A note on testing for spatial error components

Fernando Carriazo
Economic Research Service
US Department of Agriculture
Washington DC

Edward Coulson
Department of Economics
Penn State University
University Park, PA

Abstract: The two tests for spatial error components are the Kelejian-Robinson (KR) test and the LM test developed by Anselin(2001). We develop a new form for the LM test which facilitates comparisons to the KR test, and shows why the LM test is superior. We develop a second test in the spirit of KR which overcomes this particular shortcoming, but in fact incurs a second shortcoming which renders it inferior even to the original KR test. We discuss this failure as well.

We thank Joris Pinkse and Sung Jae Jun for helpful conversations, and Zhenlin Yang and other participants at the Spatial Econometrics Association meeting for helpful comments, but neither they, nor the Economic Research Service and US Department of Agriculture are to blame for anything stated herein. Comments welcome at fyj@psu.edu (Coulson).

1. Introduction

The spatial error components (SEC) model (Kelejian and Robinson (1993, 1995) is now seen as a realistic alternative to spatial autoregressions when modeling spatially dependent processes. Somewhat akin to the usual random effects models, the spatial error component model proposes that spatial observations belong to groups, just as panel data observations belong to particular cross-sections. The distinction between the two is that the use of spatial error components does not restrict observations to belong to just one group and does not even necessarily restrict the weight that an observation has in each of its groups to be equal to one. Thus it is a quite flexible and interesting method for modeling spatial dependence in the residuals of a regression equation.

There are two tests that are primarily used for detecting SEC, as discussed in Anselin and Moreno (2003). The original test, due to Kelejian and Robinson (hereafter, KR), is based on the method of moments principles while that proposed by Anselin (2001) is derived from the Lagrange Multiplier (LM) principle[1]. In the next section we develop a different form of the LM test. This second method, although it produces numerically identical results to the test in Anselin (2001), nevertheless is useful for analyzing the differences between LM and KR, and why LM typically has superior performance to KR (Anselin and Moreno, 2003). We then develop a new version of the KR test which reduces the apparent difference between these two testing principles. We also develop a modification of the LM test that is in the spirit of the original KR test. We perform Monte Carlo simulations of all of three tests. Paradoxically, the new version of KR does worse than either the LM or the original KR test, and we discuss the reasons for this as well.

2. Model formulation and testing

The generic form of the (linear) SEC is written as

---

[1] They discuss a third test also noted by Kelejian and Robinson. But this test is not based on the specification of the W matrix below and therefore is not directly comparable to those discussed in below.

$$y = X\beta + e \quad (1)$$
$$e = Wu + v \quad (2)$$

where n is the sample size, y, u, v and e are nx1 vectors, X is an nxk matrix and β is a kx1 vector of regression coefficients. W is an nxn matrix of known constants that model the spatial correlations of the error term, e. The interpretations are of the usual sort: y is a dependent variable thought to be modeled as a linear function of the k covariates in X. The parameters of interest are in the vector β. The vector e is composed of unobserved error terms, modeled as the sum of two components, u and v. We assume

$$E(u) = E(v) = 0$$

$$Cov(u) = \sigma_u^2 I$$

$$Cov(v) = \sigma_v^2 I$$

$$\mathrm{cov}(u_i, v_j) = 0 \quad \forall\, i, j$$

With these assumptions the regression error term has unconditional expectation equal to zero and its covariance matrix is

$$Cov(e) = \Omega = \sigma_u^2 P + \sigma_v^2 I \qquad (3)$$

where

$$P = WW' \qquad (4)$$

Thus, because W has non-zero off-diagonal elements, so does P, and OLS estimation of (1) is inefficient if the variance of u is non-zero[2]. The test for the absence of spatial error components is therefore based on the null hypothesis

$$H_0 : \sigma_u^2 = 0$$

Two testing principles have been suggested.

---

[2] We eschew, for present purposes, the possibility of endogenous regressors, including spatial lags of the dependent variable.

Kelejian and Robinson (1993, 1997) suggest the following test based on generalized method of moments. Since

$$E(e_i^2) = \sigma_u^2 P_{ii} + \sigma_v^2 \qquad (5)$$

where P_{ii} is the ith diagonal element of P, an OLS regression of fitted squared residuals from (1) on an intercept term and the elements of P_{ii} would generate estimates of both the variance of v, the intercept term, and the variance of u as the slope coefficient. The test for spatial error components is then simply the t-test for the slope coefficient.

Anselin proposes a test for spatial error components using the Lagrange Multiplier principle. The LM test (Godfrey (1988)) is based on the idea that the score of the likelihood function evaluated under the null is equal to zero when the null hypothesis is true, so that a χ^2 test based on the square of the score divided by the appropriate element of the information matrix (since this is the variance of the score) can be constructed. The use of the normal likelihood function requires the assumption of normality of the error term.

More formally, the likelihood function for the parameters of (1) and (3) is

$$L = K - \frac{n}{2}\ln\left|\Omega\ \right| - \frac{1}{2}e'\Omega e \qquad (6)$$

The LM test is generically of the form

$$LM = S'\Im^{-1}S' \qquad (7)$$

where S is the score (or gradient) of the likelihood and $\Im$ is its information matrix. Here, as is often the case, the information matrix is block diagonal between the parameters of the mean and of the variance so we need only concentrate on the latter. Anselin (2001) shows

$$S = -\frac{1}{2}trP + \frac{1}{2\sigma_v^2}e'Pe$$
$$\Im_{22}^{-1} = 2trPP - \frac{(trP)^2}{n} \qquad (8)$$

where the latter term is the (2,2) element of $\Im^{-1}$. Substituting (8) into (4) and rearranging, the test-statistic can be written as

$$LM = \frac{\left[\frac{e'Pe}{\hat{\sigma}^2} - trP\right]^2}{2(trPP) - \frac{(trP)^2}{n}} \qquad (9)$$

where $\hat{\sigma}^2$ is the estimate of the variance of the error term from the OLS regression (1) under the assumption of no spatial error components (i.e $\sigma_u^2 = 0$). LM is distributed as a chi-squared with one degree of freedom under that restriction.

This derivation is analogous to the derivation of the test for (uncorrelated) random effects given in Breusch and Pagan (1980) and so it is instructive to consider an alternative deriviation of LM statistics discussed in that paper, which indeed applies to any hypothesis on the nonsphericality of the error covariance matrix. From the likelihood (6) note that the score subvector for the variance components is

$$S = -\frac{1}{2}A'(\Omega^{-1} \otimes \Omega^{-1})vec(\Omega - ee')$$

where

$$A' = \frac{\partial vec\Omega}{\partial(\sigma_u^2\ \sigma_v^2)} = (vec(I)\ vec(P))$$

and

$$\Im = \frac{1}{2}A'(\Omega^{-1}\otimes\Omega^{-1})A$$

Under the null hypothesis, $\Omega = \sigma_v^2 I$, the LM test statistic becomes

$$\frac{1}{2\hat{\sigma}^4}[vec(\hat{\sigma}^2 I - ee')'A(A'A)^{-1}A'vec(\hat{\sigma}^2 I - ee') =$$

$$\frac{1}{2}[vec(I - \frac{ee'}{\hat{\sigma}^2})'A(A'A)^{-1}A'vec(I - \frac{ee'}{\hat{\sigma}^2})$$

The latter expression is easily seen to be one-half of the explained sum of squares of a regression of $vec(I - \frac{ee'}{\hat{\sigma}^2})$ on vec(I) and vec(P), and this quantity will be numerically identical to the expression in (9). We call this test LM2.

The comparison of this version of the LM test with the KR test regression is instructive. The major difference is that KR uses only the diagonal elements of ee’ as observations, while the LM test uses all of its n^2 elements. This may explain why Anselin and Moreno observed a relatively poor performance of the KR test, and in turn this suggests a natural variation on the KR test, which is of course to use all of the squares and cross-products of the OLS residuals in the KR regression. This is ostensibly only using more moments in the GMM regression. The analogous test-statistic would still be the t-ratio on the coefficient of vec(P). We call this test KR2.

The test regression for LM2 is therefore, under the null

$$vec(I - \frac{ee'}{\hat{\sigma}_u^2}) = \beta vec(I) + \gamma vec(P) \quad (10)$$

while that for KM2 is

$$vec(ee') = \sigma_v^2 \, vec(I) + \sigma_u^2 \, vec(P) \quad (11)$$

As noted, we use the asymptotic t-ratio of the coefficient of vec (P) in 11 and half of the explained sum of squares in (10) as test statistics. Finally, for purposes of comparison discussed below we also construct a test based on the LM2 test, but only using the diagonal elements; that is, we take half of the explained sum of squares from (10) in which only the diagonal elements are used. This test is not expected to have good properties, as it is not asymptotically distributed as a χ^2.

3. Monte Carlo simulations

The Monte Carlo simulations were performed as follows. We generated synthetic samples from a square grid, using sample sizes N=64, 100, 225 and 1024. The W matrices were initially constructed on the basis of first-order "queen" contiguity. We specify the underlying regression model (1) as

$$y = X + e$$

simulating normally distributed random variates for X, u, and v, where e is specified as in (2). The size of the variance of u is set at one, and we allow the variance of v to be 0, 1, 4, or 8, for each sample size. Thus there are 16 cases to consider. We perform 1000 repetitions of each of the 12 simulations, estimating

$$y = \beta_0 + \beta_1 X + e$$

in each case and constructing the four test statistics, KR, KR2, LM2 and LMD. As a safeguard we also constructed LM1; this was identical to LM2 in all cases, which verifies our programming and the derivation of the LM2 test statistic.

The results of these simulations are contained in Table 1. The first line gives the case where the variance of v is zero-- that is, the null hypothesis is true. Given that variances must be positive it is appropriate to use a one-sided rejection region in the KR and KR2 tests, so we use the 5% critical value of 1.645 in those cases. This is not a trivial concern; Anselin and Moreno (1993) and Carriazo (2007) note the problem of estimating negative values of the coefficient of P_{ii} which Anselin and Moreno attribute to "uncooperative data". In the case of the LM test, the χ^2 is perforce one-sided, since the explained sum of squares is also a positive number. We must nonetheless guard against the possibility that rejection comes from extreme negative values of both β and γ, but an examination of the simulations suggests that this is not a problem.

Consider first the two established tests, KR and LM2. Using the normal distribution, under the null (in the table, σu2σ_u^2=0) the KR test performs well, with rejection rates moderately close to the nominal rate of .05 (although surprisingly this dips below .04 for the largest sample size of 1024). The LM2 test distribution is somewhat less satisfactory; in all cases the rejection rate is lower (.02 - .04) than that suggested by the nominal rate. This is quite similar to the test performance found by Anselin and Moreno. When the lognormal distribution is used in the null model, the rejection rates for KR rise. The LM test also exhibit higher rejection rate, and indeed rise above .05 at the largest sample size. Again, this is quite similar to results displayed in Anselin and Moreno.

The relative advantage of the LM test is clearly seen in those models in which the null is false. For all values of σ_u^2 and all sample sizes, the rejection rate of LM is higher, and often substantively so, than that of KR. This relative advantage is seen most clearly for high values of σ_u^2.

When the errors are normally distributed, for all values of n greater than 64, when σ_u^2=8 the rejection rate for the LM test is universally above .9 whereas that of the KR test is never above .6. When the errors follow the lognormal distribution, these two tests behave as previously described for all values of n. It would seem that the use of the off-diagonal elements in the LM test provides information that substantially increases the power of the test to detect spatial error components.

As noted, the KR2 test attempts to rectify this defect of the KR test, but fails miserably. As Table 1 shows, when the errors terms follow the normal distribution, the test hardly ever rejects, whether the null is true or false, even at high values of σ_u^2. When the errors are lognormally distributed, the test does reject, but its power is universally lower than that of the original KR test.

Why should this be the case? Note that from (10) and (11) that we can transform (10) into (11) through a transformation of the dependent variable. Doing so, and equating the two coefficient vectors yields the fact that in the LM2 regression

$$\beta = 0$$

and

$$\gamma = \frac{\sigma_v^2}{\sigma_u^2}$$

thus under the null, both coefficients are equal to zero, which in turn demonstrates why a test-statistic that is based on the explained sum of squares- in effect a test on both coefficients- is appropriate. Thus when KR2 uses just the coefficient of vec(P) to assess the validity of the null, the t-test can be contaminated when there is collinearity between vec(I) and vec(P). When queen first-order contiguity is used to construct W, Table 2 shows that this correlation is extremely strong, and therefore using a t-test on a single coefficient can be vastly different from a joint test on both coefficients. Indeed, since the KR2 and LM2 regressions are so similar, it is unsurprising to note that in the LM2 output there are almost no large t-ratios for the coefficient of vec(P).

Note that this does not obtain with the ordinary KR test. This is because when only the diagonal elements are used in the test regression, the first regressor is now an intercept term. Therefore the two regressors are uncorrelated, although now KR suffers from a lack of information and so does poorly compared to the LM tests.

This intuition is verified in Table 3, where the W matrix is constructed according to rook contiguity. In table 2 we display the correlation of the two regressors under this formulation and it can be seen that this is higher under rook contiguity. Thus by the reasoning of the previous paragraph, we would expect the KR2 test to perform even worse. Although it is hard to perform worse than the KR2 test under normally distributed errors, there is a substantial deterioration in the power of the KR2 test when the errors are lognormally distributed. Thus the multicollinearity does indeed matter.

As an experiment, we constructed an LM test based on a regression similar to (10) but only using the diagonal elements (labeled LMD). The results were unremarkable. Using normally distributed errors the performance under the null was a rejection rate far below the expected .05, and otherwise it did somewhat worse than KR (but obviously better than KR2). With lognormal errors, the null rejection was far too high. When the null was false this LMD, actually did better than KR, but still did not have as much power as LM.

Given the differences in test performance in the normal and lognormal cases and the improved performance, we experimented with distributions with fatter tails than the lognormal. The errors were generated using the student-t with one degree of freedom, and again with the Cauchy distribution. The results, presented in Table 4, were somewhat erratic. The LMD test statistics were (as in the lognormal case) very high, which give them excellent power and poor size properties. The KR test was correspondingly low, with corresponding changes in the power and size. One result is interesting. When the errors are Student-t and the sample is large enough, the

KR2 rejects more often than KR. This is the only occurrence of this phenomenon. The most interesting result from this experiment is the remarkably good performance of LM. Even in the cases of these extreme distributions, when the null is true the LM test had rejection rates tolerably close to .05 (particularly when the sample was small) and rejection rates under the alternative which were greater than .35 and increased as the sample size increased.

4. Conclusions

The LM test is better than the KR test in assessing the likelihood of spatial error components. This is for two reasons. One is that it uses more information to construct the test. But when the information sets are equalized, the KR test does even worse, because it does not test all of the implications of the error components model. The LM test does, and performs best under nearly all the scenarios we consider here.

References

Anselin, L., 2001. Rao's socre test in spatial econometrics. J. Stat. Plan.Inference 97, 113-139.

Anselin, L., Moreno, R., 2003. Properties of tests for spatial error components. Regional Science and Urban Economics. 33, 595-618.

Breusch, T.S and Pagan, A.R., 1980. The Lagrange Multiplier Test and its Applications to Model Specification in Econometrics. The Review of Economic Studies, 47 (1) , pp.239-253

Carriazo, F., 2007. Measuring Urban Amenities and Disamenities: A Spatial Hedonic Analysis in Bogota, Colombia. Ph.D. Dissertation, The Pennsylvania State University, unpublished.

Godfrey (1988) *Misspecification Tests in Econometrics* Cambridge, UK: Cambridge University Press

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 Regional Science. 72, 297-312.

Kelejian H.H., Robinson,D.P., 1995. Spatial correlation: a suggested alternative to the autoregressive model. In: Anselin, L., Florax, R.J. (Eds), New Directions in Spatial Econometrics. Springer-Verlag, Berlin, pp. 75-95

		Normal				Lognormal			
n	σ_u^2	KR	KR2	LM	LMD	KR	KR2	LM	LMD
64	0	0.050	0.000	0.019	0.001	0.069	0.000	0.033	0.391
	1	0.113	0.000	0.128	0.023	0.100	0.001	0.212	0.341
	4	0.225	0.000	0.484	0.112	0.252	0.061	0.930	0.348
	8	0.311	0.000	0.727	0.208	0.248	0.089	0.990	0.413
100	0	0.059	0.000	0.032	0.003	0.065	0.000	0.045	0.469
	1	0.146	0.000	0.201	0.042	0.093	0.000	0.253	0.418
	4	0.146	0.000	0.712	0.157	0.093	0.092	1.000	0.801
	8	0.375	0.001	0.909	0.271	0.193	0.109	1.000	0.609
225	0	0.046	0.000	0.035	0.000	0.051	0.000	0.064	0.570
	1	0.171	0.000	0.367	0.050	0.084	0.000	0.411	0.496
	4	0.372	0.000	0.976	0.229	0.281	0.092	1.000	0.725
	8	0.502	0.000	1.000	0.390	0.228	0.163	1.000	0.833
1024	0	0.038	0.000	0.030	0.002	0.073	0.000	0.077	0.639
	1	0.199	0.000	0.924	0.048	0.297	0.146	1.000	0.891
	4	0.516	0.000	1.000	0.350	0.328	0.092	1.000	0.801
	8	0.713	0.000	1.000	0.593	0.080	0.000	0.840	0.546

Table 1: Rejection frequencies of tests of H_0: $\sigma^2_v=0$, under normal and lognormal error distributions, and queen contiguity

N	Queen	Rook
64	0.567	0.6796
100	0.5609	0.6774
225	0.5539	0.6742
1000	0.5479	0.6703

Table 2: Correlation Coefficient between Vec(I) and Vec(P) for queen and rook first order contiguity matrices.

		Normal				Lognormal			
n	σu2	KR	KR2	LM	LMD	KR	KR2	LM	LMD
64	0	0.057	0.000	0.044	0.004	0.061	0.000	0.045	0.375
	1	0.118	0.000	0.179	0.041	0.087	0.000	0.286	0.274
	4	0.210	0.000	0.745	0.120	0.087	0.000	0.286	0.274
	8	0.301	0.000	0.928	0.200	0.150	0.000	1.000	0.360
100	0	0.056	0.000	0.047	0.002	0.070	0.000	0.056	0.455
	1	0.131	0.000	0.316	0.048	0.066	0.000	0.364	0.344
	4	0.131	0.000	0.908	0.123	0.140	0.002	1.000	0.503
	8	0.275	0.000	0.995	0.176	0.116	0.006	1.000	0.590
225	0	0.050	0.000	0.062	0.000	0.066	0.000	0.070	0.539
	1	0.153	0.000	0.612	0.051	0.081	0.000	0.587	0.416
	4	0.305	0.000	1.000	0.180	0.152	0.000	1.000	0.672
	8	0.383	0.000	1.000	0.273	0.133	0.001	1.000	0.717
1024	0	0.038	0.000	0.048	0.002	0.065	0.000	0.062	0.627
	1	0.191	0.000	0.998	0.058	0.099	0.000	0.965	0.535
	4	0.417	0.000	1.000	0.257	0.177	0.000	1.000	0.806
	8	0.546	0.000	1.000	0.417	0.179	0.001	1.000	0.902

Table 3: Rejection frequencies of tests of H_0: $\sigma^2_v=0$, under normal and lognormal error distributions, and rook contiguity

		KR				**KR2**				**LM**				**LMD**			
		Student		**Cauchy**		**Student**		**Cauchy**		**Student**		**Cauchy**		**Student**		**Cauchy**	
n	σu2	Queen	Rook	Queen	Rook	Queen	Rook	Queen	Rook	Queen	Rook	Queen	Rook	Queen	Rook	Queen	Rook
64	0	0.064	0.061	0.072	0.083	0.000	0.000	0.000	0.000	0.048	0.045	0.058	0.073	0.694	0.375	0.694	0.692
	1	0.131	0.112	0.123	0.100	0.008	0.000	0.010	0.000	0.356	0.433	0.362	0.397	0.602	0.558	0.573	0.577
100	0	0.047	0.048	0.056	0.045	0.000	0.000	0.000	0.000	0.064	0.072	0.068	0.053	0.759	0.749	0.762	0.764
	1	0.098	0.363	0.077	0.073	0.015	0.382	0.012	0.000	0.408	0.455	0.397	0.463	0.731	0.698	0.694	0.689
225	0	0.035	0.034	0.039	0.039	0.000	0.000	0.000	0.000	0.130	0.090	0.137	0.081	0.859	0.866	0.865	0.864
	1	0.071	0.283	0.077	0.062	0.024	0.421	0.017	0.000	0.521	0.596	0.555	0.570	0.835	0.777	0.842	0.791
1024	0	0.092	0.104	0.105	0.120	0.000	0.000	0.000	0.000	0.154	0.132	0.151	0.140	0.934	0.944	0.941	0.939
	1	0.143	0.206	0.142	0.116	0.022	0.389	0.022	0.000	0.619	0.670	0.646	0.677	0.921	0.931	0.914	0.914

Table 4: Rejection frequencies of tests of H_0: $\boldsymbol{\sigma^2_v=0}$, under Student(1) and Cauchy, queen and rook contiguity.