big(\hat{\theta}_{2n};\hat{\theta}_{1n}\big)\big]. \tag{9}$$

As $\sqrt{n}\big(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n})\big) = O_P(1)$ by Proposition 7 in Appendix B, a second order Taylor expansion implies that

$$\frac{1}{\sqrt{n}}\big[\bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n}) - \bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big)\big] = \frac{1}{2}\big(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n})\big)'\frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\check{\theta}_{2n};\hat{\theta}_{1n})}{\partial\theta_2\partial\theta_2'}\sqrt{n}\big(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n})\big) = o_P(1),$$

where $\check{\theta}_{2n}$ is between $\hat{\theta}_{2n}$ and $\bar{\theta}_{2n}(\hat{\theta}_{1n})$. Thus, (8) and (9) are asymptotically equivalent. Note that $\bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big) \geq \bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n})$, so the expression in (8) is smaller than that in (9). The original version of the Cox test is based on (8), while (9) corresponds to Atkinson (1970)'s version.

As shown in the proof of Proposition 5, we have

$$\begin{aligned}&\frac{1}{\sqrt{n}}\big[L_{2n}(\hat{\theta}_{2n}) - \bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big)\big]\\ &= \frac{1}{\sqrt{n}}[L_{2n}(\bar{\theta}_{2n,1}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})] - C_{2n,1}'\Sigma_{1n,1}^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1} + o_P(1),\end{aligned} \tag{10}$$

where $C_{2n,1} = \frac{1}{n}\frac{\partial \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\theta_1}$. The second term on the r.h.s. of (10) appears as we estimate θ_{10} by $\hat{\theta}_{1n}$. The first term on the r.h.s. of (10) can be written as a linear-quadratic form of ϵ_{1n} and elements of $\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1}$ are also of such forms, so the asymptotic distributions of the Cox-type test statistics follow by applying the central limit theorem for linear-quadratic forms. Let $\sigma_{c,n}^2$ be the variance of $\frac{1}{\sqrt{n}}[L_{2n}(\bar{\theta}_{2n,1}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})] - C_{2n,1}'\Sigma_{1n,1}^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1}$, then

$$\sigma_{c,n}^2 = \frac{1}{n}[1, -C_{2n,1}'\Sigma_{1n,1}^{-1}]\,\mathrm{var}\Big(\big[L_{2n}(\bar{\theta}_{2n,1}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10}), \frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1'}\big]'\Big)[1, -C_{2n,1}'\Sigma_{1n,1}^{-1}]', \tag{11}$$

[14] The extended Wald test constructs an asymptotic χ^2 statistic using the asymptotic normality of $n^{1/2}[\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \hat{\theta}_{2n}]$, and the extended score test constructs an asymptotic χ^2 statistic using the asymptotic normality of the score vector $\frac{1}{\sqrt{n}}\frac{\partial L_{2n}(\bar{\theta}_{2n}(\hat{\theta}_{1n}))}{\partial\theta_2}$. Appendix B presents those tests to supplement the Cox-type tests.

[15] This can be seen from (A.1) and (A.2) with the estimators plugged in.

where var$(\cdot)$ denotes the VC matrix of a random vector. In the case that $\epsilon_{1n,i}$'s are normal, $C_{2n,1} = \mathrm{E}\big(\frac{1}{n}L_{2n}(\bar{\theta}_{2n,1})\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1}\big)$ and the information matrix equality that $\Sigma_{1n,1} = \Omega_{1n,1}$ can be applied, so

$$\sigma^2_{c,n} = \frac{1}{n}\,\mathrm{var}[L_{2n}(\bar{\theta}_{2n,1}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})] - C'_{2n,1}\Sigma^{-1}_{1n,1}C_{2n,1}. \tag{12}$$

The $\sigma^2_{c,n}$ involves $\bar{\theta}_{2n,1}$, θ_{10}, and also $\epsilon_{1n,i}$'s third and fourth moments μ_3 and μ_4 if $\epsilon_{1n,i}$ is non-normal. Let $\hat{\sigma}^2_{co,n}$ and $\hat{\sigma}^2_{ca,n}$ be, respectively, consistent estimators of $\sigma^2_{c,n}$ used in Cox and Atkinson's versions. The $\hat{\sigma}^2_{co,n}$ $\hat{\sigma}^2_{ca,n}$ may be obtained, e.g., by replacing θ_{10}'s in $\sigma_{c,n}$ with $\hat{\theta}_{1n}$'s, μ_3 and μ_4's with the third and fourth sample moments of the residuals from the quasi-maximum likelihood (QML) estimation, and $\bar{\theta}_{2n,1}$'s with either $\bar{\theta}_{2n}(\hat{\theta}_{1n})$'s or $\hat{\theta}_{2n}$'s.[16]

Proposition 5. *Under H_0 and Assumptions 1–16, the Cox-type test statistics*

$$Cox_o = n^{-1/2}\big[L_{2n}(\hat{\theta}_{2n}) - \bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big)\big]/\hat{\sigma}_{co,n}, \tag{13}$$

and

$$Cox_a = n^{-1/2}\big[L_{2n}(\hat{\theta}_{2n}) - \bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n})\big]/\hat{\sigma}_{ca,n}, \tag{14}$$

are asymptotically standard normal, if $\sigma^2_{c,n}$ is bounded away from zero.

Since $\bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big) \geq \bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n})$ as noted earlier, $Cox_o \leq Cox_a$ asymptotically under H_0. We shall digest a little bit more on the two versions of the Cox test under the alternative hypothesis. Let θ_{20} be the true parameter of the model (2) which generates the data, and $\bar{\theta}_{1n,2}$ be the pseudo true value of the model (1). Under the alternative hypothesis,

$$\begin{aligned}&\frac{1}{n}\Big[\big(L_{2n}(\hat{\theta}_{2n}) - L_{1n}(\hat{\theta}_{1n})\big) - \Big(\bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big) - \bar{L}_{1n}(\hat{\theta}_{1n};\hat{\theta}_{1n})\Big)\Big]\\ &= \frac{1}{n}\Big[\big(\bar{L}_{2n}(\theta_{20};\theta_{20}) - \bar{L}_{1n}(\bar{\theta}_{1n,2};\theta_{20})\big) - \Big(\bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big) - \bar{L}_{1n}(\hat{\theta}_{1n};\hat{\theta}_{1n})\Big)\Big] + o_P(1),\end{aligned} \tag{15}$$

and

$$\begin{aligned}&\frac{1}{n}\big[\big(L_{2n}(\hat{\theta}_{2n}) - L_{1n}(\hat{\theta}_{1n})\big) - \big(\bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n}) - \bar{L}_{1n}(\hat{\theta}_{1n};\hat{\theta}_{1n})\big)\big]\\ &= \frac{1}{n}\big[\big(\bar{L}_{2n}(\theta_{20};\theta_{20}) - \bar{L}_{1n}(\bar{\theta}_{1n,2};\theta_{20})\big) - \big(\bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n}) - \bar{L}_{1n}(\hat{\theta}_{1n};\hat{\theta}_{1n})\big)\big] + o_P(1).\end{aligned} \tag{16}$$

By Jensen's inequality (the information inequality), $\bar{L}_{2n}(\theta_{20};\theta_{20}) \geq \bar{L}_{1n}(\bar{\theta}_{1n,2};\theta_{20})$, $\bar{L}_{1n}(\hat{\theta}_{1n};\hat{\theta}_{1n}) \geq \bar{L}_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}\big)$ and $\bar{L}_{1n}(\hat{\theta}_{1n};\hat{\theta}_{1n}) \geq \bar{L}_{2n}\big(\hat{\theta}_{2n};\hat{\theta}_{1n}\big)$, so the leading order terms of (15) and (16) are non-negative. The Cox tests thus have one-sided critical regions such that we reject the null hypothesis if the Cox statistics are greater than the critical value $u_{1-\alpha}$, where $u_{1-\alpha}$ is the $(1-\alpha)$ quantile of the standard normal distribution

[16] Note that for Cox_a below, if we use $\hat{\theta}_{2n}$ for $\bar{\theta}_{2n,1}$ in $\sigma^2_{c,n}$, then there is no need to compute $\bar{\theta}_{2n}(\hat{\theta}_{1n})$. In the Monte Carlo study, for Cox_o, we use $\bar{\theta}_{2n}(\hat{\theta}_{1n})$ for $\bar{\theta}_{2n,1}$; for Cox_a, we use $\hat{\theta}_{2n}$.

for the chosen level of significance α. If the leading order terms of (15) and (16) are bounded away from zero, and $\hat{\sigma}_{co,n}$ and $\hat{\sigma}_{ca,n}$ are stochastically bounded under the alternative hypothesis, then the Cox tests are consistent. From (15) and (16), the two Cox-type test statistics are generally not asymptotically equivalent under the alternative hypothesis.

3. Relationship and Comparison between the Cox-type and *J*-type Tests

In this section, we first investigate whether there is an equivalence relationship between the Cox and J-type tests for SARAR models and then shortly compare these two types of tests.

To investigate the relationship between the Cox and J-type tests for SARAR models, we start from a short review on establishing the asymptotic equivalence of the Cox and J tests for univariate regressions under the null hypothesis, and then examine whether a similar relationship of these two types of tests for SARAR models would exist or not.

Consider the problem of testing a nonlinear univariate regression model against another one:

$$H_0: \quad y_{ni} = f_{1i}(X_{1n,i}, \beta_1) + \epsilon_{1n,i}, \quad \epsilon_{1n,i}\text{'s are } i.i.d.\, N(0, \sigma_1^2), \quad \theta_1 = (\beta_1', \sigma_1^2)', \tag{17}$$

$$H_1: \quad y_{ni} = f_{2i}(X_{2n,i}, \beta_2) + \epsilon_{2n,i}, \quad \epsilon_{2n,i}\text{'s are } i.i.d.\, N(0, \sigma_2^2), \quad \theta_2 = (\beta_2', \sigma_2^2)', \tag{18}$$

where y_{ni}'s are observations on a dependent variable, $X_{1n,i}$'s and $X_{2n,i}$'s are vectors of exogenous variables, and θ_1 and θ_2 are vectors of parameters. To test H_0 against H_1 by the J test (Davidson and MacKinnon, 1981), the following compound model is considered:

$$y_{ni} = (1-\tau) f_{1i}(X_{1n,i}, \beta_1) + \tau f_{2i}(X_{2n,i}, \beta_2) + \epsilon_{1n,i}. \tag{19}$$

As β_1 disappears from the model when $\tau = 1$ and β_2 disappears when $\tau = 0$, the compound model suffers from Davies's problem (Davies, 1977). The J test circumvents the problem by substituting an estimator $\hat{\beta}_{2n}$ of β_2 from H_1 into (19) and then estimating τ and β_1 jointly. The t statistic for $\tau = 0$, which is asymptotically standard normal, is the J test statistic. Davidson and MacKinnon (1981) has proved that the J test is asymptotically equivalent to the Cox test under H_0. Gourieroux and Monfort (1994) note that the Cox test statistic is asymptotic equivalent to a score test statistic for $\eta = 0$ under H_0, computed as if an estimator $\hat{\theta}_{2n}$ of θ_2 from H_2 was deterministic, in a model with the following probability density function

$$\begin{aligned} &\frac{l_1^{1-\eta}(y_n, X_{1n}, \theta_1) l_2^{\eta}(y_n, X_{2n}, \hat{\theta}_{2n})}{\int l_1^{1-\eta}(y_n, X_{1n}, \theta_1) l_2^{\eta}(y_n, X_{2n}, \hat{\theta}_{2n})\, dy_n} \\ &= \frac{(2\pi)^{-\frac{n}{2}} (\sigma_1^2)^{-\frac{1-\eta}{2}n} (\hat{\sigma}_{2n}^2)^{-\frac{\eta}{2}n} \exp\Big(-\frac{1-\eta}{2\sigma_1^2}\big\|y_n - f_1(X_{1n}, \beta_1)\big\|^2 - \frac{\eta}{2\hat{\sigma}_{2n}^2}\big\|y_n - f_2(X_{2n}, \hat{\beta}_{2n})\big\|^2\Big)}{\int (2\pi)^{-\frac{n}{2}} (\sigma_1^2)^{-\frac{1-\eta}{2}n} (\hat{\sigma}_{2n}^2)^{-\frac{\eta}{2}n} \exp\Big(-\frac{1-\eta}{2\sigma_1^2}\big\|y_n - f_1(X_{1n}, \beta_1)\big\|^2 - \frac{\eta}{2\hat{\sigma}_{2n}^2}\big\|y_n - f_2(X_{2n}, \hat{\beta}_{2n})\big\|^2\Big)\, dy_n}, \end{aligned} \tag{20}$$

where $y_n = (y_{n1}, \ldots, y_{nn})'$, $X_{jn} = (X_{jn,1}', \ldots, X_{jn,n}')'$, $f_j(X_{jn}, \beta_j) = \big(f_{j1}(X_{jn,1}, \beta_1), \ldots, f_{jn}(X_{jn,n}, \beta_j)\big)'$ for $j = 1, 2$; $|| \ldots ||$ denotes the Euclidean vector norm; and $l_{1n}(y_n, X_{1n}, \theta_1)$ and $l_{2n}(y_n, X_{2n}, \theta_2)$ are, respectively, the likelihood functions of H_0 and H_1. The asymptotic equivalence of the J and Cox tests is not surprising, since (20) is the likelihood function of the regression model[17]

$$y_{ni} = \frac{(1-\eta)\hat{\sigma}_{2n}^2}{\eta\sigma_1^2 + (1-\eta)\hat{\sigma}_{2n}^2} f_{1i}(X_{1n,i}, \beta_1) + \frac{\eta\sigma_1^2}{\eta\sigma_1^2 + (1-\eta)\hat{\sigma}_{2n}^2} f_{2i}(X_{2n,i}, \hat{\beta}_{2n}) + \xi_{ni}, \tag{21}$$

where ξ_{ni}'s are i.i.d. $N\big(0, \sigma_1^2\hat{\sigma}_{2n}^2/[\eta\sigma_1^2 + (1-\eta)\hat{\sigma}_{2n}^2]\big)$, which is the same as (19) after reparameterization. Given the equivalence result on the models (17) and (18), it is tempting to just use the J-type tests but ignore the Cox-type tests for other models. However, no such equivalence result exists for SARAR models.

For the SARAR models (1) and (2), the spatial J test, as described in Kelejian and Piras (2011), is obtained by augmenting the spatial Cochrane-Orcutt transformed null model

$$R_{1n}(\rho_1)y_n = \lambda_1 R_{1n}(\rho_1)W_{1n}y_n + R_{1n}(\rho_1)X_{1n}\beta_1 + \epsilon_{1n}$$

to the model

$$R_{1n}(\rho_1)y_n = \lambda_1 R_{1n}(\rho_1)W_{1n}y_n + R_{1n}(\rho_1)X_{1n}\beta_1 + \alpha R_{1n}(\rho_1)S_{2n}^{-1}(\lambda_2)X_{2n}\beta_2 + \epsilon_{1n}, \tag{22}$$

or

$$R_{1n}(\rho_1)y_n = \lambda_1 R_{1n}(\rho_1)W_{1n}y_n + R_{1n}(\rho_1)X_{1n}\beta_1 + \alpha R_{1n}(\rho_1)(\lambda_2 W_{2n}y_n + X_{2n}\beta_2) + \epsilon_{1n}, \tag{23}$$

as both $S_{2n}^{-1}(\lambda_2)X_{2n}\beta_2$ and $(\lambda_2 W_{2n}y_n + X_{2n}\beta_2)$ are predictors of y_n with some estimator for θ_2 plugged in. In the first step of the spatial J test, we can get an estimator $\hat{\rho}_{1n}$ of ρ_{10} from the null model and an estimator $\hat{\theta}_{2n}$ of θ_2 from the alternative model. Then $R_{1n}(\hat{\rho}_{1n})y_n$, $R_{1n}(\hat{\rho}_{1n})W_{1n}y_n$, $R_{1n}(\hat{\rho}_{1n})X_{1n}$, and the predictors $R_{1n}(\hat{\rho}_{1n})S_{2n}^{-1}(\hat{\lambda}_{2n})X_{2n}\hat{\beta}_{2n}$ or $R_{1n}(\hat{\rho}_{1n})(\hat{\lambda}_{2n}W_{2n}y_n + X_{2n}\hat{\beta}_{2n})$, can be computed. After that, (22) and (23) can be estimated by 2SLS in order to construct a t statistic to test whether α is equal to zero or not. We call the J test statistic based on (22) J_1 and the other J_2. The Monte Carlo study in Kelejian and Piras (2011) shows similar finite sample results for J_1 and J_2. For computational convenience, they suggest the use of J_2.

Let the likelihood functions of the models (1) and (2) still be denoted by $l_1(y_n, X_{1n}, \theta_1)$ and $l_2(y_n, X_{2n}, \theta_2)$, respectively. The compound model with a probability density function corresponding to (20) is

$$\begin{aligned}\frac{l_1^{1-\eta}(y_n, X_{1n}, \theta_1)l_2^{\eta}(y_n, X_{2n}, \hat{\theta}_{2n})}{\int l_1^{1-\eta}(y_n, X_{1n}, \theta_1)l_2^{\eta}(y_n, X_{2n}, \hat{\theta}_{2n})\,dy_n} &= c_n \cdot (\sigma_1^2)^{-\frac{1-\eta}{2}n}(\hat{\sigma}_{2n}^2)^{-\frac{\eta}{2}n}\exp\Big(-\frac{1-\eta}{2\sigma_1^2}\big\|R_{1n}(\rho_1)[S_{1n}(\lambda_1)y_n - X_{1n}\beta_1]\big\|^2 \\ &- \frac{\eta}{2\hat{\sigma}_{2n}^2}\big\|R_{2n}(\hat{\rho}_{2n})[S_{2n}(\hat{\lambda}_{2n})y_n - X_{2n}\hat{\beta}_{2n}]\big\|^2\Big)\big|S_{1n}(\lambda_1)R_{1n}(\rho_1)\big|^{1-\eta}\big|S_{2n}(\hat{\lambda}_{2n})R_{2n}(\hat{\rho}_{2n})\big|^{\eta},\end{aligned} \tag{24}$$

where c_n only depends on n. The score test for $\eta = 0$ in (24), computed as if $\hat{\theta}_{2n}$ is non-stochastic, can be shown to be asymptotically equivalent to the Cox test under H_0. The score test is based on the asymptotic

[17]See Atkinson (1970), among others.

distribution of the score

$$\frac{1}{\sqrt{n}}\Big[\ln l_2(y_n, X_{2n}, \hat{\theta}_{2n}) - \ln l_1(y_n, X_{1n}, \hat{\theta}_{1n}) - \int [\ln l_2(y_n, X_{2n}, \hat{\theta}_{2n}) - \ln l_1(y_n, X_{1n}, \hat{\theta}_{1n})] l_1(y_n, X_{1n}, \hat{\theta}_{1n})\, dy_n\Big], \tag{25}$$

where $\hat{\theta}_{1n}$ is from H_0. The asymptotic variance of (25) is computed as if $\hat{\theta}_{2n}$ were deterministic. (25) is equal to the numerator of Atkinson (1970)'s version of the Cox test statistic. To derive the asymptotic distribution of (25), as noted in (D.1) and (D.2), $\hat{\theta}_{2n}$ can be replaced by the non-stochastic pseudo true value $\bar{\theta}_{2n,1}$. Once the analytical form of the asymptotic variance for (25) is found, $\bar{\theta}_{2n,1}$ may be substituted by $\hat{\theta}_{2n}$ to approximate the asymptotic variance. Thus the score test for $\eta = 0$ deduced from (24) is asymptotically equivalent to the Cox test under H_0.

On the other hand, (24) is not equivalent to (22), (23) or any other simple combinations of the models (1) and (2). The exponent in (24) written in the quadratic form is equal to $-\frac{1}{2}(A_n^{\frac{1}{2}} y_n - A_n^{-\frac{1}{2}} b_n)'(A_n^{\frac{1}{2}} y_n - A_n^{-\frac{1}{2}} b_n)$ plus a term not involving y_n, where $A_n = \frac{1-\eta}{\sigma_1^2} S'_{1n}(\lambda_1) R'_{1n}(\rho_1) R_{1n}(\rho_1) S_{1n}(\lambda_1) + \frac{\eta}{\hat{\sigma}_{2n}^2} S'_{2n}(\hat{\lambda}_{2n}) R'_{2n}(\hat{\rho}_{2n}) R_{2n}(\hat{\rho}_{2n}) S_{2n}(\hat{\lambda}_{2n})$ and $b_n = \frac{1-\eta}{\sigma_1^2} S'_{1n}(\lambda_1) R'_{1n}(\rho_1) R_{1n}(\rho_1) X_{1n}\beta_1 + \frac{\eta}{\hat{\sigma}_{2n}^2} S'_{2n}(\hat{\lambda}_{2n}) R'_{2n}(\hat{\rho}_{2n}) R_{2n}(\hat{\rho}_{2n}) X_{2n}\hat{\beta}_{2n}$. The corresponding model with i.i.d. normal disturbances would be

$$A_n^{\frac{1}{2}} y_n = A_n^{-\frac{1}{2}} b_n + u_n, \tag{26}$$

which is not linear in parameter and does not correspond to any simple linear combination of the original models. In particular, this model is very different from the compound models (22) and (23) (or the one in Kelejian (2008)). Therefore, the Cox-type and J-type tests for SARAR models cannot be shown to be asymptotically equivalent under the null hypothesis by showing that the exponential compound model (24) is equivalent to (22) or (23). It seems not to be surprising that there is no such an equivalence relationship because of the spatial dependence.

The original J-type tests in Kelejian and Piras (2011) employ the generalized spatial 2SLS (GS2SLS) proposed in Kelejian and Prucha (1998) to estimate the null and alternative models, and the 2SLS to estimate the augmented model. Since the GS2SLS or 2SLS only uses linear instruments, which is less efficient than the QML or the GMM which uses both linear and quadratic moments, the power can be low due to the estimation method, especially when the variation in exogenous variables cannot explain much of the variation in the dependent variable. We may estimate the null, alternative and augmented models by the GMM or QML for the J-type tests, which is computational more involved. For the estimation of the null and alternative models in the J-type tests, an advantage of the generalized spatial 2SLS is that it can be robust to unknown heteroskedasticity while the QML is not.[18] The Cox-type tests are built upon the QMLEs of the null and alternative models, which involve nonlinear objective functions, thus identification conditions are needed. The J-type tests only involve the GS2SLS and 2SLS, where an identification condition is only

[18]The GMM can also be robust to unknown heteroskedasticity, see Lin and Lee (2010).

needed for the spatial error dependence parameter.[19] Also related to the nonlinear objective functions, the QML needs the compact parameter space assumption while the GS2SLS does not need that assumption.

4. Consistency of the Bootstrap for Cox-type Tests

In this section, we show that the bootstrap is consistent for Cox-type tests. The bootstrap testing procedure is as follows:[20]

(i) Compute the QML estimator $(\hat{\lambda}_{1n}, \hat{\rho}_{1n}, \hat{\beta}'_{1n})'$ and the corresponding residual vector $e_{1n} = R_{1n}(\hat{\rho}_{1n})[S_{1n}(\hat{\lambda}_{1n})y_n - X_{1n}\hat{\beta}_{1n}]$ for the model (1). Compute the Cox-type test statistics.

(ii) Draw an n-dimensional vector e^*_{1n} of random samples from the residuals in e_{1n} using sampling with replacement and generate data y^*_n according to $y^*_n = S^{-1}_{1n}(\hat{\lambda}_{1n})[X_{1n}\hat{\beta}_{1n} + R^{-1}_{1n}(\hat{\rho}_{1n})e^*_{1n}]$.

(iii) Compute various test statistics using the data y^*_n.

(iv) Repeat (ii) and (iii) s times, and obtain the bootstrapped p-values.[21]

(v) The bootstrap tests consist in rejecting the null hypothesis if the bootstrapped p-value is smaller than the chosen level of significance and not rejecting otherwise.

Using y^*_n, we have the estimators $\hat{\theta}^*_{1n}$, $\hat{\theta}^*_{2n}$ and $\bar{\theta}_{2n}(\hat{\theta}^*_{1n})$, corresponding to the estimators $\hat{\theta}_{1n}$, $\hat{\theta}_{2n}$ and $\bar{\theta}_{2n}(\hat{\theta}_{1n})$ respectively. Denote the bootstrapped versions of $\hat{\sigma}_{co,n}$, $\hat{\sigma}_{ca,n}$, Cox_o, Cox_a by, respectively, $\hat{\sigma}^*_{co,n}$, $\hat{\sigma}^*_{ca,n}$, Cox^*_o, Cox^*_a. Let P^* be the probability distribution induced by the bootstrap sampling process. From (10), the Cox-type test statistics can be approximated by a linear-quadratic form of disturbances, thus we can apply a theorem in Jin and Lee (2012), who establish that the bootstrap is consistent for spatial econometric statistics that can be approximated by a linear-quadratic form. The result is based on the uniform convergence of the distribution for a linear-quadratic form to the normal distribution. The consistency result for Cox-type test statistics needs a stronger assumption on the disturbances—namely, the existence of eighth moment—than assumed earlier, for non-normal disturbances. One reason of the stronger assumption is that the numerators for the Cox-type tests generally involve estimators of the fourth moments of the disturbances. The stronger condition is needed for the rate of convergence of the estimators.

Assumption 17. *$\{\epsilon_{1n,i}\}$'s in $\epsilon_{1n} = (\epsilon_{1n,1}, \ldots, \epsilon_{1n,n})'$, $i = 1, \ldots, n$, are i.i.d. with mean zero and variance σ^2_{10}, and the moment $\mathrm{E}(\epsilon^8_{1n,i})$ exists.*

[19]The identification condition is not explicitly stated in Kelejian and Piras (2011). They assume instead the high level condition that the limits of some matrices involving parameter estimates for the alternative model have nonsingular probability limits.

[20]The resampling procedure above has been used by Burridge and Fingleton (2010).

[21]For the Cox-type tests, as they are one-sided tests, the bootstrapped p-value is the percentage of test statistics calculated from the bootstrapped samples that are greater than the corresponding test statistic obtained in (i). For two-sided tests, the bootstrapped p-value is the equal-tail bootstrapped p-value which is equal to 2 times the smaller one of the percentages of test statistics that are greater and non-greater than the test statistic in (i) (MacKinnon, 2009).

Proposition 6. *Under H_0 and Assumptions 2–17,* $\sup_x |\mathrm{P}^*(Cox_o^* \leq x) - \mathrm{P}(Cox_o \leq x)| = o_P(1)$ *and* $\sup_x |\mathrm{P}^*(Cox_a^* \leq x) - \mathrm{P}(Cox_a \leq x)| = o_P(1)$*.*

5. Monte Carlo Study

We compare the finite sample size and power properties of the tests derived in this paper with those of the spatial J tests (Kelejian and Piras, 2011) with Monte Carlo experiments. In addition, we also compare them with a test derived from a comprehensive model. For the SARAR models (1) and (2), a natural comprehensive model for them is

$$y_n = \lambda_1 W_{1n} y_n + \lambda_2 W_{2n} y_n + X_{1n}\beta_1 + X_{2n,a}\beta_{2a} + u_{1n}, \quad u_{1n} = \rho_1 M_{1n} u_{1n} + \rho_2 M_{2n} u_{1n} + \epsilon_{1n}, \tag{27}$$

where $X_{2n,a}$ contains the variables in X_{2n} that are different from any in X_{1n}, and β_{2a} is the corresponding parameter vector. We test whether λ_2, ρ_2 and β_{2a} are jointly zero with a Lagrangian multiplier (LM) test. Denote the corresponding test statistic by Aug. In the experiments, the spatial weights matrix in the spatial error process is set to be the same as that in the spatial lag equation for the two SARAR models, and the two models either have the same spatial weights matrix or the same exogenous variable matrix. For the J test statistics J_1 and J_2, first estimate the model (1) to obtain $\hat{\rho}_{1n}$ by the generalized spatial 2SLS, as described in Kelejian and Prucha (1998), with instrumental variables $[X_{1n}, W_{1n}X_{1n}, W_{1n}^2 X_{1n}]_{LI}$, where LI denotes the linear independent columns of a matrix, then estimate the model (2) with instrumental variables $[X_{2n}, W_{2n}X_{2n}, W_{2n}^2 X_{2n}]_{LI}$ to obtain y_n's predictors, and finally (22) and (23) are estimated with the instrumental variables $[X_{1n}, W_{1n}X_{1n}, W_{2n}X_{1n}, W_{1n}^2X_{1n}, W_{2n}^2X_{1n}, W_{1n}W_{2n}X_{1n}, W_{2n}W_{1n}X_{1n}]_{LI}$ when $X_{1n} = X_{2n}$ but $W_{1n} \neq W_{2n}$; or $[X_{1n}, X_{2n}, W_{1n}X_{1n}, W_{1n}X_{2n}, W_{1n}^2X_{1n}, W_{1n}^2X_{2n}]_{LI}$ when $W_{1n} = W_{2n}$ but $X_{1n} \neq X_{2n}$. As an alternative, we first estimate the model (2) by the QML to derive the predictor $S_{2n}^{-1}(\hat{\lambda}_{2n})X_{2n}\hat{\beta}_{2n}$ or $(\hat{\lambda}_{2n}W_{2n}y_n + X_{2n}\hat{\beta}_{2n})$, and then estimate (22) and (23) by the GMM with both linear and quadratic moments.[22] Denote the J tests with the alternative estimation methods as J_{1a} and J_{2a} respectively. The linear instruments for J_{1a} and J_{2a} are the same for J_1 and J_2, and the matrices for the quadratic moments include different matrices of W_{1n}, W_{2n}, $W_{1n}^2 - \mathrm{tr}(W_{1n}^2)I_n/n$, $W_{2n}^2 - \mathrm{tr}(W_{2n}^2)I_n/n$, $W_{1n}W_{2n} - \mathrm{tr}(W_{1n}W_{2n})I_n/n$ and $W_{2n}W_{1n} - \mathrm{tr}(W_{2n}W_{1n})I_n/n$. Note that for the extended Wald and score tests, we use the asymptotic chi-square critical values with degrees of freedom equal to the number of parameters in the alternative model to evaluate the empirical size and power.

[22] Note that our GMM approach estimates ρ_1 jointly with λ_1 and β_1 in (22) and (23). This is different from the original approach in Kelejian and Piras (2011) where ρ_1 is first estimated in the model (1) and then the estimate is plugged into the augmented model. The GMM estimation of (22) and (23) involving quadratic moments with an initial estimate of ρ_1 plugged in would generate a complicated variance-covariance matrix because a part of the variance-covariance would be from the estimation error of ρ_1's estimator.

Table 1: Sets of Experiments

Experiments	H_0	H_1
Set I	W_a, X_a	W_b, X_a
Set II	W_c, X_a	W_b, X_a
Set III	W_c, X_b	W_c, X_a

The experimental design is based on former Monte Carlo studies of spatial models (see, e.g., Anselin and Florax 1995, Kelejian and Prucha 1999, Arraiz et al. 2010 and Kelejian and Piras 2011). We consider three different spatial weights matrices W_a, W_b and W_c: W_a is generated according to the rook criterion, W_b is generated according to the queen criterion and W_c is a block diagonal matrix with the diagonal blocks being the continuity matrix for 49 neighborhoods in Columbus, OH from Anselin (1988). We use row normalized matrices. Two exogenous variable matrices X_a and X_b are used: X_a contains a vector of ones and a vector of random samples drawn from the standard normal, and X_b contains a vector of ones, a variable drawn from the uniform distribution $U(0, 1)$, and a variable equal to 2 times the second variable plus $1/2$ times a variable drawn from the chi-square distribution with 2 degrees of freedom. For X_b, the correlation coefficient between the second and third variables is 0.5. The three sets of experiments considered are shown in Table 1. For each set of experiments, the disturbances are drawn from either the standard normal or a normalized chi-square $(\chi^2(3) - 3)/\sqrt{6}$ with mean zero and variance one. The true parameter vector is either $(0.5, 0.5)'$ or $(0.5, 2)'$ corresponding to X_a, and either $(0.5, -1, 0.5)'$ or $(0.5, 4, 1)'$ corresponding to X_b, leading to the ratio of the variance of $X\beta$ with the sum of the variance of $X\beta$ and that of the error terms to be equal to 0.2 and 0.8, respectively.[23] Denote this ratio by $\tilde{R}^2$. When the null and alternative models generate the data, i.e., when the empirical size and power are considered, λ_1 in the null model and λ_2 in the alternative model, or, ρ_1 in the null model and ρ_2 in the alternative model, are the same, taking value of 0.2 or 0.8. Denote the two parameters by λ and ρ respectively in the reported tables. In total, we have $3 \times 2 \times 2 \times 2 \times 2 = 48$ experiments for each sample size n. We consider a small size $n = 98$ and a large sample size $n = 1519$.[24] The nominal level of significance is set to 5% and the number of Monte Carlo repetitions is 1000. For $n = 98$, bootstrapped tests of various test statistics are also implemented.[25] We set the number of resampling s to 199, leading to a standard error of the bootstrapped p-value being equal to 1.5%.

The Monte Carlo results for $n = 98$ are reported in Tables 2–7. Using the asymptotic p-values, J_1, J_2,

[23] This kind of Monte Carlo setting for spatial models follows from Lee (2007) and Lee and Liu (2010).

[24] For $n = 98$, the W_a and W_b are first generated on a 10×10 grid, then the last two rows and last two columns are deleted, and finally they are row-normalized to have row sun 1 by dividing each element in a row by the sum of all elements in that row. The W_a and W_b for $n = 1519$ are similarly derived.

[25] For $n = 1519$, implementing bootstrap tests for all statistics with 1000 repetitions takes too long, so bootstrap tests are not implemented.

Aug and *Score* generally have small size distortions while other statistics have large size distortions in some cases. The empirical sizes of J_1 deviate from the nominal one by no more than 3 percentage points in all experiments, the empirical size of J_2 can be as large as 9.7% as shown in Table 3, *Aug* in experiment set III and *Score* in experiment sets II and III with chi-square disturbances significantly under-reject the true null hypothesis. The J_{1a} and J_{2a} almost have no size distortion in experiment set III, but have large size distortion in the first two sets of experiments. The empirical size of J_{2a} can be over 40% when $\tilde{R}^2 = 0.2$ in experiment set I. The *Wald* have empirical sizes larger than 50% in many cases. The size distortion of Cox_o and Cox_a is no more than 3.7 percentage points in experiment set I, but the size of Cox_o can be as large as 20.2% in experiment set II and 30.2% in experiment set III, and the size of Cox_a can be as large as 23.2% in experiment set II and 22.6% in experiment set III. The empirical sizes based on the bootstrapped critical values show that the bootstrap removes the size distortion of various statistics in most cases. We thus compare the empirical powers of different statistics based on the bootstrapped p-values.

Several patterns for the empirical powers of the bootstrapped tests can be summarized as follows: none of the tests can dominate the rest of tests in power in all experiments, but the Cox-type statistics usually have high powers compared to other statistics and dominate other ones in some cases; in most cases of all experiments, J_{1a} is more powerful than J_1; in most cases, J_{2a} is more powerful than J_2 in experiment sets II and III, but less powerful in experiment set I; J_2 is more powerful than J_1 in almost all cases. We now investigate the results for experiment set I with normal disturbances in some detail, and briefly summarize results for other experiments. Table 2 presents the results for experiment set I with normal disturbances. The powers of Cox_o and Cox_a are similar, which are the highest among all the test statistics, and the powers of other statistics are significantly lower in most cases. Taking the case with $\tilde{R}^2 = 0.8$, $\lambda = 0.2$ and $\rho = 0.8$ as an example, Cox_o and Cox_a have powers higher than 90%, *Aug* has a power of 73.7%, *Score* has a power of 52.5%, but the powers of the rest statistics are all below 21%. In all cases except the one with $\tilde{R}^2 = 0.2$, $\lambda = 0.2$ and $\rho = 0.2$, J_2 has a higher power than J_1. When $\tilde{R}^2 = 0.8$, $\lambda = 0.8$ and $\rho = 0.8$, J_2 has a power of 84.0%, while J_1 has a power of only 52.6%.[26] Table 3 presents the results for experiment set I with chi-square disturbances. Changing the distributions of the disturbances from normal to chi-square has not led to big changes in the results. For experiment set II, Tables 4 and 5 show that, J_{2a}, *Aug*, *Score*, Cox_o and Cox_a have similar magnitude of power, among which Cox_a has the highest power in most cases, and other statistics have significantly lower powers. For experiment set III, all statistics, except J_1 and *Wald* in some cases, have powers close or equal to 100%. The *Wald* has very low power compared to other test statistics.

The empirical size and power based on the asymptotic p-values for $n = 1519$ are reported in Tables 8–10.

[26] In the Monte Carlo study of Kelejian and Piras (2011), their Monte Carlo design has produced high powers for the J tests, where in general J_2 is also relatively more powerful than J_1, but due to their high power, their differences seem small.

Most statistics have no significant size distortion with a sample size of 1519, except for *Wald*, Cox_o and Cox_a in some cases, which have much smaller size distortion compared to that with a sample size of 98. The *Wald* still has significant size distortion for all experiments. For experiment set I, Cox_o and Cox_a have empirical sizes close to the nominal level. For experiment set II, Cox_o and Cox_a have large size distortion only when $\lambda = 0.2$ and $\rho = 0.2$. For experiment set III, Cox_o and Cox_a still have large distortion in some cases. For example, when $\tilde{R}^2 = 0.2$, $\lambda = 0.2$, $\rho = 0.8$ and the disturbances are normal, Cox_o and Cox_a with $n = 1519$ have empirical sizes equal to 17.2% and 17.3% respectively, smaller than the sizes 22.9% and 22.6% for $n = 98$. All the statistics have powers close or equal to 100% with the large sample size except for J_1, J_2 and J_{1a}. For experiment set I, when $\tilde{R}^2 = 0.2$ and $\lambda = 0.2$, J_1 and J_2 have very low powers, less than 27%, and J_{1a} has powers lower than 76% with $\rho = 0.2$ and lower than 41% with $\rho = 0.8$. For experiment set II, when $\tilde{R}^2 = 0.2$ and $\lambda = 0.2$, J_1, J_2 and J_{1a} have powers lower than 60%. All statistics in experiment set III have powers close or equal to 100%. Note that with $n = 1519$, J_{2a} may still have slightly lower power than Cox_o and Cox_a, e.g., in experiment set I with $\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.2$ and chi-square disturbances, J_{2a} has a power of 98.5%, while both Cox_o and Cox_a have a power of 100%.

The Cox-type tests are computationally more involved than the J-type tests, especially for large sample sizes.[27] First, the Cox-type tests are based on the QMLEs. However, with the development of more advanced computers and computational techniques[28], the QMLE can be efficiently computed. A further computational problem in calculating the Cox-type test statistics after deriving the QMLEs is on the traces involving the inverses $S_{1n}^{-1}(\hat{\lambda}_{1n})$ and $R_{1n}^{-1}(\hat{\rho}_{1n})$ or on the product of $S_{1n}^{-1}(\hat{\lambda}_{1n})$ and a vector (see Appendix A). LeSage and Pace (2009, pp. 110–113) have discussed some techniques in computing such terms. Those approaches may make the computation practically easier.

6. Empirical Illustration

We illustrate the use of the Cox-type tests with the housing data set in Harrison and Rubinfeld (1978). Pace and Gilley (1997) added longitude-latitude coordinates for census tracts to the data set. With the augmented data set, LeSage (1999, pp. 83–94) estimates a SARAR model, where the dependent variable is the studentized log of median housing prices for each of the 506 census tracts, the explanatory variables include 13 covariates, and the spatial weights matrix for both the spatial lag and the spatial error dependence is a first order contiguity matrix (call it W_{foc}). We create a row-normalized spatial weights matrix based on 5 nearest neighbors (call it W_{5nn}), where the elements corresponding to a census tract's five nearest

[27] For Experiment Set I with the sample size of $n = 1519$, when $\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.2$ and the disturbances are normal, Computing J_1, J_2, J_{1a}, J_{2a}, Cox_o and Cox_a once take, respectively, 0.3, 0.3, 7.8, 7.8, 101.6 and 17.6 seconds on average, using Matlab on a desktop computer with Intel Core i7-2600 processor and 8 gigabyte memory.

[28] See, e.g., Pace and LeSage (2009) and Smirnov and Anselin (2009).

Table 2: Empirical size and power for experiment set I with normal disturbances and $n = 98^{\dagger}$

	Asymptotic†		Bootstrap†		Asymptotic†		Bootstrap†	
	Size	Power	Size	Power	Size	Power	Size	Power
	$\tilde{R}^2$=0.2, λ=0.2, ρ=0.2				$\tilde{R}^2$=0.2, λ=0.2, ρ=0.8			
J_1	6.7	5.0	5.8	4.4	5.0	5.4	8.9	8.4
J_2	5.5	5.1	4.0	3.3	6.7	20.9	5.9	17.8
J_{1a}	12.8	14.7	4.7	4.9	11.5	16.6	5.3	6.7
J_{2a}	46.8	26.8	4.1	2.5	24.3	12.3	4.0	1.2
Aug	6.5	5.2	5.4	5.1	5.3	35.7	5.6	35.0
Wald	70.3	74.2	2.3	2.9	50.4	80.5	4.5	3.1
Score	6.6	3.9	6.6	4.0	4.7	15.2	4.0	17.0
Cox_o	6.7	24.7	3.1	8.8	5.3	74.2	5.0	57.4
Cox_a	7.0	19.4	4.0	10.4	3.9	72.7	4.9	55.2
	$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	2.8	3.2	7.7	5.4	2.3	1.3	6.5	5.7
J_2	5.3	24.9	4.6	22.1	4.3	35.9	5.2	35.6
J_{1a}	10.4	29.7	4.7	10.5	8.4	23.1	5.3	13.9
J_{2a}	43.2	13.6	4.7	1.2	25.5	40.0	4.5	14.2
Aug	5.9	38.2	5.3	38.6	7.3	92.4	6.2	92.2
Wald	50.5	87.5	4.1	6.3	30.6	98.8	2.8	0.6
Score	5.7	17.3	4.8	17.5	6.9	76.5	5.3	77.5
Cox_o	5.1	77.8	4.2	59.8	3.4	99.4	4.9	97.3
Cox_a	3.2	75.9	4.4	58.2	1.8	99.6	3.8	97.3
	$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.8$			
J_1	5.2	18.8	3.6	11.6	6.3	15.9	4.1	12.9
J_2	5.6	21.0	4.2	13.2	7.5	27.4	4.1	20.5
J_{1a}	14.6	31.2	4.9	12.6	11.1	30.7	4.6	16.9
J_{2a}	9.3	35.3	4.5	16.7	13.3	49.3	5.9	18.0
Aug	5.3	19.9	5.1	18.3	6.3	75.5	5.6	73.7
Wald	53.0	81.0	2.6	9.3	27.6	64.2	4.8	13.1
Score	6.6	12.7	6.3	12.2	5.7	53.4	4.9	52.5
Cox_o	8.7	57.2	4.4	31.4	7.6	93.9	5.7	92.1
Cox_a	8.7	49.3	4.5	28.8	6.5	94.3	5.0	91.5
	$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	6.1	91.8	5.8	92.3	6.1	58.3	6.3	52.6
J_2	5.2	98.9	4.7	99.0	6.3	95.2	8.2	84.0
J_{1a}	18.5	88.8	5.0	77.1	14.3	89.9	4.9	79.1
J_{2a}	9.5	83.7	5.2	57.5	13.3	73.5	5.1	63.9
Aug	5.3	97.7	4.8	97.2	5.7	99.5	5.8	99.5
Wald	26.0	97.1	3.9	88.0	34.7	96.6	4.5	81.2
Score	5.8	92.9	5.1	91.4	6.8	97.2	5.7	96.8
Cox_o	5.0	100.0	4.9	99.9	3.9	100.0	4.8	100.0
Cox_a	4.2	99.9	5.4	99.9	2.3	100.0	5.1	100.0

† All empirical sizes and powers are expressed as percentages with the sign % being omitted. The "Asymptotic" and "Bootstrap" mean that the reported empirical size and power are computed by using, respectively, the asymptotic and bootstrapped p-values.

Table 3: Empirical size and power for experiment set I with chi-square disturbances and $n = 98^{\dagger}$

	Asymptotic		Bootstrap		Asymptotic		Bootstrap	
	Size	Power	Size	Power	Size	Power	Size	Power
	$\tilde{R}^2$=0.2, λ=0.2, ρ=0.2				$\tilde{R}^2$=0.2, λ=0.2, ρ=0.8			
J_1	5.7	5.1	5.4	4.8	5.3	4.2	10.9	7.3
J_2	6.1	6.4	4.1	4.9	7.8	20.8	7.3	16.9
J_{1a}	10.3	13.7	2.8	4.7	8.8	15.3	4.3	5.8
J_{2a}	44.2	29.2	4.5	2.3	23.0	11.4	3.1	1.0
Aug	5.5	4.4	5.2	4.8	4.9	33.9	5.2	34.8
Wald	51.7	54.6	3.4	4.2	42.1	70.4	3.5	2.8
Score	1.8	0.7	5.8	3.9	2.6	9.2	5.7	22.8
Cox_o	6.0	22.1	2.9	7.6	5.0	72.3	4.9	56.3
Cox_a	5.7	17.5	4.2	10.6	4.2	74.8	5.5	60.1
	$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	3.3	3.6	7.4	4.8	2.2	1.5	8.9	5.5
J_2	6.6	25.3	5.6	22.3	4.1	36.5	5.6	36.3
J_{1a}	7.6	26.3	3.6	9.7	6.3	23.0	4.1	11.4
J_{2a}	41.8	14.3	3.4	0.8	22.8	39.5	4.1	13.6
Aug	4.9	35.6	5.1	36.9	7.4	92.7	6.5	92.5
Wald	42.2	83.6	5.6	6.4	24.7	97.0	3.5	0.2
Score	2.3	11.1	5.3	23.8	3.0	69.6	5.9	84.8
Cox_o	4.4	75.9	4.7	58.2	3.1	99.3	4.0	97.9
Cox_a	3.1	75.6	3.7	58.9	1.6	98.7	4.5	97.3
	$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.8$			
J_1	5.3	19.3	4.7	12.5	7.3	18.0	4.6	11.9
J_2	5.6	22.1	4.8	14.8	9.7	28.7	5.2	20.9
J_{1a}	11.5	34.8	4.4	13.0	9.3	33.6	4.9	14.4
J_{2a}	9.2	35.8	3.9	17.2	13.5	50.4	5.7	16.5
Aug	5.1	20.5	4.5	20.7	6.2	76.7	5.6	75.2
Wald	33.6	58.5	4.1	5.8	21.6	36.2	5.7	6.5
Score	2.2	5.2	4.8	14.4	2.5	46.7	5.2	58.6
Cox_o	8.4	56.0	4.9	31.4	6.3	95.0	4.8	92.6
Cox_a	8.2	45.8	5.0	30.1	6.9	94.7	5.7	91.7
	$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	4.2	90.7	3.4	92.6	7.5	62.2	8.0	54.6
J_2	3.9	98.7	4.0	98.4	6.7	94.2	9.4	84.5
J_{1a}	16.3	88.1	3.8	76.8	11.4	89.7	4.1	77.0
J_{2a}	9.5	81.8	4.2	56.2	12.2	73.2	6.0	58.5
Aug	4.6	98.2	4.2	97.6	5.9	99.7	5.5	99.7
Wald	20.6	93.4	3.7	85.2	26.0	93.7	4.4	72.5
Score	2.9	89.7	4.7	94.6	2.6	96.9	5.1	98.6
Cox_o	4.0	99.9	4.5	99.9	2.8	99.9	4.1	100.0
Cox_a	3.6	100.0	4.5	100.0	2.5	100.0	4.4	100.0

† All empirical sizes and powers are expressed as percentages with the sign % being omitted. The "Asymptotic" and "Bootstrap" mean that the reported empirical size and power are computed by using, respectively, the asymptotic and bootstrapped p-values.

Table 4: Empirical size and power for experiment set II with normal disturbances and $n = 98$[†]

	Asymptotic		Bootstrap		Asymptotic		Bootstrap	
	Size	Power	Size	Power	Size	Power	Size	Power
	$\tilde{R}^2$=0.2, λ=0.2, ρ=0.2				$\tilde{R}^2$=0.2, λ=0.2, ρ=0.8			
J_1	3.7	8.0	4.8	7.9	5.2	22.5	6.5	25.0
J_2	2.7	17.3	4.6	19.1	2.1	74.5	7.5	78.3
J_{1a}	7.8	13.2	4.8	10.8	7.1	26.0	4.7	29.4
J_{2a}	11.7	61.6	4.7	45.4	14.9	99.6	5.2	99.3
Aug	4.6	44.7	5.4	44.8	5.4	99.9	5.7	99.9
$Wald$	60.5	88.4	2.5	8.7	64.0	99.6	5.1	17.3
$Score$	4.0	34.3	4.4	34.9	4.3	99.5	5.3	99.5
Cox_o	18.6	83.3	2.6	37.4	3.7	99.9	4.2	99.9
Cox_a	19.0	70.0	4.8	42.4	3.8	100.0	4.3	100.0
	$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	3.8	37.1	4.5	34.9	5.0	33.5	5.6	34.3
J_2	2.0	81.7	5.9	82.1	1.5	97.4	4.9	97.6
J_{1a}	7.9	39.4	5.1	41.3	7.4	55.0	4.1	55.9
J_{2a}	15.7	99.5	5.1	99.1	15.5	100.0	4.2	100.0
Aug	5.3	99.9	5.9	99.9	7.2	100.0	6.2	100.0
$Wald$	67.3	99.3	5.3	17.4	71.2	99.4	5.0	41.1
$Score$	3.8	99.8	5.0	99.7	5.3	99.4	5.2	99.3
Cox_o	4.0	99.9	4.9	99.9	5.0	100.0	5.4	100.0
Cox_a	3.2	100.0	4.0	100.0	2.3	100.0	6.1	100.0
	$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.8$			
J_1	4.8	31.6	5.3	23.6	5.8	33.4	5.2	36.4
J_2	4.6	39.7	4.8	30.5	5.0	55.7	5.3	54.3
J_{1a}	9.8	45.8	4.3	26.5	7.8	43.7	3.8	38.6
J_{2a}	11.3	76.4	5.1	55.6	10.2	97.8	5.5	96.0
Aug	4.7	59.7	5.0	60.0	4.8	99.9	5.7	99.9
$Wald$	37.8	93.0	1.7	30.2	34.1	100.0	5.5	96.5
$Score$	5.1	49.2	5.2	48.7	5.3	99.9	5.0	99.9
Cox_o	20.2	92.4	3.1	53.0	4.5	99.9	4.2	99.9
Cox_a	23.2	83.7	4.9	58.7	4.9	100.0	3.6	100.0
	$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	4.0	99.1	5.0	98.7	6.0	62.1	6.6	62.4
J_2	3.4	99.9	5.0	99.9	4.7	98.6	6.4	98.5
J_{1a}	9.6	93.9	3.9	92.0	8.2	62.2	4.5	62.9
J_{2a}	10.7	100.0	5.1	100.0	12.8	100.0	5.5	100.0
Aug	4.2	100.0	4.0	100.0	6.2	100.0	6.1	100.0
$Wald$	57.0	100.0	4.7	98.6	66.2	100.0	5.6	87.6
$Score$	4.7	100.0	4.8	100.0	5.0	99.7	4.8	99.6
Cox_o	7.9	100.0	5.3	100.0	5.7	100.0	4.3	100.0
Cox_a	6.2	100.0	5.2	100.0	3.3	100.0	5.7	100.0

[†] All empirical sizes and powers are expressed as percentages with the sign % being omitted. The "Asymptotic" and "Bootstrap" mean that the reported empirical size and power are computed by using, respectively, the asymptotic and bootstrapped p-values.

Table 5: Empirical size and power for experiment set II with chi-square disturbances and $n = 98^\dagger$

	Asymptotic		Bootstrap		Asymptotic		Bootstrap	
	Size	Power	Size	Power	Size	Power	Size	Power
	$\tilde{R}^2$=0.2, λ=0.2, ρ=0.2				$\tilde{R}^2$=0.2, λ=0.2, ρ=0.8			
J_1	4.6	7.6	4.6	8.3	4.5	20.5	6.2	22.2
J_2	1.9	15.1	3.9	15.7	1.9	71.6	5.2	74.9
J_{1a}	7.8	17.3	5.6	13.5	7.8	27.6	4.8	18.5
J_{2a}	8.8	61.8	3.7	47.6	10.7	99.2	4.8	98.1
Aug	3.8	42.9	5.4	44.1	4.3	100.0	5.1	100.0
$Wald$	48.3	74.0	2.4	8.6	59.1	99.5	5.1	17.3
$Score$	0.2	16.7	4.6	37.7	2.0	99.0	4.9	99.9
Cox_o	15.5	80.3	2.5	37.0	3.2	100.0	4.2	99.9
Cox_a	13.8	70.1	4.8	49.2	1.4	99.9	3.6	99.9
	$\tilde{R}^2 = 0.2,\ \lambda = 0.8,\ \rho = 0.2$				$\tilde{R}^2 = 0.2,\ \lambda = 0.8,\ \rho = 0.8$			
J_1	4.6	36.1	5.4	37.2	2.0	30.4	2.8	35.9
J_2	2.1	85.2	4.8	86.1	0.8	96.0	4.0	96.9
J_{1a}	7.3	40.6	4.7	40.7	8.0	55.6	4.8	57.2
J_{2a}	14.0	99.2	4.1	99.0	14.8	100.0	5.0	100.0
Aug	4.7	100.0	5.4	100.0	5.2	100.0	4.4	100.0
$Wald$	64.2	99.2	5.4	18.7	64.8	99.8	6.4	32.6
$Score$	2.0	99.3	4.8	99.9	3.6	99.8	4.4	100.0
Cox_o	4.1	100.0	4.6	100.0	1.6	100.0	2.4	100.0
Cox_a	1.4	100.0	3.4	99.9	0.3	100.0	4.4	100.0
	$\tilde{R}^2 = 0.8,\ \lambda = 0.2,\ \rho = 0.2$				$\tilde{R}^2 = 0.8,\ \lambda = 0.2,\ \rho = 0.8$			
J_1	6.0	33.1	3.4	23.8	5.7	31.7	5.1	34.4
J_2	5.6	41.1	3.7	30.9	5.0	53.0	4.7	50.7
J_{1a}	9.8	43.7	3.1	30.9	8.4	46.0	4.6	42.6
J_{2a}	8.0	72.4	4.8	57.5	9.0	97.4	4.7	96.2
Aug	5.0	58.3	4.7	59.1	4.3	100.0	4.8	100.0
$Wald$	28.5	90.0	2.6	29.9	30.2	99.9	4.9	93.8
$Score$	0.6	30.2	5.3	57.0	2.8	99.7	4.5	100.0
Cox_o	17.6	91.8	2.3	51.9	3.6	100.0	3.6	99.8
Cox_a	17.1	85.6	4.7	63.2	2.2	99.9	3.3	99.9
	$\tilde{R}^2 = 0.8,\ \lambda = 0.8,\ \rho = 0.2$				$\tilde{R}^2 = 0.8,\ \lambda = 0.8,\ \rho = 0.8$			
J_1	6.1	99.5	5.1	99.1	4.9	67.5	5.3	65.6
J_2	4.8	99.9	5.2	99.9	3.6	98.2	4.5	98.0
J_{1a}	9.1	93.7	4.1	91.4	9.0	65.6	4.7	63.0
J_{2a}	10.9	100.0	5.6	100.0	9.9	100.0	4.7	100.0
Aug	5.6	100.0	6.2	100.0	4.4	100.0	4.3	100.0
$Wald$	53.0	100.0	3.2	96.2	60.6	99.9	4.2	81.5
$Score$	2.9	100.0	4.4	100.0	2.7	99.6	3.8	99.6
Cox_o	7.0	100.0	4.7	100.0	5.3	100.0	5.3	100.0
Cox_a	5.0	100.0	5.6	100.0	1.0	100.0	5.0	100.0

† All empirical sizes and powers are expressed as percentages with the sign % being omitted. The "Asymptotic" and "Bootstrap" mean that the reported empirical size and power are computed by using, respectively, the asymptotic and bootstrapped p-values.

Table 6: Empirical size and power for experiment set III with normal disturbances and $n = 98^{\dagger}$

	Asymptotic		Bootstrap		Asymptotic		Bootstrap	
	Size	Power	Size	Power	Size	Power	Size	Power
	$\tilde{R}^2$=0.2, λ=0.2, ρ=0.2				$\tilde{R}^2$=0.2, λ=0.2, ρ=0.8			
J_1	5.2	92.4	3.7	91.9	4.5	74.9	8.1	81.3
J_2	2.8	98.1	3.8	97.8	3.5	99.1	4.9	97.9
J_{1a}	5.5	98.8	4.1	97.8	5.2	97.9	4.2	96.8
J_{2a}	4.8	98.7	3.9	97.8	4.4	99.4	4.6	98.7
Aug	0.8	99.6	3.2	99.8	1.1	99.6	3.3	99.9
$Wald$	94.3	99.9	2.3	11.9	85.6	100.0	1.1	8.3
$Score$	5.0	98.9	4.3	98.5	5.5	99.4	5.3	99.4
Cox_o	30.2	100.0	3.8	98.8	22.9	100.0	3.4	99.3
Cox_a	18.6	100.0	4.0	99.8	22.6	100.0	5.4	100.0
	$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	2.7	69.9	5.3	75.7	2.0	48.0	5.9	53.3
J_2	5.0	99.8	6.4	99.2	5.0	99.9	5.1	99.7
J_{1a}	5.8	97.5	5.3	96.8	6.1	97.9	4.8	97.1
J_{2a}	4.7	99.7	4.5	98.9	4.7	99.9	5.0	99.8
Aug	1.7	99.7	4.6	100.0	1.5	100.0	4.5	100.0
$Wald$	93.9	99.9	2.7	12.5	93.3	99.7	0.0	9.3
$Score$	5.4	98.5	4.9	97.9	8.9	100.0	7.0	99.2
Cox_o	28.2	100.0	4.0	98.1	16.5	100.0	3.5	99.6
Cox_a	21.1	100.0	4.8	99.8	17.5	100.0	4.9	100.0
	$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.8$			
J_1	4.6	100.0	5.6	100.0	6.2	99.8	4.8	99.8
J_2	4.5	100.0	5.8	100.0	4.7	100.0	5.0	100.0
J_{1a}	5.0	100.0	4.5	100.0	4.9	100.0	5.0	100.0
J_{2a}	4.8	100.0	4.7	100.0	4.6	100.0	5.2	100.0
Aug	2.0	100.0	5.0	100.0	1.8	100.0	4.7	100.0
$Wald$	74.6	100.0	6.0	48.6	44.7	100.0	4.4	13.6
$Score$	5.2	100.0	5.2	100.0	5.8	100.0	5.2	100.0
Cox_o	14.1	100.0	4.5	100.0	16.9	100.0	5.3	100.0
Cox_a	10.4	100.0	5.4	100.0	15.6	100.0	5.2	100.0
	$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	3.5	100.0	5.9	100.0	5.9	95.3	7.2	96.4
J_2	4.6	100.0	5.2	100.0	4.5	100.0	4.7	100.0
J_{1a}	5.2	100.0	5.0	100.0	5.0	97.6	4.8	95.2
J_{2a}	5.0	100.0	5.6	100.0	4.5	100.0	4.5	100.0
Aug	1.9	100.0	4.7	100.0	1.6	100.0	4.5	100.0
$Wald$	78.3	99.9	4.9	8.3	50.8	100.0	3.6	6.4
$Score$	4.9	100.0	4.3	100.0	6.8	100.0	6.0	100.0
Cox_o	14.6	100.0	4.8	100.0	10.0	100.0	4.3	100.0
Cox_a	9.8	100.0	4.8	100.0	12.3	100.0	5.3	100.0

† All empirical sizes and powers are expressed as percentages with the sign % being omitted. The "Asymptotic" and "Bootstrap" mean that the reported empirical size and power are computed by using, respectively, the asymptotic and bootstrapped p-values.

Table 7: Empirical size and power for experiment set III with chi-square disturbances and $n = 98^{\dagger}$

	Asymptotic		Bootstrap		Asymptotic		Bootstrap	
	Size	Power	Size	Power	Size	Power	Size	Power
	$\tilde{R}^2$=0.2, λ=0.2, ρ=0.2				$\tilde{R}^2$=0.2, λ=0.2, ρ=0.8			
J_1	6.2	91.9	5.4	92.1	5.0	75.8	8.0	79.4
J_2	4.8	98.2	4.0	97.2	4.5	98.6	3.9	97.6
J_{1a}	5.8	98.3	4.8	97.4	5.1	97.8	3.6	96.6
J_{2a}	5.3	97.8	4.9	96.6	4.5	98.4	4.0	97.7
Aug	1.5	99.0	5.5	99.8	1.7	99.1	5.7	99.9
Wald	75.6	100.0	2.6	24.3	66.5	100.0	3.7	11.7
Score	0.7	95.4	5.2	99.1	0.7	96.9	4.7	99.7
Cox_o	20.8	100.0	5.1	98.1	17.7	100.0	4.9	99.1
Cox_a	13.5	99.9	4.6	99.7	14.3	100.0	5.2	99.9
	$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.2$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	3.3	69.7	6.3	76.3	3.0	47.2	6.8	55.6
J_2	5.6	99.2	4.7	98.4	5.0	99.7	4.0	99.3
J_{1a}	5.4	97.7	4.5	96.2	5.9	97.6	4.5	96.0
J_{2a}	5.8	99.3	5.6	98.6	4.9	100.0	3.6	99.4
Aug	1.5	99.0	4.9	99.7	1.7	99.7	6.0	100.0
Wald	71.7	100.0	4.1	16.1	84.4	100.0	2.1	22.5
Score	0.7	93.5	5.9	99.3	1.4	97.2	6.3	99.7
Cox_o	16.7	100.0	4.2	98.3	12.6	100.0	4.7	99.1
Cox_a	12.4	100.0	5.0	99.7	10.0	100.0	5.4	100.0
	$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.2$, $\rho = 0.8$			
J_1	4.2	100.0	3.9	100.0	7.3	100.0	6.4	100.0
J_2	5.3	100.0	4.7	100.0	4.6	100.0	4.7	100.0
J_{1a}	5.8	100.0	4.7	100.0	5.3	100.0	5.0	99.9
J_{2a}	5.9	100.0	4.7	100.0	5.2	100.0	5.5	100.0
Aug	1.7	100.0	5.1	100.0	1.9	100.0	5.6	100.0
Wald	59.5	100.0	7.1	64.2	39.3	100.0	6.8	20.2
Score	1.7	100.0	4.1	100.0	1.7	100.0	4.7	100.0
Cox_o	11.1	100.0	5.4	100.0	11.8	100.0	5.5	100.0
Cox_a	8.7	100.0	4.9	100.0	10.2	100.0	5.8	100.0
	$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.2$				$\tilde{R}^2 = 0.8$, $\lambda = 0.8$, $\rho = 0.8$			
J_1	3.6	100.0	4.4	100.0	5.6	94.4	6.1	95.8
J_2	5.2	100.0	4.8	100.0	4.8	100.0	4.9	100.0
J_{1a}	6.1	100.0	5.5	100.0	5.9	98.9	5.6	96.9
J_{2a}	5.4	100.0	4.8	100.0	5.5	100.0	5.7	100.0
Aug	1.7	100.0	4.9	100.0	1.8	100.0	5.5	100.0
Wald	48.1	100.0	5.0	9.6	47.2	100.0	4.1	8.7
Score	1.0	100.0	4.1	100.0	0.9	100.0	5.0	100.0
Cox_o	8.0	100.0	4.2	100.0	5.3	100.0	3.6	100.0
Cox_a	8.8	100.0	6.0	100.0	7.0	100.0	6.0	100.0

† All empirical sizes and powers are expressed as percentages with the sign % being omitted. The "Asymptotic" and "Bootstrap" mean that the reported empirical size and power are computed by using, respectively, the asymptotic and bootstrapped p-values.

Table 8: Empirical size and power computed using asymptotic p-values for experiment set I with $n = 1519^{\dagger}$

	Normal		Chi-square		Normal		Chi-squares	
	Size	Power	Size	Power	Size	Power	Size	Power
	$\tilde{R}^2 = 0.2,\ \lambda = 0.2,\ \rho = 0.2$				$\tilde{R}^2 = 0.2,\ \lambda = 0.2,\ \rho = 0.8$			
J_1	5.3	16.1	4.4	15.1	6.2	14.9	5.6	14.0
J_2	4.4	19.4	3.9	17.3	7.7	26.1	8.3	24.2
J_{1a}	5.3	75.2	5.0	73.5	5.6	37.2	5.1	40.6
J_{2a}	4.6	97.9	5.0	97.7	4.8	99.5	6.0	99.4
Aug	4.5	98.4	4.4	98.1	5.1	100.0	4.8	100.0
$Wald$	72.1	99.7	65.5	99.3	40.1	100.0	39.0	100.0
$Score$	4.6	95.8	1.8	91.9	3.9	100.0	3.6	100.0
Cox_o	2.8	99.8	5.6	99.9	4.4	100.0	5.4	100.0
Cox_a	4.3	99.9	3.6	99.7	5.5	100.0	6.7	100.0
	$\tilde{R}^2 = 0.2,\ \lambda = 0.8,\ \rho = 0.2$				$\tilde{R}^2 = 0.2,\ \lambda = 0.8,\ \rho = 0.8$			
J_1	5.0	85.3	4.6	85.0	6.2	47.8	7.6	47.6
J_2	3.7	97.6	4.4	97.4	6.1	92.4	6.4	92.9
J_{1a}	5.9	99.4	6.3	99.1	3.2	95.1	4.1	95.1
J_{2a}	6.1	98.8	7.2	98.5	7.6	100.0	8.0	99.9
Aug	4.0	100.0	4.6	100.0	5.0	100.0	4.2	100.0
$Wald$	31.2	100.0	31.0	100.0	10.4	100.0	9.1	100.0
$Score$	5.0	100.0	3.8	100.0	4.6	100.0	3.7	100.0
Cox_o	5.1	100.0	6.2	100.0	1.7	100.0	3.1	100.0
Cox_a	5.2	100.0	6.5	100.0	2.1	100.0	2.5	100.0
	$\tilde{R}^2 = 0.8,\ \lambda = 0.2,\ \rho = 0.2$				$\tilde{R}^2 = 0.8,\ \lambda = 0.2,\ \rho = 0.8$			
J_1	4.7	97.0	4.7	96.5	5.0	85.3	4.8	85.6
J_2	4.8	97.1	4.5	97.0	5.1	87.7	5.2	87.2
J_{1a}	5.5	98.3	4.5	97.9	5.1	97.6	4.7	98.0
J_{2a}	4.8	100.0	5.4	100.0	5.4	100.0	4.6	100.0
Aug	4.9	100.0	4.8	100.0	4.7	100.0	5.5	100.0
$Wald$	20.6	100.0	17.9	100.0	11.2	100.0	10.9	100.0
$Score$	4.0	100.0	2.8	100.0	4.9	100.0	5.0	100.0
Cox_o	5.4	100.0	7.8	100.0	3.9	100.0	6.1	100.0
Cox_a	4.0	100.0	5.2	100.0	4.9	100.0	4.4	100.0
	$\tilde{R}^2 = 0.8,\ \lambda = 0.8,\ \rho = 0.2$				$\tilde{R}^2 = 0.8,\ \lambda = 0.8,\ \rho = 0.8$			
J_1	4.6	100.0	5.0	100.0	4.7	100.0	4.1	100.0
J_2	3.9	100.0	5.0	100.0	5.2	100.0	4.7	100.0
J_{1a}	4.1	100.0	6.6	100.0	5.1	100.0	5.0	100.0
J_{2a}	4.3	99.9	6.3	99.6	4.6	100.0	6.5	100.0
Aug	3.1	100.0	4.6	100.0	5.0	100.0	5.3	100.0
$Wald$	6.2	100.0	6.8	100.0	15.9	100.0	13.5	100.0
$Score$	5.3	100.0	4.4	100.0	4.3	100.0	3.5	100.0
Cox_o	5.1	100.0	5.1	100.0	3.1	100.0	5.5	100.0
Cox_a	4.4	100.0	4.5	100.0	3.1	100.0	3.8	100.0

† All empirical sizes and powers are expressed as percentages with the sign % being omitted.

Table 9: Empirical size and power computed using asymptotic p-values for experiment set II with $n = 1519^{\dagger}$

	Normal		Chi-square		Normal		Chi-squares	
	Size	Power	Size	Power	Size	Power	Size	Power
	$\tilde{R}^2 = 0.2,\ \lambda = 0.2,\ \rho = 0.2$				$\tilde{R}^2 = 0.2,\ \lambda = 0.2,\ \rho = 0.8$			
J_1	5.6	29.2	5.5	28.3	6.8	34.4	5.8	35.4
J_2	4.6	37.3	5.3	35.9	5.9	58.9	5.0	56.0
J_{1a}	6.4	42.3	6.8	40.2	5.5	45.8	4.7	47.1
J_{2a}	5.4	100.0	5.7	99.9	5.4	100.0	6.9	100.0
Aug	5.0	100.0	5.6	100.0	5.0	100.0	6.2	100.0
Wald	69.4	100.0	63.5	100.0	16.8	100.0	16.1	100.0
Score	4.8	100.0	2.8	100.0	5.8	100.0	5.0	100.0
Cox_o	12.7	100.0	12.6	100.0	5.5	100.0	5.4	100.0
Cox_a	11.8	100.0	12.9	100.0	4.5	100.0	4.9	100.0
	$\tilde{R}^2 = 0.2,\ \lambda = 0.8,\ \rho = 0.2$				$\tilde{R}^2 = 0.2,\ \lambda = 0.8,\ \rho = 0.8$			
J_1	5.5	99.5	4.5	99.3	5.1	69.8	5.0	70.4
J_2	4.6	100.0	4.6	99.9	3.4	99.1	3.0	99.0
J_{1a}	5.0	56.6	5.3	54.4	6.0	48.2	4.5	47.8
J_{2a}	6.1	100.0	5.4	100.0	5.7	100.0	6.0	100.0
Aug	5.0	100.0	4.7	100.0	4.5	100.0	4.9	100.0
Wald	16.8	100.0	16.6	100.0	33.5	100.0	34.2	100.0
Score	5.9	100.0	4.8	100.0	4.2	100.0	2.9	100.0
Cox_o	5.0	100.0	6.5	100.0	7.3	100.0	9.2	100.0
Cox_a	4.9	100.0	4.8	100.0	7.1	100.0	6.5	100.0
	$\tilde{R}^2 = 0.8,\ \lambda = 0.2,\ \rho = 0.2$				$\tilde{R}^2 = 0.8,\ \lambda = 0.2,\ \rho = 0.8$			
J_1	5.4	100.0	5.4	99.8	6.4	82.7	5.2	83.1
J_2	4.7	100.0	6.1	99.8	6.2	88.2	5.5	88.0
J_{1a}	5.8	99.8	5.8	99.8	6.0	93.2	5.6	90.8
J_{2a}	4.9	100.0	6.7	100.0	4.8	100.0	6.7	100.0
Aug	4.8	100.0	6.3	100.0	4.9	100.0	6.3	100.0
Wald	9.6	100.0	9.7	100.0	16.7	100.0	16.8	100.0
Score	5.4	100.0	3.6	100.0	5.7	100.0	4.5	100.0
Cox_o	9.4	100.0	9.3	100.0	6.0	100.0	5.1	100.0
Cox_a	8.4	100.0	7.4	100.0	5.1	100.0	5.5	100.0
	$\tilde{R}^2 = 0.8,\ \lambda = 0.8,\ \rho = 0.2$				$\tilde{R}^2 = 0.8,\ \lambda = 0.8,\ \rho = 0.8$			
J_1	5.5	100.0	5.5	100.0	5.4	100.0	5.0	100.0
J_2	5.0	100.0	5.9	100.0	4.7	100.0	5.8	100.0
J_{1a}	5.6	100.0	5.6	100.0	4.8	97.1	4.9	97.6
J_{2a}	5.6	100.0	6.0	100.0	5.2	100.0	6.8	100.0
Aug	4.8	100.0	5.2	100.0	4.9	100.0	5.6	100.0
Wald	15.5	100.0	16.6	100.0	12.9	100.0	14.0	100.0
Score	5.6	100.0	5.2	100.0	5.2	100.0	4.0	100.0
Cox_o	5.7	100.0	6.1	100.0	5.2	100.0	7.0	100.0
Cox_a	6.6	100.0	6.0	100.0	5.7	100.0	5.9	100.0

† All empirical sizes and powers are expressed as percentages with the sign % being omitted.

Table 10: Empirical size and power computed using asymptotic p-values for experiment set III with $n = 1519^{\dagger}$

	Normal		Chi-square		Normal		Chi-squares	
	Size	Power	Size	Power	Size	Power	Size	Power
	$\tilde{R}^2 = 0.2,\ \lambda = 0.2,\ \rho = 0.2$				$\tilde{R}^2 = 0.2,\ \lambda = 0.2,\ \rho = 0.8$			
J_1	9.2	100.0	11.4	100.0	8.8	99.9	10.5	99.9
J_2	5.8	100.0	5.7	100.0	5.1	100.0	5.3	100.0
J_{1a}	6.4	100.0	6.4	100.0	7.1	100.0	7.7	100.0
J_{2a}	5.8	100.0	5.5	100.0	5.4	100.0	5.6	100.0
Aug	1.6	100.0	2.2	100.0	1.0	100.0	2.2	100.0
Wald	95.9	88.3	92.7	100.0	66.6	83.1	62.4	100.0
Score	4.8	100.0	1.5	100.0	5.5	99.5	1.7	100.0
Cox_o	21.7	100.0	22.1	100.0	17.2	100.0	14.3	100.0
Cox_a	18.3	100.0	12.3	100.0	17.3	100.0	12.4	100.0
	$\tilde{R}^2 = 0.2,\ \lambda = 0.8,\ \rho = 0.2$				$\tilde{R}^2 = 0.2,\ \lambda = 0.8,\ \rho = 0.8$			
J_1	5.2	100.0	6.6	100.0	4.2	99.7	5.4	99.5
J_2	5.6	100.0	5.8	100.0	5.6	100.0	5.7	100.0
J_{1a}	5.4	100.0	6.8	100.0	6.3	100.0	7.5	100.0
J_{2a}	5.7	100.0	5.8	100.0	5.6	100.0	5.9	100.0
Aug	1.8	100.0	2.1	100.0	1.0	100.0	2.1	100.0
Wald	89.6	87.7	60.0	100.0	79.6	88.5	70.4	100.0
Score	6.1	100.0	2.0	100.0	5.6	100.0	2.8	100.0
Cox_o	19.9	100.0	13.3	100.0	5.4	100.0	5.2	100.0
Cox_a	21.3	100.0	11.0	100.0	6.7	100.0	4.3	100.0
	$\tilde{R}^2 = 0.8,\ \lambda = 0.2,\ \rho = 0.2$				$\tilde{R}^2 = 0.8,\ \lambda = 0.2,\ \rho = 0.8$			
J_1	4.9	100.0	5.7	100.0	6.0	100.0	6.9	100.0
J_2	5.7	100.0	5.7	100.0	5.4	100.0	5.7	100.0
J_{1a}	5.2	100.0	5.2	100.0	5.1	100.0	5.8	100.0
J_{2a}	5.5	100.0	5.4	100.0	5.5	100.0	5.9	100.0
Aug	1.7	100.0	2.1	100.0	1.0	100.0	2.1	100.0
Wald	71.6	98.1	73.1	100.0	46.7	99.5	48.5	100.0
Score	5.8	100.0	3.7	100.0	5.7	100.0	4.2	100.0
Cox_o	9.2	100.0	8.6	100.0	8.6	100.0	7.5	100.0
Cox_a	8.4	100.0	7.1	100.0	7.6	100.0	8.4	100.0
	$\tilde{R}^2 = 0.8,\ \lambda = 0.8,\ \rho = 0.2$				$\tilde{R}^2 = 0.8,\ \lambda = 0.8,\ \rho = 0.8$			
J_1	4.9	100.0	5.2	100.0	4.7	100.0	5.4	100.0
J_2	5.7	100.0	5.6	100.0	5.4	100.0	5.5	100.0
J_{1a}	1.2	100.0	1.3	100.0	4.8	100.0	5.1	100.0
J_{2a}	0.7	100.0	0.9	100.0	5.5	100.0	5.7	100.0
Aug	1.7	100.0	2.1	100.0	1.1	100.0	2.0	100.0
Wald	36.0	99.3	25.0	100.0	32.8	98.9	36.6	100.0
Score	6.0	100.0	2.9	100.0	7.0	99.9	5.1	100.0
Cox_o	7.2	100.0	8.1	100.0	6.1	100.0	4.2	100.0
Cox_a	7.1	100.0	6.8	100.0	6.2	100.0	5.5	100.0

† All empirical sizes and powers are expressed as percentages with the sign % being omitted.

neighbors are 0.2 and other elements are zero. The matrix is then used to re-estimate the SARAR model and we test the SARAR model with W_{foc} against the one with W_{5nn} and vice versa.

The estimation of the SARAR model with W_{5nn} generates similar parameter estimates and inference to that of the SARAR model with W_{foc}, with the exception of the parameter for proportion of owner-occupied units built prior to 1940, which becomes significant at the 5% level. The coefficient of determination[29] and log likelihood with W_{5nn} are, respectively, 0.888 and -18.9, higher than the corresponding values 0.866 and -56.0 for the SARAR model with W_{foc}.

We compute various test statistics for the SARAR models with the two different spatial weights matrices. To compute the Cox-type test statistics, (3) and (A.2) can be used for the numerators and (A.14)–(A.17) can be used for the denominators. The testing results at the 5% level are reported in Table 11. For the test of the SARAR model with W_{foc} against that with W_{5nn}, the results with asymptotic and bootstrapped p-values are the same: H_0 is rejected for all tests except J_1. For the test of the SARAR model with W_{5nn} against that with W_{foc}, J_1 and J_{1a} generate different results with asymptotic and bootstrapped p-values while other test statistics generate the same results. Based on the bootstrapped p-values, the null hypothesis with W_{5nn} cannot be rejected for all test statistics except *Wald* and *Score*. For the J-type and Cox-type tests based on the bootstrapped p-values, J_2, J_{1a}, J_{2a}, Cox_o and Cox_a are in favor of W_{5nn}, but J_1 is not able to distinguish the two matrices with the given data. In conclusion, most tests are in favor of W_{5nn}.

7. Conclusion

In this paper, we derive the Cox-type tests of non-nested hypotheses for SARAR models. We show that they are not asymptotically equivalent to the spatial J tests under the null hypothesis. We also prove that the bootstrap is consistent for Cox-type tests. The bootstrap may be used to remove the possible size distortion of the Cox-type tests in finite samples.

The performances of the Cox-type tests, spatial J tests, a LM test from a simple augmented model, the extended Wald and extended score tests (derived in the appendices) are compared in a Monte Carlo study. The extended Wald and Cox-type test statistics have large size distortions in some cases. But a simple bootstrap procedure essentially removes the size distortions of all tests. Using bootstrapped p-values, the Cox tests have relatively high power in all experiments and can outperform other tests in some cases. For the J-type tests, it turns out that alternative estimation methods may significantly improve the power over the ones based on spatial 2SLS estimation methods. With alternative estimation methods to implement the J test procedure, the Cox-type and such J-type tests can be complimentary to each other for some

[29]The coefficient of determination is defined as usual, i.e., one minus the ratio of the residual sum of squares over the total sum of squares.

Table 11: Testing results with a housing data set (Whether H_0 is rejected or not)[†]

	W_{foc} against W_{5nn}		W_{5nn} against W_{foc}	
Statistic	Asymptotic	Bootstrap	Asymptotic	Bootstrap
J_1	No	No	Yes	No
J_2	Yes	Yes	No	No
J_{1a}	Yes	Yes	Yes	No
J_{2a}	Yes	Yes	No	No
Aug	Yes	Yes	No	No
Wald	Yes	Yes	Yes	Yes
Score	Yes	Yes	Yes	Yes
Cox_o	Yes	Yes	No	No
Cox_a	Yes	Yes	No	No

[†] The "Asymptotic" and "Bootstrap" mean that test statistics are computed by using, respectively, the asymptotic and bootstrapped p-values. The "Yes" and "No" mean that H_0 is, respectively, rejected and not rejected at the 5% level of significance.

cases. For the two versions of the Cox test, we suggest the use of Atkinson's version (in (14)) because of its computational simplicity.

Appendix A. Notations and Expressions

For $j = 1, 2$, $\phi_j = (\lambda_j, \rho_j)'$, $\theta_j = (\phi_j', \beta_j', \sigma_j^2)'$, $R_{jn}(\rho_j) = I_n - \rho_j M_{jn}$, $S_{jn}(\lambda_j) = I_n - \lambda_j W_{jn}$, $L_{jn}(\theta_j)$ is the log likelihood function of the model (j), $\bar{L}_{jn}(\theta_j;\theta_i)$ is the expected value of $L_{jn}(\theta_j)$ when the model (i) with the parameter θ_i generates the data, and θ_{j0} is the true parameter vector of the model (j) when it generates the data. The $\bar{\theta}_{jn}(\theta_i)$ is the pseudo true value of the model (j) when the DGP is the model (i) with the parameter θ_i, and $\bar{\theta}_{jn,i} = \bar{\theta}_{jn}(\theta_{i0})$. Denote $R_{jn} = R_{jn}(\bar{\rho}_{jn,1})$, $S_{jn} = S_{jn}(\bar{\lambda}_{jn,1})$, $Q_{1n} = W_{1n}S_{1n}^{-1}$, $Q_{2n} = W_{2n}S_{1n}^{-1}$ and $T_{1n} = M_{1n}R_{1n}^{-1}$. For any square matrix A, $A^s = A + A'$.

As many identical terms appear in various matrices needed for the computation of test statistics in the paper, we define the following expressions:

$$RX_{1n} = R_{1n}X_{1n}, \qquad RSSR_n = R_{2n}S_{2n}S_{1n}^{-1}R_{1n}^{-1}, \qquad RD_n = R_{2n}(S_{2n}S_{1n}^{-1}X_{1n}\beta_{10} - X_{2n}\bar{\beta}_{2n,1}),$$

$$RX_{2n} = R_{2n}X_{2n}, \qquad MSSR_n = M_{2n}S_{2n}S_{1n}^{-1}R_{1n}^{-1}, \qquad MD_n = M_{2n}(S_{2n}S_{1n}^{-1}X_{1n}\beta_{10} - X_{2n}\bar{\beta}_{2n,1}),$$

$$RQR_{1n} = R_{1n}Q_{1n}R_{1n}^{-1}, \qquad RQX\beta_{1n} = R_{1n}Q_{1n}X_{1n}\beta_{10}, \qquad RSSQR_n = R_{2n}S_{2n}S_{1n}^{-1}Q_{1n}R_{1n}^{-1},$$

$$RQR_{2n} = R_{2n}Q_{2n}R_{1n}^{-1}, \qquad RQX\beta_{2n} = R_{2n}Q_{2n}X_{1n}\beta_{10}, \qquad RSSQX\beta_n = R_{2n}S_{2n}S_{1n}^{-1}Q_{1n}X_{1n}\beta_{10},$$

$$RSSX_n = R_{2n}S_{2n}S_{1n}^{-1}X_{1n}.$$

The concentrated quasi log likelihood function $L_{jn}(\phi_j) = \max_{\beta_j,\sigma_j^2} L_{jn}(\theta_j)$ for $j = 1, 2$ is equal to

$$L_{jn}(\phi_j) = -\frac{n}{2}[\ln(2\pi) + 1] - \frac{n}{2}\ln\hat{\sigma}_{jn}^2(\phi_j) + \ln|S_{jn}(\lambda_j)| + \ln|R_{jn}(\rho_j)|, \tag{A.1}$$

where $\hat{\sigma}_{jn}^2(\phi_j) = n^{-1}y_n'S_{jn}'(\lambda_j)R_{jn}'(\rho_j)H_{jn}(\rho_j)R_{jn}(\rho_j)S_{jn}(\lambda_j)y_n$ with
$H_{jn}(\rho_j) = I_n - R_{jn}(\rho_j)X_{jn}[X_{jn}'R_{jn}'(\rho_j)R_{jn}(\rho_j)X_{jn}]^{-1}X_{jn}'R_{jn}'(\rho_j)$. The $\bar{L}_{jn}(\theta_j;\theta_{10}) = \mathrm{E}\,L_{jn}(\theta_j)$ is

$$\begin{aligned}\bar{L}_{jn}(\theta_j;\theta_{10}) = &-\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\sigma_j^2 + \ln|S_{jn}(\lambda_j)| + \ln|R_{jn}(\rho_j)| \\ &- \frac{\sigma_{10}^2}{2\sigma_j^2}tr[R_{1n}'^{-1}S_{1n}'^{-1}S_{jn}'(\lambda_j)R_{jn}'(\rho_j)R_{jn}(\rho_j)S_{jn}(\lambda_j)S_{1n}^{-1}R_{1n}^{-1}] \\ &- \frac{1}{2\sigma_j^2}[S_{jn}(\lambda_j)S_{1n}^{-1}X_{1n}\beta_{10} - X_{jn}\beta_j]'R_{jn}'(\rho_j)R_{jn}(\rho_j)[S_{jn}(\lambda_j)S_{1n}^{-1}X_{1n}\beta_{10} - X_{jn}\beta_j].\end{aligned} \tag{A.2}$$

By the maximization of $\bar{L}_{2n}(\theta_2;\theta_{10})$ for a given ϕ_2, we have

$$\bar{\beta}_{2n}(\phi_2;\theta_{10}) = [X_{2n}'R_{2n}'(\rho_2)R_{2n}(\rho_2)X_{2n}]^{-1}X_{2n}'R_{2n}'(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S_{1n}^{-1}X_{1n}\beta_{10}, \tag{A.3}$$

$$\begin{aligned}\bar{\sigma}_{2n}^2(\phi_2;\theta_{10}) = &\frac{\sigma_{10}^2}{n}\mathrm{tr}[R_{1n}'^{-1}S_{1n}'^{-1}S_{2n}'(\lambda_2)R_{2n}'(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S_{1n}^{-1}R_{1n}^{-1}] \\ &+ \frac{1}{n}(X_{1n}\beta_{10})'S_{1n}'^{-1}S_{2n}'(\lambda_2)R_{2n}'(\rho_2)H_{2n}(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S_{1n}^{-1}X_{1n}\beta_{10}.\end{aligned} \tag{A.4}$$

Then $\bar{L}_{jn}(\phi_j;\theta_{10}) = \max_{\beta_j,\sigma_j^2} \bar{L}_{jn}(\theta_j;\theta_{10})$ is

$$\bar{L}_{jn}(\phi_j;\theta_{10}) = -\frac{n}{2}[\ln(2\pi)+1] - \frac{n}{2}\ln\bar{\sigma}_{jn}^2(\phi_j;\theta_{10}) + \ln|S_{jn}(\lambda_j)| + \ln|R_{jn}(\rho_j)|. \tag{A.5}$$

The first order derivatives of $L_{1n}(\theta_1)$ at θ_{10} are

$$\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial\lambda_1} = \frac{1}{\sqrt{n}\sigma_{10}^2}[\epsilon_{1n}' RQR_{1n}\epsilon_{1n} - \sigma_{10}^2\operatorname{tr}(Q_{1n})] + \frac{1}{\sqrt{n}\sigma_{10}^2} RQX\beta_{1n}'\epsilon_{1n}, \tag{A.6}$$

$$\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial\rho_1} = \frac{1}{\sqrt{n}\sigma_{10}^2}[\epsilon_{1n}' T_{1n}\epsilon_{1n} - \sigma_{10}^2\operatorname{tr}(T_{1n})], \tag{A.7}$$

$$\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial\beta_1} = \frac{1}{\sqrt{n}\sigma_{10}^2} RX_{1n}'\epsilon_{1n}, \tag{A.8}$$

$$\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial\sigma_1^2} = \frac{1}{2\sqrt{n}\sigma_{10}^4}(\epsilon_{1n}'\epsilon_{1n} - n\sigma_{10}^2), \tag{A.9}$$

The first order derivatives of $L_{2n}(\theta_2)$ at $\bar{\theta}_{2n,1}$ with y_n expressed as the model (1) being the DGP are

$$\begin{aligned}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\lambda_2} &= \frac{1}{\bar{\sigma}_{2n,1}^2}(RD_n' RQR_{2n} + RQX\beta_{2n}' RSSR_n)\epsilon_{1n} \\ &\quad + \frac{1}{\bar{\sigma}_{2n,1}^2}[\epsilon_{1n}' RQR_{2n}' RSSR_n\epsilon_{1n} - \sigma_{10}^2\operatorname{tr}(RQR_{2n}' RSSR_n)],\end{aligned} \tag{A.10}$$

$$\begin{aligned}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\rho_2} &= \frac{1}{\bar{\sigma}_{2n,1}^2}(MD_n' RSSR_n + RD_n' MSSR_n)\epsilon_{1n} \\ &\quad + \frac{1}{\bar{\sigma}_{2n,1}^2}[\epsilon_{1n}' MSSR_n' RSSR_n\epsilon_{1n} - \sigma_{10}^2\operatorname{tr}(MSSR_n' RSSR_n)],\end{aligned} \tag{A.11}$$

$$\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\beta_2} = \frac{1}{\bar{\sigma}_{2n,1}^2} RX_{2n}' RSSR_n\epsilon_{1n}, \tag{A.12}$$

$$\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\sigma_2^2} = \frac{1}{\bar{\sigma}_{2n,1}^4} RD_n' RSSR_n\epsilon_{1n} + \frac{1}{2\bar{\sigma}_{2n,1}^4}[\epsilon_{1n}' RSSR_n' RSSR_n\epsilon_{1n} - \sigma_{10}^2\operatorname{tr}(RSSR_n' RSSR_n)]. \tag{A.13}$$

For any n-dimensional square matrices A_n and B_n, and n-dimensional vectors a_n and b_n, let $\Pi_1(A_n, a_n, B_n, b_n) = \mathrm{E}\big[\big(\epsilon_{1n}' A_n\epsilon_{1n} - \sigma_0^2\operatorname{tr}(A_n) + a_n'\epsilon_{1n}\big)\big(\epsilon_{1n}' B_n\epsilon_{1n} - \sigma_0^2\operatorname{tr}(B_n) + b_n'\epsilon_{1n}\big)\big]$, which is the covariance of two linear-quadratic forms. The detailed expression for $\Pi_1(A_n, a_n, B_n, b_n)$ is given in Lemma 1. Denote $\Pi_1(A_n, a_n) = \Pi_1(A_n, a_n, A_n, a_n)$ for short. Let μ_{31} be the third moment of ϵ_{1n}, $0_{i\times j}$ be an $i\times j$ matrix of zeros, and $\operatorname{vec}_{\mathrm{D}}(A_n)$ be a column vector consisting of the diagonal elements of A_n. Then according to (A.6)–(A.9), the symmetric matrix $\Omega_{1n,1}$ in (5) is[30]

$$\Omega_{1n,1} = \frac{1}{n\sigma_{10}^4}\cdot \begin{pmatrix} \Pi_1(RQR_{1n}, RQX\beta_{1n}) & * & * & * \\ \Pi_1(T_{1n}, 0_{n\times1}, RQR_{1n}, RQX\beta_{1n}) & \Pi_1(T_{1n}, 0_{n\times1}) & * & * \\ RX_{1n}'[\mu_{31}\operatorname{vec}_{\mathrm{D}}(RQR_{1n}) + \sigma_{10}^2 RQX\beta_{1n}] & \mu_{31}RX_{1n}'\operatorname{vec}_{\mathrm{D}}(T_{1n}) & \sigma_{10}^2 RX_{1n}' RX_{1n} & * \\ \frac{1}{2\sigma_{10}^2}\Pi_1(I_n, 0_{n\times1}, RQR_{1n}, RQX\beta_{1n}) & \frac{1}{2\sigma_{10}^2}\Pi_1(I_n, 0_{n\times1}, T_{1n}, 0_{n\times1}) & \frac{\mu_{31}}{2\sigma_{10}^2}\operatorname{vec}_{\mathrm{D}}'(I_n) RX_{1n} & \frac{1}{4\sigma_{10}^4}\Pi_1(I_n, 0_{n\times1}) \end{pmatrix}.$$

[30]When $\epsilon_{1n,i}$'s are normal, as $\Omega_{1n,1} = \Sigma_{1n,1}$, only $\Omega_{1n,1}$ or $\Sigma_{1n,1}$ needs to be estimated.

According to (A.10)–(A.13), the symmetric matrix $\Omega_{2n,1}$ in (7) may be written as a 4×4 block matrix, where the $(1,1)$th block is

$$\frac{1}{n\bar{\sigma}_{2n,1}^4}\Pi_1(RQR_{2n}'RSSR_n, RQR_{2n}'RD_n + RSSR_n'RQX\beta_{2n}),$$

the $(2,1)$th block is

$$\frac{1}{n\bar{\sigma}_{2n,1}^4}\Pi_1(MSSR_n'RSSR_n, MSSR_n'RD_n + RSSR_n'MD_n, RQR_{2n}'RSSR_n, RQR_{2n}'RD_n + RSSR_n'RQX\beta_{2n}),$$

the $(2,2)$th block is

$$\frac{1}{n\bar{\sigma}_{2n,1}^4}\Pi_1(MSSR_n'RSSR_n, MSSR_n'RD_n + RSSR_n'MD_n),$$

the $(3,1)$th, $(3,2)$th and $(3,3)$th blocks form the vector

$$\begin{aligned}&\frac{1}{n\bar{\sigma}_{2n,1}^4}RX_{2n}'RSSR_n[\mu_{31}\operatorname{vec}_{\mathrm{D}}(RQR_{2n}'RSSR_n) + \sigma_{10}^2(RQR_{2n}'RD_n + RSSR_n'RQX\beta_{2n}),\\ &\mu_{31}\operatorname{vec}_{\mathrm{D}}(MSSR_n'RSSR_n) + \sigma_{10}^2(MSSR_n'RD_n + RSSR_n'MD_n), \sigma_{10}^2 RSSR_n'RX_{2n}],\end{aligned}$$

and, the $(4,1)$th, $(4,2)$th, $(4,3)$th and $(4,4)$th blocks form the vector

$$\begin{aligned}&\frac{1}{n\bar{\sigma}_{2n,1}^6}\big[\Pi_1\big(\frac{1}{2}RSSR_n'RSSR_n, RSSR_n'RD_n, RQR_{2n}'RSSR_n, RQR_{2n}'RD_n + RSSR_n'RQX\beta_{2n}\big),\\ &\Pi_1\big(\frac{1}{2}RSSR_n'RSSR_n, RSSR_n'RD_n, MSSR_n'RSSR_n, MSSR_n'RD_n + RSSR_n'MD_n\big),\\ &\big(\frac{\mu_{31}}{2}\operatorname{vec}_{\mathrm{D}}'(RSSR_n'RSSR_n) + \sigma_{10}^2 RD_n'RSSR_n\big)RSSR_n'RX_{2n}, \Pi_1\big(\frac{1}{2}RSSR_n'RSSR_n, RSSR_n'RD_n\big)\big].\end{aligned}$$

For computational simplicity, $\Sigma_{1n,1}$ in (5) may be estimated by $\frac{1}{n}\frac{\partial^2 L_{1n}(\hat{\theta}_{1n})}{\partial\theta_1\partial\theta_1'}$, and $\Sigma_{2n,1}$ in (7) may be estimated by $\frac{1}{n}\frac{\partial^2 L_{2n}(\hat{\theta}_{2n})}{\partial\theta_2\partial\theta_2'}$, as shown in the proof of Proposition 4. We thus only give the expressions for $\frac{\partial^2 L_{jn}(\theta_j)}{\partial\theta_j\partial\theta_j'}$. For $j=1,2$,

$$\begin{aligned}
\frac{\partial^2 L_{jn}(\theta_j)}{\partial\lambda_j^2} &= -\operatorname{tr}[W_{jn}S_{jn}^{-1}(\lambda_j)W_{jn}S_{jn}^{-1}(\lambda_j)] - \frac{1}{\sigma_j^2}y_n'W_{jn}'R_{jn}'(\rho_j)R_{jn}(\rho_j)W_{jn}y_n,\\
\frac{\partial^2 L_{jn}(\theta_j)}{\partial\lambda_j\partial\rho_j} &= -\frac{1}{\sigma_j^2}y_n'W_{jn}'[M_{jn}'R_{jn}(\rho_j) + R_{jn}'(\rho_j)M_{jn}][S_{jn}(\lambda_j)y_n - X_{jn}\beta_j],\\
\frac{\partial^2 L_{jn}(\theta_j)}{\partial\lambda_j\partial\beta_j} &= -\frac{1}{\sigma_j^2}X_{jn}'R_{jn}'(\rho_j)R_{jn}(\rho_j)W_{jn}y_n,\\
\frac{\partial^2 L_{jn}(\theta_j)}{\partial\lambda_j\partial\sigma_j^2} &= -\frac{1}{\sigma_j^4}y_n'W_{jn}'R_{jn}'(\rho_j)R_{jn}(\rho_j)[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j],\\
\frac{\partial^2 L_{jn}(\theta_j)}{\partial\rho_j\partial\rho_j} &= -\operatorname{tr}[M_{jn}R_{jn}^{-1}(\rho_j)M_{jn}R_{jn}^{-1}(\rho_j)] - \frac{1}{\sigma_j^2}[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j]'M_{jn}'M_{jn}[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j],\\
\frac{\partial^2 L_{jn}(\theta_j)}{\partial\rho_j\partial\beta_j} &= -\frac{1}{\sigma_j^2}X_{jn}'[M_{jn}'R_{jn}(\rho_j) + R_{jn}'(\rho_j)M_{jn}][S_{jn}(\lambda_j)y_n - X_{jn}\beta_j],\\
\frac{\partial^2 L_{jn}(\theta_j)}{\partial\rho_j\partial\sigma_j^2} &= -\frac{1}{\sigma_j^4}[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j]'M_{jn}'R_{jn}(\rho_j)[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j],
\end{aligned}$$

$$\frac{\partial^2 L_{jn}(\theta_j)}{\partial\beta_j\partial\beta_j'} = -\frac{1}{\sigma_j^2}X_{jn}'R_{jn}'(\rho_j)R_{jn}(\rho_j)X_{jn},$$

$$\frac{\partial^2 L_{jn}(\theta_j)}{\partial\beta_j\partial\sigma_j^2} = -\frac{1}{\sigma_j^4}X_{jn}'R_{jn}'(\rho_j)R_{jn}(\rho_j)[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j],$$

$$\frac{\partial^2 L_{jn}(\theta_j)}{\partial(\sigma_j^2)^2} = \frac{n}{2\sigma_j^4} - \frac{1}{\sigma_j^6}[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j]'R_{jn}'(\rho_j)R_{jn}(\rho_j)[S_{jn}(\lambda_j)y_n - X_{jn}\beta_j].$$

In (11) which gives the expression for $\sigma_{c,n}^2$, $C_{2n,1}$ is equal to

$$\begin{aligned} C_{2n,1} = -\frac{1}{n}\big[&\frac{1}{\bar{\sigma}_{2n,1}^2}RD_n'RSSQX\beta_n + \frac{\sigma_{10}^2}{\bar{\sigma}_{2n,1}^2}\operatorname{tr}(RSSQR_n'RSSR_n), \\ &\frac{\sigma_{10}^2}{\bar{\sigma}_{2n,1}^2}\operatorname{tr}(RSSR_n'RSSR_nT_{1n}), \frac{1}{\bar{\sigma}_{2n,1}^2}RD_n'RSSX_n, \frac{1}{2\bar{\sigma}_{2n,1}^2}\operatorname{tr}(RSSR_n'RSSR_n)\big]', \end{aligned} \tag{A.14}$$

$$\frac{1}{n}\operatorname{var}\begin{pmatrix}[L_{2n}(\bar{\theta}_{2n,1}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})] \\ \frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1}\end{pmatrix} = \begin{pmatrix}\frac{1}{n}\operatorname{var}\big(L_{2n}(\bar{\theta}_{2n,1}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})\big) & \frac{1}{n}\operatorname{E}\big(L_{2n}(\bar{\theta}_{2n,1})\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1'}\big) \\ \frac{1}{n}\operatorname{E}\big(L_{2n}(\bar{\theta}_{2n,1})\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1}\big) & \Omega_{1n,1}\end{pmatrix}, \tag{A.15}$$

where

$$\operatorname{var}\big(L_{2n}(\bar{\theta}_{2n,1}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})\big) = \frac{1}{\bar{\sigma}_{2n,1}^4}\Pi_1\big(\frac{1}{2}RSSR_n'RSSR_n, RSSR_n'RD_n\big), \tag{A.16}$$

and

$$\begin{aligned} &\operatorname{E}\big(L_{2n}(\bar{\theta}_{2n,1})\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1'}\big) \\ &= -\frac{1}{\sigma_{10}^2\bar{\sigma}_{2n,1}^2}\big[\Pi_1\big(\frac{1}{2}RSSR_n'RSSR_n, RSSR_n'RD_n, RQR_{1n}, RQX\beta_{1n}\big), \Pi_1\big(\frac{1}{2}RSSR_n'RSSR_n, RSSR_n'RD_n, T_{1n}, 0_{n\times1}\big), \\ &\quad\big(\frac{\mu_{31}}{2}vec_D'(RSSR_n'RSSR_n) + RD_n'RSSR_n\big)RX_{1n}, \frac{1}{2\sigma_{10}^2}\Pi_1\big(\frac{1}{2}RSSR_n'RSSR_n, RSSR_n'RD_n, I_n, 0_{n\times1}\big)\big]. \end{aligned} \tag{A.17}$$

For $P_{2n,1} = \frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\theta_2\partial\theta_1'}$ in (B.2), we have

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\lambda_2\partial\lambda_1} = \frac{1}{\bar{\sigma}^2_{2n,1}}(RD_n' R_{2n}Q_{2n}Q_{1n}X_{1n}\beta_{10} + RSSQX\beta_n' RQX\beta_{2n}) + \frac{\sigma^2_{10}}{\bar{\sigma}^2_{2n,1}}\operatorname{tr}\big(RSSQR_n' RQR_{2n} + R_{1n}'^{-1}Q_{1n}'Q_{2n}'R_{2n}'RSSR_n\big),$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\lambda_2\partial\rho_1} = \frac{\sigma^2_{10}}{\bar{\sigma}^2_{2n,1}}\operatorname{tr}\big(T_{1n}'(RSSR_n' RQR_{2n})^s\big),$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\lambda_2\partial\beta_1} = \frac{1}{\bar{\sigma}^2_{2n,1}}(RSSX_n' RQX\beta_{2n} + X_{1n}'Q_{2n}'R_{2n}'RD_n),$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\lambda_2\partial\sigma_1^2} = \frac{1}{\bar{\sigma}^2_{2n,1}}\operatorname{tr}(RSSR_n' RQR_{2n}),$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\rho_2\partial\lambda_1} = \frac{1}{\bar{\sigma}^2_{2n,1}}(MD_n' R_{2n} + RD_n' M_{2n})S_{2n}S_{1n}^{-1}Q_{1n}X_{1n}\beta_{10} + \frac{\sigma^2_{10}}{\bar{\sigma}^2_{2n,1}}\operatorname{tr}\big(R_{1n}'^{-1}Q_{1n}'S_{1n}'^{-1}S_{2n}'(M_{2n}'RSSR_n + R_{2n}'MSSR_n)\big),$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\rho_2\partial\rho_1} = \frac{\sigma^2_{10}}{\bar{\sigma}^2_{2n,1}}\operatorname{tr}\big(T_{1n}'(MSSR_n' RSSR_n)^s\big), \quad \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\rho_2\partial\beta_1} = \frac{1}{\bar{\sigma}^2_{2n,1}}X_{1n}'S_{1n}'^{-1}S_{2n}'(M_{2n}'RD_n + R_{2n}'MD_n),$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\rho_2\partial\sigma_1^2} = \frac{1}{\bar{\sigma}^2_{2n,1}}\operatorname{tr}(RSSR_n' MSSR_n), \quad \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\beta_2\partial\lambda_1} = \frac{1}{\bar{\sigma}^2_{2n,1}}RX_{2n}'RSSQX\beta_n,$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\beta_2\partial\rho_1} = 0_{k_2\times 1}, \quad \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\beta_2\partial\beta_1'} = \frac{1}{\bar{\sigma}^2_{2n,1}}RX_{2n}'RSSX_n,$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\beta_2\partial\sigma_1^2} = 0_{k_2\times 1}, \quad \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\sigma_2^2\partial\rho_1} = \frac{\sigma^2_{10}}{\bar{\sigma}^4_{2n,1}}\operatorname{tr}(T_{1n}'RSSR_n' RSSR_n),$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\sigma_2^2\partial\beta_1} = \frac{1}{\bar{\sigma}^4_{2n,1}}RSSX_n' RD_n, \quad \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\sigma_2^2\partial\sigma_1^2} = \frac{1}{2\bar{\sigma}^4_{2n,1}}\operatorname{tr}(RSSR_n' RSSR_n),$$

$$\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\sigma_2^2\partial\lambda_1} = \frac{1}{\bar{\sigma}^4_{2n,1}}[RD_n' RSSQX\beta_n + \sigma^2_{10}\operatorname{tr}(RSSR_n' RSSQR_n)].$$

The $V_{2n,1}$ in (B.3) is

$$V_{2n,1} = [I_{k_2}, -P_{2n,1}\Sigma^{-1}_{1n,1}]\begin{pmatrix} \Omega_{2n,1} & \frac{1}{n}\operatorname{E}\big(\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2}\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1'}\big) \\ \frac{1}{n}\operatorname{E}\big(\frac{\partial L_{1n}(\bar{\theta}_{10})}{\partial\theta_1}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2'}\big) & \Omega_{1n,1}\end{pmatrix}\begin{pmatrix} I_{k_2} \\ -\Sigma^{-1}_{1n,1}P_{2n,1}'\end{pmatrix},$$

where the expression for $\frac{1}{n}\operatorname{E}\big(\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2}\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1'}\big)$ can be derived from (A.6)–(A.13).

Appendix B. The Extended Wald and Extended Score Tests

Appendix B.1. The Extended Wald Test

Under the null hypothesis, both $\hat{\theta}_{2n}$ and $\bar{\theta}_{2n}(\hat{\theta}_{1n})$ are estimators of the pseudo-true value $\bar{\theta}_{2n,1}$ and their difference can be shown to converge to zero in probability. We would like to test whether this difference, after being properly scaled, is significantly different from zero, i.e., whether the null hypothesis could explain the alternative model significantly well. This gives rise to the extended Wald test, which is based on

$$\sqrt{n}\big(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n})\big) = \sqrt{n}\big(\hat{\theta}_{2n} - \bar{\theta}_{2n,1}\big) - \sqrt{n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \bar{\theta}_{2n,1}\big). \tag{B.1}$$

The first term on the right hand side of the above equation has been shown to be asymptotically normal with mean zero by using (6). The second term is also asymptotically normal. Jointly, the asymptotical distribution of $\sqrt{n}\big(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n})\big)$ can be obtained. By the mean value theorem,

$$0 = \frac{\partial \bar{L}_{2n}(\bar{\theta}_{2n}(\hat{\theta}_{1n}); \hat{\theta}_{1n})}{\partial \theta_2} = \frac{\partial \bar{L}_{2n}(\bar{\theta}_{2n,1}; \hat{\theta}_{1n})}{\partial \theta_2} + \frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}_{2n,1}; \hat{\theta}_{1n})}{\partial \theta_2 \partial \theta_2'}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \bar{\theta}_{2n,1}\big),$$

where $\tilde{\theta}_{2n,1}$ is between $\bar{\theta}_{2n,1}$ and $\bar{\theta}_{2n}(\hat{\theta}_{1n})$. Thus,

$$\begin{aligned}
\sqrt{n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \bar{\theta}_{2n,1}\big) &= \Big(-\frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}_{2n,1}; \hat{\theta}_{1n})}{\partial \theta_2 \partial \theta_2'}\Big)^{-1} \frac{1}{\sqrt{n}}\frac{\partial \bar{L}_{2n}(\bar{\theta}_{2n,1}; \hat{\theta}_{1n})}{\partial \theta_2} \\
&= \Sigma_{2n,1}^{-1} P_{2n,1}\sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) + o_P(1) \\
&= \Sigma_{2n,1}^{-1} P_{2n,1}\Sigma_{1n,1}^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1} + o_P(1),
\end{aligned}$$

where $P_{2n,1} = \frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial \theta_2 \partial \theta_1'}$. Therefore,

$$\sqrt{n}\big(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n})\big) = \Sigma_{2n,1}^{-1}\Big(\frac{1}{\sqrt{n}}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial \theta_2} - P_{2n,1}\Sigma_{1n,1}^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1}\Big) + o_P(1). \tag{B.2}$$

The partial derivatives of the log-likelihood functions at the true or pseudo-true values have been shown to be linear-quadratic forms of ϵ_{1n}, so $\sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1})$ is asymptotically normal.

Proposition 7. *Under H_0 and Assumptions 1–4, 9—16,*

$$\sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n})) \xrightarrow{d} N\big(0, \lim_{n\to\infty}(\Sigma_{2n,1}^{-1} V_{2n,1}\Sigma_{2n,1}^{-1})\big), \tag{B.3}$$

where $V_{2n,1} = \mathrm{var}\big(\frac{1}{\sqrt{n}}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial \theta_2} - P_{2n,1}\Sigma_{1n,1}^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1}\big)$. When $\epsilon_{1n,i}$'s are normally distributed, $V_{2n,1} = \Omega_{2n,1} - P_{2n,1}\Sigma_{1n,1}^{-1}P_{2n,1}'$.

When $\epsilon_{1n,i}$'s are normally distributed, $L_{1n}(\theta_{10})$ is the true probability density function. Then $P_{2n,1} = \mathrm{E}\big(\frac{1}{n}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial \theta_2}\frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1'}\big)$ and the information matrix equality holds for $L_{1n}(\theta_{10})$. Similar to the case of non-spatial models (Gourieroux et al., 1983), $\mathrm{avar}\big(\sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n}))\big) = \mathrm{avar}\big(\sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1})\big) - \mathrm{avar}\big(\sqrt{n}(\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \bar{\theta}_{2n,1})\big)$, where $\mathrm{avar}(\cdot)$ denotes the asymptotic VC matrix. Thus $\bar{\theta}_{2n}(\hat{\theta}_{1n})$ as an estimator for $\bar{\theta}_{2n,1}$ is more efficient than $\hat{\theta}_{2n}$.

Let $\hat{\Sigma}_{2n,1}$ and $\hat{V}_{2n,1}$ be, respectively, estimators of $\Sigma_{2n,1}$ and $V_{2n,1}$ such that $\hat{\Sigma}_{2n,1} - \Sigma_{2n,1} = o_P(1)$ and $\hat{V}_{2n,1} - V_{2n,1} = o_P(1)$, and $\hat{V}_{2n,1}^{+}$ be a generalized inverse of $\hat{V}_{2n,1}$. If $\lim_{n\to\infty}\Pr\big(\mathrm{rk}(\hat{V}_{2n,1}) = \mathrm{rk}(\lim_{n\to\infty} V_{2n,1})\big) =$

1 (Andrews, 1987), where $\mathrm{rk}(\cdot)$ denotes the rank of a matrix, then under the null hypothesis, the extended Wald test statistic

$$Wald = n\big(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n})\big)'\hat{\Sigma}_{2n,1}\hat{V}_{2n,1}^{+}\hat{\Sigma}_{2n,1}\big(\hat{\theta}_{2n} - \bar{\theta}_{2n}(\hat{\theta}_{1n})\big) \tag{B.4}$$

is asymptotically distributed as a chi-square with degrees of freedom df given by the rank of $\lim_{n\to\infty} V_{2n,1}$. The extended Wald test of H_0 against H_1 rejects H_0 if $Wald_e > \chi^2_{1-\alpha}(df)$, where $\chi^2_{1-\alpha}(df)$ is the $(1-\alpha)$ quantile of a chi-square distribution with df degrees of freedom for the chosen level of significance α, and does not reject otherwise.

The $V_{2n,1}$, even in the case where the DGP has normal i.i.d. disturbances, has a complicated form and the rank of $\lim_{n\to\infty} V_{2n,1}$ is hard to check. One solution is to test rank constraints and estimate the rank via a series of tests. But this kind of procedure often fails when the estimated matrix is positive semidefinite.[31] Another solution is to modify the test statistic to make sure that the involved matrix has full rank.[32]

Appendix B.2. The Extended Score Test

Under the null hypothesis,

$$\frac{1}{\sqrt{n}}\frac{\partial L_{2n}\big(\bar{\theta}_{2n}(\hat{\theta}_{1n})\big)}{\partial \theta_2} = \Sigma_{2n,1}\sqrt{n}(\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \hat{\theta}_{2n}) + o_P(1), \tag{B.5}$$

which is asymptotically normal with mean zero and limiting VC matrix $\lim_{n\to\infty} V_{2n,1}$. Then the extended score test statistic

$$Score = \frac{1}{n}\Big(\frac{\partial L_{2n}(\bar{\theta}_{2n}(\hat{\theta}_{1n}))}{\partial \theta_2'}\Big)\hat{V}_{2n,1}^{+}\Big(\frac{\partial L_{2n}(\bar{\theta}_{2n}(\hat{\theta}_{1n}))}{\partial \theta_2}\Big) \tag{B.6}$$

is asymptotically chi-square distributed with degrees of freedom df, if $\lim_{n\to\infty} P\big(\mathrm{rk}(\hat{V}_{2n,1}) = \mathrm{rk}(\lim_{n\to\infty} V_{2n,1})\big) = 1$. From (B.2) and (B.4)–(B.6), it is clear that the extended Wald and score statistics are asymptotically equivalent under the null hypothesis.

Appendix C. Lemmas

Lemma 1. *Suppose that A_n and B_n are n-dimensional square matrices, a_n and b_n are n-dimensional vectors, and ϵ_{ni}'s in $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})'$ are i.i.d. with mean zero, variance σ_0^2, third moment μ_3 and finite fourth moment μ_4. Then,*

i) $\mathrm{E}(\epsilon_n \cdot \epsilon_n' A_n \epsilon_n) = \mu_3 \,\mathrm{vec_D}(A_n)$,

ii) $\mathrm{E}(\epsilon_n' A_n \epsilon_n \cdot \epsilon_n' B_n \epsilon_n) = (\mu_4 - 3\sigma_0^4) vec_D'(A_n)\,\mathrm{vec_D}(B_n) + \sigma_0^4 \,\mathrm{tr}(A_n)\,\mathrm{tr}(B_n) + \sigma_0^4 \,\mathrm{tr}(A_n B_n^s)$.

iii) $\mathrm{E}\big[\big(\epsilon_n' A_n \epsilon_n - \sigma_0^2 \,\mathrm{tr}(A_n) + a_n'\epsilon_n\big)\big(\epsilon_n' B_n \epsilon_n - \sigma_0^2 \,\mathrm{tr}(B_n) + b_n'\epsilon_n\big)\big] = (\mu_4 - 3\sigma_0^4) vec_D'(A_n)\,\mathrm{vec_D}(B_n) + \sigma_0^4 \,\mathrm{tr}(A_n B_n^s) + \mu_3\big(a_n' \,\mathrm{vec_D}(B_n) + b_n' \,\mathrm{vec_D}(A_n)\big) + \sigma_0^2 a_n' b_n$.

[31] See, e.g., Donald et al. (2007, 2010) and the cited references therein.
[32] See Lütkepohl and Burda (1997).

Proof. For i) and ii), see Lin and Lee (2010). Compared to Lin and Lee (2010), the additional terms $\mu_3 \operatorname{vec}_{\mathrm{D}}(A_n)$, $(\mu_4 - 3\sigma_0^4) vec'_D(A_n) \operatorname{vec}_{\mathrm{D}}(B_n)$ and $\sigma_0^4 \operatorname{tr}(A_n) \operatorname{tr}(B_n)$ appear because we do not assume that A_n and B_n have zero diagonals. iii) is a direct result of i) and ii). □

Lemmas 2–6 are elementary, and can be found, for example, in Lee (2004b). Lemma 7 is from Kelejian and Prucha (2001).

Lemma 2. *Suppose that A_n is uniformly bounded in either row or column sum norm, elements of the $n \times k$ matrices X_n are uniformly bounded, and $\lim_{n\to\infty} \frac{1}{n} X_n' X_n$ exists and is nonsingular. Then $\operatorname{tr}(M_{X_n} A_n) = \operatorname{tr}(A_n) + O(1)$, where $M_{X_n} = I_n - X_n(X_n'X_n)^{-1}X_n'$.*

Lemma 3. *Suppose that n-dimensional square matrices $\{A_n\}$ are bounded in either row or column sum norm and ϵ_{ni}'s in $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})'$ are i.i.d. with mean zero, variance σ_0^2 and finite fourth moment. Then, $\mathrm{E}(\epsilon_n' A_n \epsilon_n) = O(n)$, $\operatorname{var}(\epsilon_n' A_n \epsilon_n) = O(n)$, $\epsilon_n' A_n \epsilon_n = O_P(n)$ and $\frac{1}{n}\epsilon_n' A_n \epsilon_n - \frac{1}{n}\mathrm{E}(\epsilon_n' A_n \epsilon_n) = o_P(1)$.*

Lemma 4. *Suppose that A_n is an $n \times n$ matrix with its column sum norm being bounded, elements of the $n \times k$ matrix C_n are uniformly bounded, and elements ϵ_{ni}'s of $\epsilon_n = (\epsilon_{n1}, \ldots, \epsilon_{nn})'$ are i.i.d. $(0, \sigma_0^2)$. Then $\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n = O_P(1)$. Furthermore, if the limit of $\frac{1}{n} C_n' A_n A_n' C_n$ exists and is positive definite, then $\frac{1}{\sqrt{n}} C_n' A_n \epsilon_n \xrightarrow{d} N(0, \sigma_0^2 \lim_{n\to\infty} \frac{1}{n} C_n' A_n A_n' C_n)$.*

Lemma 5. *Suppose that the elements of the sequences of n-dimensional vectors P_n and Q_n are uniformly bounded, and n-dimensional square matrices $\{A_n\}$ are bounded in either row or column sum norm, then $|Q_n' A_n P_n| = O(n)$.*

Lemma 6. *Suppose that $n \times n$ matrices $\{||W_n||\}$ and $\{||S_n^{-1}(\lambda_0)||\}$ are bounded, where $||\cdot||$ is a matrix norm and $S_n(\lambda) = I_n - \lambda W_n$. Then the sequence $\{||S_n^{-1}(\lambda)||\}$ is uniformly bounded in a neighborhood of λ_0.*

Lemma 7. *Suppose that $n \times n$ symmetric matrices $\{A_n = [a_{n,ij}]\}$ are UB, $b_n = (b_{n1}, \ldots, b_{nn})'$ is a vector such that $\sup_n n^{-1} \sum_{i=1}^n |b_{ni}|^{2+\eta_1} < \infty$ for some $\eta_1 > 0$, and ϵ_{ni}'s in $\epsilon_n = (\epsilon_{n1}, \cdots, \epsilon_{nn})'$ are mutually independent, with mean zero, variance σ_{ni}^2 and finite moment of order higher than four such that $\mathrm{E}(|\epsilon_{ni}|^{4+\eta_2})$ for some $\eta_2 > 0$ are uniformly bounded for all n and i. Let $\sigma_{Q_n}^2$ be the variance of Q_n where $Q_n = \epsilon_n' A_n \epsilon_n + b_n' \epsilon_n - \sum_{i=1}^n a_{n,ii}\sigma_{ni}^2$. Assume that $\sigma_{Q_n}^2/n$ is bounded away from zero. Then, $Q_n/\sigma_{Q_n} \xrightarrow{d} N(0,1)$.*

Lemma 8. *Suppose that $n \times n$ matrices $\{M_n\}$ are UB. The smallest eigenvalue of $R_n'(\rho)R_n(\rho)$ is bounded away from zero uniformly over the interval $[-\delta, \delta]$, where $R_n(\rho) = I_n - \rho M_n$. Elements of the $n \times k$ matrix X_n are uniformly bounded. The limit of $\frac{1}{n} X_n' R_n'(\rho) R_n(\rho) X_n$ exists and is nonsingular for any $\rho \in [-\delta, \delta]$. Then elements of $(\frac{1}{n} X_n' R_n'(\rho) R_n(\rho) X_n)^{-1}$ are uniformly bounded in $[-\delta, \delta]$, and $H_n(\rho) = I_n - R_n(\rho) X_n (X_n' R_n'(\rho) R_n(\rho) X_n)^{-1} X_n' R_n'(\rho)$ is UB uniformly in $\rho \in [-\delta, \delta]$.*

Proof. As the smallest eigenvalue of $R_n'(\rho)R_n(\rho)$ is bounded away from zero uniformly on $[-\delta,\delta]$, there exists a constant $\kappa > 0$ such that the smallest eigenvalue of $R_n'(\rho)R_n(\rho)$ is greater or equal to κ for any n and $\rho \in [-\delta,\delta]$. Write $R_n'(\rho)R_n(\rho) = \Gamma_n'(\rho)\Lambda_n(\rho)\Gamma_n(\rho)$, where $\Gamma_n(\rho)$ is an $n\times n$ orthonormal matrix and $\Lambda_n(\rho)$ is a diagonal matrix with the diagonal elements being the eigenvalues of $R_n'(\rho)R_n(\rho)$. Then $R_n'(\rho)R_n(\rho) - \kappa I_n = \Gamma_n'(\rho)[\Lambda_n(\rho) - \kappa I_n]\Gamma_n(\rho)$ is positive semi-definite, which implies that $(\frac{1}{n}\kappa X_n'X_n)^{-1} - (\frac{1}{n}X_n'R_n'(\rho)R_n(\rho)X_n)^{-1}$ is also positive semi-definite. Thus, elements of $(\frac{1}{n}X_n'R_n'(\rho)R_n(\rho)X_n)^{-1}$ and

$$X_n(\frac{1}{n}X_n'R_n'(\rho)R_n(\rho)X_n)^{-1}X_n'$$

are uniformly bounded in $\rho \in [-\delta,\delta]$. It follows that $\frac{1}{n}X_n(\frac{1}{n}X_n'R_n'(\rho)R_n(\rho)X_n)^{-1}X_n'$ is UB uniformly in $\rho \in [-\delta,\delta]$. As $R_n(\rho)$ is UB uniformly in $\rho \in [-\delta,\delta]$, $H_n(\rho)$ is also UB uniformly in $\rho \in [-\delta,\delta]$. □

Lemma 9. *Let W_n, M_n and A_n be $n\times n$ matrices that are UB, b_n be an n-dimensional vector with uniformly bounded elements, X_n be an $n\times k$ matrix with uniformly bounded elements, and $\epsilon_n = (\epsilon_{n1},\ldots,\epsilon_{nn})'$ be a random vector with i.i.d. elements that have mean zero, variance σ_0^2 and finite fourth moment. Assume that $\lim_{n\to\infty}\frac{1}{n}X_n'R_n'(\rho)R_n(\rho)X_n$ exists and is nonsingular for any $\rho \in [-\delta,\delta]$, where $R_n = I_n - \rho M_n$. Let $S_n(\lambda) = I_n - \lambda W_n$, and $T_n(\phi) = G_n'(\phi)H_n(\rho)G_n(\phi)$ with $\phi = (\lambda,\rho)'$, $G_n(\phi) = R_n(\rho)S_n(\lambda)$ and $H_n(\rho) = I_n - R_n(\rho)X_n(X_n'R_n'(\rho)R_n(\rho)X_n)^{-1}X_n'R_n'(\rho)$. Then $\frac{1}{n}b_n'T_n(\phi)A_n\epsilon_n = o_P(1)$ uniformly on the parameter space $\Phi = [-\delta,\delta]\times[-\delta,\delta]$, $\frac{1}{n}[\epsilon_n'A_n'T_n(\phi)A_n\epsilon_n - \sigma_0^2\operatorname{tr}(A_n'T_n(\phi)A_n)] = o_P(1)$ uniformly on Φ, and $\frac{1}{n}\operatorname{tr}[A_n'(G_n'(\phi)G_n(\phi) - T_n(\phi))A_n] = o(1)$ uniformly on Φ.*

Proof. By 21.9 Theorem on p. 337 of Davidson (1994), the uniform convergence of a sequence of stochatic functions $\{f_n(\phi)\}$ on Φ follows from the pointwise convergence in probability $f_n(\phi) = o_P(1)$ for every $\phi \in \Phi$ and the stochastic equicontinuity of $\{f_n(\phi)\}$. For the stochastic equicotinuity, by 21.10 Theorem on p. 339 of Davidson (1994), a sufficient condition is that $|f_n(\phi^*) - f_n(\phi)| \le e_n h(||\phi^* - \phi||)$, for any $\phi^*, \phi \in \Phi$, where $\{e_n\}$ is a stochastically bounded sequence not depending on ϕ, $h(x)$ is nonstochastic which goes down to 0 as x goes down to 0, and $||\cdot||$ denotes the Euclidean vector norm. By Lemma 8, $H_n(\rho)$ is UB uniformly over the parameter space. Then $\frac{1}{n}b_n'T_n(\phi)A_n\epsilon_n = o_P(1)$ for any $\phi = (\lambda,\rho)'$ in Φ and $\frac{1}{n}[\epsilon_n'A_n'T_n(\phi)A_n\epsilon_n - \sigma_0^2\operatorname{tr}(A_n'T_n(\phi)A_n)] = o_P(1)$ for any $\phi \in \Phi$ by Lemma 4, and $\frac{1}{n}\operatorname{tr}[A_n'G_n'(\phi)P_n(\rho)G_n(\phi)A_n] = o(1)$ for any $\phi \in \Phi$ by Lemma 2, where $P_n(\rho) = I_n - H_n(\rho)$. It remains to show the stochastic equicontinuity of the sequences $\{\frac{1}{n}b_n'T_n(\phi)A_n\epsilon_n\}$, $\{\frac{1}{n}[\epsilon_n'A_n'T_n(\phi)A_n\epsilon_n - \sigma_0^2\operatorname{tr}(A_n'T_n(\phi)A_n)]\}$ and $\{\frac{1}{n}\operatorname{tr}[A_nG_n'(\phi)P_n(\rho)G_n(\phi)A_n]\}$.

By the mean value theorem,

$$\frac{1}{n}b_n'T_n(\phi^*)A_n\epsilon_n - \frac{1}{n}b_n'T_n(\phi)A_n\epsilon_n = \frac{1}{n}b_n'\frac{\partial T_n(\tilde{\phi})}{\partial\lambda}A_n\epsilon_n(\lambda^* - \lambda) + \frac{1}{n}b_n'\frac{\partial T_n(\tilde{\phi})}{\partial\rho}A_n\epsilon_n(\rho^* - \rho),$$

where $\frac{\partial T_n(\phi)}{\partial\lambda} = -G_n'(\phi)H_n(\rho)R_n(\rho)W_n - W_n'R_n'(\rho)H_n(\rho)G_n(\phi)$,

$$\frac{\partial T_n(\phi)}{\partial\rho} = -G_n'(\phi)H_n(\rho)M_nS_n(\lambda) - S_n'(\lambda)M_n'H_n(\rho)G_n(\phi) + G_n'(\phi)\frac{\partial H_n(\rho)}{\partial\rho}G_n(\phi)$$

with

$$\frac{\partial H_n(\rho)}{\partial \rho} = M_n X_n[X_n' R_n'(\rho) R_n(\rho) X_n]^{-1} X_n' R_n'(\rho) + R_n(\rho) X_n [X_n' R_n'(\rho) R_n(\rho) X_n]^{-1} X_n' M_n'$$
$$- R_n(\rho) X_n [X_n' R_n'(\rho) R_n(\rho) X_n]^{-1} [X_n' M_n R_n(\rho) X_n + X_n' R_n'(\rho) M_n X_n][X_n' R_n'(\rho) R_n(\rho) X_n]^{-1} X_n' R_n'(\rho),$$

and $\tilde{\phi}$ is between ϕ^* and ϕ. Since $X_n[X_n' R_n'(\rho) R_n(\rho) X_n]^{-1} X_n'$ and $H_n(\rho)$ are UB uniformly in ρ by Lemma 8, $R_n(\rho)$ is linear in ρ and $S_n(\lambda)$ is linear in λ, there exists a finite constant c such that all elements of $|\frac{1}{n} b_n' \frac{\partial T_n(\tilde{\phi})}{\partial \lambda} A_n|$ and $|\frac{1}{n} b_n' \frac{\partial T_n(\tilde{\phi})}{\partial \rho} A_n|$ are bounded by c. Hence,

$$|\frac{1}{n} b_n' T_n(\phi^*) A_n \epsilon_n - \frac{1}{n} b_n' T_n(\phi) A_n \epsilon_n| \leq \frac{2c}{n} \sum_{i=1}^{n} |\epsilon_{ni}| \cdot ||\phi^* - \phi||,$$

where $\frac{1}{n}\sum_{i=1}^{n} |\epsilon_{ni}| = O_P(1)$ by Markov's inequality. Then $\{\frac{1}{n} b_n' T_n(\phi) A_n \epsilon_n\}$ is stochastically equicontinuous.

For $\{\frac{1}{n}[\epsilon_n' A_n' T_n(\phi) A_n \epsilon_n - \sigma_0^2 \operatorname{tr}(A_n' T_n(\phi) A_n)]\}$, by the mean value theorem,

$$\begin{aligned}
&\frac{1}{n}[\epsilon_n' A_n' T_n(\phi^*) A_n \epsilon_n - \sigma_0^2 \operatorname{tr}(A_n' T_n(\phi^*) A_n)] - \frac{1}{n}[\epsilon_n' A_n' T_n(\phi) A_n \epsilon_n - \sigma_0^2 \operatorname{tr}(A_n' T_n(\phi) A_n)] \\
&= \frac{1}{n} \epsilon_n' A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \lambda} A_n \epsilon_n (\lambda^* - \lambda) + \frac{1}{n} \epsilon_n' A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \rho} A_n \epsilon_n (\rho^* - \rho) \\
&\quad - \frac{\sigma_0^2}{n} \operatorname{tr}[A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \lambda} A_n](\lambda^* - \lambda) - \frac{\sigma_0^2}{n} \operatorname{tr}[A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \rho} A_n](\rho^* - \rho) \\
&\leq [\frac{1}{n} |\epsilon_n' A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \lambda} A_n \epsilon_n| + \frac{1}{n} |\epsilon_n' A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \rho} A_n \epsilon_n| \\
&\quad + \frac{\sigma_0^2}{n} |\operatorname{tr}[A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \lambda} A_n]| + \frac{\sigma_0^2}{n} |\operatorname{tr}[A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \rho} A_n]] ||\phi^* - \phi||,
\end{aligned}$$

where $\tilde{\phi}$ lies in between ϕ^* and ϕ. As $A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \lambda} A_n$ is symmetric, by the eigenvalue-eigenvector decomposition, there exists othornormal matrix Γ_n and eigenvalue matrix $\Lambda_n = \operatorname{Diag}\{\lambda_{n1}, \cdots, \lambda_{nn}\}$ such that

$$\begin{aligned}
\frac{1}{n} |\epsilon_n' A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \lambda} A_n \epsilon_n| &= \frac{1}{n} |\epsilon_n' \Gamma_n \Lambda_n \Gamma_n' \epsilon_n| \leq \frac{1}{n} \max_{i=1,\cdots,n} |\lambda_{ni}| \cdot \epsilon_n' \epsilon_n \\
&\leq \frac{1}{n} ||A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \lambda} A_n||_\infty \cdot \epsilon_n' \epsilon_n \leq \frac{c_1}{n} \epsilon_n' \epsilon_n = O_P(1),
\end{aligned}$$

by the spectral radius theorem, for some constant c_1, because $A_n' \frac{\partial T_n(\phi)}{\partial \lambda} A_n$ is UB uniformly in $\phi \in \Phi$. Similarly, $\frac{1}{n} |\epsilon_n' A_n' \frac{\partial T_n(\tilde{\phi})}{\partial \rho} A_n \epsilon_n| \leq \frac{c_1}{n} \epsilon_n' \epsilon_n = O_P(1)$. Furthermore, $\frac{1}{n} \operatorname{tr}[A_n' \frac{\partial T_n(\phi)}{\partial \lambda} A_n]$ and $\frac{1}{n} \operatorname{tr}[A_n' \frac{\partial T_n(\phi)}{\partial \rho} A_n]$ are bounded uniformly on Φ. Then $\{\frac{1}{n}[\epsilon_n' A_n' T_n(\phi) A_n \epsilon_n - \operatorname{tr}(A_n' T_n(\phi) A_n \Sigma_n)]\}$ is stochastically equicontinuous.

For $\{\frac{1}{n} \operatorname{tr}(A_n' G_n'(\phi) P_n(\rho) G_n(\phi) A_n)\}$, its derivative is $\frac{1}{n} \frac{\partial}{\partial \phi} \operatorname{tr}(A_n' G_n'(\phi) G_n(\phi) A_n) - \frac{1}{n} \frac{\partial}{\partial \phi} \operatorname{tr}(A_n' T_n(\phi) A_n)$, which is bounded by a constant not depending on ϕ in absolute value. Then by the mean value theorem, $\frac{1}{n} \operatorname{tr}(A_n' G_n'(\phi) P_n(\rho) G_n(\phi) A_n)$ is equicontinuous.

The results in the lemma follow from the pointwise convergence and stochastic equicontinuity. □

The following lemmas are for the consistency of the bootstrap for Cox-type tests. Let $\hat{\epsilon}_{1n}^*$ be the residual vector from the QML estimation of the the model (1) with the bootstrapped data y_n^*, E^* be the expectation induced by the bootstrap sampling process and $||\cdot||$ be the Euclidean matrix norm.

Lemma 10. *For any integer* r*, if* $\mathrm{E}\,|\epsilon_{1n,i}|^r < \infty$, $\mathrm{E}^*\,\epsilon^{*r}_{1n,i} = \mathrm{E}\,\epsilon^r_{1n,i} + o_P(1)$, $n^{-1}\sum_{i=1}^n \hat{\epsilon}^r_{1n,i} = \mathrm{E}\,\epsilon^r_{1n,i} + o_P(1)$, $\mathrm{E}^*\,|\epsilon_{1n,i}|^{*r} = \mathrm{E}\,|\epsilon_{1n,i}|^r + o_P(1)$ *and* $n^{-1}\sum_{i=1}^n |\hat{\epsilon}_{1n,i}|^r = \mathrm{E}\,|\epsilon_{1n,i}|^r + o_P(1)$*. If* $\mathrm{E}\,\epsilon^{2r}_{1n,i} < \infty$, $n^{1/2}[\mathrm{E}^*\,\epsilon^{*r}_{1n,i} - \mathrm{E}\,\epsilon^r_{1n,i}] = O_P(1)$ *and* $n^{1/2}[n^{-1}\sum_{i=1}^n \hat{\epsilon}^r_{1n,i} - \mathrm{E}\,\epsilon^r_{1n,i}] = O_P(1)$.

Proof. This is Lemma 5 in Jin and Lee (2012). □

Lemma 11. *For* $\eta > 0$ *and an integer* r, $\mathrm{P}^*(|n^{-1}\sum_{i=1}^n \hat{\epsilon}^{*r}_{1n,i} - \mathrm{E}^*\,\epsilon^{*r}_{1n,i}| > \eta) = o_P(1)$ *if* $\mathrm{E}\,|\epsilon_{ni}|^r < \infty$.

Proof. This is Lemma 7 in Jin and Lee (2012). □

Lemma 12. $\frac{1}{\sqrt{n}}\|\frac{\partial^2 L_{1n}(\tilde{\theta}_{1n})}{\partial\theta_1\partial\theta_1'} - \frac{\partial^2 \bar{L}_{1n}(\theta_{10};\theta_{10})}{\partial\theta_1\partial\theta_1'}\| = O_P(1)$, $\frac{1}{\sqrt{n}}\|\frac{\partial^2 L_{2n}(\tilde{\theta}_{2n})}{\partial\theta_2\partial\theta_2'} - \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\theta_2\partial\theta_2'}\| = o_P(1)$, $\frac{1}{n}\|\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}_{2n};\tilde{\theta}_{1n})}{\partial\theta_1\partial\theta_1'} - \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\theta_1\partial\theta_1'}\| = o_P(1)$, $\frac{1}{n}\|\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}_{2n};\tilde{\theta}_{1n})}{\partial\theta_1\partial\theta_2'} - \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\theta_1\partial\theta_2'}\| = o_P(1)$ *and* $\frac{1}{n}\|\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}_{2n};\tilde{\theta}_{1n})}{\partial\theta_2\partial\theta_2'} - \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\theta_2\partial\theta_2'}\| = o_P(1)$*, where* $\tilde{\theta}_{1n}$ *is between* $\hat{\theta}_{1n}$ *and* θ_{10}*, and* $\tilde{\theta}_{2n}$ *is between* $\hat{\theta}_{2n}$ *and* $\bar{\theta}_{2n,1}$.

Proof. We prove the first result by showing that (i) $n^{-1/2}\|\frac{\partial^2 L_{1n}(\tilde{\theta}_{1n})}{\partial\theta_1\partial\theta_1'} - \frac{\partial^2 L_{1n}(\theta_{10})}{\partial\theta_1\partial\theta_1'}\| = O_P(1)$ and (ii) $n^{-1/2}\|\frac{\partial^2 L_{1n}(\theta_{10})}{\partial\theta_1\partial\theta_1'} - \mathrm{E}\,\frac{\partial^2 L_{1n}(\theta_{10})}{\partial\theta_1\partial\theta_1'}\| = O_P(1)$. To prove (i), apply the mean value theorem to each term in the second order derivative. Specifically, we investigate $n^{-1/2}\|\frac{\partial^2 L_{1n}(\tilde{\theta}_{1n})}{\partial\lambda_1^2} - \mathrm{E}\,\frac{\partial^2 L_{1n}(\theta_{10})}{\partial\lambda_1^2}\|$. Results for other terms can be derived similarly. By the mean value theorem,

$$\frac{1}{\sqrt{n}}(\frac{\partial^2 L_{1n}(\tilde{\theta}_{1n})}{\partial\lambda_1^2} - \frac{\partial^2 L_{1n}(\theta_{10})}{\partial\lambda_1^2}) = B_{1n} + \frac{2}{\breve{\sigma}^2_{1n}}B_{2n}\sqrt{n}(\tilde{\rho}_{1n} - \breve{\rho}_{1n}) + B_{3n},$$

where $B_{1n} = -2n^{-1}\,\mathrm{tr}\big[\big(W_{1n}S^{-1}_{1n}(\check{\lambda}_{1n})\big)^3\big]n^{1/2}(\tilde{\lambda}_{1n} - \hat{\lambda}_{1n})$, $B_{2n} = n^{-1}y_n'W_{1n}'M_{1n}'R_{1n}(\breve{\rho}_{1n})W_{1n}y_n$ and $B_{3n} = (n\breve{\sigma}^4_{1n})^{-1}y_n'W_{1n}'R_{1n}'(\breve{\rho}_{1n})R_{1n}(\breve{\rho}_{1n})W_{1n}y_n n^{1/2}(\tilde{\sigma}^2_{1n} - \hat{\sigma}^2_{1n})$ with $\check{\theta}_{1n}$ being between $\tilde{\theta}_{1n}$ and θ_{10}. By the uniform boundedness of $S^{-1}_{1n}(\lambda_1)$ in $\lambda_1 \in \Lambda_1$, $B_{1n} = O_P(1)$. Note that $B_{2n} = B_{2n,1} + B_{2n,2}(\rho_{10} - \breve{\rho}_{1n})$, where $B_{2n,1} = n^{-1}y_{1n}'W_{1n}'M_{1n}'R_{1n}W_{1n}y_n = O_P(1)$ and $B_{2n,2} = n^{-1}y_{1n}'W_{1n}'M_{1n}'M_{1n}W_{1n}y_n = O_P(1)$, then $2\breve{\sigma}^{-2}_{1n}B_{2n}n^{1/2}(\tilde{\rho}_{1n} - \breve{\rho}_{1n}) = O_P(1)$. Similarly, $B_{3n} = O_P(1)$. Hence (i) holds. (ii) follows from Chebyshev's inequality.

The proof of the second result resembles the above proof and the rest results are proved by a similar use of the mean value theorem. □

Lemma 13. *For* $\eta > 0$, $\mathrm{P}^*(||\hat{\theta}^*_{1n} - \hat{\theta}_{1n}|| > \eta) = o_P(1)$, $\mathrm{P}^*(||\hat{\theta}^*_{2n} - \hat{\theta}_{2n}|| > \eta) = o_P(1)$ *and* $\mathrm{P}^*(||\bar{\theta}_{2n}(\hat{\theta}^*_{1n}) - \bar{\theta}_{2n}(\hat{\theta}_{1n})|| > \eta) = o_P(1)$.

Proof. We first prove the result on $\hat{\theta}^*_{1n}$. Let $\bar{L}_{1n}(\phi_1;\theta_{10}) = \max_{\beta_1,\sigma_1^2}\bar{L}_{1n}(\theta_1;\theta_{10})$, $L^*_{1n}(\theta_1)$ be the log likelihood function of the the model (1) with the dependent variable y^*_n, and $\bar{L}_{1n}(\phi_1;\hat{\theta}_{1n,a}) = \max_{\beta_1,\sigma_1^2}\mathrm{E}^*\,L^*_{1n}(\theta_1)$, where $\hat{\theta}_{1n,a} = (\hat{\lambda}_{1n}, \hat{\rho}_{1n}, \hat{\beta}_{1n}', \mathrm{E}^*\,\epsilon^{*2}_{1n,i})'$, then

$$\bar{L}_{1n}(\phi_1;\theta_{10}) = -\frac{n}{2}[\ln(2\pi) + 1] - \frac{n}{2}\ln\bar{\sigma}^2_{1n}(\phi_1) + \ln|S_{1n}(\lambda_1)| + \ln|R_{1n}(\rho_1)|,$$
$$\bar{L}_{1n}(\phi_1;\hat{\theta}_{1n,a}) = -\frac{n}{2}[\ln(2\pi) + 1] - \frac{n}{2}\ln\bar{\sigma}^{*2}_{1n}(\phi_1) + \ln|S_{1n}(\lambda_1)| + \ln|R_{1n}(\rho_1)|,$$

where

$$\bar{\sigma}_{1n}^2(\phi_1) = \frac{1}{n}\sigma_{10}^2 \operatorname{tr}\big(R_{1n}'^{-1}S_{1n}'^{-1}S_{1n}'(\lambda_1)R_{1n}'(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S_{1n}^{-1}R_{1n}^{-1}\big)$$
$$+ \frac{1}{n}(X_{1n}\beta_{10})'S_{1n}'^{-1}S_{1n}'(\lambda_1)R_{1n}'(\rho_1)H_{1n}(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S_{1n}^{-1}X_{1n}\beta_{10},$$
$$\bar{\sigma}_{1n}^{*2}(\phi_1) = \frac{1}{n}(\mathrm{E}^*\,\epsilon_{1n,i}^{*2}) \operatorname{tr}\big(R_{1n}'^{-1}(\hat{\rho}_{1n})S_{1n}'^{-1}(\hat{\rho}_{1n})S_{1n}'(\lambda_1)R_{1n}'(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S_{1n}^{-1}(\hat{\lambda}_{1n})R_{1n}^{-1}(\hat{\rho}_{1n})\big)$$
$$+ \frac{1}{n}(X_{1n}\hat{\beta}_{1n})'S_{1n}'^{-1}(\hat{\lambda}_{1n})S_{1n}'(\lambda_1)R_{1n}'(\rho_1)H_{1n}(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S_{1n}^{-1}(\hat{\lambda}_{1n})X_{1n}\hat{\beta}_{1n},$$

with $H_{1n}(\rho_1) = I_n - R_{1n}(\rho_1)X_{1n}[X_{1n}'R_{1n}'(\rho_1)R_{1n}(\rho_1)X_{1n}]^{-1}X_{1n}'R_{1n}'(\rho_1)$ being UB uniformly on ϱ_1 (see the proof of Proposition 3). By the mean value theorem,

$$\frac{1}{n}[\bar{L}_{1n}(\phi_1;\hat{\theta}_{1n,a}) - \bar{L}_{1n}(\phi_1;\theta_{10})] = -\frac{1}{2}\frac{\bar{\sigma}_{1n}^{*2}(\phi_1) - \bar{\sigma}_{1n}^2(\phi_1)}{\tilde{\sigma}_{1n}^2},$$

where $\tilde{\sigma}_{1n}^2$ is between $\bar{\sigma}_{1n}^2(\phi_1)$ and $\bar{\sigma}_{1n}^{*2}(\phi_1)$, and

$$\begin{aligned}
&\bar{\sigma}_{1n}^{*2}(\phi_1) - \bar{\sigma}_{1n}^2(\phi_1)\\
&= \frac{1}{n}(\mathrm{E}^*\,\epsilon_{1n,i}^{*2} - \sigma_0^2)\operatorname{tr}\big(R_{1n}'^{-1}(\check{\rho}_{1n})S_{1n}'^{-1}(\check{\rho}_{1n})S_{1n}'(\lambda_1)R_{1n}'(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S_{1n}^{-1}(\check{\lambda}_{1n})R_{1n}^{-1}(\check{\rho}_{1n})\big)\\
&\quad + \frac{2}{n}(X_{1n}\check{\beta}_{1n})'S_{1n}'^{-1}(\check{\lambda}_{1n})S_{1n}'(\lambda_1)R_{1n}'(\rho_1)H_{1n}(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S_{1n}^{-1}(\check{\lambda}_{1n})X_{1n}(\hat{\beta}_{1n} - \beta_{10})\\
&\quad + \frac{2\check{\sigma}_{1n}^2}{n}\operatorname{tr}\big(R_{1n}'^{-1}(\check{\rho}_{1n})S_{1n}'^{-1}(\check{\rho}_{1n})S_{1n}'(\lambda_1)R_{1n}'(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S_{1n}^{-1}(\check{\lambda}_{1n})R_{1n}^{-1}(\check{\rho}_{1n})M_{1n}R_{1n}^{-1}(\check{\rho}_{1n})\big)(\hat{\rho}_{1n} - \rho_{10})\\
&\quad + \frac{2\check{\sigma}_{1n}^2}{n}\operatorname{tr}\big(R_{1n}'^{-1}(\check{\rho}_{1n})S_{1n}'^{-1}(\check{\rho}_{1n})S_{1n}'(\lambda_1)R_{1n}'(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S_{1n}^{-1}(\check{\lambda}_{1n})W_{1n}S_{1n}^{-1}(\check{\lambda}_{1n})R_{1n}^{-1}(\check{\rho}_{1n})\big)(\hat{\lambda}_{1n} - \lambda_{10})\\
&\quad + \frac{2}{n}(X_{1n}\check{\beta}_{1n})'S_{1n}'^{-1}(\check{\lambda}_{1n})S_{1n}'(\lambda_1)R_{1n}'(\rho_1)H_{1n}(\rho_1)R_{1n}(\rho_1)S_{1n}(\lambda_1)S_{1n}^{-1}(\check{\lambda}_{1n})W_{1n}S_{1n}^{-1}(\check{\lambda}_{1n})X_{1n}\check{\beta}_{1n}(\hat{\lambda}_{1n} - \lambda_{10}),
\end{aligned}$$

with $\check{\gamma}_{1n} = (\check{\lambda}_{1n}, \check{\rho}_{1n}, \check{\beta}_{1n})'$ being between γ_{10} and $\hat{\gamma}_{1n}$, and $\check{\sigma}_{1n}^2$ being between σ_{10}^2 and $\mathrm{E}^*\,\epsilon_{1n,i}^{*2}$. By Lemma 10, $\sup_{\phi_1\in\varphi_1}|\bar{\sigma}_{1n}^{*2}(\phi_1) - \bar{\sigma}_{1n}^2(\phi_1)| = o_P(1)$. As $\bar{\sigma}_{1n}^2(\phi_1)$ is bounded away from zero uniformly on Φ_1 (see the proof of Proposition 3 for a similar result on $\bar{\sigma}_{2n}^2(\phi_2;\theta_{10})$), $\sup_{\phi_1\in\varphi_1}|n^{-1}[\bar{L}_{1n}(\phi_1;\hat{\theta}_{1n,a}) - \bar{L}_{1n}(\phi_1;\theta_{10})]| = o_P(1)$.

If $||\phi_1 - \hat{\phi}_{1n}|| > \eta$, $||\phi_1 - \phi_{10}|| \geq ||\phi_1 - \hat{\phi}_{1n}|| - ||\hat{\phi}_{1n} - \phi_{10}|| > \eta/2$ with probability $1 - o(1)$. Note that

$$\frac{1}{n}\big(\bar{L}_{1n}(\hat{\phi}_{1n};\hat{\theta}_{1n,a}) - \bar{L}_{1n}(\phi_1;\hat{\theta}_{1n,a})\big) = \frac{1}{n}\big(\bar{L}_{1n}(\hat{\phi}_{1n};\hat{\theta}_{1n,a}) - \bar{L}_{1n}(\hat{\phi}_{1n};\theta_{10})\big) - \frac{1}{n}\big(\bar{L}_{1n}(\phi_1;\hat{\theta}_{1n,a}) - \bar{L}_{1n}(\phi_1;\theta_{10})\big)$$
$$+ \frac{1}{n}\big(\bar{L}_{1n}(\phi_{10};\theta_{10}) - \bar{L}_{1n}(\phi_1;\theta_{10})\big) - \frac{1}{n}\big(\bar{L}_{1n}(\phi_{10};\theta_{10}) - \bar{L}_{1n}(\hat{\phi}_{1n};\theta_{10})\big),$$

given $\eta > 0$, there exists a $\kappa > 0$, such that $||\phi_1 - \hat{\phi}_{1n}|| > \eta$ implies that $n^{-1}\big(\bar{L}_{1n}(\hat{\phi}_{1n};\hat{\theta}_{1n,a}) - \bar{L}_{1n}(\phi_1;\hat{\theta}_{1n,a})\big) \geq \kappa$ with probability $1 - o(1)$. Then

$$\begin{aligned}
&\mathrm{P}^*(||\hat{\phi}_{1n}^* - \hat{\phi}_{1n}|| > \eta)\\
&\leq \mathrm{P}^*\big(n^{-1}\big(\bar{L}_{1n}(\hat{\phi}_{1n};\hat{\theta}_{1n,a}) - \bar{L}_{1n}(\hat{\phi}_{1n}^*;\hat{\theta}_{1n,a})\big) \geq \kappa\big) + o(1)\\
&\leq \mathrm{P}^*\big(n^{-1}\big(\bar{L}_{1n}(\hat{\phi}_{1n};\hat{\theta}_{1n,a}) - L_{1n}^*(\hat{\phi}_{1n}) + L_{1n}^*(\hat{\phi}_{1n}^*) - \bar{L}_{1n}(\hat{\phi}_{1n}^*;\hat{\theta}_{1n,a})\big) \geq \kappa\big) + o(1)\\
&\leq \mathrm{P}^*\big(2n^{-1}\sup_{\phi_1\in\varphi_1}\big|L_{1n}^*(\phi_1) - \bar{L}_{1n}(\phi_1;\hat{\theta}_{1n,a})\big| \geq \kappa\big) + o(1),
\end{aligned}$$

where φ_1 is the parameter space of ϕ_1, $L^*_{1n}(\phi_1) = \max_{\beta_1,\sigma_1^2} L^*_{1n}(\theta_1)$, and

$$\frac{1}{n}\big(L^*_{1n}(\phi_1) - \bar{L}_{1n}(\phi_1;\hat{\theta}_{1n,a})\big) = -\frac{\hat{\sigma}^{*2}_{1n}(\phi_1) - \bar{\sigma}^{*2}_{1n}(\phi_1)}{2\check{\sigma}^2_{1n}(\phi_1)},$$

with $\check{\sigma}^{*2}_{1n}(\phi_1)$ being between $\hat{\sigma}^{*2}_{1n}(\phi_1)$ and $\bar{\sigma}^{*2}_{1n}(\phi_1)$, and

$$\begin{aligned}\hat{\sigma}^{*2}_{1n}(\phi_1) - \bar{\sigma}^{*2}_{1n}(\phi_1) = {}& \frac{1}{n}\epsilon^{*\prime}_{1n} R'^{-1}_{1n}(\hat{\rho}_{1n}) S'^{-1}_{1n}(\hat{\lambda}_{1n}) S'_{1n}(\lambda_1) R'_{1n}(\rho_1) H_{1n}(\rho_1) R_{1n}(\rho_1) S_{1n}(\lambda_1) S^{-1}_{1n}(\hat{\lambda}_{1n}) R^{-1}_{1n}(\hat{\rho}_{1n})\epsilon^*_{1n} \\ & - \frac{\mathrm{E}^*\,\epsilon^{*2}_{1n,i}}{n}\operatorname{tr}[R'^{-1}_{1n}(\hat{\rho}_{1n}) S'^{-1}_{1n}(\hat{\lambda}_{1n}) S'_{1n}(\lambda_1) R'_{1n}(\rho_1) R_{1n}(\rho_1) S_{1n}(\lambda_1) S^{-1}_{1n}(\hat{\lambda}_{1n}) R^{-1}_{1n}(\hat{\rho}_{1n})] \\ & + \frac{2}{n}(X_{1n}\hat{\beta}_{1n})' S'^{-1}_{1n}(\hat{\lambda}_{1n}) S'_{1n}(\lambda_1) R'_{1n}(\rho_1) H_{1n}(\rho_1) R_{1n}(\rho_1) S_{1n}(\lambda_1) S^{-1}_{1n}(\hat{\lambda}_{1n}) R^{-1}_{1n}(\hat{\rho}_{1n})\epsilon^*_{1n}.\end{aligned}$$

The $\hat{\sigma}^{*2}_{1n}(\phi_1) - \bar{\sigma}^{*2}_{1n}(\phi_1)$ is equal to a LQ form plus $n^{-1}(\mathrm{E}^*\,\epsilon^{*2}_{1n,i})\operatorname{tr}[R'^{-1}_{1n}(\hat{\rho}_{1n}) S'^{-1}_{1n}(\hat{\lambda}_{1n}) S'_{1n}(\lambda_1) R'_{1n}(\rho_1)(H_{1n}(\rho_1) - I_n) R_{1n}(\rho_1) S_{1n}(\lambda_1) S^{-1}_{1n}(\hat{\lambda}_{1n}) R^{-1}_{1n}(\hat{\rho}_{1n})]$. Since $R_{1n}(\rho_1)$ is linear in ρ_1, $S_{1n}(\lambda_1)$ is linear in λ_1 and $H_{1n}(\rho_1)$ is UB uniformly in $\rho_1 \in \varrho_1$, Chebyshev's inequality implies that $n\,\mathrm{P}^*(\sup_{\phi_1\in\varphi_1}|\hat{\sigma}^{*2}_{1n}(\phi_1) - \bar{\sigma}^{*2}_{1n}(\phi_1)| > \eta)$ for $\eta > 0$ is bounded by a term depending only on $\hat{\beta}_{1n}$, $\mathrm{E}\,\epsilon^{*2}_{1n,i}$, $\mathrm{E}\,\epsilon^{*3}_{1n,i}$ and $\mathrm{E}\,\epsilon^{*4}_{1n,i}$, which has the order $O_P(1)$ by Lemma 10. Then $\mathrm{P}^*(\sup_{\phi_1\in\varphi_1}|\hat{\sigma}^{*2}_{1n}(\phi_1) - \bar{\sigma}^{*2}_{1n}(\phi_1)| > \eta) = o_P(1)$. It has been shown above that $\sup_{\phi_1\in\varphi_1}|\bar{\sigma}^{*2}_{1n}(\phi_1) - \bar{\sigma}^2_{1n}(\phi_1)| = o_P(1)$ with $\bar{\sigma}^2_{1n}(\phi_1)$ being bounded away from zero uniformly on Φ_1, then $\mathrm{P}^*(||\hat{\phi}^*_{1n} - \hat{\phi}_{1n}|| > \eta) = o_P(1)$. Now the mean value theorem and the formulas of $\hat{\beta}^*_{1n}$ and $\hat{\sigma}^{*2}_{1n}$ as functions of $\hat{\phi}^*_{1n}$ can be used to show that we also have $\mathrm{P}^*(||\hat{\beta}^*_{1n} - \hat{\beta}_{1n}|| > \eta) = o_P(1)$ and $\mathrm{P}^*(||\hat{\sigma}^{*2}_{1n} - \hat{\sigma}^2_{1n}|| > \eta) = o_P(1)$.

The result on $\hat{\theta}^*_{2n}$ can be similarly proved. For the result on $\bar{\theta}_{2n}(\hat{\theta}^*_{1n})$, some modifications are needed. First, by the mean value theorem, we can show that $\mathrm{P}^*(\sup_{\phi_2\in\varphi_2} n^{-1}|\bar{L}_{2n}(\phi_2;\hat{\theta}^*_{1n}) - \bar{L}_{2n}(\phi_2;\hat{\theta}_{1n})| > \eta) = o_P(1)$ for $\eta > 0$, where $\bar{L}_{2n}(\phi_2;\theta_1) = \max_{\beta_2,\sigma_2^2}\bar{L}_{2n}(\theta_2;\theta_1)$. Given $\eta > 0$, there exists a $\kappa > 0$, such that $||\phi_2 - \bar{\phi}_{2n}(\hat{\theta}_{1n})|| > \eta$ implies that $n^{-1}\big(\bar{L}_{2n}(\bar{\phi}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}) - \bar{L}_{2n}(\phi_2;\hat{\theta}_{1n})\big) \geq \kappa$ with probability $1 - o(1)$, where $\bar{\phi}_{2n}(\hat{\theta}_{1n}) = \max_{\beta_2,\sigma_2^2}\bar{L}_{2n}(\theta_2;\hat{\theta}_{1n})$. Then

$$\begin{aligned}& \mathrm{P}^*(||\bar{\phi}_{2n}(\hat{\theta}^*_{1n}) - \bar{\phi}_{2n}(\hat{\theta}_{1n})|| > \eta) \\ & \leq \mathrm{P}^*\big(n^{-1}\big(\bar{L}_{2n}(\bar{\phi}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}) - \bar{L}_{2n}(\bar{\phi}_{2n}(\hat{\theta}^*_{1n});\hat{\theta}_{1n})\big) \geq \kappa\big) + o(1) \\ & \leq \mathrm{P}^*\big(n^{-1}\big(\bar{L}_{2n}(\bar{\phi}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n}) - \bar{L}_{2n}(\bar{\phi}_{2n}(\hat{\theta}_{1n});\hat{\theta}^*_{1n}) + \bar{L}_{2n}(\bar{\phi}_{2n}(\hat{\theta}^*_{1n});\hat{\theta}^*_{1n}) - \bar{L}_{2n}(\bar{\phi}_{2n}(\hat{\theta}^*_{1n});\hat{\theta}_{1n})\big) \geq \kappa\big) + o(1) \\ & \leq \mathrm{P}^*\big(\sup_{\phi_2\in\varphi_2} n^{-1}|\bar{L}_{2n}(\phi_2;\hat{\theta}_{1n}) - \bar{L}_{2n}(\phi_2;\hat{\theta}^*_{1n})| > \kappa\big) + o(1) = o_P(1).\end{aligned}$$

The rest proof is similar to that for $\hat{\theta}^*_{1n}$. □

Lemma 14. *For $\eta > 0$, $\mathrm{P}^*\big(n^{-1}\|\frac{\partial^2 L^*_{1n}(\tilde{\theta}^*_{1n})}{\partial\theta_1\partial\theta_1'} - \mathrm{E}^*\,\frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\theta_1\partial\theta_1'}\| > \eta\big) = o_P(1)$, $\mathrm{P}^*\big(n^{-1}\|\frac{\partial^2 L^*_{2n}(\tilde{\theta}^*_{2n})}{\partial\theta_2\partial\theta_2'} - \mathrm{E}^*\,\frac{\partial^2 L^*_{2n}(\hat{\theta}_{2n})}{\partial\theta_2\partial\theta_2'}\| > \eta\big) = o_P(1)$, $\mathrm{P}^*\big(n^{-1}\|\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}^*_{2n};\tilde{\theta}_{1n})}{\partial\theta_1\partial\theta_1'} - \frac{\partial^2 \bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n})}{\partial\theta_1\partial\theta_1'}\| > \eta\big) = o_P(1)$, $\mathrm{P}^*\big(n^{-1}\|\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}^*_{2n};\tilde{\theta}_{1n})}{\partial\theta_1\partial\theta_2'} - \frac{\partial^2 \bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n})}{\partial\theta_1\partial\theta_2'}\| > \eta\big) = o_P(1)$ and $\mathrm{P}^*\big(n^{-1}\|\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}^*_{2n};\tilde{\theta}_{1n})}{\partial\theta_2\partial\theta_2'} - \frac{\partial^2 \bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n})}{\partial\theta_2\partial\theta_2'}\| > \eta\big) = o_P(1)$, where $\tilde{\theta}_{1n}$ is between $\hat{\theta}_{1n}$ and θ_{10}, $\tilde{\theta}^*_{1n}$ is between $\hat{\theta}^*_{1n}$ and $\hat{\theta}_{1n}$, and $\tilde{\theta}^*_{2n}$ is between $\hat{\theta}^*_{2n}$ and $\hat{\theta}_{2n}$.*

Proof. We prove the first result in the lemma by showing that (i) $\mathrm{P}^*\big(n^{-1}\|\frac{\partial^2 L^*_{1n}(\tilde{\theta}^*_{1n})}{\partial\theta_1\partial\theta_1'} - \frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\theta_1\partial\theta_1'}\| > \eta\big) = o_P(1)$ and (ii) $\mathrm{P}^*\big(n^{-1}\|\frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\theta_1\partial\theta_1'} - \frac{\partial^2\,\mathrm{E}^*\,L^*_{1n}(\hat{\theta}_{1n})}{\partial\theta_1\partial\theta_1'}\| > \eta\big) = o_P(1)$. As in the proof of Lemma 12, use the mean value theorem for each term in the second order derivative to prove (i). Here we only investigate $n^{-1}\left|\frac{\partial^2 L^*_{1n}(\tilde{\theta}_{1n})}{\partial\lambda_1^2} - \frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\lambda_1^2}\right|$. Results for other terms are similarly derived. By the mean value theorem,

$$\frac{1}{n}\Big(\frac{\partial^2 L^*_{1n}(\tilde{\theta}^*_{1n})}{\partial\lambda_1^2} - \frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\lambda_1^2}\Big) = B^*_{1n} + \frac{2}{\check{\sigma}^2_{1n}} B^*_{2n}(\tilde{\rho}^*_{1n} - \hat{\rho}_{1n}) + B^*_{3n},$$

where $B^*_{1n} = -2n^{-1}\operatorname{tr}\big((W_{1n}S^{-1}_{1n}(\check{\lambda}^*_{1n}))^3\big)(\tilde{\lambda}^*_{1n} - \hat{\lambda}_{1n})$, $B^*_{2n} = n^{-1}y^{*'}_n W'_{1n}M'_{1n}R_{1n}(\check{\rho}^*_{1n})W_{1n}y^*_n$ and $B^*_{3n} = (n\check{\sigma}^{*4}_{1n})^{-1}y^{*'}_n W'_{1n}R'_{1n}(\check{\rho}^*_{1n})R_{1n}(\check{\rho}^*_{1n})W_{1n}y^*_n(\tilde{\sigma}^{*2}_{1n} - \hat{\sigma}^2_{1n})$ with $\check{\theta}^*_{1n}$ being between $\tilde{\theta}^*_{1n}$ and $\hat{\theta}_{1n}$. By Lemma 13 and the uniform boundedness of $S^{-1}_{1n}(\lambda_1)$, $\mathrm{P}^*(|B^*_{1n}| > \eta) = o_P(1)$. Let $B^*_{2n,1} = n^{-1}y^{*'}_n W'_{1n}M'_{1n}R_{1n}(\hat{\rho}_{1n})W_{1n}y^*_n$ and $B^*_{2n,2} = n^{-1}y^{*'}_n W'_{1n}M'_{1n}M_{1n}W_{1n}y^*_n$. Then $\mathrm{P}^*(|B^*_{1n,1} - \mathrm{E}^*\, B^*_{1n,1}| > \eta) = o_P(1)$ and $\mathrm{P}^*(|B^*_{2n,1} - \mathrm{E}^*\, B^*_{2n,1}| > \eta) = o_P(1)$. Since $B^*_{2n} = B^*_{2n,1} + B^*_{2n,2}(\hat{\rho}_{1n} - \check{\rho}^*_{1n})$, $\mathrm{P}^*\big(\big|2\check{\sigma}^{*-2}_{1n}B^*_{2n}(\tilde{\rho}^*_{1n} - \hat{\rho}_{1n})\big| > \eta\big) = o_P(1)$. Similarly, $\mathrm{P}^*(|B^*_{3n}| > \eta) = o_P(1)$. Therefore, $\mathrm{P}^*\big(n^{-1}\big|\frac{\partial^2 L^*_{1n}(\tilde{\theta}^*_{1n})}{\partial\lambda_1^2} - \frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\lambda_1^2}\big| > \eta\big) = o_P(1)$. (ii) is proved by using Chebyshev's inequality.

The proof of the second result is almost the same. The rest of results are proved by using the mean value theorem. □

Lemma 15. *$n^{-1}\big\|\mathrm{E}^*\, \frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\theta_1\partial\theta'_1} - \mathrm{E}\, \frac{\partial^2 L_{1n}(\theta_{10})}{\partial\theta_1\partial\theta'_1}\big\| = o_P(1)$, $n^{-1}\big\|\mathrm{E}^*\, \frac{\partial^2 L^*_{2n}(\hat{\theta}_{2n})}{\partial\theta_2\partial\theta'_2} - \mathrm{E}\, \frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta'_2}\big\| = o_P(1)$ and $n^{-1}\big\|\frac{\partial\bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n})}{\partial\theta_1} - \frac{\partial\bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\theta_1}\big\| = o_P(1)$.*

Proof. The lemma is proved by using the mean value theorem and Lemma 10. □

Lemma 16. *For $\eta > 0$ and $0 \le a < \frac{1}{2}$, $\mathrm{P}^*(n^a||\hat{\theta}^*_{1n} - \hat{\theta}_{1n}|| > \eta) = o_P(1)$, $\mathrm{P}^*(n^a||\hat{\theta}^*_{2n} - \hat{\theta}_{2n}|| > \eta) = o_P(1)$ and $\mathrm{P}^*(n^a||\bar{\theta}_{2n}(\hat{\theta}^*_{1n}) - \bar{\theta}_{2n}(\hat{\theta}_{1n})|| > \eta) = o_P(1)$.*

Proof. We only prove the result on $\hat{\theta}^*_{1n}$, as the proofs for the rest of results are similar. By the mean value theorem,

$$n^a(\hat{\theta}^*_{1n} - \hat{\theta}_{1n}) = \Big(-\frac{1}{n}\frac{\partial^2 L^*_{1n}(\tilde{\theta}^*_{1n})}{\partial\theta_1\partial\theta'_1}\Big)^{-1} n^{a-1}\frac{\partial L^*_{1n}(\hat{\theta}_{1n})}{\partial\theta_1},$$

where $\tilde{\theta}^*_{1n}$ is between $\hat{\theta}^*_{1n}$ and $\hat{\theta}_{1n}$. Then

$$\mathrm{P}^*(n^a||\hat{\theta}^*_{1n} - \hat{\theta}_{1n}|| > \eta) \le \mathrm{P}^*\Big(\Big\|\frac{1}{n}\frac{\partial^2 L^*_{1n}(\tilde{\theta}^*_{1n})}{\partial\theta_1\partial\theta'_1} - \frac{1}{n}\mathrm{E}^*\,\frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\theta_1\partial\theta'_1}\Big\| > \eta\Big)$$
$$+ \mathrm{P}^*\Big(n^a||\hat{\theta}^*_{1n} - \hat{\theta}_{1n}|| > \eta, \Big\|\frac{1}{n}\frac{\partial^2 L^*_{1n}(\tilde{\theta}^*_{1n})}{\partial\theta_1\partial\theta'_1} - \frac{1}{n}\mathrm{E}^*\,\frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\theta_1\partial\theta'_1}\Big\| \le \eta\Big).$$

The result follows from Lemmas 10–15 and Chebyshev's inequality. □

Lemma 17. *For $\eta > 0$ and $0 \le a < \frac{1}{2}$, $\mathrm{P}^*\big(n^{a-1}\big\|\frac{\partial^2 L^*_{1n}(\tilde{\theta}^*_{1n})}{\partial\theta_1\partial\theta'_1} - \mathrm{E}^*\,\frac{\partial^2 L^*_{1n}(\hat{\theta}_{1n})}{\partial\theta_1\partial\theta'_1}\big\| > \eta\big) = o_P(1)$, where $\tilde{\theta}^*_{1n}$ is between $\hat{\theta}^*_{1n}$ and $\hat{\theta}_{1n}$.*

Proof. The proof is similar to that for Lemma 14 except for the adjustments of orders and the application of Lemma 16. □

Appendix D. Proofs

Propositions 1 and 2 present the consistency and asymptotic normality of the QMLE for the null model. Their proofs are similar to those of Theorems 3.1 and 3.2 in Lee (2004a), except for some modifications to allow for SAR disturbances. Thus we omit their proofs, but focus on proving the results on the QMLE for the alternative model (Propositions 3 and 4), where necessary conditions and modifications will be pointed out.

Proof of Proposition 3. The convergence of $\hat{\theta}_{2n}-\bar{\theta}_{2n,1}$ to zero in probability will follow from the uniform convergence of $\frac{1}{n}[L_{2n}(\phi_2)-\bar{L}_{2n}(\phi_2;\theta_{10})]$ to zero on Φ_2 and the unique identification condition (White, 1994, Theorem 3.4).

We first show that $\sup_{\phi_2\in\Phi_2}|\frac{1}{n}L_{2n}(\phi_2)-\frac{1}{n}\bar{L}_{2n}(\phi_2;\theta_{10})|=o_P(1)$. For any $\phi_2\in\Phi_2$, $\frac{1}{n}\big(\bar{L}_{2n}(\phi_2;\theta_{10})-\bar{L}_{2n}(\bar{\phi}_{2n,1};\theta_{10})\big)\leq 0$ implies that

$$\frac{1}{2}\ln\big(\bar{\sigma}^2_{2n}(\phi_2;\theta_{10})\big)\geq\frac{1}{2}\ln(\bar{\sigma}^2_{2n,1})-\frac{1}{n}\big(\ln|S_{2n}|-\ln|S_{2n}(\lambda_2)|\big)-\frac{1}{n}\big(\ln|R_{2n}|-\ln|R_{2n}(\rho_2)|\big).$$

As in the proof of Theorem 3.1 in Lee (2004a), $\frac{1}{n}\big(\ln|S_{2n}|-\ln|S_{2n}(\lambda_2)|\big)$ is bounded uniformly in $\lambda_2\in\Lambda_2$ and $\frac{1}{n}\big(\ln|R_{2n}|-\ln|R_{2n}(\rho_2)|\big)$ is bounded uniformly in $\rho_2\in\varrho_2$. Since $\bar{\sigma}^2_{2n,1}$ is bounded away from zero by Assumption 15 and (A.4), $\ln(\bar{\sigma}^2_{2n,1})$ is also bounded. Thus, $\bar{\sigma}^2_{2n}(\phi_2;\theta_{10})$ is bounded away from zero uniformly in $\phi_2\in\Phi_2$. By the mean value theorem,

$$\frac{1}{n}[L_{2n}(\phi_2)-\bar{L}_{2n}(\phi_2;\theta_{10})]=-\frac{\hat{\sigma}^2_{2n}(\phi_2)-\bar{\sigma}^2_{2n}(\phi_2;\theta_{10})}{2\tilde{\sigma}^2_{2n}(\phi_2;\theta_{10})},$$

where $\tilde{\sigma}^2_{2n}(\phi_2;\theta_{10})$ is between $\hat{\sigma}^2_{1n}(\phi_2)$ and $\bar{\sigma}^2_{1n}(\phi_2;\theta_{10})$, and $\hat{\sigma}^2_{2n}(\phi_2)-\bar{\sigma}^2_{2n}(\phi_2;\theta_{10})=n^{-1}[\epsilon'_{1n}G_{1n}\epsilon_{1n}-\sigma^2_{10}\operatorname{tr}(G_{2n})+G_{3n}]$ with

$$\begin{aligned}
G_{1n}&=R'^{-1}_{1n}S'^{-1}_{1n}S'_{2n}(\lambda_2)R'_{2n}(\rho_2)H_{2n}(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S^{-1}_{1n}R^{-1}_{1n},\\
G_{2n}&=R'^{-1}_{1n}S'^{-1}_{1n}S'_{2n}(\lambda_2)R'_{2n}(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S^{-1}_{1n}R^{-1}_{1n},\\
G_{3n}&=2(X_{1n}\beta_{10})'S'^{-1}_{1n}S'_{2n}(\lambda_2)R'_{2n}(\rho_2)H_{2n}(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S^{-1}_{1n}R^{-1}_{1n}\epsilon_{1n}.
\end{aligned}$$

By Lemma 8, $X_{2n}(X'_{2n}R'_{2n}(\rho_2)R_{2n}(\rho_2)X_{2n})^{-1}X'_{2n}$ and $H_{2n}(\rho_2)$ are UB uniformly on ϱ_2.[33] By Lemma 9, $n^{-1}[\epsilon'_{1n}G_{1n}\epsilon_{1n}-\sigma^2_{10}(G_{2n})]=o_P(1)$ uniformly on Φ_2 and $n^{-1}G_{3n}=o_P(1)$ uniformly on Φ_2. Then $\hat{\sigma}^2_{2n}(\phi_2)-\bar{\sigma}^2_{2n}(\phi_2;\theta_{10})=o_P(1)$ uniformly on Φ_2. Consequently, $\sup_{\phi_2\in\Phi_2}\frac{1}{n}|L_{2n}(\phi_2)-\bar{L}_{2n}(\phi_2;\theta_{10})|=o_P(1)$, as $\tilde{\sigma}^2_{2n}(\phi_2;\theta_{10})$ is bounded away from zero uniformly on Φ_2 in probability.

With the uniform boundedness in both row and column sum norms of $H_{2n}(\rho_2)$, $\frac{1}{n}\bar{L}_{2n}(\phi_2;\theta_{10})$ is uniformly equicontinuous on Φ_2 as in the proof of Theorem 3.1 in Lee (2004a). The identification unique condition is

[33] Similarly, $H_{1n}(\rho_1)$ is UB uniformly on ϱ_1 for the proof of Proposition 1.

guaranteed by Assumption 14.[34] It follows that $\hat{\theta}_{2n} - \bar{\theta}_{2n,1} = o_P(1)$. □

Proof of Proposition 4. The proof is based on (6) obtained from the mean value theorem. We first prove that $\frac{1}{n}\frac{\partial^2 L_{2n}(\tilde{\theta}_{2n})}{\partial\theta_2\partial\theta_2'} - \mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'}\big) = o_P(1)$, which is done by showing that i) $\frac{1}{n}\frac{\partial^2 L_{2n}(\tilde{\theta}_{2n})}{\partial\theta_2\partial\theta_2'} - \frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'} = o_P(1)$ and ii) $\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'} - \mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'}\big) = o_P(1)$. After that, $\lim_{n\to\infty}\mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'}\big)$ is shown to be nonsingular in Step iii). Finally, applying the central limit theorem in Lemma 7 to $\frac{1}{\sqrt{n}}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2}$ and using Slutsky's Lemma, we obtain the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1})$.

i) Prove that $\frac{1}{n}\frac{\partial^2 L_{2n}(\tilde{\theta}_{2n})}{\partial\theta_2\partial\theta_2'} - \frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'} = o_P(1)$. By the mean value theorem and Assumption 12, $n^{-1}\operatorname{tr}\big[\big(W_{2n}S_{2n}^{-1}(\tilde{\lambda}_{2n})\big)^2\big] - n^{-1}\operatorname{tr}[(W_{2n}S_{2n}^{-1})^2] = 2n^{-1}(\tilde{\lambda}_{2n} - \bar{\lambda}_{2n,1})\operatorname{tr}\big[\big(W_{2n}S_{2n}^{-1}(\check{\lambda}_{2n})\big)^3\big] = o_P(1)$, where $\check{\lambda}_{2n}$ is between $\tilde{\lambda}_{2n}$ and $\bar{\lambda}_{2n,1}$. Similarly, $n^{-1}\operatorname{tr}\big[\big(M_{2n}R_{2n}^{-1}(\tilde{\rho}_{2n})\big)^2\big] - n^{-1}\operatorname{tr}[(M_{2n}R_{2n}^{-1})^2] = o_P(1)$. For the other terms in $\frac{1}{n}\frac{\partial^2 L_{2n}(\tilde{\theta}_{2n})}{\partial\theta_2\partial\theta_2'}$, we may first rewrite $S_{2n}(\tilde{\lambda}_{2n}) = S_{2n} + (\bar{\lambda}_{2n,1} - \tilde{\lambda}_{2n})W_{2n}$, $R_{2n}(\tilde{\rho}_{2n}) = R_{2n} + (\bar{\rho}_{2n,1} - \tilde{\rho}_{2n})M_{2n}$ and $\tilde{\beta}_{2n} = \bar{\beta}_{2n,1} + (\tilde{\beta}_{2n} - \bar{\beta}_{2n,1})$, and then expand these terms. Noting that $\tilde{\sigma}_{2n}^2$ is bounded away from zero in probability, $y_n = S_{1n}^{-1}X_{1n}\beta_{10} + S_{1n}^{-1}R_{1n}^{-1}\epsilon_{1n}$ and $S_{2n}y_n - X_{2n}\bar{\beta}_{2n,1} = [I_n - X_{2n}(X_{2n}'R_{2n}'R_{2n}X_{2n})^{-1}X_{2n}'R_{2n}'R_{2n}]S_{2n}S_{1n}^{-1}X_{1n}\beta_{10} + S_{1n}^{-1}R_{1n}^{-1}\epsilon_{1n}$, where $X_{2n}(X_{2n}'R_{2n}'R_{2n}X_{2n})^{-1}X_{2n}'$ is UB as shown in the proof of Proposition 3, we have $\frac{1}{n}\frac{\partial^2 L_{2n}(\tilde{\theta}_{2n})}{\partial\theta_2\partial\theta_2'} - \frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\phi}_{2n,1},\bar{\beta}_{2n,1},\tilde{\sigma}_{2n}^2)}{\partial\theta_2\partial\theta_2'} = o_P(1)$, by Lemmas (3)–(5). In addition, $\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\phi}_{2n,1},\bar{\beta}_{2n,1},\tilde{\sigma}_{2n}^2)}{\partial\theta_2\partial\theta_2'} - \frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'} = o_P(1)$ by the mean value theorem. Therefore, $\frac{1}{n}\frac{\partial^2 L_{2n}(\tilde{\theta}_{2n})}{\partial\theta_2\partial\theta_2'} - \frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'} = o_P(1)$.

ii) Prove that $\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'} - \mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'}\big) = o_P(1)$. Terms in $\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'} - \mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'}\big)$ have the form $\frac{1}{n}[\epsilon_{1n}'A_n\epsilon_{1n} - \operatorname{tr}(A_n)] + \frac{1}{n}c_n'B_n\epsilon_{1n}$ or $\frac{1}{n}X_{2n}'B_n\epsilon_{1n}$, where the n-dimensional square matrices A_n and B_n are UB, and elements of n-dimensional vector c_n are uniformly bounded. By Lemmas (3) and (4), $\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'} - \mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'}\big) = o_P(1)$.

iii) Prove the non-singularity of $\lim_{n\to\infty}\mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2\partial\theta_2'}\big)$. Let $\psi_2 = (\beta_2', \sigma_2^2)'$. Then $\lim_{n\to\infty}\mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\psi_2\partial\psi_2'}\big) =$

$\begin{pmatrix} -\lim_{n\to\infty}\frac{1}{n\bar{\sigma}_{2n,1}^2}X_{2n}'R_{2n}'R_{2n}X_{2n} & 0 \\ 0 & -\lim_{n\to\infty}\frac{1}{2\bar{\sigma}_{2n,1}^4}\end{pmatrix}$ is nonsingular. Suppose that we have a block matrix $G = \begin{pmatrix} A & B \\ C & D\end{pmatrix}$, where A is a square matrix and D is invertible, then it is sufficient to show that $A - BD^{-1}C$ is nonsingular to prove the nonsingularity of G, because $\begin{pmatrix} A & B \\ C & D\end{pmatrix} = \begin{pmatrix} A - BD^{-1}C & B \\ 0 & D\end{pmatrix}\begin{pmatrix} I_l & 0 \\ D^{-1}C & I_m\end{pmatrix}$.

In the current situation, we need to show that

$\lim_{n\to\infty}\mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\phi_2\partial\phi_2'}\big) - \lim_{n\to\infty}\mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\phi_2\partial\psi_2'}\big)\big[\lim_{n\to\infty}\mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\psi_2\partial\psi_2'}\big)\big]^{-1}\lim_{n\to\infty}\mathrm{E}\big(\frac{1}{n}\frac{\partial^2 L_{2n}(\bar{\theta}_{2n,1})}{\partial\psi_2\partial\phi_2'}\big)$

[34] For the identification uniqueness condition of the null model, note that $\frac{1}{n}[\bar{L}_{1n}(\phi_1;\theta_{10}) - \bar{L}_{1n}(\phi_{10};\theta_{10})]$ can be decomposed as the sum of $\frac{1}{2n}[\ln|\sigma_{10}^2 S_{1n}^{-1}R_{1n}^{-1}R_{1n}'^{-1}S_{1n}'^{-1}| - \ln|\bar{\sigma}_{1n,a}^2(\phi_1)S_{1n}^{-1}(\lambda_1)R_{1n}^{-1}(\rho_1)R_{1n}'^{-1}(\rho_1)S_{1n}'^{-1}(\lambda_1)|]$ and $-\frac{1}{n}(\lambda_{10} - \lambda_1)^2(Q_{1n}X_{1n}\beta_{10})'R_{1n}'(\rho_1)H_{1n}(\rho_1)R_{1n}(\rho_1)Q_{1n}X_{1n}\beta_{10}/\tilde{\sigma}_{1n}^2$ with both terms being non-positive, where $\tilde{\sigma}_{1n}^2$ is between $\bar{\sigma}_{1n}^2(\phi_1;\theta_{10})$ and $\bar{\sigma}_{1n,a}^2(\phi_1)$, by the method in the proof of Theorem 3.1 in Lee (2004a). Then Assumption 7 provides sufficient conditions for global identification.

is nonsingular. Let $\psi_{2n}(\phi_2)$ satisfy $\frac{\partial \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \psi_2} = 0$ and let $g_n(\phi_2) = \operatorname{E} L_{2n}(\phi_2, \psi_{2n}(\phi_2))$. Taking the derivative of $\frac{\partial \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \psi_2} = 0$ with respective to ϕ_2, we have $\frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \psi_2 \partial \phi_2'} + \frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \psi_2 \partial \psi_2'} \frac{\partial \psi_{2n}(\phi_2)}{\partial \phi_2'} = 0$. So $\lim_{n\to\infty} \frac{\partial \psi_{2n}(\phi_2)}{\partial \phi_2'} = -\big(\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \psi_2 \partial \psi_2'}\big)^{-1} \lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \psi_2 \partial \phi_2'}$. Since $\frac{\partial g_n(\phi_2)}{\partial \phi_2} = \frac{\partial \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \phi_2}$, $\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g_n(\phi_2)}{\partial \phi_2 \partial \phi_2'} = \lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \phi_2 \partial \phi_2'} + \lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \phi_2 \partial \psi_2'} \lim_{n\to\infty} \frac{\partial \psi_{2n}(\phi_2)}{\partial \phi_2'} = \lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \phi_2 \partial \phi_2'} - \lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \phi_2 \partial \psi_2'} \big(\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \psi_2 \partial \psi_2'}\big)^{-1} \lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \operatorname{E} L_{2n}(\phi_2,\psi_{2n}(\phi_2))}{\partial \psi_2 \partial \phi_2'}$. As $\psi_{2n}(\bar{\phi}_{2n,1}) = \bar{\psi}_{2n,1}$, Assumption 16 implies that $\lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 g_n(\bar{\phi}_{2n,1})}{\partial \phi_2 \partial \phi_2'} = \lim_{n\to\infty} \frac{1}{n} \frac{\partial^2 \bar{L}_{2n}(\bar{\phi}_{2n,1};\theta_{10})}{\partial \phi_2 \partial \phi_2'}$ is nonsingular.[35]

The asymptotic distribution of $\sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1})$ follows from the expansion in (6) by using the central limit theorem in Lemma 7. □

Proof of Proposition 5. We only check that (10) holds, as other details are in the text. By a second order Taylor expansion,

$$\frac{1}{\sqrt{n}}[L_{2n}(\bar{\theta}_{2n,1}) - L_{2n}(\hat{\theta}_{2n})] = \frac{1}{2}(\bar{\theta}_{2n,1} - \hat{\theta}_{2n})' \frac{1}{n} \frac{\partial^2 L_{2n}(\check{\theta}_{2n,1})}{\partial \theta_2 \partial \theta_2'} \sqrt{n}(\bar{\theta}_{2n,1} - \hat{\theta}_{2n}) = o_P(1), \tag{D.1}$$

where $\check{\theta}_{2n,1}$ is between $\hat{\theta}_{2n}$ and $\bar{\theta}_{2n,1}$, and $\frac{1}{n} \frac{\partial^2 L_{2n}(\check{\theta}_{2n,1})}{\partial \theta_2 \partial \theta_2'} = O_P(1)$ can be seen from the proof of Proposition 4. Similarly,

$$\begin{aligned} &\frac{1}{\sqrt{n}}[\bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\hat{\theta}_{1n})] \\ &= \frac{1}{n} \frac{\partial \bar{L}_{2n}(\bar{\theta}_{2n,1};\hat{\theta}_{1n})}{\partial \theta_2'} \sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1}) + \frac{1}{2}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1})' \Big(\frac{1}{n} \frac{\partial^2 \bar{L}_{2n}(\check{\theta}_{2n,1};\hat{\theta}_{1n})}{\partial \theta_2 \partial \theta_2'}\Big) \sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1}) \\ &= (\hat{\theta}_{1n} - \theta_{10})' \frac{1}{n} \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\check{\theta}_{1n})}{\partial \theta_1 \partial \theta_2'} \sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1}) + o_P(1) = o_P(1), \end{aligned} \tag{D.2}$$

where $\check{\theta}_{2n,1}$ is between $\hat{\theta}_{2n}$ and $\bar{\theta}_{2n,1}$, and $\check{\theta}_{1n}$ is between $\hat{\theta}_{1n}$ and θ_{10}, since $\frac{1}{n} \frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\check{\theta}_{1n})}{\partial \theta_1 \partial \theta_2'} = O_P(1)$ and $\frac{1}{n} \frac{\partial^2 \bar{L}_{2n}(\check{\theta}_{2n,1};\hat{\theta}_{1n})}{\partial \theta_2 \partial \theta_2'} = O_P(1)$. Furthermore,

$$\begin{aligned} \frac{1}{\sqrt{n}}[\bar{L}_{2n}(\bar{\theta}_{2n,1};\hat{\theta}_{1n}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})] &= \frac{1}{n} \frac{\partial \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial \theta_1'} \sqrt{n}(\hat{\theta}_{1n} - \theta_{10}) + o_P(1) \\ &= C_{2n,1}' \Sigma_{1n,1}^{-1} \frac{1}{\sqrt{n}} \frac{\partial L_{1n}(\theta_{10})}{\partial \theta_1} + o_P(1). \end{aligned} \tag{D.3}$$

Combining (D.1)–(D.3) yields (10). □

Proof of Proposition 6. We prove the result for Cox_a. The result for Cox_o can be proved similarly. Rewrite $L_{2n}(\hat{\theta}_{2n}) - \bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n})$ as

$$\begin{aligned} &L_{2n}(\hat{\theta}_{2n}) - \bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n}) \\ &= \big(L_{2n}(\bar{\theta}_{2n,1}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})\big) - \big(\bar{L}_{2n}(\bar{\theta}_{2n,1};\hat{\theta}_{1n}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})\big) \\ &\quad - \big(\bar{L}_{2n}(\hat{\theta}_{2n};\hat{\theta}_{1n}) - \bar{L}_{2n}(\bar{\theta}_{2n,1};\hat{\theta}_{1n})\big) - \big(L_{2n}(\bar{\theta}_{2n,1}) - L_{2n}(\hat{\theta}_{2n})\big) \end{aligned}$$

[35] For the estimation of the null model, Assumption 8 is needed instead for the non-singularity of $\frac{1}{n} \frac{\partial \bar{L}_{1n}(\theta_{10};\theta_{10})}{\partial \theta_1 \partial \theta_1'}$ in the limit.

$$= \big(L_{2n}(\bar\theta_{2n,1}) - \bar L_{2n}(\bar\theta_{2n,1};\theta_{10})\big) - \Big(\frac{\partial \bar L_{2n}(\bar\theta_{2n,1};\theta_{10})}{\partial\theta_1'} + \frac{1}{2}(\hat\theta_{1n}-\theta_{10})'\frac{\partial^2 \bar L_{2n}(\bar\theta_{2n,1};\tilde\theta_{1n})}{\partial\theta_1\partial\theta_1'}\Big)(\hat\theta_{1n}-\theta_{10})$$
$$- \Big((\hat\theta_{1n}-\theta_{10})'\frac{\partial^2 \bar L_{2n}(\ddot\theta_{2n};\ddot\theta_{1n})}{\partial\theta_1\partial\theta_2'} + (\dot\theta_{2n}-\bar\theta_{2n,1})'\frac{\partial^2 \bar L_{2n}(\ddot\theta_{2n};\ddot\theta_{1n})}{\partial\theta_2\partial\theta_2'}\Big)(\hat\theta_{2n}-\bar\theta_{2n,1})$$
$$- \frac{1}{2}(\hat\theta_{2n}-\bar\theta_{2n,1})'\frac{\partial^2 L_{2n}(\check\theta_{2n})}{\partial\theta_2\partial\theta_2'}(\hat\theta_{2n}-\bar\theta_{2n,1}),$$

where $\tilde\theta_{1n}$ and $\ddot\theta_{1n}$ are both between $\hat\theta_{1n}$ and θ_{10}, $\dot\theta_{2n}$ and $\check\theta_{2n}$ are both between $\hat\theta_{2n}$ and $\bar\theta_{2n,1}$, and $\ddot\theta_{2n}$ is between $\dot\theta_{2n}$ and $\bar\theta_{2n,1}$. By the mean value theorem, $\hat\theta_{1n}-\theta_{10} = \Sigma_{1n,1}^{-1}\frac{1}{n}\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1} + \Sigma_{1n,1}^{-1}\big(\frac{1}{n}\frac{\partial^2 L_{1n}(\check\theta_{1n})}{\partial\theta_1\partial\theta_1'} + \Sigma_{1n,1}\big)(\hat\theta_{1n}-\theta_{10})$, where $\Sigma_{1n,1} = -\frac{1}{n}\frac{\partial \bar L_{1n}(\theta_{10};\theta_{10})}{\partial\theta_1\partial\theta_1'}$ and $\check\theta_{1n}$ is between $\hat\theta_{1n}$ and θ_{10}. Let $Cox_a = (D_n + E_n)/\hat\sigma_{ca,n}$, where $D_n = \frac{1}{\sqrt{n}}[L_{2n}(\bar\theta_{2n,1}) - \bar L_{2n}(\bar\theta_{2n,1};\theta_{10})] - C_{2n,1}'\Sigma_{1n,1}^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1}$. Then

$$n^{1/4}E_n = n^{1/4}\Big(\frac{1}{\sqrt{n}}\big(L_{2n}(\hat\theta_{2n}) - \bar L_{2n}(\hat\theta_{2n};\hat\theta_{1n})\big) - D_n\Big)$$
$$= -\frac{1}{n}\frac{\partial \bar L_{2n}(\bar\theta_{2n,1};\theta_{10})}{\partial\theta_1'}\Sigma_{1n,1}^{-1}n^{3/8}\Big(\frac{1}{n}\frac{\partial^2 L_{1n}(\check\theta_{1n})}{\partial\theta_1\partial\theta_1'} + \Sigma_{1n,1}\Big)n^{3/8}(\hat\theta_{1n}-\theta_{10})$$
$$- \frac{1}{2}n^{3/8}(\hat\theta_{1n}-\theta_{10})'\frac{1}{n}\frac{\partial^2 \bar L_{2n}(\bar\theta_{2n,1};\tilde\theta_{1n})}{\partial\theta_1\partial\theta_1'}n^{3/8}(\hat\theta_{1n}-\theta_{10})$$
$$- \Big(n^{3/8}(\hat\theta_{1n}-\theta_{10})'\frac{1}{n}\frac{\partial^2 \bar L_{2n}(\ddot\theta_{2n};\ddot\theta_{1n})}{\partial\theta_1\partial\theta_2'} + n^{3/8}(\dot\theta_{2n}-\bar\theta_{2n,1})'\frac{1}{n}\frac{\partial^2 \bar L_{2n}(\ddot\theta_{2n};\ddot\theta_{1n})}{\partial\theta_2\partial\theta_2'}\Big)n^{3/8}(\hat\theta_{2n}-\bar\theta_{2n,1})$$
$$- \frac{1}{2}n^{3/8}(\hat\theta_{2n}-\bar\theta_{2n,1})'\frac{1}{n}\frac{\partial^2 L_{2n}(\check\theta_{2n})}{\partial\theta_2\partial\theta_2'}n^{3/8}(\hat\theta_{2n}-\bar\theta_{2n,1}).$$

By Propositions (2), (4) and Lemma 12, $n^{1/4}E_n = o_P(1)$. Then $n^{1/4}E_n/\sigma_{c,n} = o_P(1)$, as $\sigma_{c,n}$ is bounded away from zero. Since $\sigma_{c,n}$ is the standard deviation of the LQ form D_n, we can easily show that $n^{1/2}(\hat\sigma_{ca,n}^2 - \sigma_{c,n}^2) = O_P(1)$ by the mean value theorem, Propositions (2), (4) and Lemma 10. Note that $n^{1/4}(Cox_a - D_n/\sigma_{c,n}) = n^{1/4}\frac{E_n}{\sigma_{c,n}} + n^{3/8}\frac{\sigma_{c,n}-\hat\sigma_{ca,n}}{\hat\sigma_{ca,n}}n^{-1/8}\frac{D_n}{\sigma_{c,n}} + \frac{\sigma_{c,n}-\hat\sigma_{ca,n}}{\hat\sigma_{ca,n}}n^{1/4}\frac{E_n}{\sigma_{c,n}}$, then $n^{1/4}(Cox_a - D_n/\sigma_{c,n}) = o_P(1)$. Let E_n^* be the bootstrapped E_n. An expression for $n^{1/4}E_n^*$ can be derived from $n^{1/4}E_n$ by replacing some terms:

$$n^{1/4}E_n^* = n^{1/4}\Big(\frac{1}{\sqrt{n}}\big(L_{2n}^*(\hat\theta_{2n}^*) - \bar L_{2n}(\hat\theta_{2n}^*;\hat\theta_{1n}^*)\big) - D_n^*\Big)$$
$$= \frac{1}{n}\frac{\partial \bar L_{2n}(\hat\theta_{2n};\hat\theta_{1n})}{\partial\theta_1'}\frac{1}{n}\frac{\partial^2\, \mathrm{E}^*\, L_{1n}^*(\hat\theta_{1n})}{\partial\theta_1\partial\theta_1'}n^{3/8}\Big(\frac{1}{n}\frac{\partial^2 L_{1n}^*(\check\theta_{1n}^*)}{\partial\theta_1\partial\theta_1'} - \frac{1}{n}\frac{\partial^2\, \mathrm{E}^*\, L_{1n}^*(\hat\theta_{1n})}{\partial\theta_1\partial\theta_1'}\Big)n^{3/8}(\hat\theta_{1n}^*-\hat\theta_{1n})$$
$$- \frac{1}{2}n^{3/8}(\hat\theta_{1n}^*-\hat\theta_{1n})'\frac{1}{n}\frac{\partial^2 \bar L_{2n}(\hat\theta_{2n};\tilde\theta_{1n}^*)}{\partial\theta_1\partial\theta_1'}n^{3/8}(\hat\theta_{1n}^*-\hat\theta_{1n})$$
$$- \Big(n^{3/8}(\hat\theta_{1n}^*-\hat\theta_{1n})'\frac{1}{n}\frac{\partial^2 \bar L_{2n}(\ddot\theta_{2n}^*;\ddot\theta_{1n}^*)}{\partial\theta_1\partial\theta_2'} + n^{3/8}(\dot\theta_{2n}^*-\hat\theta_{2n})'\frac{1}{n}\frac{\partial^2 \bar L_{2n}(\ddot\theta_{2n}^*;\ddot\theta_{1n}^*)}{\partial\theta_2\partial\theta_2'}\Big)n^{3/8}(\hat\theta_{2n}^*-\hat\theta_{2n})$$
$$- \frac{1}{2}n^{3/8}(\hat\theta_{2n}^*-\hat\theta_{2n})'\frac{1}{n}\frac{\partial^2 L_{2n}^*(\check\theta_{2n}^*)}{\partial\theta_2\partial\theta_2'}n^{3/8}(\hat\theta_{2n}^*-\hat\theta_{2n}),$$

where $\tilde\theta_{1n}^*$, $\ddot\theta_{1n}^*$ and $\check\theta_{1n}^*$ are between $\hat\theta_{1n}^*$ and $\hat\theta_{1n}$, $\dot\theta_{2n}^*$ and $\check\theta_{2n}^*$ are both between $\hat\theta_{2n}^*$ and $\hat\theta_{2n}$, and $\ddot\theta_{2n}^*$ is between $\dot\theta_{2n}^*$ and $\hat\theta_{2n}$. By Lemmas 14–17, $\mathrm{P}^*(n^{1/4}|E_n^*| > \eta) = o_P(1)$. Since $\mathrm{P}^*(n^{3/8}|\hat\sigma_{ca,n}^* - \sigma_{c,n}^*| >$

$\eta) = O_P(n^{-1/4})$ and $n^{1/4}(Cox_a^* - D_n^*/\sigma_{c,n}^*) = n^{1/4}\frac{E_n^*}{\sigma_{c,n}^*} + n^{3/8}\frac{\sigma_{c,n}^*-\hat{\sigma}_{ca,n}^*}{\hat{\sigma}_{ca,n}^*}n^{-1/8}\frac{D_n^*}{\sigma_{c,n}^*} + \frac{\sigma_{c,n}^*-\hat{\sigma}_{ca,n}^*}{\hat{\sigma}_{ca,n}^*}n^{1/4}\frac{E_n^*}{\sigma_{c,n}^*}$, $\mathrm{P}^*(n^{1/4}|Cox_a^* - D_n^*/\sigma_{c,n}^*| > \eta) = o_P(1)$. The consistency result on Cox_a in the proposition holds by Theorem 1 in Jin and Lee (2012) with $\delta = 1/2$. □

Proof of Proposition 7. We first prove that $\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \bar{\theta}_{2n,1} = o_P(1)$. For a fixed ϕ_2, the maximization of $\bar{L}_{2n}(\theta_2;\hat{\theta}_{1n})$ yields $\bar{\beta}_{2n}(\phi_2;\hat{\theta}_{1n})$ and $\bar{\sigma}_{2n}^2(\phi_2;\hat{\theta}_{1n})$, whose expressions are given in (A.3) and (A.4). Then by the mean value theorem,

$$
\begin{aligned}
&\bar{\beta}_{2n}(\phi_2;\hat{\theta}_{1n}) - \bar{\beta}_{2n}(\phi_2;\theta_{10})\\
&= [X_{2n}'R_{2n}'(\rho_2)R_{2n}(\rho_2)X_{2n}]^{-1}X_{2n}'R_{2n}'(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S_{1n}^{-1}(\tilde{\lambda}_{1n})[W_{1n}S_{1n}^{-1}(\tilde{\lambda}_{1n})X_{1n}\tilde{\beta}_{1n}(\hat{\lambda}_{1n}-\lambda_{10})\\
&\quad + X_{1n}(\hat{\beta}_{1n}-\beta_{10})],\\
&\bar{\sigma}_{2n}^2(\phi_2;\hat{\theta}_{1n}) - \bar{\sigma}_{2n}^2(\phi_2;\theta_{10})\\
&= \frac{1}{n}\operatorname{tr}[R_{1n}'^{-1}(\tilde{\rho}_{1n})S_{1n}'^{-1}(\tilde{\lambda}_{1n})S_{2n}'(\lambda_2)R_{2n}'(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S_{1n}^{-1}(\tilde{\lambda}_{1n})R_{1n}^{-1}(\tilde{\rho}_{1n})](\hat{\sigma}_{1n}^2-\sigma_{10}^2)\\
&\quad + \frac{2\tilde{\sigma}_{1n}^2}{n}\operatorname{tr}[R_{1n}'^{-1}(\tilde{\rho}_{1n})M_{1n}'R_{1n}'^{-1}(\tilde{\rho}_{1n})S_{1n}'^{-1}(\tilde{\lambda}_{1n})S_{2n}'(\lambda_2)R_{2n}'(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S_{1n}^{-1}(\tilde{\lambda}_{1n})R_{1n}^{-1}(\tilde{\rho}_{1n})](\hat{\rho}_{1n}-\rho_{10})\\
&\quad + \frac{2\tilde{\sigma}_{1n}^2}{n}\operatorname{tr}[R_{1n}'^{-1}(\tilde{\rho}_{1n})S_{1n}'^{-1}(\tilde{\lambda}_{1n})W_{1n}'S_{1n}'^{-1}(\tilde{\lambda}_{1n})S_{2n}'(\lambda_2)R_{2n}'(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S_{1n}^{-1}(\tilde{\lambda}_{1n})R_{1n}^{-1}(\tilde{\rho}_{1n})](\hat{\lambda}_{1n}-\lambda_{10})\\
&\quad + \frac{2}{n}(X_{1n}\tilde{\beta}_{1n})'S_{1n}'^{-1}(\tilde{\lambda}_{1n})W_{1n}'S_{1n}'^{-1}(\tilde{\lambda}_{1n})S_{2n}'(\lambda_2)R_{2n}'(\rho_2)H_{2n}(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S_{1n}^{-1}(\tilde{\lambda}_{1n})X_{1n}\tilde{\beta}_{1n}(\hat{\lambda}_{1n}-\lambda_{10}),\\
&\quad + \frac{2}{n}(X_{1n}\tilde{\beta}_{1n})'S_{1n}'^{-1}(\tilde{\lambda}_{1n})S_{2n}'(\lambda_2)R_{2n}'(\rho_2)H_{2n}(\rho_2)R_{2n}(\rho_2)S_{2n}(\lambda_2)S_{1n}^{-1}(\tilde{\lambda}_{1n})X_{1n}(\hat{\beta}_{1n}-\beta_{10}),
\end{aligned}
$$

where $\tilde{\theta}_{1n} = (\tilde{\phi}_{1n}, \tilde{\beta}_{2n}', \tilde{\sigma}_{1n}^2)'$ is between $\hat{\theta}_{1n}$ and θ_{10}. Elements of $\left(n^{-1}X_{2n}'R_{2n}'(\rho_2)R_{2n}(\rho_2)X_{2n}\right)^{-1}$ are bounded uniformly on ϱ_2 and $H_{2n}(\rho_2)$ is UB uniformly on ϱ_2 as in the proof of Proposition 3. Writing $\tilde{\beta}_{1n} = \beta_{10} + (\tilde{\beta}_{1n}-\beta_{10})$, then by Lemmas 5 and 6, $\bar{\beta}_{2n}(\phi_2;\hat{\theta}_{1n}) - \bar{\beta}_{2n}(\phi_2;\theta_{10})$ and $\bar{\sigma}_{2n}^2(\phi_2;\hat{\theta}_{1n}) - \bar{\sigma}_{2n}^2(\phi_2;\theta_{10})$ both converge to zero in probability uniformly on Φ_2. To verify that $\bar{\lambda}_{2n}(\hat{\theta}_{1n}) - \bar{\lambda}_{2n,1} = o_P(1)$ and $\bar{\rho}_{2n}(\hat{\theta}_{1n}) - \bar{\rho}_{2n,1} = o_P(1)$, we only need to show that $n^{-1}[\bar{L}_{2n}(\phi_2;\hat{\theta}_{1n}) - \bar{L}_{2n}(\phi_2;\theta_{10})]$ converges in probability to zero uniformly on Φ_2, as the unique identification is guaranteed by Assumption 14. By the mean value theorem,

$$
\frac{1}{n}[\bar{L}_{2n}(\phi_2;\hat{\theta}_{1n}) - \bar{L}_{2n}(\phi_2;\theta_{10})] = -\frac{1}{2\tilde{\sigma}_{2n,1}^2}[\bar{\sigma}_{2n}^2(\phi_2;\hat{\theta}_{1n}) - \bar{\sigma}_{2n}^2(\phi_2;\theta_{10})],
$$

where $\tilde{\sigma}_{2n,1}^2$ is between $\bar{\sigma}_{2n}^2(\phi_2;\hat{\theta}_{1n})$ and $\bar{\sigma}_{10}^2(\phi_2;\theta_{10})$. Since $\bar{\sigma}_{2n}^2(\phi_2;\theta_{10})$ is bounded away from zero and $\bar{\sigma}_{2n}^2(\phi_2;\hat{\theta}_{1n}) - \bar{\sigma}_{2n}^2(\phi_2;\theta_{10}) = o_P(1)$ uniformly on Φ_2, $\sup_{\phi_2\in\Phi_2}|\frac{1}{n}[\bar{L}_{2n}(\phi_2;\hat{\theta}_{1n}) - \bar{L}_{2n}(\phi_2;\theta_{10})]| = o_P(1)$.

An expression for $\sqrt{n}[\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \bar{\theta}_{2n,1}]$ can be derived from the expansion of the first order condition $\frac{\partial\bar{L}_{2n}(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n})}{\partial\theta_2} = 0$ at $\bar{\theta}_{2n,1}$:

$$
0 = \frac{\partial\bar{L}_{2n}(\bar{\theta}_{2n}(\hat{\theta}_{1n});\hat{\theta}_{1n})}{\partial\theta_2} = \frac{\partial\bar{L}_{2n}(\bar{\theta}_{2n,1};\hat{\theta}_{1n})}{\partial\theta_2} + \frac{\partial^2\bar{L}_{2n}(\tilde{\theta}_{2n,1};\hat{\theta}_{1n})}{\partial\theta_2\partial\theta_2'}[\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \bar{\theta}_{2n,1}],
$$

where $\tilde{\theta}_{2n,1}$ is between $\bar{\theta}_{2n,1}$ and $\bar{\theta}_{2n}(\hat{\theta}_{1n})$. Then we have

$$\begin{aligned}\sqrt{n}[\bar{\theta}_{2n}(\hat{\theta}_{1n}) - \bar{\theta}_{2n,1}] &= \Big(-\frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}_{2n,1};\hat{\theta}_{1n})}{\partial\theta_2\partial\theta_2'}\Big)^{-1}\frac{1}{\sqrt{n}}\frac{\partial \bar{L}_{2n}(\bar{\theta}_{2n,1};\hat{\theta}_{1n})}{\partial\theta_2} \\ &= \Big(-\frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}_{2n,1};\hat{\theta}_{1n})}{\partial\theta_2\partial\theta_2'}\Big)^{-1}\frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\tilde{\theta}_{1n})}{\partial\theta_2\partial\theta_1'}\sqrt{n}(\hat{\theta}_{1n}-\theta_{10}),\end{aligned} \tag{D.4}$$

where $\tilde{\theta}_{1n}$ is between $\hat{\theta}_{1n}$ and θ_{10}. We can show that

$$-\frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\tilde{\theta}_{2n,1};\hat{\theta}_{1n})}{\partial\theta_2\partial\theta_2'} = \Sigma_{2n,1} + o_P(1) \quad \text{and} \quad \frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\tilde{\theta}_{1n})}{\partial\theta_2\partial\theta_1'} = \frac{1}{n}\frac{\partial^2 \bar{L}_{2n}(\bar{\theta}_{2n,1};\theta_{10})}{\partial\theta_2\partial\theta_1'} + o_P(1),$$

by writing $\tilde{\theta}_{2n,1} = \bar{\theta}_{2n,1} + (\tilde{\theta}_{2n,1} - \bar{\theta}_{2n,1})$, $\hat{\theta}_{1n} = \theta_{10} + (\hat{\theta}_{1n} - \theta_{10})$ and $\tilde{\theta}_{1n} = \theta_{10} + (\tilde{\theta}_{1n} - \theta_{10})$, and then expanding the expressions. Using (6), (B.1) and (D.4), we obtain (B.2). The asymptotic distribution of $\sqrt{n}(\hat{\theta}_{2n} - \bar{\theta}_{2n,1})$ follows from applying the central limit theorem in Lemma 7.

In the case that $\epsilon_{1n,i}$'s are normally distributed, We note that $P_{2n,1} = \mathrm{E}\big(\frac{1}{n}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2}\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1'}\big)$ and $\Sigma_{1n,1} = \Omega_{1n,1}$. Then the covariance matrix between $\frac{1}{\sqrt{n}}\frac{\partial L_{2n}(\bar{\theta}_{2n,1})}{\partial\theta_2}$ and $P_{2n,1}\Sigma_{1n,1}^{-1}\frac{1}{\sqrt{n}}\frac{\partial L_{1n}(\theta_{10})}{\partial\theta_1}$ is just equal to the VC matrix of the latter, and we have $V_{2n,1} = \Omega_{2n,1} - P_{2n,1}\Sigma_{1n,1}^{-1}P_{2n,1}'$. □

References

Aguirre-Torres, V., Gallant, A.R., 1983. The null and non-null asymptotic distribution of the Cox test for multivariate nonlinear regression: Alternatives and a new distribution-free Cox test. Journal of Econometrics 21, 5–33.

Amemiya, T., 1985. Advanced econometrics. Harvard University Press, Cambridge, MA.

Andrews, D.W.K., 1987. Asymptotic results for generalized Wald tests. Econometric Theory 3, 348–358.

Anselin, L., 1984. Specification tests on the structure of interaction in spatial econometric models. Papers in Regional Science 54, 165–182.

Anselin, L., 1986. Non-nested tests of the weight structure in spatial autoregressive models: Some Monte Carlo results. Journal of Regional Science 26, 267–284.

Anselin, L., 1988. Spatial Econometrics: Methods and Models. Boston: Kluwer Academic Publishers.

Anselin, L., Florax, R.J., 1995. Small sample properties of tests for spatial dependence in regression models: Some further results, in: Anselin, L., Florax, R.J. (Eds.), New Directions in Spatial Econometrics. Springer-Verlag, Berlin, pp. 21–74.

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. Regional Science and Urban Economics 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.

Cliff, A., Ord, J.K., 1981. Spatial Process: Models and Applications. Pion, London.

Cox, D., 1961. Tests of separate families of hypotheses. Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability 1, 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, J., 1994. Stochastic Limit Theory: An Introduction for Econometricians. Oxford University Press, USA.

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

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

Davies, R.B., 1977. Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 62, 247–254.

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. New York: Wiley.

Donald, S.G., Fortuna, N., Pipiras, V., 2007. On rank estimation in symmetric matrices: the case of indefinite matrix estimators. Econometric Theory 23, 1217–1232.

Donald, S.G., Fortuna, N., Pipiras, V., 2010. On rank estimation in semidefinite matrices. CEF.UP working paper.

Fan, Y., Li, Q., 1995. Bootstrapping J-type tests for non-nested regression models. Economics Letters 48, 107–112.

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. Amsterdam: North-Holland Elsevier.

Gourieroux, C., Monfort, A., 1995. Testing, encompassing and simulating dynamic econometric models. Econometric Theory 11, 195–228.

Gourieroux, C., Monfort, A., Trognon, A., 1983. Testing nested or non-nested hypotheses. Journal of Econometrics 21, 83–115.

Harrison, David, J., Rubinfeld, D.L., 1978. Hedonic housing prices and the demand for clean air. Journal of Environmental Economics and Management 5, 81–102.

Horn, R., Johnson, C., 1985. Matrix Analysis. New York: Cambridge University Press.

Horowitz, J.L., 1994. Bootstrap-based critical values for the information matrix test. Journal of Econometrics 61, 395–411.

Jin, F., Lee, L.F., 2012. On the bootstrap for Moran's I test for spatial dependence. Working paper, Department of Economics, The Ohio State University.

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., 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509–533.

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., Robinson, D.P., 1993. A suggested method of estimation for spatial interdependent models with autocorrelated errors, and an application to a county expenditure model. Papers in Regional Science 72, 297–312.

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

Lee, L.F., 2004b. A supplement to asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models. Department of Economics, The Ohio State University, Columbus, OH.

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

Lee, L.F., Liu, X., 2010. Efficient GMM estimation of high order spatial autoregressive models with autoregressive disturbances. Econometric Theory 26, 187–230.

LeSage, J., Pace, R.K., 2009. Introduction to spatial econometrics. volume 196. Chapman & Hall/CRC.

LeSage, J.P., 1999. Spatial Econometrics. Web book of Regional Science. Available at: http://www.rri.wvu.edu/WebBook/LeSage/spatial/wbook.pdf.

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., 2013. Social multiplier or social norms? Working paper.

Lütkepohl, H., Burda, M.M., 1997. Modified Wald tests under nonregular conditions. Journal of Econometrics 78, 315–332.

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.

McAleer, M., 1995. The significance of testing empirical non-nested models. Journal of Econometrics 67, 149–171.

McAleer, M., Pesaran, M.H., 1986. Statistical inference in non-nested econometric models. Applied Mathematics and Computation 20, 271–311.

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

Pace, R.K., Gilley, O.W., 1997. Using the spatial configuration of the data to improve estimation. Journal of the Real Estate Finance and Economics 14, 333–340.

Pace, R.K., LeSage, J.P., 2009. A sampling approach to estimate the log determinant used in spatial likelihood problems. Journal of Geographical Systems 11, 209–225.

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. Oxford: Basil Blackwell.

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

Sawa, T., 1978. Information criteria for discriminating among alternative regression models. Econometrica 46, 1273–1291.

Smirnov, O., Anselin, L., 2009. An $O(N)$ parallel method of computing the log-Jacobian of the variable transformation for models with spatial interaction on a lattice. Computational Statistics and Data Analysis 53, 2980–2988.

White, H., 1982. Maximum likelihood estimation of misspecified models. Econometrica 50, 1–25.

White, H., 1994. Estimation, inference and specification analysis. New York: Cambridge University Press.