JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **31**: 387–402 (2016)
Published online 5 January 2015 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/jae.2425

AN EXTENSION OF THE *J*-TEST TO A SPATIAL PANEL DATA FRAMEWORK

HARRY H. KELEJIAN[a] AND GIANFRANCO PIRAS[b]*
[a] *University of Maryland, College Park, MD, USA*
[b] *Regional Research Institute, West Virginia University, Morgantown, WV, USA*

SUMMARY

Kelejian (*Letters in Spatial and Resources Sciences* 2008; **1**: 3–11) extended the *J*-test procedure to a spatial framework. Although his suggested test was computationally simple and intuitive, it did not use the available information in an efficient manner. Kelejian and Piras (*Regional Science and Urban Economics* 2011; **41**: 281–292) generalized and modified Kelejian's test to account for all the available information. However, neither Kelejian (2008) nor Kelejian and Piras (2011) considered a panel data framework. In this paper we generalize these earlier works to a panel data framework with fixed effects and additional endogenous variables. We give theoretical as well as Monte Carlo results relating to our suggested tests. An empirical application on a crime model for North Carolina is also estimated. Copyright © 2015 John Wiley & Sons, Ltd.

Received 6 February 2014; Revised 30 July 2014

1. INTRODUCTION

The *J*-test is a procedure for testing a null model against non-nested alternatives.[1] As described in Kelejian and Piras (2011), the *J*-test is based on whether or not predictions of the dependent variable based on the alternative models add significantly to the explanatory power of the null model.

Kelejian (2008) extended the *J*-test procedure to a spatial framework but the suggested test was not based on all of the available information. This was pointed out by Kelejian and Piras (2011) who, among other things, generalized Kelejian's assumptions. However, neither Kelejian (2008) nor Kelejian and Piras (2011) considered a panel data framework. This is unfortunate because a great many studies in recent years have been in a panel data framework.[2]

In this paper we generalize these earlier works on the *J*-test to a fixed-effects panel data framework. We specify a null model which contains spatial lags in both the dependent variable and the disturbance term, as well as additional (other than the spatial lag) endogenous variables. We allow for *G* alternative models which can also, but need not, have such 'complications'. The error terms in these alternative models are specified in a general way so that spatial correlation of various sorts, as well as general patterns of heteroskedasticity, are special cases. However, if there is a single alternative model which is

* Correspondence to: Gianfranco Piras, Regional Research Institute, West Virginia University, 886 Chestnut Ridge Road, PO Box 6825, Morgantown, WV 26506, USA. E-mail: gpiras@mac.com

[1] There is, of course, a large literature relating to the *J*-test. For example, see Davidson and MacKinnon (1981); MacKinnon *et al.* (1983); Godfrey (1983); Pesaran and Deaton (1978); Dastoor (1983); Pesaran (1974, 1982); Delgado and Stengos (1994), and the reviews given in Greene (2003, pp. 153–155, 178–180) and Kmenta (1986, pp. 593–600). A nice overview of issues relating to non-nested models is given in Pesaran and Weeks (2001).

[2] For example, see Anselin *et al.* (2008); Kapoor *et al.* (2007); Baltagi *et al.* (2007a,b, 2003, 2009); Baltagi and Liu (2008); Piras (2013); Debarsy and Ertur (2010); Elhorst and Freret (2009); Elhorst (2003, 2008, 2009, 2010); Elhorst *et al.* (2010); Lee and Yu (2010a,b,c,d); Mutl and Pfaffermayr (2011); Pesaran and Tosetti (2011); Yu and Lee (2010); Yu *et al.* (2008); Parent and LeSage (2010).

identical to the null except that it has a random effect structure, the J-test framework for discriminating between these two models should not be used.[3]

As in Kelejian and Piras (2011) we show that, given a critical assumption, the full information J-test in a panel is computationally simple, and indeed, simpler than those suggested in Kelejian (2008). We also demonstrate that this 'crucial' assumption would typically be satisfied in most spatial models. Finally, we design an extensive Monte Carlo experiment and give results that suggest that our proposed J-tests have decent power and proper size even for relatively small samples.

Finally, an empirical application to a crime model for North Carolina is also considered. The results of the J-test applied to this situation suggest that counties take into account policies implemented by their immediate neighbors.

In Section 2 we specify the null and alternative models. Section 3 contains a discussion of the J-test. In Section 4 we introduce the Monte Carlo model and present the results in Section 5. The empirical application is discussed in Section 6. Conclusions are given in Section 7. Technical details are relegated to the Appendix.

2. THE NULL AND ALTERNATIVE MODELS

2.1. The Null Model

Let e_T be a $T \times 1$ vector of unit elements. Consider the null panel data model with fixed effects

$$\begin{aligned} y &= X\beta_0 + \lambda_0(I_T \otimes W)y + Yd_0 + (e_T \otimes I_N)\mu_0 + u \\ &\equiv Z\gamma_0 + (e_T \otimes I_N)\mu_0 + u;\ Z = (X, (I_T \otimes W)y, Y) \\ \gamma_0 &= (\beta_0', \lambda_0, d_0')';\ u = \rho_0(I_T \otimes W)u + \varepsilon \end{aligned} \tag{1}$$

where y is the $NT \times 1$ vector of values of the dependent variable corresponding to N cross-sectional units over T time periods; X is an $NT \times k$ matrix of observations on exogenous variables; W is an $N \times N$ nonstochastic weighting matrix; Y is an $NT \times h$ matrix of observations on h endogenous variables; μ_0 is the $N \times 1$ fixed effects vector; u is the corresponding disturbance vector which is specified as a panel SAR process in the third line of equation (1) where ε is the random innovation vector; β_0 and d_0 are, respectively, $k \times 1$ and $h \times 1$ parameter vectors; and λ_0 and ρ_0 are scalar parameters. To avoid unnecessary complications, we have assumed that the weighting matrix in the regression model is the same as that in the error process. This assumption is typically made in practice. Our results can easily be extended to the case in which these two weighting matrices are not equal.

We allow for triangular arrays but do not index the variables in equation (1) with the sample size, in order to simplify the notation. We also assume that the system determining the endogenous variables in Y is not known. Therefore, we do not consider maximum likelihood estimation. In a typical limited information setting underlying instrumental variable estimation, we assume that the researcher has observations on $r \geq h$ exogenous variables that appear in that unknown system. For future reference, let the $NT \times r$ matrix of observations on these r exogenous variables be S.

Using evident notation, let the elements of ε be $\varepsilon_{it}, i = 1, \ldots, n;\ t = 1, \ldots, T$. Then, we assume the ε_{it} is i.i.d. over both i and t with mean and variance $(0, \sigma_\varepsilon^2)$. This intuitive specification is more than adequate to understand our presentation below, but it does not allow for triangular arrays. A formal specification which does is given as Assumption 1 in Kapoor *et al.* (2007). In order for the model to be complete, we also assume that $(I_N - aW)$ is nonsingular for all $|a| < 1$.

[3] Mutl and Pfaffermayr (2011) suggest and give large-sample results for a Hausman test to discriminate between these two models.

DOI: 10.1002/jae

Let $Q_0 = \left(\left(I_T - \frac{J_T}{T}\right) \otimes I_N\right)$, $J_T = e_T e_T'$, and note that $Q_0(e_T \otimes I_N) = 0$. Thus pre-multiplying the second and third lines in equation (1) by Q_0 yields

$$\begin{aligned} Q_0 y &= Q_0 Z\gamma_0 + Q_0 u \\ Q_0 u &= \rho_0(I_T \otimes W)Q_0 u + Q_0\varepsilon \end{aligned} \tag{2}$$

since $Q_0(I_T \otimes W) = (I_T \otimes W)Q_0$. Finally, consistent with the spatial Cochrance–Orcutt procedure, let

$$\begin{aligned} y^*(\rho_0) &= [I_{NT} - \rho_0(I_T \otimes W)]Q_0\mathbf{y} \\ Z^*(\rho_0) &= [I_{NT} - \rho_0(I_T \otimes W)]Q_0 Z \end{aligned} \tag{3}$$

and note from equation (2) that

$$y^*(\rho_0) = Z^*(\rho_0)\gamma_0 + Q_0\varepsilon \tag{4}$$

2.2. The G Alternative Models Under H_1

We assume the researcher believes that the true alternative to the null is one of $J = 1, \ldots, G$ models, namely

$$H_{1,J} : y_t = P_{t,J}\beta_J + \lambda_J M_J\ y_t + Y_{t,J} d_J + \mu_J + \psi_{t,J};\ J = 1, \ldots, G;\ t = 1, \ldots, T \tag{5}$$

where y_t is the $N \times 1$ vector of observations corresponding to the dependent variable at time t, $P_{J,t}$ is the $N \times k_J$ matrix of observations on the exogenous variables in the Jth model at time t, M_J is the corresponding weighting matrix, $Y_{t,J}$ is an $N \times h_J$ matrix of observations on h_J endogenous variables at time t, μ_J is a fixed-effects vector, $\psi_{t,J}$ is the corresponding disturbance vector, etc. As for the null model, we assume that the only information on the system determining the elements of $Y_{t,J}$ are observations on r_J exogenous variables. At this point, note that unlike for the null model we have specified these alternative models for each time period t.

Stacking the model in equation (5) over $t = 1, \ldots, T$, and using evident notation, we have

$$\begin{aligned} H_{1,J} : y &= P_J\beta_J + \lambda_J(I_T \otimes M_J)y + Y_J d_J + (e_T \otimes I_N)\mu_J + \psi_J \\ &\equiv Z_J\ \gamma_J + (e_T \otimes I_N)\mu_J + \psi_J;\ Z_J = [P_J, (I_T \otimes M_J)y, Y_J];\gamma_J = \left(\beta_J', \lambda_J, d_J'\right)' \end{aligned} \tag{6}$$

Multiplying the second line in equation (6) across by Q_0 we have

$$Q_0 y = Q_0 Z_J\ \gamma_J + Q_0\psi_J;\ J = 1, \ldots, G \tag{7}$$

since $Q_0(I_T \otimes M_J) = (I_T \otimes M_J)Q_0$. For future reference, let S_J be the $NT \times r_J$ matrix of observations on the r_J exogenous variables which appear in the system determining the elements of Y under $H_{1,J}$.

Our assumptions concerning the disturbance terms in the alternative models are quite general. Specifically, we only assume

$$E(\psi_J | H_{1,J}, \Phi_{1,J}) = 0 \text{ and } E(\psi_J\psi_J' | H_{1,J}, \Phi_{1,J}) = \Omega_{\psi_J} \tag{8}$$

where $\Phi_{1,J} = (P_J, M_J, S_J)$, and the row and column sums of Ω_{ψ_J} and $\Omega_{\psi_J}^{-1}$ are uniformly bound in absolute value. This assumption is consistent with general patterns of spatial correlation and heteroskedasticity. It is also typically assumed in large-sample analyses of spatial models (see, for example, Kelejian and Prucha, 1999, 2004, 2007; Yu *et al.*, 2008; Mutl and Pfaffermayr, 2011).

DOI: 10.1002/jae

3. THE FORM OF THE J-TEST

3.1. The Augmented Equation

Our J-test is based on augmenting equation (4) with the G predictions of $y^*(\rho_0)$ based on the G alternative models, namely

$$E\left(y^*(\rho_0)|H_{1,J}, \text{INFO}_J\right) = [I_{NT} - \rho_0(I_T \otimes W)]\ [Q_0 E(\mathbf{y}|H_{1,J}, \text{INFO}_J)];\ J = 1, \ldots, G \quad (9)$$

where INFO_J is the information set the predictions are based on. In essence, our test of H_0 versus H_1 is a test of the joint significance of the G augmenting variables.

We consider two information sets and so there are two forms of our augmented model. The first set is just $\Phi_{1,J}$. Let $y_{i,t}$ be the ith value of y_t and let $Y_{i,t,J}$ be the ith row of $Y_{t,J}$. Then, the second information set is a full information set, say $\Phi_{2,J}$, which relates to the prediction of each $y_{i,t}$ based on the Jth alternative model. Let $M_{i,J}$ be the ith row of M_J. Then, as discussed in more detail in the Appendix, our second information set, say $\Phi_{2,J}$, augments $\Phi_{1,J}$ for each $y_{i,t}$ with $M_{i,J}y_t$ and $Y_{i,t,J}$, which would be regressors in the Jth model under H_1 for $y_{i,t}$—see equation (5). Since weighting matrices have zeroes on the diagonal, $M_{i,J}y_t$ does not include $y_{i,t}$ but it is obviously correlated with the error term, as would be $Y_{i,t,J}$ in the Jth model. Let $Q_0 y_J^{(1)} = Q_0 E(y|H_{1,J}, \Phi_{1,J})$ and, in equation (10) below, let $Q_0 y_J^{(2)} = Q_0 E(y|H_{1,J}, \Phi_{2,J}) - Q_0\eta_J$. Again, as demonstrated in the Appendix:

$$\begin{aligned} Q_0 E(y|H_{1,J}, \Phi_{1,J}) &= [I_T \otimes (I_N - \lambda_J M_J)^{-1}][Q_0 P_J \beta_J + Q_0 E(Y_J|H_{1,J}, \Phi_{1,J}) d_J] \\ Q_0 E(y|H_{1,J}, \Phi_{2,J}) &= Q_0 P_J \beta_J + \lambda_J (I_T \otimes M_J) Q_0 y + Y_J d_J + Q_0 \eta_J \end{aligned} \quad (10)$$

where $\eta_J = E(\psi_J|H_{1,J}, \Phi_{2,J}) \neq 0$ but, as shown in the Appendix, $E(\eta_J|H_{1,J}, \Phi_{1,J}) = 0$.

Under reasonable conditions, we show in the Appendix that if $H_{1,J}$ is the only alternative under H_1, i.e. $G = 1$, the term involving η_J in equation (10) is asymptotically negligible even if it is observed, and so $Q_0\eta_J$ can be ignored. In addition, a demonstration virtually identical to that in Kelejian and Piras (2011) will show that the power of the J-test based on the predictor $Q_0 P_{J,t}\beta_J + \lambda_J(I_T \otimes M_J)Q_0 y + Y_J d_J$ is asymptotically equivalent to the power of the J-test based on $[I_T \otimes (I_N - \lambda_J M_J)^{-1}][Q_0 P_J \beta_J + Q_0 E(Y_J|H_{1,J}, \Phi_{1,J})d_J]$, again when $G = 1$. This predictor corresponds to the mean of the dependent variable via the reduced form of equation (6). Our Monte Carlo results are consistent with these asymptotic results even if $G > 1.0$. Since the predictor on the second line of equation (10), minus the term $Q_0\eta_J$, does not involve an inverse, we suggest its use.

Let

$$Y_J^{(s)} = [I_{NT} - \rho_0(I_T \otimes W)]Q_0 y_J^{(s)}\ s = 1, 2;\ J = 1, \ldots, G \quad (11)$$

Our suggested augmented equation is based on $y_J^{(2)}$, $J = 1, \ldots, G$:

$$\begin{aligned} y^*(\rho_0) &= Z^*(\rho_0)\gamma_0 + Y^{(2)}\alpha + Q_0\varepsilon \\ Y^{(2)} &= \left(Y_1^{(2)}, \ldots, Y_G^{(2)}\right);\ \alpha' = (\alpha_1, \ldots, \alpha_G) \end{aligned} \quad (12)$$

where α is a $G \times 1$ parameter vector. For future reference we define $Y^{(1)} = \left(Y_1^{(1)}, \ldots, Y_G^{(1)}\right)$.

To implement the J-test, the parameters ρ_0, β_J, d_J, and λ_J defining $Y_J^{(2)}$ in equation (11) must be estimated, $J = 1, \ldots, G$. Our Monte Carlo results suggest that even in moderately sized samples the power of the J-test is not very sensitive to the efficiency with which these parameters are estimated as

DOI: 10.1002/jae

long as, under $H_{1,J}$, these estimators are consistent. This implies that in implementing the J -test the alternative models can simply be estimated by an IV procedure which need not account for particular 'complications' the researcher may assume for the error term. The parameter ρ_0 is estimated in terms of the null model, as described below.

3.2. The Necessary Calculations and the Test

To implement the J-test, the null model and all G alternative models must be estimated in a preliminary, but consistent fashion. Given these estimations, the augmented model must then be estimated. The instrument matrices we use to estimate the null model, and the alternative models are, respectively,

$$\begin{aligned} H_{(0)} &= \left[Q_0X, Q_0S, (I_T \otimes W)Q_0X, \left(I_T \otimes W^2\right) Q_0X\right]_{\mathrm{LI}} \\ H_{(J)} &= \left[Q_0P_J, Q_0S_J, (I_T \otimes M_J)Q_0P_J, \left(I_T \otimes M_J^2\right) Q_0P_J\right]_{\mathrm{LI}} \\ J &= 1, \ldots, G \end{aligned} \tag{13}$$

where LI denotes the linearly in dependent columns of the matrix in brackets. The augmented model is estimated using the union of these instrument matrices:

$$H_{(A)} = [H_{(0)}, H_{(1)}, \ldots, H_{(G)}]_{\mathrm{LI}} \tag{14}$$

Step 1. We follow Mutl and Pfaffermayr (2011), Kapoor *et al.* (2007), and Piras (2013) in estimating the fixed-effects null model in equation (2) by 2SLS based on $H_{(0)}$ and use the residuals to obtain the generalized method of moments (GMM) estimator of ρ_0, say $\hat{\rho}_0$. This estimator is based on the first three equations of the GMM procedure in Kapoor *et al.* (2007).

Step 2. Estimate the regression parameter vector γ_J of the alternative models in equation (7) by any consistent method under $H_{1,J}$. One such method would be 2SLS based on the instrument matrix $H_{(J)}$.

Step 3. Let $\hat{\gamma}_J$, $J = 1, \ldots, G$ be a consistent estimator of γ_J. Let $\hat{Y}^{(2)} = \left(\hat{Y}_1^{(2)}, \ldots, \hat{Y}_G^{(2)}\right)$, where $\hat{Y}_J^{(2)}$ is identical to $Y_J^{(2)}$ in (11) except that ρ_0, β_J, d_J, and λ_J are replaced by $\hat{\rho}_0$, $\hat{\beta}_J, \hat{d}_J$, and $\hat{\lambda}_J$, $J = 1, \ldots, G$. Similarly, let $y^*(\hat{\rho}_0)$ and $Z^*(\hat{\rho}_0)$ be identical to $y^*(\rho_0)$ and $Z^*(\rho_0)$ in equation (3) except that ρ_0 is replaced by $\hat{\rho}_0$. Then, the empirical form of our augmented model is

$$\begin{aligned} y^*(\hat{\rho}_0) &= Z^*(\hat{\rho}_0)\gamma_0 + \hat{Y}\alpha + \zeta \\ &= F\delta + \zeta \end{aligned} \tag{15}$$

where $F = (Z^*(\hat{\rho}_0), \hat{Y})$, and $\delta' = (\gamma_0', \alpha')$, and ζ is an error term.

Step 4. Let $\hat{F} = H_{(A)}(H_{(A)}'H_{(A)})^{-1}H_{(A)}'F$, where the instrument matrix $H_{(A)}$ is defined in equation (14). Then, our test is based on the 2SLS estimator of δ in equation (15), namely

$$\hat{\delta} = \left(\hat{F}'\hat{F}\right)^{-1} \hat{F}'y^*(\hat{\rho}_0) \tag{16}$$

DOI: 10.1002/jae

Assuming the standard conditions in Mutl and Pfaffermayr (2011) with their regressor matrix expanded to include $\hat{Y}$, and those in Kapoor *et al.* (2007), $\hat{\rho}_0$ is consistent and

$$(NT)^{1/2}[\hat{\delta}-\delta] \xrightarrow{D} N\left(0, \sigma_\varepsilon^2 \; p \lim_{N\to\infty}(NT)\left[\hat{\Gamma}\left(H'_{(A)}H_{(A)}\right)^{-1}\hat{\Gamma}'\right]\right)$$
$$\hat{\Gamma} = (\hat{F}'\hat{F})^{-1}\left[F'H_{(A)}\right] \tag{17}$$

Small-sample inferences can be based on the approximation

$$\hat{\delta} \dot{\sim} N\left(\delta, \hat{\sigma}_\varepsilon^2\hat{\Gamma}\left(H'_{(A)}H_{(A)}\right)^{-1}\hat{\Gamma}'\right) \tag{18}$$

where[4]

$$\hat{\sigma}_\varepsilon^2 = \frac{1}{N(T-1)}\left[y^*(\hat{\rho}_0) - F\hat{\delta}\right]'\left[y^*(\hat{\rho}_0) - F\hat{\delta}\right] \tag{19}$$

Our test of H_0 against H_1 is to reject H_0 at the 5% level if

$$\hat{\alpha}'\left[\mathrm{VC}_{\hat{\alpha}}^{-1}\right]\alpha > \chi_G^2(0.95)$$

where $\hat{\alpha}$ is the last G elements of $\hat{\delta}$ and $\mathrm{VC}_{\hat{\alpha}}$ is the lower diagonal $G \times G$ block of $\hat{\sigma}_\varepsilon^2\hat{\Gamma}(H'_{(A)}H_{(A)})^{-1}\hat{\Gamma}'$.

4. THE MONTE CARLO MODEL

The experimental design for the Monte Carlo simulation is based on a format extensively used in studies on spatial panel regression models (e.g. Kapoor *et al.*, 2007; Baltagi *et al.*, 2003, 2007b; Debarsy and Ertur, 2010).

Specifically, we generate two sets of data corresponding to two regular grids of dimension 7×7 and 10×10, leading to sample sizes of, respectively, $N = 49$ and $N = 100$. We only consider one value of the time dimension, namely $T = 4$. For each sample size we construct three row-normalized weights matrices. Following Kelejian and Prucha (1999), the first of these three matrices (W_1) is defined in a circular world and is generally referred to as a '5 ahead and 5 behind' spatial weights matrix. Our second matrix (W_2) is a distance matrix based on the 10 nearest neighbors.[5] Finally, the third matrix (W_3) is a contiguity matrix based on the rook criterion (i.e. only borders but not vertex).

For each sample size, we design six sets of experiments. In the first two sets, the null and alternative models only differ in terms of the weighting matrix. In the third set, the null and alternative models differ only in terms of the regressor matrix. In the fourth set they differ in terms of both their regressors and weighting matrices. In the last two sets of experiments, we consider two models under the alternative. In the fifth set, the null and alternative models only differ in terms of the weighting matrix employed. In the sixth set, they differ in terms of both the regressors and the weighting matrix.

[4] Note that the expression for $\hat{\sigma}_\varepsilon^2$ is divided by $\frac{1}{N(T-1)}$. This is actually in line with the findings of Lee and Yu (2010a). In a different context related to maximum likelihood estimation they suggest a transformation, i.e. based on the orthonormal matrix of the eigenvectors of J_T, to avoid linear dependence of the disturbances over the time dimension. Furthermore, when the model is specified only in terms of individual effects, both the transformation and the direct approach lead to the same coefficients estimate except for σ_ε^2. In fact, the estimation of $\hat{\sigma}_\varepsilon^2$ from the direct approach will be $\frac{T-1}{T}$ times the estimate from the transformation approach.

[5] We used Euclidean distance to determine the 10 nearest neighbors.

DOI: 10.1002/jae

Using evident notation, in all of the experiments the null model is of the form

$$\begin{aligned} y &= \lambda(I_T \otimes W)y + X\beta + u \\ u &= \rho(I_T \otimes W)u + \varepsilon \end{aligned} \tag{20}$$

where, for simplicity, we assume that the weighting matrix in the regression model is the same as that in the disturbance process.[6]

We consider two distributions for ε. In the first case the elements of ε are i.i.d. $N(0, 1)$. As in Kelejian and Prucha (1999), the second distribution considered is a normalized version of the log-normal, henceforth abbreviated as N log N. In this case the elements of ε are specified as

$$\varepsilon_i = [\exp(\xi_i) - \exp(0.5)]/[\exp(2) - \exp(1)]^{0.5} \tag{21}$$

where ξ_i is i.i.d. $N(0, 1)$. This normalization implies that ε is i.i.d. $(0, 1)$ but the distributions in equation 21 is not symmetric.[7]

In all experiments, except those in the fourth set, the alternative models are defined only in terms of the first line of equation (20).[8]

There are two regressor matrices: one for the null (X_0) and one for the alternative model (X_1).[9] The regressor matrix X_0 is taken as $X_0 = (x_0, x_1)$. Following Debarsy and Ertur (2010), the (i, t)th value of x_0, as well as that of x_1, is an independent draw from $N(\mu_t, 1)$ where μ_t is an independent draw from $U(0, 1), t = 1, \ldots, 4; i = 1, \ldots, N$. The regressor matrix X_1 is taken as $X_1 = (z_1, z_2)$, where the NT values of z_1 are generated in the same way as x_0 (and x_1) and

$$z_2 = ax_0 + \xi \tag{22}$$

where $\xi = N(0, I_{NT})$ and $a = .5$. As in Kelejian and Piras (2011), we choose this value of a because it leads to a correlation between z_1 and x_0 of approximately 0.5. In all experiments, once generated, the values of the regressors are held fixed in the Monte Carlo trials. Finally, the elements of β are taken as 1.0 or 0.5. These values of the parameters lead to the ratio of the variance of $X\beta$ to the sum of the variance of $X\beta$ and the variance of the error term of approximately 0.35 and 0.68. We refer to this ratio as $\tilde{R}^2$ in our tables in the next section.

In all sets of experiments, we consider six values for λ, namely -0.6, -0.4, -0.2, 0.2, 0.4 and 0.6; and two values for ρ, namely -0.4 and 0.4.

For all experiments, 2000 replications were performed. This is roughly the number of replications needed to obtain a 95% confidence interval of length 0.019 on the size of a test statistic; i.e. an estimate of the size of the test in the tables below would be viewed as being significantly different from the 0.05 theoretical level if it was not in the interval (0.041, 0.060).

5. MONTE CARLO RESULTS

Our Monte Carlo results are summarized in Tables I–IV. These tables give the frequency of rejection of the null hypothesis at the 5% level. These figures relate to the estimated size of the test, as well as to its estimated power. As already mentioned, all the results are based on the disturbance distribution $\varepsilon_n =$

[6] Although restrictive, this assumption is generally made in many spatial econometrics applications (e.g. Donovan *et al.*, 2007; Arraiz *et al.*, 2010; Piras and Lozano-Gracia, 2012).

[7] Since the results for the N log N are virtually identical, they are not reported in the paper. However, they can be obtained from the authors.

[8] In the fourth set of experiments we also consider the same two distributions for the disturbance terms. However, unlike for the other experiments, the alternative model has the same structure as in equation (20).

[9] The use of these matrices in the various sets of experiments is described in the tables.

DOI: 10.1002/jae

Table I. Frequency of rejection of the null hypothesis: two predictors (2000 replications)

		First set of experiments: H_0: X_0, W_1, H_1: X_0, W_2							
		$\varepsilon_n = N(0, I_n)$ $n = 49$				$\varepsilon_n = N(0, I_n)$ $n = 100$			
		Size		Power		Size		Power	
		$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$
$\tilde{R}^2 = 0.35$									
ρ=−0.4	λ=−0.6	0.0535	0.0430	0.2830	0.1825	0.0555	0.0515	0.5660	0.5115
	λ=−0.4	0.0580	0.0425	0.1860	0.0745	0.0580	0.0460	0.3230	0.2435
	λ=−0.2	0.0650	0.0460	0.0990	0.0335	0.0605	0.0460	0.1410	0.0815
	λ=0.2	0.0600	0.0525	0.0955	0.1880	0.0555	0.0535	0.1335	0.2400
	λ=0.4	0.0790	0.0495	0.2450	0.4670	0.0615	0.0525	0.4025	0.6215
	λ=0.6	0.0850	0.0470	0.4655	0.7700	0.0625	0.0485	0.7275	0.9210
ρ=0.4	λ=−0.6	0.0665	0.0900	0.3155	0.1890	0.0565	0.0590	0.5820	0.5075
	λ=−0.4	0.0720	0.0890	0.2065	0.0955	0.0560	0.0545	0.3445	0.2580
	λ=−0.2	0.0660	0.0805	0.1125	0.0460	0.0490	0.0500	0.1185	0.0700
	λ=0.2	0.0810	0.0660	0.1020	0.1785	0.0610	0.0505	0.1255	0.2110
	λ=0.4	0.0925	0.0575	0.2420	0.4755	0.0760	0.0470	0.4050	0.6140
	λ=0.6	0.1000	0.0465	0.4920	0.7845	0.0815	0.0440	0.7345	0.9225
Averages		0.0732	0.0592	0.2370	0.2904	0.0611	0.0503	0.3836	0.4335
$\tilde{R}^2 = 0.68$									
ρ=−0.4	λ=−0.6	0.0530	0.0555	0.8035	0.7500	0.0510	0.0515	0.9905	0.9900
	λ=−0.4	0.0555	0.0590	0.4985	0.4040	0.0535	0.0490	0.8465	0.8185
	λ=−0.2	0.0565	0.0535	0.1765	0.1165	0.0575	0.0470	0.3605	0.3100
	λ=0.2	0.0540	0.0540	0.2300	0.3175	0.0510	0.0550	0.4120	0.5050
	λ=0.4	0.0575	0.0540	0.7100	0.8205	0.0590	0.0560	0.9395	0.9700
	λ=0.6	0.0580	0.0445	0.9570	0.9870	0.0540	0.0510	0.9995	1.0000
ρ=0.4	λ=−0.6	0.0510	0.0655	0.7985	0.7460	0.0535	0.0560	0.9940	0.9910
	λ=−0.4	0.0505	0.0635	0.4935	0.4215	0.0440	0.0520	0.8680	0.8370
	λ=−0.2	0.0525	0.0695	0.1815	0.1315	0.0430	0.0480	0.3450	0.2855
	λ=0.2	0.0460	0.0525	0.2305	0.3215	0.0455	0.0495	0.3960	0.4930
	λ=0.4	0.0510	0.0590	0.7100	0.8270	0.0535	0.0505	0.9305	0.9685
	λ=0.6	0.0520	0.0440	0.9580	0.9875	0.0540	0.0515	0.9990	1.0000
Averages		0.0531	0.0562	0.5623	0.5692	0.0516	0.0514	0.7568	0.7640

$N(0, I_n)$. Except for Table III leave a space (discussed later), the results in the first four columns of the tables correspond to the sample size of $N = 49$ observations, whereas the results in the remaining columns relate to larger sample size (i.e. $N = 100$). Additionally, the top parts of the tables refer to the lower value of the $\tilde{R}^2$ statistic whereas the bottom parts refer to the higher $\tilde{R}^2$. Results are given for each of the two predictors $y^{(1)}$ and $y^{(2)}$.

Owing to space limitations, we only report on some experiments. Since experiments one and two are similar (i.e. only the spatial weighting matrix are different between the null and alternative model in both experiments) we only show the complete set of results (for the normal distribution of the error term) for the first experiment (Table I). The results reported in Tables II, III and IV refer respectively to the third, fourth and fifth set of experiments.[10]

Consider the results in Table I, which are based on the first set of experiments. In these experiments, the matrix of regressors in the null and alternative model is the same and the only difference between

[10] We do not show any evidence for the sixth experiment. Of course, the results are available from the authors.

DOI: 10.1002/jae

Table II. Frequency of rejection of the Null Hypothesis: two predictors (2000 replications)

		Third set of experiments: H_0: X_0, W_1, H_1: X_1, W_1							
		$\varepsilon_n = N(0, I_n)$ $n = 49$				$\varepsilon_n = N(0, I_n)$ $n = 100$			
		Size		Power		Size		Power	
		$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$
$\tilde{R}^2 = 0.35$									
ρ=−0.4	λ=−0.6	0.0600	0.0510	1.0000	1.0000	0.0645	0.0555	1.0000	1.0000
	λ=−0.4	0.0640	0.0515	1.0000	1.0000	0.0630	0.0565	1.0000	1.0000
	λ=−0.2	0.0440	0.0340	1.0000	1.0000	0.0545	0.0500	1.0000	1.0000
	λ=0.2	0.0435	0.0495	1.0000	1.0000	0.0440	0.0505	1.0000	1.0000
	λ=0.4	0.0510	0.0510	0.9995	1.0000	0.0460	0.0535	1.0000	1.0000
	λ=0.6	0.0470	0.0455	0.9990	1.0000	0.0365	0.0540	1.0000	1.0000
ρ=0.4	λ=−0.6	0.0800	0.0570	1.0000	1.0000	0.0620	0.0475	1.0000	1.0000
	λ=−0.4	0.0895	0.0600	1.0000	1.0000	0.0600	0.0545	1.0000	1.0000
	λ=−0.2	0.0760	0.0550	1.0000	1.0000	0.0600	0.0585	1.0000	1.0000
	λ=0.2	0.0695	0.0490	1.0000	1.0000	0.0545	0.0540	1.0000	1.0000
	λ=0.4	0.0770	0.0525	1.0000	1.0000	0.0500	0.0525	0.9995	1.0000
	λ=0.6	0.0625	0.0470	0.9980	1.0000	0.0440	0.0545	0.9995	1.0000
Averages		0.0637	0.0503	0.9997	1.0000	0.0533	0.0535	0.9999	1.0000
$\tilde{R}^2 = 0.68$									
ρ=−0.4	λ=−0.6	0.0525	0.0510	1.0000	1.0000	0.0610	0.0575	1.0000	1.0000
	λ=−0.4	0.0570	0.0545	1.0000	1.0000	0.0605	0.0580	1.0000	1.0000
	λ=−0.2	0.0445	0.0380	1.0000	1.0000	0.0505	0.0510	1.0000	1.0000
	λ=0.2	0.0520	0.0510	1.0000	1.0000	0.0465	0.0460	1.0000	1.0000
	λ=0.4	0.0605	0.0550	1.0000	1.0000	0.0500	0.0515	1.0000	1.0000
	λ=0.6	0.0555	0.0550	1.0000	1.0000	0.0410	0.0465	1.0000	1.0000
ρ=0.4	λ=−0.6	0.0680	0.0595	1.0000	1.0000	0.0455	0.0470	1.0000	1.0000
	λ=−0.4	0.0770	0.0675	1.0000	1.0000	0.0490	0.0485	1.0000	1.0000
	λ=−0.2	0.0665	0.0530	1.0000	1.0000	0.0580	0.0560	1.0000	1.0000
	λ=0.2	0.0705	0.0595	1.0000	1.0000	0.0475	0.0515	1.0000	1.0000
	λ=0.4	0.0700	0.0545	1.0000	1.0000	0.0450	0.0490	1.0000	1.0000
	λ=0.6	0.0635	0.0535	1.0000	1.0000	0.0400	0.0450	1.0000	1.0000
Averages		0.0615	0.0543	1.0000	1.0000	0.0495	0.0506	1.0000	1.0000

the null and alternative pertains to the spatial weighting matrices. We will focus first on the size of the test and then move to the power.

Let us concentrate first on the top left part of Table I, where $\tilde{R}^2 = 0.35$ and $N = 49$. Looking at column averages, the empirical size of the test is quite close to the theoretical 5% level only for the predictor $y^{(2)}$. In general, there are many cases in which the empirical size of the test exceeds the theoretical level. However, the 'size of test problem' highlighted in the top part of the table mitigates when the value of the $\tilde{R}^2 = 0.68$ (bottom left part of the table). In this case, the empirical size of the test based on both specifications of the error term, is, on average, quite close to the theoretical 5% level. There are only a few values outside of the acceptance interval. In particular, when the distribution of the errors is normal, there are only three cases and all of them are related to the predictor $y^{(2)}$. This is quite impressive given that the sample size is very small.

When the sample size increases but the value of $\tilde{R}^2 = 0.35$ (as in the top right part of Table I), the empirical size of the test is still close to the theoretical 5% level only for the predictor $y^{(2)}$. However, the average size for the predictor $y^{(1)}$ is much closer to the acceptance interval (0.0611). Also, there are fewer cases in the table where the estimate of the size is significantly different from the theoretical level. Fortunately, if $\tilde{R}^2$ increases as well, the estimate of the size is always close to the theoretical 5%

DOI: 10.1002/jae

Table III. Frequency of rejection of the null hypothesis: two predictors (2000 replications)

		Fourth set of experiments: H_0: X_0, W_1, H_1: X_1, W_3							
		$\varepsilon_n = N(0, I_n)$							
		Size 2sls		Size 2sls + gm		Power 2sls		Power 2sls + gm	
		$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$
$N = 49, \tilde{R}^2 = 0.35$									
ρ=−0.4	λ=−0.6	0.0680	0.0840	0.0780	0.0740	0.9965	0.9980	1.0000	1.0000
	λ=−0.4	0.0815	0.1000	0.0880	0.0925	0.9995	0.9995	1.0000	1.0000
	λ=−0.2	0.0730	0.0895	0.0765	0.0795	1.0000	1.0000	1.0000	1.0000
	λ=0.2	0.0860	0.1005	0.0855	0.0890	1.0000	1.0000	1.0000	1.0000
	λ=0.4	0.0705	0.0925	0.0790	0.0870	1.0000	1.0000	1.0000	1.0000
	λ=0.6	0.0620	0.0700	0.0675	0.0790	1.0000	1.0000	1.0000	1.0000
ρ=0.4	λ=−0.6	0.0730	0.0975	0.0775	0.0805	1.0000	1.0000	1.0000	1.0000
	λ=−0.4	0.0900	0.1150	0.0950	0.0995	1.0000	1.0000	1.0000	1.0000
	λ=−0.2	0.0740	0.1005	0.0850	0.0950	1.0000	1.0000	1.0000	1.0000
	λ=0.2	0.0705	0.0935	0.0735	0.0805	1.0000	1.0000	1.0000	1.0000
	λ=0.4	0.0615	0.0755	0.0730	0.0810	0.9985	0.9975	1.0000	1.0000
	λ=0.6	0.0565	0.0670	0.0670	0.0745	0.9890	0.9885	1.0000	1.0000
Averages		0.0722	0.0905	0.0788	0.0843	0.9986	0.9986	1.0000	1.0000
$N = 49, \tilde{R}^2 = 0.68$									
ρ=−0.4	λ=−0.6	0.0560	0.0630	0.0545	0.0560	1.0000	1.0000	1.0000	1.0000
	λ=−0.4	0.0695	0.0755	0.0710	0.0745	1.0000	1.0000	1.0000	1.0000
	λ=−0.2	0.0585	0.0610	0.0585	0.0545	1.0000	1.0000	1.0000	1.0000
	λ=0.2	0.0675	0.0755	0.0645	0.0720	1.0000	1.0000	1.0000	1.0000
	λ=0.4	0.0605	0.0645	0.0660	0.0685	1.0000	1.0000	1.0000	1.0000
	λ=0.6	0.0555	0.0555	0.0610	0.0605	1.0000	1.0000	1.0000	1.0000
ρ=0.4	λ=−0.6	0.0610	0.0650	0.0595	0.0610	1.0000	1.0000	1.0000	1.0000
	λ=−0.4	0.0745	0.0820	0.0755	0.0750	1.0000	1.0000	1.0000	1.0000
	λ=−0.2	0.0685	0.0735	0.0665	0.0690	1.0000	1.0000	1.0000	1.0000
	λ=0.2	0.0560	0.0640	0.0585	0.0670	1.0000	1.0000	1.0000	1.0000
	λ=0.4	0.0540	0.0595	0.0620	0.0625	1.0000	1.0000	1.0000	1.0000
	λ=0.6	0.0515	0.0580	0.0580	0.0640	1.0000	1.0000	1.0000	1.0000
Averages		0.0611	0.0664	0.0630	0.0654	1.0000	1.0000	1.0000	1.0000

level, as revealed by the figures at the bottom-right part of the table. In fact, there is not a single case in which the estimate of the size is significantly different from 5%.

Consider now the power estimates in Table I. Note that in all cases considered, consistent with the theoretical development, the power increases on average both with sample size and with the value of $\tilde{R}^2$. For example, the average power of the test when $n = 49$ and $\tilde{R}^2 = 0.35$ (top left) is 0.2370 (for the first predictor) and 0.2904 (for the second predictor). The same average increases to 0.5623 (for the first predictor) and 0.5692 (for the second predictor) when the sample size is still $n = 49$ but $\tilde{R}^2 = 0.68$ (bottom left). Additionally, when the sample size $n = 100$ and $\tilde{R}^2 = 0.35$ (top right) the average power of the test for the two predictors is 0.3836 and 0.4335. Finally, when the sample size $n = 100$ and $\tilde{R}^2 = 0.68$ (bottom right), the power of the test is close to 75% with both predictors.

Of course, the power of the test also depends on the combination of ρ and λ. Specifically, for small values of λ (e.g. $-0.2, 0.2$) the power is always consistently lower than in other cases. As an example, when $n = 100$ and $\tilde{R}^2 = 0.68$ the value of the power for small values of λ ranges between 0.28 and 0.50. When λ is large, the value of the power ranges between 0.81 and 1.00.

Finally, it should also be noted that the power of the test corresponding to the use of $y^{(1)}$ is quite close, on average, to the power corresponding to the use of $y^{(2)}$.

DOI: 10.1002/jae

Table IV. Frequency of rejection of the null hypothesis: two predictors (2000 replications)

		Fifth set of experiments: H_0: X_0, W_3, H_1: X_0, W_1, W_2							
		$\varepsilon_n = N(0, I_n)$ $n = 49$				$\varepsilon_n = N(0, I_n)$ $n = 100$			
		Size		Power		Size		Power	
		$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$	$y_i^{(1)}$	$y_i^{(2)}$
$\tilde{R}^2 = 0.35$									
ρ=−0.4	λ=−0.6	0.0515	0.0465	0.1315	0.1650	0.0455	0.0455	0.3865	0.4195
	λ=−0.4	0.0535	0.0420	0.0860	0.0820	0.0490	0.0495	0.1930	0.1705
	λ=−0.2	0.0500	0.0525	0.0490	0.0315	0.0580	0.0565	0.0920	0.0760
	λ=0.2	0.0500	0.0425	0.0610	0.0990	0.0665	0.0560	0.1195	0.1670
	λ=0.4	0.0605	0.0485	0.1430	0.3150	0.0725	0.0495	0.3745	0.5465
	λ=0.6	0.0695	0.0405	0.2720	0.6300	0.0940	0.0435	0.7335	0.9155
ρ=0.4	λ=−0.6	0.0620	0.0760	0.1250	0.1590	0.0545	0.0590	0.3940	0.4065
	λ=−0.4	0.0530	0.0840	0.0810	0.0780	0.0490	0.0595	0.2060	0.1805
	λ=−0.2	0.0585	0.0825	0.0500	0.0365	0.0505	0.0675	0.0825	0.0595
	λ=0.2	0.0505	0.0450	0.0630	0.1090	0.0515	0.0405	0.1095	0.1645
	λ=0.4	0.0605	0.0340	0.1485	0.3125	0.0540	0.0345	0.3890	0.5490
	λ=0.6	0.0620	0.0210	0.2880	0.6670	0.0670	0.0300	0.7345	0.9075
Averages		0.0568	0.0513	0.1248	0.2237	0.0593	0.0493	0.3179	0.3802
$\tilde{R}^2 = 0.68$									
ρ=−0.4	λ=−0.6	0.0505	0.0455	0.5655	0.6095	0.0505	0.0500	0.9675	0.9675
	λ=−0.4	0.0580	0.0520	0.3055	0.2885	0.0505	0.0530	0.7200	0.7085
	λ=−0.2	0.0570	0.0535	0.1020	0.0870	0.0595	0.0575	0.2335	0.2255
	λ=0.2	0.0520	0.0510	0.1575	0.2015	0.0635	0.0565	0.3480	0.4170
	λ=0.4	0.0685	0.0555	0.4965	0.6390	0.0605	0.0475	0.9285	0.9615
	λ=0.6	0.0750	0.0430	0.8540	0.9480	0.0750	0.0535	0.9995	1.0000
ρ=0.4	λ=−0.6	0.0545	0.0585	0.5375	0.5800	0.0535	0.0515	0.9465	0.9505
	λ=−0.4	0.0585	0.0710	0.2925	0.2990	0.0545	0.0565	0.7055	0.6945
	λ=−0.2	0.0590	0.0750	0.0975	0.0900	0.0565	0.0595	0.2380	0.2135
	λ=0.2	0.0470	0.0555	0.1405	0.1960	0.0455	0.0510	0.3580	0.4300
	λ=0.4	0.0565	0.0515	0.4990	0.6490	0.0500	0.0450	0.9315	0.9615
	λ=0.6	0.0710	0.0310	0.8810	0.9625	0.0610	0.0415	0.9995	1.0000
Averages		0.0590	0.0536	0.4108	0.4625	0.0567	0.0519	0.6980	0.7108

Table II is based on the third set of experiments. The structure of the table is identical to that of Table I. In these experiments the null and alternative models differ in their regressor matrices whereas the spatial weighting matrix is the same for both the null and alternative model.

Concentrating on Table II, it is possible to note that only the predictor $y^{(2)}$, on average, has the expected size. For the predictor $y^{(1)}$ when $N = 49$ (left part of the table), there are many cases in which the empirical size of the test exceeds the theoretical level (with both values of $\tilde{R}^2$). Interestingly, many of these cases are related to the positive value of $\rho = 0.4$. When the sample size increases (right part), both $y^{(1)}$ and $y^{(2)}$ have, on average, the expected size even if $\tilde{R}^2$ is only 0.35. When the difference between the null and alternative model pertains to the matrix of regressors, our J-test presents very high power even for relatively small sample size.

Table III is based on the fourth set of experiments and the sample size is $N = 49$. The null and alternative models differ in terms of both the matrix of regressors and the spatial weighting matrix. This set of experiments is designed in such a way that the error term of the model under the alternative follows the same specification of the error term under the null. Aside from parameter specifications, each section of Table III is made up of four columns. The results in the first two columns (labeled '2sls') are obtained by estimating the alternative model with a simple two-stage least square procedure

DOI: 10.1002/jae

which ignores the details of the error specification, e.g. spatial correlation. The last two columns (labeled '2sls + gm') report results obtained when the alternative model is estimated in the same way as the null: a procedure which accounts for spatial correlation. Both methods of estimation are consistent, but clearly the '2sls + gm' procedure is more efficient than the simple '2sls'.

The column averages of the empirical size of the test for both values of $\tilde{R}^2$ and estimation procedures are quite far away from the theoretical 5% level.[11]

Moving to the power of the test, we note from the table that, for all combinations of model parameters and estimation procedures considered, the power of the test is almost always equal to 1.0. In summary, the results in Table III suggest that there is almost no loss of power in the J-test if the alternative model is estimated by an inefficient procedure as long as that procedure is consistent. Consistent with the results in Table II, the results in Table III also suggest that if the null and alternative models differ in terms of their regressor matrices, the power of the J-test will be high.

The last two sets of experiments are specified in such a way that the null model is tested against two possible alternatives. In what follows, we only present the fifth set of experiments (Table IV), where the models under the alternative differ from the null only in terms of the spatial weighting matrix.

Looking at the column averages in Table 5, all of the empirical size of the test are close to the theoretical 0.05 level. However, the results corresponding to the individual size estimates are not always close to the 0.05 level. Comparing these results with the results in Table I, while there seems to be an improvement in term of size the values for the power of the test are lower when the alternative specification includes more than one possible model.

Summarizing our results, we can conclude that the size of the test is not always close to the nominal level when either the sample size is small or the error variance is large. The power of the test increases with sample size and there is not much difference if the alternative model is estimated efficiently other than consistently. However, the power of the test is lower when the alternative specification includes more than one possible alternative.

6. EMPIRICAL APPLICATION

The empirical application in this section is based on a well-known economic model of crime estimated by Cornwell and Trumbull (1994). They use panel data on 90 counties in North Carolina over the period 1981–1987. [12]

The empirical model relates the crime rate[13] to a series of variables which control for the return to legal opportunities, as well as a set of deterrent variables (such as probability of arrest, probability of conviction conditional on arrest and probability of imprisonment conditional on conviction).

The ratio of arrests to offenses is a proxy for the probability of arrest; the ratio of convictions to arrest is a proxy for the probability of conviction and, finally, the proportion of total convictions resulting in prison sentences is a proxy for the probability of imprisonment. The model also includes a measure of sanction severity measured by the average prison sentence length in days. All of the other variables are either observable county characteristics or controls for the relative return to legal activities.[14]

Cornwell and Trumbull (1994) estimate the model both by the between and the within estimators. In their estimation, they account for the endogeneity of the police per capita and the probability of arrest

[11] However, when the sample size increases ($N = 100$) the problem improves and the empirical size of the test is, on average, close to the nominal value when $\tilde{R}^2 = 0.68$. These results are available from the authors.

[12] This dataset is well known because it is one of the datasets in Baltagi (2008). The data are available from the website associated with Baltagi's book.

[13] The crime rate variable is the ratio between an FBI index that measures the number of crimes and county population (i.e. crime per capita in the county).

[14] For greater detail on the data see Baltagi (2008).

DOI: 10.1002/jae

variables. They use two instruments for these variables, namely an offense mix variable (the ratio of crimes involving face-to-face contacts to those that do not) and per capita tax revenue.[15]

As suggested by Brueckner (2003), a crime model is a strategic interactions model and, in particular, a spillover model. In this framework each county 'chooses' the level of crime by choosing the level of enforcement, e.g. the extent of the police force. At the same time, however, the county will be affected by the decision process made by other counties, thus indicating the presence of spillover.[16]

The aim of our empirical application is twofold. First, we want to take a spatial explicit approach in estimating the crime model of Cornwell and Trumbull (1994). Additionally, we want to test the hypothesis that the spatial weights matrix should be based on a contiguity criterion and not on a distance-based approach. The explanation behind this second hypothesis is that county i may only be concerned of the decisions that immediate neighbors are taking and ignore counties that are more distant.

With our J-test we wish to test two competing non-nested alternatives. The null model is a modification of the one estimated by Cornwell and Trumbull (1994), which includes a spatial lag of the dependent variable (i.e. the spatial lag of the crime rate) as well as spatial correlation in the errors. This null model is specified in terms of a spatial weights matrix based on the 10 nearest neighbors. On the other hand, the alternative model is identical to the null except that the spatial weights matrix is specified in terms of contiguity.[17]

As indicated, the J-test is based on augmenting the null model by predictions based on the alternative model. The procedure then is to test for the significance of the augmenting variable.

At the 5% level the J-test rejects the null model since the chi-squared variable $= 6.373 > \chi^2_1 = 3.841$.[18] We conclude, then, that counties only consider policies implemented in boarder counties and do not consider those of counties that are more distant.

7. CONCLUSIONS

In the present paper we have extended the J-test to a panel data framework. Our J-test is suitable for testing a null spatial model against one or more alternative spatial models. These alternative models may differ from the null either in their regressor matrix, their weighting matrix or both. The J-test is not appropriate for testing an alternative model that only differs from the null in its error specifications.

Our suggested test is computationally simple. Its size in relatively small panel data samples is reasonably close to the theoretical size, except when the sample size is very small. On the other hand, the power of the test is high and seems to be only mildly affected by the number of models in the alternative.

A suggestion for future research would be an extension of our results to panel data models that has fixed effects, as well as both spatially and time-lagged dependent variables. Another suggestion for future research would be an extension to a nonlinear spatial panel framework. Among others, such a framework would arise in a qualitative or a limited dependent variable setting.

REFERENCES

Anselin L, Le Gallo J, Jayet H. 2008. Spatial panel econometrics. In *The Econometrics of Panel Data: Fundamentals and Recent Developments in Theory and Practice* (3rd edn), Matyas L, Sevestre P (eds). Springer: Berlin; 624–660.

[15] See Baltagi (2008) for further explanation on the instruments.

[16] In other terms we are estimating a reaction function which gives county i 's best reaction to the choices of other counties.

[17] The results of the estimation of the null, alternative and augmented models are available from the authors.

[18] This result is obtained with the first predictor. The chi-squared variable calculated with the second predictor is 3.906.

DOI: 10.1002/jae

Arraiz I, Drukker DM, Kelejian HH, Prucha IR. 2010. A spatial Cliff–Ord-type model with heteroskedastic innovations: small and large sample results. *Journal of Regional Science* **50**(2): 592–614.
Baltagi BH. 2008. *Econometric Analysis of Panel Data* (4th edn). Wiley: New York.
Baltagi BH, Liu L. 2008. Testing for random effects and spatial lag dependence in panel data models. *Statistics and Probability Letters* **78**: 3304–3306.
Baltagi BH, Song SH, Koh W. 2003. Testing panel data regression models with spatial error correlation. *Journal of Econometrics* **117**: 123–150.
Baltagi BH, Egger P, Pfaffermayr M. 2007a. Estimating models of complex FDI: are there third-country effects? *Journal of Econometrics* **140**: 260–281.
Baltagi BH, Song SH, Jung BC, Koh W. 2007b. Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics* **140**(1): 5–51.
Baltagi BH, Egger P, Pfaffermayr M. 2009. A generalized spatial panel data model with random effects. CPR Working Paper Series.
Brueckner JK. 2003. Strategic interactions among governments: an overview of empirical studies. *International Regional Science Review* **26**(2): 175–188.
Cornwell C, Trumbull WN. 1994. Estimating the economic model of crime with panel data. *Review of Economics and Statistics* **76**(2): 360–366.
Dastoor N. 1983. Some aspects of testing non-nested hypothesis. *Journal of Econometrics* **21**: 213–228.
Davidson R, MacKinnon J. 1981. Several tests for model specification in the presence of alternative hypotheses. *Econometrica* **49**: 781–794.
Debarsy N, Ertur C. 2010. Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science and Urban Economics* **40**: 453–470.
Delgado M, Stengos T. 1994. Semiparametric specification testing of non-nested econometric models. *Review of Economic Studies* **61**: 291–303.
Donovan GH, Champ PA, Butry DT. 2007. Wildfire risk and housing prices: a case study from Colorado Springs. *Land Economics* **83**: 217–233.
Elhorst JP. 2003. Specification and estimation of spatial panel data models. *International Regional Sciences Review* **26**(3): 244–268.
Elhorst JP. 2008. Serial and spatial error correlation. *Economics Letters* **100**: 422–424.
Elhorst JP. 2009. Spatial panel data models. In *Handbook of Applied Spatial Analysis*, Fischer MM, Getis A (eds). Springer: Berlin; 377–407.
Elhorst JP. 2010. Dynamic panels with endogenous interactions effects when T is small. *Regional Science and Urban Economics* **40**: 272–282.
Elhorst JP, Freret S. 2009. Yardstick competition among local governments: French evidence using a two-regimes spatial panel data model. *Journal of Regional Science* **49**: 931–951.
Elhorst JP, Piras G, Arbia G. 2010. Growth and convergence in a multi-regional model with space-time dynamics. *Geographical Analysis* **42**: 338–355.
Godfrey L. 1983. Testing non-nested models after estimation by instrumental variables or least squares. *Econometrica* **51**: 355–366.
Greene WH. 2003. *Econometric Analysis* (5th edn). Prentice Hall: Upper Saddle River, NJ.
Kapoor M, Kelejian HH, Prucha IR. 2007. Panel data model with spatially correlated error components. *Journal of Econometrics* **140**(1): 97–130.
Kelejian HH. 2008. A spatial J-test for model specification against a single or a set of nonnested alternatives. *Letters in Spatial and Resources Sciences* **1**(1): 3–11.
Kelejian HH, Piras G. 2011. An extension of Kelejian?s J-test for non-nested spatial models. *Regional Science and Urban Economics* **41**: 281–292.
Kelejian HH, Prucha IR. 1999. A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review* **40**(2): 509–533.
Kelejian HH, Prucha IR. 2004. Estimation of systems of spatially interrelated cross sectional equations. *Journal of Econometrics* **118**: 27–50.
Kelejian HH, Prucha IR. 2007. HAC estimation in a spatial framework. *Journal of Econometrics* **140**: 131–154.
Kmenta J. 1986. *Elements of Econometrics* (2nd edn). Macmillan: New York.
Lee LF, Yu J. 2010a. Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics* **154**: 165–185.
Lee LF, Yu J. 2010b. Some recent development in spatial panel data models. *Regional Science and Urban Economics* **40**: 255–271.
Lee LF, Yu J. 2010c. A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory* **26**: 564–597.

DOI: 10.1002/jae

Lee LF, Yu J. 2010d. A unified transformation approach to the estimation of spatial dynamic panel data models: stability, spatial cointegration and explosive roots. In *Handbook of Empirical Economics and Finance*, Ullah A, Giles DEA (eds). Chapman & Hall/CRC: London; 397–434.
MacKinnon J, White H, Davidson R. 1983. Test for model specification in the presence of alternative hypotheses: some further results. *Journal of Econometrics* **21**(1): 53–70.
Mutl J, Pfaffermayr M. 2011. The Hausman test in a Cliff and Ord panel model. *Econometrics Journal* **14**: 48–76.
Parent O, LeSage J. 2010. A spatial dynamic panel model with random effects applied to commuting times. *Transportation Research Part B* **44**: 633–645.
Pesaran HM. 1974. On the general problem of model selection. *Review of Economic Studies* **41**: 153–171.
Pesaran HM. 1982. Comparison of local power of alternative tests of non-nested regression models. *Econometrica* **50**: 1287–1305.
Pesaran HM, Deaton AS. 1978. Testing non-nested non-linear regression models. *Econometrica* **46**: 677–694.
Pesaran HM, Tosetti E. 2011. Large panels with common factors and spatial correlations. *Journal of Econometrics* **161**(2): 182–202.
Pesaran HM, Weeks M. 2001. Non-nested hypothesis testing: an overview. In *A Companion to Theoretical Econometrics*, Baltagi B (ed). Blackwell: Malden, MA; 279–309.
Piras G. 2013. Efficient GM estimation of a Cliff and Ord panel data model with random effects. *Spatial Economic Analysis* **8**(3): 370–388.
Piras G, Lozano-Gracia N. 2012. Spatial J-test: some Monte Carlo evidence. *Statistics and Computing* **22**(1): 169–183.
Yu J, Lee LF. 2010. Estimation of unit root spatial dynamic panel data models. *Econometric Theory* **26**: 1332–1362.
Yu J, de Jong R, Lee LF. 2008. Quasi maximum likelihood estimators for spatial dynamic panel data with fixed effects when both n and T are large. *Journal of Econometrics* **146**: 118–134.

APPENDIX

The VC matrix of η_J. The model for $y_{i,t}$ under $H_{1,J}$ in (5) is

$$y_{i,t} = P_{i,t,J}\beta_J + \lambda_J M_{i,J} y_t + Y_{i,t,J} d_J + \mu_J + \psi_{i,t,J} \tag{A.1}$$

where $P_{i,t,J}$ and $M_{i,J}$ are, respectively, the ith rows of $P_{t,J}$ and M_J, and $\psi_{i,t,J}$ is the ith element of $\psi_{t,J}$. The mean of $y_{i,t}$ given $\Phi_{2,i,t,J} = [\Phi_{1,J}, M_{i,J} y_t, Y_{i,t,J}]$ is

$$E(y_{i,t}|H_{1,J}, \Phi_{2,i,t,J}) = P_{i,t,J}\beta_J + \lambda_J M_{i,J} y_t + Y_{i,t,J} d_J + \mu_{i,J} + \eta_{i,t,J} \tag{A.2}$$

where $\eta_{i,t,J} = E(\psi_{i,t}|H_{1,J}, \Phi_{2,i,t,J}) \neq 0$, because of the correlation of $M_{i,J} y_t$ and $Y_{i,t,J}$ with $\psi_{i,t,J}$. However, using iterated expectations,

$$\begin{aligned} E(\eta_{i,t,J}|H_{1,J}, \Phi_{1,J}) &= E\left[E(\psi_{i,t,J}|H_{1,J}, \Phi_{2,i,t,J})|H_{1,J}, \Phi_{1,J}\right] \\ &= E\left[E\left(\psi_{i,t,J}|H_{1,J}, \Phi_{1,J}, M_{i,J} y_t, Y_{i,t,J}\right)|H_{1,J}, \Phi_{1,J}\right] \\ &= E\left(\psi_{i,t,J}|H_{1,J}, \Phi_{1,J}\right) = 0 \end{aligned} \tag{A.3}$$

Since $\eta_{i,t,J}$ is the i,tth element of η_J, it follows from equation (A.3) that $E(\eta_J|H_{1,J}, \Phi_{1,J}) = 0$. Since $E(\psi_{i,t,J}|H_{1,J}, \Phi_{2,i,t,J}) = \eta_{i,t,J}$, given $H_{1,J}$, $\eta_{i,t,J}$ is a function of $\Phi_{2,i,t,J}$, say $f(\Phi_{2,i,t,J})$, since a conditional mean is a function of the conditioning variables. It also follows that $\psi_{i,t,J}$ can be expressed as

$$\psi_{i,t,J} = \eta_{i,t,J} + \varphi_{i,t,J};\ E\left(\varphi_{i,t,J}|H_{1,J}, \Phi_{2,i,t,J}\right) = 0 \tag{A.4}$$

and so

$$E(\varphi_{i,t,J}\ \eta_{i,t,J}|H_{1,J}, \Phi_{2,i,t,J}) = \eta_{i,t,J} E(\varphi_{i,t,J}|H_{1,J}, \Phi_{2,i,t,J}) = 0 \tag{A.5}$$

DOI: 10.1002/jae

Again, via iterated expectations,

$$E(\varphi_{i,t,J}\ \eta_{i,t,J}|H_{1,J},\Phi_{1,J})=0 \tag{A.6}$$

Using evident notation, let φ_J and η_J be the $NT\times 1$ vectors of $\varphi_{i,t,J}$ and $\eta_{i,t,J}, i=1,\ldots,N; t=1,\ldots,T;\ J=1,\ldots,G$. Then, from equations (8) and (A.3)–(A.6):

$$\begin{aligned}E(\psi_J\psi_J'|H_{1,J},\Phi_{1,J}) &= \Omega_{\psi_J}\\ &= \Omega_{\varphi_J}+\Omega_{\eta_J}\end{aligned}$$

where Ω_{φ_J} and Ω_{η_J} are the VC matrices, respectively, of φ_J and η_J. Since the row and column sums of Ω_{ψ_J} are uniformly bounded in absolute value so are the row and column sums of Ω_{φ_J} and Ω_{η_J}.

The asymptotic negligibility of η_J when G=1. Since $Q_0(e_T\otimes\mu_J)=0$, the second line of equation (10) follows from the stacked version of equation (A.2). If the researcher observed η_J and so took used $Q_0\eta_j$ as part of the predictor $Q_0E(y|H_{1,J},\Phi_{2,J})$ in equation (10), the only term involving $Q_0\eta_j$ in $(NT)^{1/2}[\hat{\delta}-\delta]$ based on equation (16) and (17) would be Δ_J where

$$\begin{aligned}\Delta_J &= (NT)^{-1}\left[H'\left(I_{NT}-\hat{\rho}_0(I_T\otimes W)\right)\right]Q_0\eta_J\\ &= (NT)^{-1}H'Q_0\eta_J-\hat{\rho}_0(NT)^{-1}\left[H'(I_T\otimes W)\right]Q_0\eta_J\end{aligned}$$

Given the results in Mutl and Pfaffermayr (2011) and Kapoor *et al.* (2007), $\hat{\rho}_0\xrightarrow{P}\rho_0$. Assume as in Mutl and Pfaffermayr (2011) and Kapoor *et al.* (2007), that the row and column sums of W are uniformly bound in absolute value, that the elements of H are uniformly bounded in absolute value and $(NT)^{-1}H'H\to Q_{HH}$, where Q_{HH} is finite and nonsingular.

Let $F_{J,1}=(NT)^{-1}H'Q_0\eta_J$ and $F_{J,2}=(NT)^{-1}\left[H'(I_T\otimes W)\right]Q_0\eta_J$ so that $\Delta_J=F_{J,1}+\hat{\rho}_0F_{J,2}$. From equation (A.3) we have $E(F_{J,q}|H_{1,J},\Phi_{1,J})=0$, $q=1,2$. Also, the VC matrices of $F_{J,1}$ and $F_{J,2}$ are, respectively, $\Omega_{F_{J,1}}=(NT)^{-2}H'Q_0\Omega_{\eta_J}Q_0H$ and $\Omega_{F_{J,2}}=(NT)^{-2}\left[H'(I_T\otimes W)\right]Q_0\Omega_{\eta_J}Q_0(I_T\otimes W')H]$. Since the row and column sums of W, Q_0 and Ω_{η_J} are uniformly bounded in absolute value, and the elements of H are uniformly bounded, the elements of $\Omega_{F_{J,1}}$ and $\Omega_{F_{J,1}}$ are all $0(N^{-1})\to 0$. Since $\hat{\rho}_0\xrightarrow{P}\rho_0$, $\Omega_{F_{J,q}}\xrightarrow{P}0$ and so, by Chebyshev's inequality, $\Delta_J\xrightarrow{P}0$.

DOI: 10.1002/jae