Bootstrap inference in spatial econometrics: the J test

Peter Burridge
Department of Economics and Related Studies, University of York
Bernard Fingleton
Department of Economics, University of Strathclyde
This version is for the SEA conference Barcelona, July 2009,
and lacks figures that will be presented at the meeting.

May 28, 2009

Abstract

Kelejian (2008) introduces a J-type test for the situation in which a null linear regression model, Model_0, is to be tested against one or more rival non-nested alternatives, Model_1, ..,Model_g, where typically the competing models possess endogenous spatial lags and spatially autoregressive error processes. The original J test of Davidson and MacKinnon (1981), in which $g = 1$, is constructed by augmenting the right-hand side of Model_0 with a vector of fitted values from Model_1. Under appropriate conditions the conventional t - ratio on the estimated coefficient on these fitted values converges in distribution to $N(0, 1)$ when Model_0 is true. In simulations of the finite-sample distribution of this t ratio, it is common to observe that Model_0 is over-rejected, for the reasons studied in detail in Davidson and MacKinnon (2002a). If the tests are to be relied upon, any such size-distortion must be controlled either by modifying the test statistic or by correcting the reference distribution, for example by the use of resampling. Concentrating on the case, $g = 1$, in this paper we examine the finite sample properties of a spatial J statistic that is asymptotically χ^2_2 under the null, and an alternative version that is conjectured to be approximately χ^2_1 both introduced by Kelejian (2008). We demonstrate numerically that the tests are excessively liberal in some leading cases using the relevant Chi-square asymptotic approximations, and explore how far this may be corrected using a simple bootstrap resampling method. In our experiments we vary the degree of correlation between the regressors of rival models, the strength of lag and error dependence, the spatial weight matrices, and the choice of instruments. We find the asymptotic tests often perform very well, but they are either very liberal or lack power in some parts of the parameter space. The bootstrap approach, whilst not perfect, is demonstrated to be clearly superior to reliance on the asymptotic χ^2_1 or χ^2_2 critical values in most cases.

1 Introduction

In Kelejian (2008), spatial extensions of the J-test of Davidson and MacKinnon (1981) are introduced for testing a null model, Model_0 against g alternative models in the situation in which the $g+1$ models are non-nested. This testing problem is an important and long-standing issue in spatial econometrics, rendered difficult by the widely acknowledged need to capture in such models patterns of spatial interaction that are not directly observed. One of Kelejian's test statistics has $2g$ degrees of freedom, and he shows that under suitable conditions, it has an asymptotic χ^2_{2g} distribution under the null hypothesis that Model_0 is true. A second form of the test has g degrees of freedom, and can be conjectured to have an asymptotic χ^2_g distribution under the same or similar conditions yet to be established. Experience in the non-spatial setting, such as in the studies by Godfrey (1998) and by Davidson and MacKinnon (2002a), suggests on the one hand that use of critical values based on large-sample approximations may make J-type tests too liberal, and on the other that this problem may be cured by the use of resampling. Further, to implement the tests in the form proposed by Kelejian, which requires intrumental variable estimation of the competing models, users will need to choose which instruments to use, and also decide whether to use a test with g or $2g$ degrees of freedom. We therefore explore these issues. In the following we focus on the case of a single alternative, $g = 1$, and we will refer to the spatial J tests implemented using χ^2_1 or χ^2_2 critical values as the *asymptotic* tests, and tests using $p-$ values obtained by a simple parametric bootstrap, described below, as *simple bootstrap* tests.

We find that in some leading cases the empirical significance level of the asymptotic tests is affected by the choice of instruments, while the simple bootstrap tests are not affected to the same extent. Further, there appears to be no general loss of power in using the simple bootstrap beyond that attributable to bringing the significance level down to its nominal value. However, the asymptotic tests can be too liberal to be relied upon in applications, while the simple bootstrap corrects the empirical significance level of the tests for most but not all regions of the parameter space. Whilst it is possible in many settings to improve upon the performance of the simple bootstrap by implementing a fast double bootstrap of the kind studied by Davidson and MacKinnon (2002b) we find that the extreme cases remain problematic in the present application. However the bootstrap is implemented, its feasibility depends on the properties of the parameter estimates obtained, in particular when the alternative model is true. This sensitivity, which is not shared by the asymptotic tests, may limit the value of the parametric bootstrap in this context, and suggests a line of enquiry for future work.

The next section describes the spatial models between which the tests are designed to discriminate and defines the test statistics, while in Section 3 previous related work is reviewed and in Section 4 the bootstrap is introduced. Section 5 presents our experimental evidence, and Section 6 concludes.

2 The models and the J-type tests

2.1 Null and alternative models

Following Kelejian (2008), we use the SARAR(1,1) model set-up. Under the null hypothesis, Model_0 is true:

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}_0\boldsymbol{\beta}_0+\boldsymbol{\lambda}_0\mathbf{W}_0\mathbf{Y}+\mathbf{u}_0 \\ \mathbf{u}_0 &= \rho_0\mathbf{M}_0\mathbf{u}_0+\mathbf{v}_0. \end{aligned} \tag{1}$$

Here, the $n \times k_0$ matrix of observations of exogenous variables, $\mathbf{X}_0$, and the $n \times 1$ vector of observations of the dependent variable, $\mathbf{Y}$, are each measured without error, the $n \times n$ matrices of fixed weights, $\mathbf{W}_0$ and $\mathbf{M}_0$ are assumed known, and the unobserved shock vector, $\mathbf{v}_0 \sim IID(0, \sigma_0^2\mathbf{I}_n)$ independent of the regressors, $\mathbf{X}_0$. The parameters to be estimated are the slope coefficients, $\boldsymbol{\beta}_0$, the spatial lag and error coefficients, $\boldsymbol{\lambda}_0$, and ρ_0, and the variance, σ_0^2. Under the alternative, the data are generated by a similar structure, Model_1:

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}_1\boldsymbol{\beta}_1+\boldsymbol{\lambda}_1\mathbf{W}_1\mathbf{Y}+\mathbf{u}_1 \\ \mathbf{u}_1 &= \rho_1\mathbf{M}_1\mathbf{u}_1+\mathbf{v}_1. \end{aligned} \tag{2}$$

Kelejian's detailed assumptions about this specification are reproduced in the Appendix. In fact, Kelejian allows for some finite number, $g \geq 1$, of non-nested alternatives of the same type.

2.2 Kelejian's J-tests for the case g=1

Because of the number of modelling choices that must be made in order to implement the tests, it is essential to describe them in some detail here. We follow Kelejian, but comment on some features of the estimation problem on the way.

Step 1.

Define $\mathbf{Z}_0 = [\mathbf{X}_{01} \vdots \mathbf{X}_{02} \vdots \mathbf{W}_0\mathbf{Y}]$ and $\boldsymbol{\gamma}_0 = [\boldsymbol{\beta}_0^{'}, \lambda_0]'$ to write the null model as

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}_0\boldsymbol{\beta}_0+\lambda_0\mathbf{W}_0\mathbf{Y}+\mathbf{u}_0 \\ &= \mathbf{Z}_0\boldsymbol{\gamma}_0 + \mathbf{u}_0. \end{aligned} \tag{3}$$

Since ordinary least squares (OLS) would be inconsistent applied to (3), we define the matrix,

$$\mathbf{L}_{0,r} = [\mathbf{X}_{01} \vdots \mathbf{X}_{02} \vdots \mathbf{W}_0\mathbf{X}_{02} .. \vdots \mathbf{W}_0^r\mathbf{X}_{02}]$$

for some small integer r, and construct a matrix of instruments,

$$\mathbf{H}_{0,r} = [\mathbf{L}_{0,r} \vdots \mathbf{M}_0\mathbf{L}_{0,r}]_{LI}$$

in which the subscript, LI denotes a spanning set of linearly independent columns from $\mathbf{H}_{0,r}$. This instrument set yields the projection matrix,

$$\mathbf{P}_{0,r} = \mathbf{H}_{0,r}(\mathbf{H}'_{0,r}\mathbf{H}_{0,r})^{-1}\mathbf{H}'_{0,r}$$

leading to the instrumental variable (IV) estimator,

$$\hat{\boldsymbol{\gamma}}_{0r} = [\mathbf{Z}'_0\mathbf{P}_{0,r}\mathbf{Z}_0]^{-1}\mathbf{Z}'_0\mathbf{P}_{0,r}\mathbf{Y}.$$

Similarly, for Model$_1$ we have $\mathbf{Z}_1 = [\mathbf{X}_{11}\vdots\mathbf{X}_{12}\vdots\mathbf{W}_1\mathbf{Y}]$, $\boldsymbol{\gamma}_1 = [\boldsymbol{\beta}'_1, \lambda_1]'$,

$$\mathbf{L}_{1,r} = [\mathbf{X}_{11}\vdots\mathbf{X}_{12}\vdots\mathbf{W}_1\mathbf{X}_{12}..\vdots\mathbf{W}_1^r\mathbf{X}_{12}]$$

and $\mathbf{H}_{1,r} = [\mathbf{L}_{1,r}\vdots\mathbf{M}_1\mathbf{L}_{1,r}]_{LI}$, giving the projector, $\mathbf{P}_{1,r}$, and IV estimator,

$$\hat{\boldsymbol{\gamma}}_{1r} = [\mathbf{Z}'_1\mathbf{P}_{1,r}\mathbf{Z}_1]^{-1}\mathbf{Z}'_1\mathbf{P}_{1,r}\mathbf{Y}.$$

For convenience, we also define the hybrid matrices

$$\mathbf{L}_{01,r} = [[\mathbf{X}_0\vdots\mathbf{X}_1]\vdots\mathbf{W}_0[\mathbf{X}_0\vdots\mathbf{X}_1]..\vdots\mathbf{W}_0^r[\mathbf{X}_0\vdots\mathbf{X}_1]]_{LI}$$

$$\mathbf{H}_{01,r} = [\mathbf{L}_{01,r}\vdots\mathbf{M}_0\mathbf{L}_{01,r}]_{LI}$$

As always with IV estimators, there will in general be other instrument choices available; Kelejian comments that "typically one would take $r \leq 2$", so we explore the relationship, if any, between r and J-test size and power in the numerical experiments below.

Step 2.

Define the vector of residuals from the null model,

$$\hat{\mathbf{u}}_0 = \mathbf{Y} - \mathbf{Z}_0\hat{\boldsymbol{\gamma}}_0$$

and use this to estimate ρ_0 in (1) by some consistent method, such as the GMM procedure of Kelejian and Prucha (1999), outlined below.

We implement Kelejian and Prucha's non-linear GMM method for estimating ρ from the residuals, $\hat{\mathbf{u}}_0$, in preference to QML because of the numerical difficulty of calculating accurately the eigenvalues of even moderately sized matrices, as discussed by those authors. It is defined as follows. In (1) the disturbances satisfy the following moment conditions:

$$\begin{aligned} E\{n^{-1}\mathbf{v}'_0\mathbf{v}_0\} &= \sigma^2 \\ E\{n^{-1}\mathbf{v}'_0\mathbf{M}'_0\mathbf{M}_0\mathbf{v}_0\} &= \sigma^2 n^{-1}Tr\{\mathbf{M}'_0\mathbf{M}_0\} \\ E\{n^{-1}\mathbf{v}'_0\mathbf{M}'_0\mathbf{v}_0\} &= \sigma^2 n^{-1}Tr\{\mathbf{M}'_0) = 0 \end{aligned}$$

which are easily converted into statements about the moments of $\mathbf{u}_0$ using the fact that

$$\mathbf{v}_0 = \mathbf{u}_0 - \rho_0\mathbf{M}_0\mathbf{u}_0.$$

Thus we have

$$\begin{aligned} E\{n^{-1}\mathbf{u}_0'(\mathbf{I}-\rho_0\mathbf{M}_0')(\mathbf{I}-\rho_0\mathbf{M_0})\mathbf{u}_0\} &= \sigma^2 \\ E\{n^{-1}\mathbf{u}_0'(\mathbf{I}-\rho_0\mathbf{M}_0')\mathbf{M_0'}\mathbf{M_0}(\mathbf{I}-\rho_0\mathbf{M_0})\mathbf{u}_0\} &= \sigma^2 n^{-1} Tr\{\mathbf{M}_0'\mathbf{M}_0\} \\ E\{n^{-1}\mathbf{u}_0'(\mathbf{I}-\rho_0\mathbf{M}_0')\mathbf{M_0'}(\mathbf{I}-\rho_0\mathbf{M_0})\mathbf{u}_0\} &= 0 \end{aligned}$$

which is a set of three non-linear equations in σ^2, ρ_0 and ρ_0^2. Re-arranging these into a more convenient form, we have, imitating Kelejian and Prucha's notation,

$$E\{\mathbf{G}\begin{bmatrix}\rho\\ \rho^2\\ \sigma^2\end{bmatrix} - \mathbf{g}\} = \mathbf{0}$$

where

$$\begin{aligned} \mathbf{G} &= \begin{bmatrix} 2\mathbf{u}_0'\mathbf{M_0}\mathbf{u}_0 & -\mathbf{u}_0'\mathbf{M}_0'\mathbf{M_0}\mathbf{u}_0 & n \\ 2\mathbf{u}_0'\mathbf{M}_0'\mathbf{M_0^2}\mathbf{u}_0 & -\mathbf{u}_0'[\mathbf{M}_0']^2\mathbf{M_0^2}\mathbf{u}_0 & Tr\{\mathbf{M}_0'\mathbf{M}_0\} \\ \mathbf{u}_0'[\mathbf{M}_0'\mathbf{M_0} + \mathbf{M}_0'\mathbf{M_0^2}]\mathbf{u}_0 & -\mathbf{u}_0'\mathbf{M}_0'\mathbf{M_0^2}\mathbf{u}_0 & 0 \end{bmatrix} \\ \mathbf{g} &= \begin{bmatrix} \mathbf{u}_0'\mathbf{u}_0 \\ \mathbf{u}_0'\mathbf{M}_0'\mathbf{M_0}\mathbf{u}_0 \\ \mathbf{u}_0'\mathbf{M_0}\mathbf{u}_0 \end{bmatrix} \end{aligned}$$

To obtain estimates $\hat{\rho}$ and $\hat{\sigma}^2$ we can therefore replace the expectations with sample averages and take

$$(\hat{\rho}, \hat{\sigma}^2) = \arg\min\{\mathbf{G}\begin{bmatrix}\hat{\rho}\\ \hat{\rho}^2\\ \hat{\sigma}^2\end{bmatrix} - \mathbf{g}\}'\{\mathbf{G}\begin{bmatrix}\hat{\rho}\\ \hat{\rho}^2\\ \hat{\sigma}^2\end{bmatrix} - \mathbf{g}\}. \tag{4}$$

Kelejian and Prucha show that this estimator is consistent under suitable conditions, and report experiments suggesting that its efficiency is similar to that of the computationally much more expensive QMLE. In the experiments we have used (4) with $\mathbf{u}_0$ replaced by the residual, $\hat{\mathbf{u}}_0$.

Defining the vector of residuals from the alternative,

$$\hat{\mathbf{u}}_1 = \mathbf{Y} - \mathbf{Z}_1\hat{\boldsymbol{\gamma}}_1$$

we estimate ρ in similar fashion to get $\hat{\rho}_1$, say.

Step 3.

Using $\hat{\rho}_0$ from Step 2, construct the spatially lag-transformed regression

$$\begin{aligned} (I - \hat{\rho}_0\mathbf{M}_0)Y &= (I - \hat{\rho}_0\mathbf{M}_0)(\mathbf{Z}_0\boldsymbol{\gamma}_0 + \mathbf{u}_0) \qquad (5) \\ Y^*(\hat{\rho}_0) &= \mathbf{Z}_0^*(\hat{\rho}_0)\boldsymbol{\gamma}_0 + \mathbf{v}^*(\hat{\rho}_0) \quad \text{say} \end{aligned}$$

and estimate this equation by IV using the same instruments as before, $\mathbf{H}_0$; the result is the generalised spatial 2SLS procedure suggested in Kelejian and Prucha (1998) that yields, say,

$$Y^*(\hat{\rho}_0) = \mathbf{Z}_0^*(\hat{\rho}_0)\hat{\boldsymbol{\gamma}}_0(\hat{\rho}_0) + \hat{\mathbf{v}}^*(\hat{\rho}_0). \tag{6}$$

Use the residual vector, $\hat{\mathbf{v}}^*(\hat{\rho}_0)$, to estimate the variance of the shocks, $\hat{\sigma}_0^2 = \hat{\mathbf{v}}^*(\hat{\rho}_0)'\hat{\mathbf{v}}^*(\hat{\rho}_0)/n$.

Similarly, using $\hat{\rho}_1$ from Step 2, construct the alternative spatially lag-transformed regression

$$\begin{aligned}(I - \hat{\rho}_1\mathbf{M}_1)Y &= (I - \hat{\rho}_1\mathbf{M}_1)(\mathbf{Z}_1\boldsymbol{\gamma}_1 + \mathbf{u}_1) \\ \mathbf{Y}^*(\hat{\rho}_1) &= \mathbf{Z}_1^*(\hat{\rho}_1)\boldsymbol{\gamma}_1 + \mathbf{v}^*(\hat{\rho}_1) \quad \text{say}\end{aligned} \tag{7}$$

and estimate it by IV using the instruments, $\mathbf{H}_1$ to obtain

$$\mathbf{Y}^*(\hat{\rho}_1) = \mathbf{Z}_1^*(\hat{\rho}_1)\hat{\boldsymbol{\gamma}}_1(\hat{\rho}_1) + \hat{\mathbf{v}}^*(\hat{\rho}_1). \tag{8}$$

Let $\hat{\mathbf{Y}}^*(\hat{\rho}_1)$ denote the fitted value from (8). At this point we are in possession of an approximation to the forecast value of $(I - \rho_1\mathbf{M}_1)\mathbf{Y}$ obtained from the alternative model. We can now augment the RHS of (5) to generate a test of the hypothesis that Model_0 is true. Kelejian defines two tests:

Step 4a.(conjectured χ_1^2 version)

Using the fitted value from (8), set up the augmented equation

$$\begin{aligned}\mathbf{Y}^*(\hat{\rho}_0) &= \mathbf{Z}_0^*(\hat{\rho}_0)\boldsymbol{\gamma}_0 + \hat{\mathbf{Y}}^*(\hat{\rho}_1)\delta + \mathbf{v}^*(\hat{\rho}_0) \\ &= \mathbf{Z}^{**}\boldsymbol{\gamma}^{**} + \mathbf{v}^{**}\end{aligned} \tag{9}$$

and the augmented matrix of instruments

$$\mathbf{H}_r^{**} = [\mathbf{H}_{0,r} \vdots \mathbf{H}_{01,r}]_{LI} \tag{10}$$

with projection matrix $\mathbf{P}_r^{**}$, say, obtaining the IV estimator

$$\hat{\boldsymbol{\gamma}}^{**} = (\mathbf{Z}'^{**}\mathbf{P}_r^{**}\mathbf{Z}^{**})^{-1}\mathbf{Z}'^{**}\mathbf{P}_r^{**}\mathbf{Y}^*(\hat{\rho}_0)$$

with estimated asymptotic covariance matrix,

$$\hat{\mathbf{V}} = \hat{\sigma}_0^2(\mathbf{Z}'^{**}\mathbf{P}_r^{**}\mathbf{Z}^{**})^{-1}$$

which is used to extract a Wald test statistic for $\delta = 0$ in (9) in the usual way. That is, when Model_0 is true, letting l denote the number of elements in $\boldsymbol{\gamma}^{**}$, so that $\hat{\boldsymbol{\gamma}}^{**}(l)$ is the last estimated coefficient, and $\hat{\mathbf{V}}(l,l)$ its estimated variance, we conjecture that,

$$\frac{(\hat{\boldsymbol{\gamma}}^{**}(l))^2}{\hat{\mathbf{V}}(l,l)} \to^d \chi^2_{(1)}. \tag{11}$$

This remains a conjecture, but a proof that $\hat{\rho}_1$ from Step 2 converges to a constant under the null under appropriate conditions would be sufficient - see the first Remark under Kelejian's Equation 9. The specification of $\mathbf{H}_{01,r}$ in the instrument set, (10), is as given by Kelejian (2008); however, in our experiments we found that test power improved dramatically in some cases if $\mathbf{H}_{1,r}$ was used in place of $\mathbf{H}_{01,r}$ at this point. See discussion of "Case 2" below.

Step 4b.(χ^2_2 version)

Use the first step estimates, $\hat{\boldsymbol{\gamma}}_1$, to augment the RHS of (5) with both $\mathbf{Z}_1\hat{\boldsymbol{\gamma}}_1$ and $\mathbf{M}_1\mathbf{Z}_1\hat{\boldsymbol{\gamma}}_1$, in place of the single forecast value, $\hat{\mathbf{Y}}^*(\hat{\rho}_1)$, and augment the instrument vector as before. Following the same line of development leads to a statistic that is asymptotically $\chi^2_{(2)}$. That is, now estimate the equation

$$\begin{aligned} \mathbf{Y}^*(\hat{\rho}_0) &= \mathbf{Z}^*_0(\hat{\rho}_0)\boldsymbol{\gamma}_0 + \mathbf{Z}_1\hat{\boldsymbol{\gamma}}_1\delta_1 + \mathbf{M}_1\mathbf{Z}_1\hat{\boldsymbol{\gamma}}_1\delta_2 + \mathbf{v}^\dagger(\hat{\rho}_0) \\ &= \mathbf{Z}^\dagger\boldsymbol{\gamma}^\dagger + \mathbf{v}^\dagger \end{aligned} \tag{12}$$

using the instruments, $\mathbf{H}^{**}_r$, as above, obtaining the IV estimator

$$\hat{\boldsymbol{\gamma}}^\dagger = (\mathbf{Z}'^\dagger\mathbf{P}^{**}_r\mathbf{Z}^\dagger)^{-1}\mathbf{Z}'^\dagger\mathbf{P}^{**}_r\mathbf{Y}^*(\hat{\rho}_0)$$

with estimated asymptotic covariance matrix

$$\hat{\mathbf{V}}^\dagger = \hat{\sigma}^2_0(\mathbf{Z}'^\dagger\mathbf{P}^{**}_r\mathbf{Z}^\dagger)^{-1}.$$

Writing the matrix that selects the final two elements of $\boldsymbol{\gamma}^\dagger$ in the usual way as

$$\mathbf{R} = \begin{bmatrix} 0 & \cdots & 0 & 1 & 0 \\ 0 & \cdots & 0 & 0 & 1 \end{bmatrix}$$

the hypothesis to be tested becomes $H_0 : \mathbf{R}\boldsymbol{\gamma}^\dagger = \mathbf{0}$ and a Wald test statistic is

$$\hat{\boldsymbol{\gamma}}'^\dagger\mathbf{R}'[\mathbf{R}\hat{\mathbf{V}}^\dagger\mathbf{R}']^{-1}\mathbf{R}\hat{\boldsymbol{\gamma}}^\dagger \to^d \chi^2_{(2)}. \tag{13}$$

Kelejian proves the asymptotic Normality of $\hat{\boldsymbol{\gamma}}^\dagger$ that is sufficient, together with convergence of $\hat{\mathbf{V}}^\dagger$, for (13) while the alternative one degree-of-freedom form, (11), is introduced in a remark that also raises the question of the relative efficiency of the two tests. We give some evidence on this issue in the experiments reported below.

3 History of the problem

Anselin (1986) discusses the problem of discriminating between competing spatial interaction matrices in linear regression models, reporting some experimental results for the simple model,

$$Y_i = \alpha + \rho\sum_{j=1}^{j=n} w_{ij}Y_j + u_i \tag{14}$$

where the disturbance, u_i is either independent Normal, log-Normal, or spatially correlated and Normal, and in which the $n \times n$ weight matrix, $\mathbf{W} = \{w_{ij}\}$, takes one of three different forms, $\mathbf{W}_A, \mathbf{W}_B$ or $\mathbf{W}_C$ say, and $\rho = 0.25$ or 0.75.

Taking $\mathbf{W}_A$ as null hypothesis, he considers J-type statistics based on separately augmenting the right-hand side of (14) with the fitted value from the model with weights $\mathbf{W}_B$ or fitted value from the model with weights $\mathbf{W}_C$. This results in 12 comparisons in which each model plays the role of null or alternative. The sample sizes considered are quite small, $n = 24$, $n = 38$, the $\mathbf{W}$ matrices relating to planning districts in Columbus, Ohio. While his experimental design is thus quite narrow, nonetheless the most striking feature of the results reported by Anselin is that the tests he implements are severely over-sized in all cases. Evidently some means would have to be found to control significance levels were such tests to be used more widely, as was immediately recognised: Haining (1986 p.795) commenting, "... each researcher is probably going to have to carry out his own simulation experiments anyway in order to identify such things as critical values...". Haining also doubts whether the model with spatially autoregressive lag structure is necessarily the best way to fit the presumed pattern of autocorrelation in the disturbance. Suppose we write (14) in an obvious matrix notation as,

$$(\mathbf{I}-\rho\mathbf{W})\mathbf{Y} = \mathbf{1}\alpha + \mathbf{u} \tag{15}$$

and assume that $(\mathbf{I}-\rho\mathbf{W})$ is non-singular, so that

$$\begin{aligned} \mathbf{Y} &= (\mathbf{I}-\boldsymbol{\rho}\mathbf{W})^{-1}(\mathbf{1}\alpha + \mathbf{u}) \\ &= \boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ say.} \end{aligned} \tag{16}$$

Then we see that if $\mathbf{u} \sim N(\mathbf{0}, \sigma^2\mathbf{I}_n)$ as often assumed, the covariance matrix of $\boldsymbol{\varepsilon}$ in (16) is

$$\boldsymbol{\Omega} = E\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\} = \sigma^2\{(\mathbf{I}-\boldsymbol{\rho}\mathbf{W}')(\mathbf{I}-\boldsymbol{\rho}\mathbf{W})\}^{-1}$$

while the mean function of $\mathbf{Y}$ is obviously

$$\boldsymbol{\beta} = E\{\mathbf{Y}\} = (\mathbf{I}-\boldsymbol{\rho}\mathbf{W})^{-1}\mathbf{1}\alpha.$$

The point is that it is far from obvious that we should impose so much structure on $\boldsymbol{\beta}$ and $\boldsymbol{\Omega}$ at the outset, and that procedures designed to test one form of $\mathbf{W}$ against another may be applied to seriously misspecified models. Supposing, on the contrary, that there really were several competing models that differed solely in the form taken by $\mathbf{W}$, then Haining advocates use of an information criterion to choose among them as this would remove the need to interpret $2g!$ model comparisons; he reports no evidence that such an approach would lead to better model choices than the J-type tests however.

More recently, Leenders (2002) has placed the specification of $\mathbf{W}$ or $\mathbf{M}$ in a model of the form (1) at the heart of a discussion of the formulation of network autocorrelation models for social interaction. Differing influence mechanisms lead to distinct forms for $\mathbf{W}$ or $\mathbf{M}$ leading to a need for discrimination between these alternatives. As Leenders notes, there may be a material difference between the situation in which one competing model, Model_0, is being subjected to specification testing using the remaining models in the set to generate suitable diagnostics, and the situation in which all the competing models have the same status and we wish to select one. He advocates a non-nested hypothesis test for the former and the use of an information criterion for the latter, in neither case giving any evidence about the properties such procedures will have in finite samples. For the specification testing problem, he takes (1) as null and (2) as alternative, with $\rho_0 = \rho_1 = 0$ maintained apriori. To extend this to g alternatives he writes down an auxiliary regression of the form,

$$\mathbf{Y} = (1-\sum_{i=1}^{g} \alpha_i)(\lambda_0 \mathbf{W}_0 \mathbf{Y} + \mathbf{X}_0 \boldsymbol{\beta}_0) + \sum_{i=1}^{g} \alpha_i (\hat{\lambda}_i \mathbf{W}_i \mathbf{Y} + \mathbf{X}_i \hat{\boldsymbol{\beta}}_i) + \mathbf{e} \tag{17}$$

in which $\hat{\lambda}_i$ and $\hat{\boldsymbol{\beta}}_i$ $(i = 1, ..., g)$ are maximum likelihood estimates of the parameters of Models 1 to g and the test has null hypothesis $\alpha_i = 0$, $i = 1, ..., g$ (Leenders 2002, p.40). If we ignore $(1-\sum_{i=1}^{g} \alpha_i)$, the first factor on the right-hand side, then (17) is essentially the augmented regression that would replace (9) if it were maintained a priori that $\rho_j = 0$, $j = 0, ..., g$ and all parameters were estimated by maximum likelihood rather than via instrumental variables. For the model selection problem, the superficial appeal of using an information criterion is obvious. Once the criterion has been chosen, so the argument runs, the procedure delivers a unique best model. Unfortunately different information criteria will in general lead to the selection of different models, so that the uniqueness of the chosen model relies on the investigator first selecting which criterion to adopt. Leenders does not discuss such issues.

A recent formal implementation of a J-type test on a spatial model is Fingleton (2007) in which two non-nested wage equations are estimated. Fingleton's approach is intermediate in the sense that the J-tests he calculates by instrumental variable regression are applied to models without a spatially correlated disturbance but which have endogenous "potential"-type variables present, and he uses a bootstrap to evaluate the significance of the resulting statistics. In a parallel exercise he estimates and tests an artificial nesting model that does have a spatially correlated disturbance, and which is estimated by the 2sls/GMM methods of Kelejian and Prucha, again using resampling to generate reference distributions for the test statistics. There may be other studies implementing such tests on spatial models of which we are unaware, but it seems to us that formal testing of non-nested models should be more widespread than it is, and so it is desirable to investigate the issues that arise in the practical implementation of such tests in order that they may be used with confidence in empirical studies.

4 The bootstrap

The purpose of our resampling experiments is to establish whether bootstrap $p-$ values may be used to match nominal and empirical significance levels of the non-nested hypothesis tests described in Section 2, given that in many cases, as we shall demonstrate, such tests are either conservative or liberal if asymptotic

χ^2 critical values are used. In the case of the original J-tests of Davidson and MacKinnon (1981), a great deal is known about the properties of resampling-based $p-$ values and, in particular, Davidson and MacKinnon (2002a p.168) are able to conclude that "the bootstrap J test works extraordinarily well in almost every case in which a non-nested test is worth doing". This strong conclusion, relating to the closeness of the nominal and empirical sizes of the tests, rests heavily on the exact analysis of the J test statistic made possible by the linearity of the underlying family of models and the estimators defined on them. In the leading case considered by Davidson and MacKinnon the null and alternative models may be written

$$\mathbf{Y} = \mathbf{X}_0\boldsymbol{\beta}_0+\mathbf{u}_0 \text{ (null)} \tag{18}$$
$$and \tag{19}$$
$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1+\mathbf{u}_1 \text{ (alternative)}$$

and the disturbances are taken to be independent $N(\mathbf{0},\sigma^2\mathbf{I})$ in each model. Define the projection matrices, $\mathbf{P}_i = \mathbf{X}_i[\mathbf{X}_i'\mathbf{X}_i]^{-1}\mathbf{X}_i'$ $i = 0, 1$ and $\mathbf{Q}_i = \mathbf{I} - \mathbf{P}_i$ then (see Davidson and MacKinnon 2002a, p. 169) the J statistic for testing (18) is the ordinary t statistic for $\alpha = 0$ in the regression

$$\mathbf{Y}=\mathbf{X}_0\mathbf{b}_0+\hat{\alpha}\mathbf{P}_1\mathbf{Y} + \mathbf{e}$$

and they show that a key quantity is $||\boldsymbol{\theta}||^2 = ||\mathbf{Q}_0\mathbf{P}_1\mathbf{X}_0\boldsymbol{\beta}_0||^2/\sigma^2$. When this is held at zero, for example when $\mathbf{P}_1\mathbf{X}_0$ lies in the space spanned by the columns of $\mathbf{X}_0$ or, indeed, when $\mathbf{X}_1'\mathbf{X}_0 = \mathbf{0}$, or $\boldsymbol{\beta}_0/\sigma^2 = \mathbf{0}$ then the J statistic does not have the same distribution as it does elsewhere. The significance of this is that the sampling variability in $||\hat{\boldsymbol{\theta}}_0||^2$, the estimate of this quantity under the null, will prevent the bootstrap distribution from coinciding with the true sampling distribution. However, Davidson and MacKinnon are able to show that in the situation of (18) the size distortions will be quite small, of the order of a few percentage points even in the worst cases. In more general models, for example a time series model with a lagged dependent variable, the size distortions can be very large (see for example Davidson and MacKinnon 2002b Figure 1). Unfortunately, the presence of the spatial dependence in our models and the non-linear estimator of ρ that results, makes an exact analysis of any generality intractable. However, we will be able to demonstrate numerically conditions under which a simple bootstrap is relatively reliable or unreliable.

Our implementation of the simple bootstrap is as follows. Under Kelejian's assumptions, the "spatial 2SLS" estimator, (5), is consistent under the null, and so therefore is (7) when the alternative is true.

4.1 The simple resampling scheme:

Compute the J test statistics as above, then

(i) Use $\hat{\mathbf{v}}^*$ from estimation of (5) as the building block. Draw a random sample from $\hat{\mathbf{v}}^*$ using sampling with replacement; call this random sample, $\mathbf{e}^*$

(ii) Using $\hat{\rho}_0$ from Step 2, generate

$$\mathbf{u}^* = [\mathbf{I}-\hat{\rho}_0\mathbf{M}_0]^{-1}\mathbf{e}^*$$

(iii) Recall that in (5)

$$\boldsymbol{\gamma}_0 = [\boldsymbol{\beta}_0^{'}, \lambda_0]'$$

and generate

$$\mathbf{Y}^* = [\mathbf{I} - \hat{\boldsymbol{\lambda}}_0\mathbf{W}_0]^{-1}(\mathbf{X}_0\hat{\boldsymbol{\beta}}_0 + \mathbf{u}^*)$$

(iv) Calculate the J statistic using the $\mathbf{Y}^*$ sample

(v) Repeat (ii)-(iv) the designated number of times, m, to create a sample from the bootstrap distribution of the relevant J statistic.

(vi) If the proportion of the m bootstrap replicates that exceed the observed J statistic is less than the chosen significance level, reject the null hypothesis at that level.

4.2 Evaluating empirical size or power

In the experiments, we create a large number, s, of samples from each data generation process, either Model_0 to investigate test empirical size, or Model_1 to study power, and for each sample we perform steps (i)-(vi) above with the chosen value of m. The empirical size of the test at nominal significance level, α, is then the proportion of the s samples on which the null is true, but is rejected at step (vi) above. Similarly the empirical power at nominal significance level α for some specified alternative is the proportion of the s samples on which that alternative is true and on which the null is rejected at step (vi). We note that the need for resampling from the fitted model residuals, rather than importing Normal pseudo-random numbers, isn't established here - if the tests' behaviour were robust to variations in the distribution of the shocks, then the bootstrap could as well use imported pseudo-random numbers. However, we would still need to simulate the distributions of the test statistics under Model_0 using the estimated parameters and spatial structure, and it is no harder to do that by resampling from the residuals.

5 Experimental Results

Evidently, a great variety of cases are possible in this framework, such as (a) different regressors, same spatial weights, (b) same regressors, different spatial weights, and (c) both different regressors and different spatial weights. We have some results for (a), preliminary indications for (b) and as yet no numerical results for (c). We designate the first as Case 1, the second as Case 2, below.

Case 1. We implement this with a single explanatory variable other than the constant, that is, we have $\mathbf{X}_0 = [\mathbf{X}_{01} \vdots \mathbf{X}_{02}]$ where $\mathbf{X}_{01} = \mathbf{1}$, the constant vector, and $\mathbf{X}_{02}$ is, a draw from $N(\mathbf{0}, \mathbf{I}_n)$, and the two spatial weight matrices are equal, $\mathbf{M}_0 = \mathbf{W}_0$, while for the alternative, we have the same spatial structure,

$\mathbf{M}_1 = \mathbf{W}_1 = \mathbf{W}_0$, but the explanatory variable, $\mathbf{X}_{02}$ is replaced by another that is in general correlated with it, constructed in our experiments as, $\mathbf{X}_{12} = \rho_x \mathbf{X}_{02} + (1 - \rho_x^2)^{1/2} \times N(0, \mathbf{I})$ for various ρ_x values, including zero.

Case 2. We implement by having the explanatory variables, $\mathbf{X}_0$ and $\mathbf{X}_1$, the same in the two models, but the spatial structures differ, so that $\mathbf{W}_1 \neq \mathbf{W}_0$ and/or $\mathbf{M}_1 \neq \mathbf{M}_0$; for simplicity we have set, as before, $\mathbf{M}_0 = \mathbf{W}_0$ and $\mathbf{M}_1 = \mathbf{W}_1$.

Obviously these two cases do not exhaust the possibilities, but it is important to establish whether or not the procedure can provide good discrimination in these quite straightforward circumstances before letting it loose on wider model comparisons.

5.1 The set up

We consider two spatial frameworks, the 26 counties of Ireland, with weight matrix as employed in Cliff and Ord (1973, p. 164), and a set of 200 EU NUTS-2 regions with weight matrix W_0 based on a matrix of 1s and 0s denoting contiguous and non-contiguous regions respectively, subsequently normalised so that rows sum to 1,as used by Fingleton (2007). Under Case 2, the 26 county alternative weight matrix, W_1, is defined by replacing the non-zero elements of row i of the corresponding W_0 by $n_{i.}^{-1}$ the reciprocal of the number of non-zero entries in the i^{th} row. For the 200 EU regions, with $w_{ij} = d_{ij}^{-2}$ for $d_{ij} \leq 300km$ and $d_{ij} = 0$ otherwise, where d_{ij} is the straight line (Euclidean) distance between regions i and j, $W_{1ij} = \frac{w_{ij}}{\sum_j w_{ij}}$. Thus the tests are being asked to discriminate between really quite similar weight matrices. For Case 1, we restrict attention to the original weight matrices, but of course vary the explanatory regressors, as described further below. For Case 1 we also vary the instrument set, using $r \in (0, 1, 2)$ that is, a minimal, intermediate, and a rich set. In Case 2 only $r \in (1, 2)$ is relevant. We will eventually use $s = 40000$ replications of each model comparison, using $m = 399$ bootstrap samples formed as described above. We report results for a nominal significance level of 5%. With this number of replications, $s = 40000$, the standard error of an estimate, $\hat{\pi}$, of a true π equal to 0.05 is 0.001. However, the non-linear nature of the GMM estimation step means that the algorithm takes significant time to run, and so in the indicative results reported below we have used either a smaller m (99) or s (as noted where relevant). Setting the number of simple bootstrap replications equal to a quite small number, 99, had no detectable effect on test power. In a non-experimental setting one would of course use a very much larger number to secure the maximum possible power, as discussed by Davidson and MacKinnon (2000). We have evaluated the J tests at the following parameter values $(\rho_0, \lambda_0) \in (0.0, 0.3, 0.6, 0.9, 0.95) \times (0.0, 0.3, 0.6, 0.9, 0.95)$ and set $(\rho_0, \lambda_0) = (\rho_1, \lambda_1)$ so that empirical significance levels and powers reflect solely differences between the explanatory variables (Case 1) or weight matrices (Case 2); in Case 1 the explanatory variables observed for region i, $\mathbf{X}_{02i}$ and $\mathbf{X}_{12i}$ $(i = 1, ..., n)$ are drawn from a bivariate Normal distribution with variances unity and correlation $\rho_x \in (-0.5, 0.0, 0.5, 0.9, 0.95)$; the shocks are independent standard Normal, $\mathbf{v}_0 \sim IIDN(0, \mathbf{I}_n)$ and similarly $\mathbf{v}_1$ in each case. We experimented with other shock distributions, centred χ^2_1, lognormal, and a Student t with 5 degrees of freedom but found no major differences from the results we report below. In all cases the $\mathbf{W}$ and $\mathbf{M}$ matrices have non-negative elements satisfying $w_{ii} = 0$, $m_{ii} = 0$, $\sum_{j=1}^{j=n} w_{ij} = \sum_{j=1}^{j=n} m_{ij} = 1$, and the real constants λ_i and ρ_i satisfy $0 \leq |\lambda_i|, |\rho_i| < 1$ so the matrices, $(\mathbf{I} - \lambda_i \mathbf{W}_i)$ and $(\mathbf{I} - \rho_i \mathbf{M}_i)$ are non-singular, $i = 0, 1$.

5.2 Overall performance using critical values from the relevant Chisquared distributions

We first describe the performance of the two forms of test statistic, which here have either 1 or 2 degrees of freedom, when referred to critical values from the relevant χ^2 distribution, by which we mean the asymptotic sampling distribution that the 2 degree-of-freedom test statistic is shown by Kelejian (2008) to have under the conditions he gives (χ^2_2), and the asymptotic sampling distribution that the 1 degree-of-freedom test statistic is conjectured to have under similar but as yet unspecified conditions (χ^2_1).

5.2.1 Case 1

For the small spatial lattice our parameter settings for Case 1 generate 375 sets of empirical size and power estimates. For this reason we first give some summary results, then describe in more detail the situations in which the tests run into difficulties. Summary statistics for the small lattice, with sample size, $n = 26$, and $s = 40,000$ appear in Table 1.

Table 1
Case 1 Small Lattice χ^2 Critical Values

Empirical Size	r	Mean	Median	Max	Min
$1d.f.$	0	.043	.040	.076	.019
$1d.f.$	1	.072	.067	.188	.043
$1d.f.$	2	.092	.079	.251	.053
$2d.f.$	0	.030	.027	.047	.016
$2d.f.$	1	.057	.054	.144	.030
$2d.f.$	2	.072	.063	.205	.043
Empirical Power	r	Mean	Median	Max	Min
$1d.f.$	0	.41	.39	.76	.10
$1d.f.$	1	.58	.51	.87	.31
$1d.f.$	2	.67	.60	.93	.33
$2d.f.$	0	.33	.30	.67	.06
$2d.f.$	1	.53	.53	.80	.21
$2d.f.$	2	.65	.71	.89	.22

Looking at the upper panel we see that there are some cases in which the tests are quite seriously too liberal, and that in both cases empirical size increases with the number of instruments used. The presence of the very liberal cases makes it imperative to look more closely at the pattern of test sizes before drawing any conclusions. It turns out that the majority of empirical sizes are close to nominal size, while the excessively liberal cases arise for parameter combinations in which ρ_x is small or zero, and $\lambda_0 = \lambda_1$ is also small $(0, 0.3$ or $0.6)$. We experimented with estimating a response surface to describe test size variations, but while it proved relatively easy to account for much of the variation in empirical size, fitting the extremely liberal cases proved difficult, most likely because of the influence of the spatial lattice and weight matrix we are using; we have therefore adopted a slightly different approach. Table 2 lists the upper 10 and lower 10 order statistics of the sample of empirical sizes of the 1 degree-of-freedom test and the corresponding values of r, ρ_x, ρ_0 and λ_0 and Table 3 gives the corresponding information for the 2 degree-of-freedom test, in each case using the 95% quantiles of the respective Chisquare distribution as critical values; evidently these are poor approximations to the true 95% quantiles of the sampling distributions of the J-type statistics for these parameter values and this particular lattice/weight matrix.

Table 2: Extreme empirical sizes, small lattice, 1 degree-of-freedom test, Case 1
using χ^2_1 critical value, $s = 40,000$
using bootstrap critical value, $s = 5,000$, $m = 99$

Upper extreme sizes						Lower extreme sizes					
χ^2_1	BS	r	ρ_x	λ_0	ρ_0	χ^2_1	BS	r	ρ_x	λ_0	ρ_0
.251	.09	2	0.0	0.0	.95	.024	n.a.	0	.95	.95	.90
.248	.08	2	0.0	0.0	.90	.024	n.a.	0	.95	.90	.95
.187	.05	2	0.0	0.3	.90	.023	n.a.	0	-.5	.95	.90
.184	.08	1	0.0	0.0	.95	.023	.05	0	-.50	.90	.95
.178	.05	2	0.0	0.3	.95	.022	.05	0	.50	.90	.95
.177	.07	1	0.0	0.0	.90	.021	n.a.	0	.95	.95	.95
.175	n.a.	2	0.0	0.0	.60	.020	n.a.	0	.00	.95	.95
.170	.07	2	0.5	0.0	.95	.020	n.a.	0	.50	.95	.95
.170	.07	2	-0.5	0.0	.95	.019	n.a.	0	-.50	.95	.95
.163	.06	2	0.5	0.0	.90	.019	n.a.	0	.90	.95	.95

Table 3: Extreme empirical sizes, small lattice, 2 degree-of-freedom test, Case 1
using χ^2_2 critical value, $s = 40,000$
using bootstrap critical value, $s = 5,000$, $m = 99$

Upper extreme sizes						Lower extreme sizes					
χ^2_2	BS	r	ρ_x	λ_0	ρ_0	χ^2_2	BS	r	ρ_x	λ_0	ρ_0
.202	.09	2	0.0	0.0	0.95	.018	n.a	0	0.95	0.95	0.90
.187	.08	2	0.0	0.0	0.90	.018	n.a	0	0.95	0.90	0.95
.151	.08	2	0.5	0.0	0.95	.018	.05	0	-0.50	0.90	0.95
.151	.08	2	-0.5	0.0	0.95	.018	.05	0	0.50	0.90	0.95
.145	.05	2	0.0	0.3	0.90	.017	n.a	0	0.00	0.95	0.95
.143	.08	1	0.0	0.0	0.95	.017	n.a	0	0.50	0.95	0.95
.142	.07	2	0.5	0.0	0.90	.016	n.a	0	-0.50	0.95	0.90
.142	.05	2	0.0	0.3	0.95	.016	n.a	0	0.95	0.95	0.95
.136	.08	2	-0.5	0.0	0.90	.016	n.a	0	-0.50	0.95	0.95
.132	.07	1	0.0	0.0	0.90	.016	n.a	0	0.90	0.95	0.95

For the larger lattice, with sample size, $n = 200$, we find the corresponding results in Tables 4 and 5. Again, this is one particular lattice and weight matrix, but the overall pattern is quite similar to that observed for the smaller lattice, except that the most liberal cases are now much worse.

Table 4: Extreme empirical sizes, $n = 200$ lattice, 1 degree-of-freedom test, Case 1
using χ^2_1 critical value, $s = 40,000$
using bootstrap critical value, $s = 3,000$, $m = 99$

Upper extreme sizes							Lower extreme sizes					
χ^2_1	BS	r	ρ_x	λ_0	ρ_0		χ^2_1	BS	r	ρ_x	λ_0	ρ_0
.639	.45	2	0.0	0.0	0.95		.030	.045	0	-0.5	.90	.90
.557	.41	2	0.0	0.3	0.95		.030	n.a.	0	-0.5	.95	.90
.507	.37	1	0.0	0.0	0.95		.030	n.a.	0	0.5	.90	.90
.467	.30	2	0.0	0.0	0.90		.029	n.a.	0	0.0	.95	.90
.440	.35	1	0.0	0.3	0.95		.046	n.a.	0	-0.5	.90	.95
.410	.29	2	0.0	0.3	0.90		.028	n.a.	0	-0.5	.95	.95
.352	.23	1	0.0	0.0	0.90		.028	n.a.	0	0.0	.90	.95
.342	.27	2	-0.5	0.0	0.95		.028	n.a.	0	0.0	.95	.95
.322	.28	2	0.5	0.0	0.95		.027	n.a.	0	0.5	.90	.95
.313	.22	1	0.0	0.3	0.90		.027	n.a.	0	0.5	.95	.95

Table 5: Extreme empirical sizes, $n = 200$ lattice, 2 degree-of-freedom test, Case 1
using χ^2_2 critical value, $s = 40,000$
using bootstrap critical value, $s = 3,000$, $m = 99$

Upper extreme sizes							Lower extreme sizes					
χ^2_2	BS	r	ρ_x	λ_0	ρ_0		χ^2_2	BS	r	ρ_x	λ_0	ρ_0
.598	.47	2	0.0	0.0	.95		.024	n.a.	0	0.9	.95	0.90
.510	.40	2	-0.5	0.0	.95		.023	n.a.	0	-0.5	.90	0.95
.500	.41	2	0.0	0.3	.95		.023	n.a.	0	-0.5	.95	0.95
.486	.41	2	0.5	0.0	.95		.023	n.a.	0	0.0	.90	0.95
.470	.38	1	0.0	0.0	.95		.023	n.a.	0	0.5	.95	0.90
.410	.35	2	-0.5	0.3	.95		.023	n.a.	0	0.95	.90	0.95
.410	.34	2	0.0	0.0	.90		.023	n.a.	0	0.95	.95	0.95
.403	.34	2	0.5	0.3	.95		.022	n.a.	0	0.5	.95	0.95
.402	.34	1	-0.5	0.0	.95		.022	n.a.	0	0.9	.90	0.95
.400	.34	1	0.5	0.0	.95		.021	n.a.	0	0.9	.95	0.95

The upper extreme sizes in Tables 4 and 5 are much more liberal than is the case for the smaller sample size, suggesting the possibility that the test statistics may not in fact have limiting Chisquare distributions for these cases. For this to be so, at least one of the conditions introduced by Kelejian must be violated, of course. We conjecture that the answer is to be found in the analysis of Davidson and MacKinnon (2002a) described above. The key feature of our excessively liberal cases is that both ρ_x and λ_0 are small while ρ_0 is large; together, these parameter values act to keep us close to the exact condition, $||\theta|| = 0$, studied by Davidson and MacKinnon, and, if maintained while the sample size increases without limit, the condition, $\rho_x = 0$ would also violate Kelejian's Assumption A6(b). Of course, the J-type tests we are implementing lack the exact representation that Davidson and MacKinnon studied, and so the implications of the condition, $||\theta|| = 0$, must be taken as at best indicative pending further analysis.

5.2.2 Case 2

Tables 6a and 6b are based on 3000 replications of the same ρ and λ combinations as were used in Tables 1a and 1b, with the same explanatory variable in both null and alternative models, that is, a single draw from the n-dimensional IID $\mathbf{N}(\mathbf{0}, \mathbf{I})$ distribution for each replication, but with the weight matrices differing as described above. It is evident that the test performs quite poorly here.

Table 6a
Case 2 $n = 26$. χ^2 Critical Values

Empirical Size	r	Mean	Median	Max	Min
$1d.f.$	1	.08	.07	.19	.04
$1d.f.$	2	.09	.08	.19	.05
$2d.f.$	1	.10	.10	.14	.06
$2d.f.$	2	.09	.07	.21	.04
Empirical Power	r	Mean	Median	Max	Min
$1d.f.$	1	.05	.04	.08	.03
$1d.f.$	2	.13	.15	.17	.07
$2d.f.$	1	.04	.04	.06	.03
$2d.f.$	2	.09	.09	.16	.05

Table 6b

Case 2 $n = 200$. χ^2 Critical Values

Empirical Size	r	Mean	Median	Max	Min
$1d.f.$	1	.13	.05	.43	.04
$1d.f.$	2	.13	.06	.42	.04
$2d.f.$	1	.13	.08	.36	.04
$2d.f.$	2	.15	.07	.48	.04
Empirical Power	r	Mean	Median	Max	Min
$1d.f.$	1	.09	.09	.18	.04
$1d.f.$	2	.23	.23	.40	.05
$2d.f.$	1	.07	.06	.13	.03
$2d.f.$	2	.20	.16	.42	.04

Even for the larger sample size, the tests appear to be biased at many parameter combinations. As yet we have no full explanation for this, but replacing $\mathbf{H}_{01,r}$ in the augmented instrument set, (10), by $\mathbf{H}_{1,r}$ dramatically improves performance, as illustrated for thesample size, $n = 26$ in Table 6c and for sample size, $n = 200$, in Table 6d. Subsequent tables and discussion are therefore based on the use of $\mathbf{H}_{1,r}$.

Table 6c

Case 2 $n = 26$. χ^2 Critical Values

Use of $\mathbf{H}_{1,r}$ in place of $\mathbf{H}_{01,r}$ in augmented instrument set

Empirical Size	r	Mean	Median	Max	Min
$1d.f.$	1	.06	.06	.10	.04
$1d.f.$	2	.08	.08	.13	.05
$2d.f.$	1	.06	.05	.11	.04
$2d.f.$	2	.07	.07	.13	.04
Empirical Power	r	Mean	Median	Max	Min
$1d.f.$	1	.21	.22	.33	.06
$1d.f.$	2	.32	.33	.50	.08
$2d.f.$	1	.15	.15	.23	.05
$2d.f.$	2	.26	.26	.43	.07

Table 6d

Case 2 $n = 200$. χ^2 Critical Values
Use of $\mathbf{H}_{1,r}$ in place of $\mathbf{H}_{01,r}$ in augmented instrument set

Empirical Size	r	Mean	Median	Max	Min
$1d.f.$	1	.08	.05	.23	.04
$1d.f.$	2	.09	.06	.28	.03
$2d.f.$	1	.10	.05	.34	.04
$2d.f.$	2	.11	.06	.38	.04
Empirical Power	r	Mean	Median	Max	Min
$1d.f.$	1	.45	.44	.97	.05
$1d.f.$	2	.53	.57	.99	.06
$2d.f.$	1	.47	.50	.96	.05
$2d.f.$	2	.56	.67	.97	.06

Now let's turn to the extreme cases. Because the regressor is the same under null and alternative, these no longer arise from the same source as in Case 1. The upper and lower 10 empirical significance levels for the sample of size $n = 26$ are given in Tables 7 and 8 (in these tables $\lambda_0 = \lambda_1 \in (0.0, 0.3, 0.6, 0.9, .95)$, while $\rho_0 = \rho_1 \in (0.0, 0.3, 0.6, 0.9, 0.95)$). Similarly, the upper and lower 10 empirical significance levels for the sample of size $n = 200$ are given in Tables 9 and 10 (in these tables $\lambda_0 = \lambda_1 \in (0.0, 0.3, 0.6, 0.9)$, while $\rho_0 = \rho_1 \in (0.0, 0.3, 0.6, 0.9, 0.95)$). It is notable that the size distortions of the asymptotic tests are much smaller than for Case 1; however, it remains true that such distortions are greater for the *larger* sample size, which is unexpected. Of course, the particular weight matrices used play a part here, but in each case the pattern is similar, with the tests being liberal when λ is small and ρ large, and more-or-less correctly sized when λ is large, or when both λ and ρ are small. For the smaller sample the empirical sizes of the 1 and 2 $d.f.$ tests are similar, while the 2 $d.f.$ tests are more liberal for the larger sample.

Table 7: Extreme empirical sizes, $n = 26$ lattice, 1 degree-of-freedom test, Case 2

Upper extreme sizes						Lower extreme sizes				
χ_1^2	BS	r	λ_0	ρ_0		χ_1^2	BS	r	λ_0	ρ_0
.13	.09	2	0.0	.95		.05	.05	1	.90	0.6
.11	.08	2	0.0	.90		.05	.05	1	.90	0.0
.11	.09	2	0.3	.95		.05	.05	1	.90	0.3
.10	.08	2	0.3	.90		.05	.05	1	.95	0.6
.10	.09	1	0.0	.95		.04	.06	1	.95	0.9
.10	.08	1	0.0	.90		.04	.05	1	.95	0.3
.09	.08	1	0.3	.95		.04	.05	1	.90	.95
.09	.08	1	0.3	.90		.04	.05	1	.95	0.0
.09	.07	2	0.6	.90		.04	.05	1	.90	0.9
.09	.06	2	0.3	.60		.04	.05	1	.95	.95

Table 8: Extreme empirical sizes, $n = 26$ lattice, 2 degree-of-freedom test, Case 2

Upper extreme sizes						Lower extreme sizes				
χ_2^2	BS	r	λ_0	ρ_0		χ_2^2	BS	r	λ_0	ρ_0
.13	.10	2	0.0	.95		.04	.04	2	.90	0.0
.12	.10	2	0.0	.90		.04	.06	1	.95	.90
.11	.11	1	0.0	.95		.04	.06	1	.90	.90
.11	.10	1	0.0	.90		.04	.06	1	.90	.95
.11	.09	2	0.3	.95		.04	.05	2	.95	0.0
.11	.09	2	0.3	.90		.04	.05	1	.90	.30
.10	.10	1	0.3	.95		.04	.05	1	.95	0.0
.09	.09	1	0.3	.90		.04	.05	1	.95	.95
.08	.06	2	0.0	.60		.04	.05	1	.95	.60
.08	.06	2	0.3	.60		.04	.04	1	.95	.30

Table 9: Extreme empirical sizes, $n = 200$ lattice, 1 degree-of-freedom test, Case 2

Upper extreme sizes						Lower extreme sizes				
χ_1^2	BS	r	λ_0	ρ_0		χ_1^2	BS	r	λ_0	ρ_0
.28	.23	2	0.0	.95		.05	.05	1	0.6	0.0
.23	.20	1	0.0	.95		.05	.04	2	0.0	0.6
.22	.18	2	0.3	.95		.05	.05	1	0.9	0.3
.20	.17	1	0.3	.95		.05	.05	1	0.9	0.6
.18	.15	2	0.3	.90		.04	.05	2	0.9	0.9
.17	.14	2	0.0	.90		.04	.05	1	0.9	0.0
.14	.13	1	0.3	.90		.04	.05	1	0.9	0.9
.14	.12	1	0.0	.90		.04	.05	1	0.9	.95
.12	.11	1	0.6	.95		.04	.04	2	0.9	0.6
.11	.10	2	0.6	.90		.03	.05	2	0.9	.95

Table 10: Extreme empirical sizes, $n = 200$ lattice, 2 degree-of-freedom test, Case 2

Upper extreme sizes						Lower extreme sizes				
χ_2^2	BS	r	λ_0	ρ_0		χ_2^2	BS	r	λ_0	ρ_0
.38	.31	2	0.0	.95		.05	.05	1	0.6	0.6
.34	.29	1	0.0	.95		.05	.05	1	0.3	0.0
.29	.24	2	0.3	.95		.05	.05	1	0.9	0.6
.28	.25	1	0.3	.95		.05	.05	1	0.3	0.6
.26	.20	2	0.0	.90		.05	.05	1	0.9	0.3
.23	.18	2	0.3	.90		.04	.06	2	0.9	0.9
.22	.19	1	0.0	.90		.04	.04	2	0.9	0.6
.20	.18	1	0.3	.90		.04	.05	1	0.9	0.9
.14	.14	2	0.6	.95		.04	.05	2	0.9	.95
.14	.14	1	0.6	.95		.04	.05	1	0.9	.95

5.3 Overall performance of the simple bootstrap

5.3.1 Case 1

Notice immediately from the second columns of Tables 2-5 that the simple bootstrap tests have sizes much closer to nominal 5% than do either of the asymptotic tests in the cases in which the latter are too liberal. Elsewhere in the parameter space, when the asymptotic tests are closer to their nominal sizes we still find the bootstrap empirical sizes are often closer to the nominal values, though not uniformly so. With better size control comes a reduction in empirical power, but if we restrict attention to cases in which empirical size is close to nominal size, there is no loss of power associated with the bootstrap as shown in the figures (see presentation). However, the simple bootstrap does not control test size everywhere in the parameter space, and so some refinement is required.

5.3.2 Case 2

That the simple bootstrap is superior to the asymptotic tests in this setting is not obvious from the rather marginal improvements visible in Tables 7 - 10. However, we anticipate that a refinement could prove effective here also.

5.4 Power of 1 vs 2 degree-of-freedom tests

See figures to be presented at the meeting

5.5 Choice of instruments

See figures to be presented at the meeting

5.6 Implementation problems associated with the bootstrap

Because the parametric bootstrap resampling scheme requires both $[\mathbf{I}-\hat{\rho}_0\mathbf{M}_0]$ and $[\mathbf{I}-\hat{\boldsymbol{\lambda}}_0\mathbf{W}_0]$ to be non-singular, we need to respond appropriately when one or other of these conditions fails. We have considered three approaches in the experiments, (i) discard the sample, and ignore the effect of doing so on estimated test size or power, (ii) constrain $\hat{\rho}$ to lie in a suitable interval, such as $[-.97, +.97]$, and similarly constrain $\hat{\boldsymbol{\lambda}}$ so that the matrix, $[\mathbf{I}-\hat{\boldsymbol{\lambda}}_0\mathbf{W}_0]$ was never singular, (iii) discard the sample but record it as a failure to reject the null hypothesis, thus lowering the reported empirical power of the bootstrap test. The second approach has the merit that it leads to a test outcome on each sample, which is what the user cares about. We found that the estimator of $\hat{\boldsymbol{\lambda}}_0$ was very badly behaved, and so the values at the end of the range were often imposed, even when the true value was zero. This seems an obvious area for improvement, and seems to be the reason why the fast double bootstrap will not solve the size-distortion problem in the present setting.

In each of Cases 1 and 2, we believe the performance of the J tests would most likely improve if better estimators are developed, particularly of λ. It is possible that quasi-maximum likelihood estimators would be more robust here, but we have no evidence as yet.

6 Conclusions

We have experimented with the J type tests introduced by Kelejian (2008) only in rather limited cases, namely a single alternative model, so $g = 1$, and a single non-constant explanatory variable, and either different weight matrices or different regressors, but not both. The weights were taken from real examples that have been used in empirical research. The tests are clearly subject to implementation problems that are not entirely eliminated by use of a simple bootstrap resampling procedure, and merit further study.

Appendix

We now list the assumptions, adapted from Kelejian (2008 Appendix A) to the case of a single alternative model..

A.1 All diagonal elements of the spatial weight matrices, $\mathbf{W}_i$ and $\mathbf{M}_i$ are zero, $i = 0, 1$.

A.2 The matrices ($\mathbf{I}-\lambda\mathbf{W}_0$) and ($\mathbf{I}-\rho\mathbf{M}_0$) are non-singular for all $|\lambda| < 1$ and $|\rho| < 1$.

A.2a The matrices ($\mathbf{I}-\lambda\mathbf{W}_1$) and ($\mathbf{I}-\rho\mathbf{M}_1$) are non-singular for all $|\lambda| < 1$ and $|\rho| < 1$.

A.3 The row and column sums of the matrices $\mathbf{W}_0$ $\mathbf{M}_0$ $\mathbf{W}_1$ $\mathbf{M}_1$ $(\mathbf{I}-\lambda\mathbf{W}_0)^{-1}$ and $(\mathbf{I}-\rho\mathbf{M}_0)^{-1}$ are bounded uniformly in absolute value.

A.4 The regressor matrices, $\mathbf{X}_0$ and $\mathbf{X}_1$ have full column rank for large enough n and their elements are uniformly bounded in absolute value.

A.5 The vector $\mathbf{v}_0$ has elements that are independently and identically distributed with mean 0 variance σ_0^2 and finite fourth moment.

A.6 The following limiting sample second moment matrices exist and are finite and non-singular:

(i) $\mathbf{Q}_{\mathbf{H}_{0r}\mathbf{H}_{0r}} = \lim_{n\to\infty} n^{-1}\mathbf{H}'_{0,r}\mathbf{H}_{0,r}$

(ii) $\mathbf{Q}_{\mathbf{H}_{1r}\mathbf{H}_{1r}} = \lim_{n\to\infty} n^{-1}\mathbf{H}'_{1,r}\mathbf{H}_{1,r}$

(iii) $\mathbf{Q}_{\mathbf{H}_r^{**}\mathbf{H}_r^{**}} = \lim_{n\to\infty} n^{-1}\mathbf{H}_r^{**\prime}\mathbf{H}_r^{**}$

The following exist and are finite

(iv) $\mathbf{Q}_{\mathbf{H}_{0r}\mathbf{Z}_0} = p\lim_{n\to\infty} n^{-1}\mathbf{H}'_{0,r}\mathbf{Z}_0$

(v) $\mathbf{Q}_{\mathbf{H}_{1r}\mathbf{Z}_1} = p\lim_{n\to\infty} n^{-1}\mathbf{H}'_{1,r}\mathbf{Z}_1$

(vi) $\mathbf{Q}_{\mathbf{H}_{1r}\mathbf{Z}_0} = p\lim_{n\to\infty} n^{-1}\mathbf{H}'_{1,r}\mathbf{Z}_0$

(vii) $\mathbf{Q}_{\mathbf{H}_{0r}\mathbf{M}_0\mathbf{Z}_0} = p\lim_{n\to\infty} n^{-1}\mathbf{H}'_{0,r}\mathbf{M}_0\mathbf{Z}_0$

(viii) $\mathbf{Q}_{\mathbf{H}_r^{**}\mathbf{Z}^{\dagger*}} = \lim_{n\to\infty} n^{-1}\mathbf{H}_r^{**\prime}\mathbf{Z}^{\dagger*}$

where $\mathbf{Z}_r^{\dagger*} = [(\mathbf{I}-\rho\mathbf{M}_0)\mathbf{Z}_0 \vdots \mathbf{Z}_1\boldsymbol{\phi}_{1r} \vdots \mathbf{M}_1\mathbf{Z}_1\boldsymbol{\phi}_{1r}]$

in which

$\boldsymbol{\phi}_{1r} = p\lim_{n\to\infty}\hat{\gamma}_{1r} = [\mathbf{Q}'_{\mathbf{H}_{1r}\mathbf{Z}_1}\mathbf{Q}^{-1}_{\mathbf{H}_{1r}\mathbf{H}_{1r}}\mathbf{Q}_{\mathbf{H}_{1r}\mathbf{Z}_1}]^{-1}\mathbf{Q}'_{\mathbf{H}_{1r}\mathbf{Z}_1}\mathbf{Q}^{-1}_{\mathbf{H}_{1r}\mathbf{H}_{1r}}\mathbf{Q}_{\mathbf{H}_{1r}\mathbf{Z}_0}\gamma$

and $\mathbf{Q}_{\mathbf{H}_{0r}\mathbf{Z}_0}$ $\mathbf{Q}_{\mathbf{H}_{1r}\mathbf{Z}_1}$ $\mathbf{Q}_{\mathbf{H}_r^{**}\mathbf{Z}^{\dagger*}}$ and $\mathbf{Q}_{\mathbf{H}_{0r}\mathbf{M}_0\mathbf{Z}_0}$ have full column rank.

It is further assumed that the matrix,

$\mathbf{Q}_{\mathbf{H}_{0r}\mathbf{Z}_0} - \rho\mathbf{Q}_{\mathbf{H}_{0r}\mathbf{M}_0\mathbf{Z}_0} = p\lim n^{-1}\mathbf{H}'_{0,r}(\mathbf{I}-\rho\mathbf{M}_0)\mathbf{Z}_0$

has full column rank for all $|\rho| < 1$.

(ix) $\mathbf{\Phi}_H = \lim_{n\to\infty} n^{-1}\mathbf{H}'_{0,r}(\mathbf{I}-\rho\mathbf{M}_0)^{-1}(\mathbf{I}-\rho\mathbf{M}'_0)^{-1}\mathbf{H}_{0,r}$

and $\mathbf{\Phi}_{\mathbf{H}_r^{**}} = \lim_{n\to\infty} n^{-1}\mathbf{H}_r^{\prime**}(\mathbf{I}-\rho\mathbf{M}_0)^{-1}(\mathbf{I}-\rho\mathbf{M}'_0)^{-1}\mathbf{H}_r^{**}$

are finite and nonsingular for all $|\rho| < 1$.

A.7 The smallest eigenvalue of the matrix, $\mathbf{\Gamma}'\mathbf{\Gamma}$ is bounded away from zero, where $\mathbf{\Gamma} = E\{n^{-1}\mathbf{G}\}$. The matrix, $\mathbf{G}$ is as defined in the discussion of the non-linear GMM estimator of ρ.

References

[1] Anselin, L. 1986. Non-nested tests on the weight structure in spatial autoregressive models: some Monte Carlo results. Journal of Regional Science 26, 2, 267-284.

[2] Box, G.E.P., Watson, G.S. 1962. Robustness to non-Normality of regression tests. Biometrika 49 93-106.

[3] Cliff, A.D., Ord, J.K. 1973. Spatial Autocorrelation. Pion. London.

[4] Fingleton, B. 2007. Testing the "new economic geography" : a comparative analysis based on EU regional data, Chapter 3 pp 70-97 in New Directions in Economic Geography (Ed B Fingleton)

[5] Davidson, R., MacKinnon, J.G. 1981. Several tests for model specification in the presence of alternative hypotheses. Econometrica 49, 781-793.

[6] Davidson, R., MacKinnon, J.G. 2000. Bootstrap tests: how many bootstraps? Econometric Reviews 19, 55-68.

[7] Davidson, R., MacKinnon, J.G. 2002a. Bootstrap J tests of nonnested linear regression models. Journal of Econometrics 109, 167-193.

[8] Davidson, R., MacKinnon, J.G. 2002b. Fast double bootstrap tests of nonnested linear regression models. Econometric Reviews 21, 417-427.

[9] 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.

[10] Haining, R.1986. Spatial models and regional science: a comment on Anselin's paper and research directions. Journal of Regional Science 26, 4, 793-798.

[11] 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.

[12] Kelejian, H.H., Prucha, I.R. 1998. A generalised 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

[13] Kelejian, H.H., Prucha, I.R. 1999. A generalised moments estimator for the autoregressive parameter in a spatial model. International Economic Review 40, 509-533.

[14] Leenders, R.Th.A.J. 2002. Modeling social influence through network autocorrelation: constructing the weight matrix. Social Networks 24, 21-47.

[15] MacKinnon, J.G., White, H. and Davidson, R. 1983. Tests for model specification in the presence of alternative hypotheses: some further results. Journal of Econometrics 21, 53-70.