# Regional Research Institute

## Working Paper Series

## A *J*-test for Panel Models with Fixed Effects, Spatial and Time Dependence

By: Harry H. Kelejian, Professor of Economics, University of Maryland and

Gianfranco Piras, Research Assistant Professor, West Virginia University

Working Paper Number 2013-03

Website address: rri.wvu.edu

# Regional Research Institute

## Working Paper Series

## A *J*-test for Panel Models with Fixed Effects, Spatial and Time Dependence

By: Harry H. Kelejian, Professor of Economics, University of Maryland and

Gianfranco Piras, Research Assistant Professor, West Virginia University

Working Paper Number 2013-01

Website address: rri.wvu.edu

# A $J$-test for panel models with fixed effects, spatial and time dependence*

Harry H. Kelejian
University of Maryland
College Park, MD 20742, USA

Gianfranco Piras†
West Virginia University
Morgantown, WV 26506, USA

May 13, 2013

*We would like to thank the referees and the editor Hashem Pesaran for helpful comments on an earlier version of this paper. They are not, however, responsible for any shortcoming remaining in this paper.

†Piras is the contacting author: Regional Research Institute, West Virginia University, 886 Chestnut Ridge Road, Room 510, P.O. Box 6825, Morgantown, WV 26506-6825; E-mail: gpiras@mac.com; Tel: 304-293-2643; Fax: 304-293-6699.

**Abstract**

In this paper we suggest a $J$-test in a spatial panel framework of a null model against one or more alternatives. The null model we consider has fixed effects, along with spatial and time dependence. The alternatives can have either fixed or random effects. We implement our procedure to test the specifications of a demand for cigarette model. We find that the most appropriate specification is one that contains the average price of cigarettes in neighboring states, as well as the spatial lag of the dependent variable. Along with formal large sample results, we also give small sample Monte Carlo results. Our large sample results are based on the assumption $N \to \infty$ and $T$ is fixed. Our Monte Carlo results suggest that our proposed $J$-test has good power, and proper size even for small to moderately sized samples.

*JEL classification: C01, C12*

*Key Words: Spatial Panel Models, Fixed Effects, Time and Spatial Lags, Non-nested J-test*

## 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]

[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] See, e.g., Anselin et al. (2008); Kapoor et al. (2007); Baltagi et al. (2007c, 2003); Baltagi and Liu (2008); Baltagi et al. (2007a, 2013); Debarsy and Ertur (2010); Elhorst (2003); Elhorst and Freret (2009); Elhorst

In this paper we generalize these earlier works on the $J$-test to a panel data framework. We specify a null model containing fixed effects, a spatially lagged dependent variable, and a time lagged dependent variable. The disturbance term is specified non-parametrically and allows for general patterns of spatial and time correlation, as well as heteroskedasticity. We allow for $G$ alternative models which are specified in such a way that both spatial and time correlation of various sorts, as well as general patterns of heteroskedasticity are special cases. However, we note that our $J$-test procedure is not suitable for testing two models which *only* differ in that one has fixed effects while the other has random effects.[3]

As in Kelejian and Piras (2011) we show that, given reasonable assumptions, the full information $J$-test in a panel is computationally simple, and indeed, simpler than the tests suggested in Kelejian (2008). We also illustrate that these assumptions would typically be satisfied in most spatial models.

We give large sample results, as well as small sample Monte Carlo results. Our Monte Carlo results suggest that our proposed $J$-test has good power, and proper size even for small to moderately sized samples. We also implement our procedure to test the specifications of a demand for cigarette model which has appeared numerous times in the literature (Baltagi and Levin, 1986, 1992). Using our $J$-test, we find that the most appropriate specification is one that includes, along with the average price of cigarettes in neighboring states, the spatial lag of the dependent variable.

In Section 2 we specify the null model, while the alternative models are specified in Section 3. Section 4 is devoted to a discussion of the underlying features of the $J$-test. Formal model specifications are given in Section 5, along with their interpretations. The $J$-test is given in Section 6. Section 7 describes a dynamic demand for cigarettes model which we estimate and then test using our proposed $J$-test procedure. Empirical results relating to this demand for cigarette model are also given in this section. Section 8 describes the Monte Carlo model used to study the small sample behavior of our proposed test, while the results of our Monte

(2008, 2009, 2010); Elhorst et al. (2010); Lee and Yu (2010c,a,b,d); Mutl and Pfaffermayr (2011); Pesaran and Tosetti (2011); Yu and Lee (2010); Yu et al. (2008); Parent and LeSage (2010).

[3] In a slightly different context Mutl and Pfaffermayr (2011) suggest and give large sample results for a Hausman test to discriminate between such models.

Carlo study are given in Section 9. Conclusions and suggestions for further work are given in Section 10. Technical details are relegated to the appendices.

## 2 The null model

Consider the model corresponding to $N$ cross sectional units at time $t$

$$\begin{aligned} H_0 &: \\ y_t &= X_t\beta_1 + P_t\beta_2 + \lambda_0 W y_t + \alpha_0 y_{t-1} + \mu + u_t; \\ t &= 1, ..., T \end{aligned} \tag{1}$$

where $y_t$ is the $N \times 1$ vector of observations on the dependent variable at time $t$; $X_t$ is a $N \times k_x$ matrix of observations at time $t$ on $k_x$ exogenous variables which vary with respect to both time and cross sectional units; $P_t$ is an $N \times (T-1)$ matrix of observations on $T-1$ time dummy variables; $W$ is an $N \times N$ observed exogenous weighting matrix, $\mu$ is an $N \times 1$ vector of fixed effects, and $u_t$ is the corresponding $N \times 1$ disturbance vector.[4] The available data are from $t = 0, ..., T$, so that $y_t$, $y_{t-1}, X_t$, and $P_t$ are observed for all $t = 1, ..., T$.

The regression parameters of the model are $\beta_1$, and $\beta_2$ which are, respectively, $k_x \times 1$ and $(T-1) \times 1$ vectors, and $\lambda_0$ and $\alpha_0$ which are both scalars. Our formal list of assumptions is given below; at this point we note that our large sample theory relates to $N \to \infty$, with $T$ fixed. We allow for triangular arrays but do not index the variables in (1) with the sample size in order to simplify the notation.

Let $e_T$ be a $T \times 1$ vector of unit elements. Then, stacking the model in (1) over $t = 1, ..., T$ yields

$$\begin{aligned} y = {} & X\beta_1 + P\beta_2 + \lambda_0(I_T \otimes W)y + \\ & \alpha_0 y_{-1} + (e_T \otimes I_N)\mu + u \end{aligned} \tag{2}$$

[4] Clearly if the model also had regressors which only varied cross sectionally, the corresponding coefficients would not be identified. This was the case in a paper by Kelejian et al. (2013).

where $y' = [y'_1, ..., y'_T]$, $X^{'} = [X'_1, ..., X'_T]$, $P' = (P'_1, ..., P'_T)$, $y_{-1}$ is identical to $y$ except all of its elements are lagged one time period, and $u' = [u'_1, ..., u'_T]$.

Let $\varepsilon' = [\varepsilon'_1, ..., \varepsilon'_T]$, where $\varepsilon_t$, $t = 1, ..., T$, is a $N \times 1$ vector of random elements. At this point we assume that $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = I_{NT}$. More formal assumptions are given below. Given this notation, we assume the following structure for the $NT \times 1$ disturbance vector $u$

$$u = R\varepsilon \tag{3}$$

where $R$ is an $NT \times NT$ unknown lower triangular block nonstochastic matrix with $N \times N$ blocks $R_{ij}, i, j = 1, ..., T$ where $R_{ij} = 0$ if $j > i$. Clearly, the form of $R$, namely

$$R = \begin{bmatrix} R_{11} & 0 & . & . & . & 0 \\ R_{21} & R_{22} & 0 & . & . & 0 \\ . & . & . & 0 & . & 0 \\ . & . & . & . & 0 & 0 \\ . & . & . & . & . & 0 \\ R_{T1} & R_{T2} & . & . & . & R_{TT} \end{bmatrix} \tag{4}$$

restricts each element of $u$ from depending upon future elements of $\varepsilon$. It also permits the elements of $u$ to be spatially and time autocorrelated, as well as heteroskedastic.

# 3 Alternatives under $H_1$

The alternatives under $H_1$ correspond to (2) in that they have the same structure but they may have different regressors, weighting matrices, or error term specifications. These models may have fixed or random effects, and a disturbance term which may be spatially and time correlated, as well as heteroskedastic. However, we do not allow for alternatives which only differ from the null in their error term specification.

Using evident notation, we assume the alternatives

$$\begin{aligned} H_1 &: \\ y_t &= M_{J,t}\phi_{1,J} + P_t\ \phi_{2,J} + \lambda_J W_J y_t + \alpha_J y_{t-1} + \theta_J + \nu_{J,t} \\ J &= 1, ..., G;\ t = 1, ..., T \end{aligned} \tag{5}$$

where for the $J^{th}$ model under $H_1$, $M_{J,t}$ is an $N \times k_{M_J}$ exogenous regressor matrix whose elements vary with respect to both time and cross sectional units; $P_t$ is defined in (1) above; $W_J$ is an $N \times N$ matrix of observations on an exogenous weighting matrix, $\theta_J$ is the vector of fixed or random effects, and $\nu_{J,t}$ is the $N \times 1$ error vector. Finally, $\phi_{1,J}$ and $\phi_{2,J}$, are conformably defined parameter vectors, and $\lambda_J$, and $\alpha_J$ are scalar parameters. Stacking (5) over $t = 1, ..., T$ yields

$$\begin{aligned} y &= M_J\phi_{1,J} + P\phi_{2,J} + \lambda_J(I_T \otimes W_J)y + \\ &\quad \alpha_J y_{-1} + (e_T \otimes I_N)\theta_J + \nu_J \\ J &= 1, ..., G; \end{aligned} \tag{6}$$

where, $M_J' = [M_{J,1}', ..., M_{J,T}']$, $\nu_J' = (\nu_{J,1}', ..., \nu_{J,T}')$, and $y, P$, and $y_{-1}$ where defined in (2). Denote the $J^{th}$ model under $H_1$ as $H_{1,J}$. At this point we assume $E(\nu_J|H_{1,J}, \Upsilon_J) = 0$ and $E(\nu_J\nu_J'|H_{1,J}, \Upsilon_J) = V_J$, where $\Upsilon_J = (M_J, P, W_J)$. Further assumptions are given in Section 5.

**A suggestion relating to nested models and the $J$-test.**

As indicated above, the $J$-test is a procedure for testing a null model against a non-nested alternative. Although the procedure can be applied to certain nested models, we recommend against doing so. As an example, using evident notation, the following hypotheses may be of interest to some researchers

$$H_0 \quad : \quad y = XB_1 + (I_T \otimes W)XB_2 + \lambda_1(I_T \otimes W)y + u$$

$$H_1 \quad : \quad y = XB_3 + \lambda_2(I_T \otimes W)y + u$$

The null model is commonly referred to as a Durbin model; the alternative is often called a spatial lag model. Note that $H_1$ is nested in $H_0$. A simple test of $H_0$ against $H_1$ would be a $\chi^2$ test relating to the significance of the estimator of $B_2$. On the other hand, the $J$-test can not be applied in this case because the predicted value of the dependent variable based on $H_1$ would be perfectly collinear in the variables under $H_0$. However, if $H_1$ were the null model, and $H_0$ were the alternative model, the $J$-test could be applied because the predicted value of the dependent variable based on the alternative, which is $H_0$ in this case, would not be perfectly collinear with the variables of the null. Despite this, most researchers would just estimate the full model under $H_0$ above and use a $\chi^2$ test for the significance of $\beta_2$.

# 4 The components of the $J$-tests

In this section we describe the components underlying our suggested $J$-test. Specifically, after some preliminaries, we develop the augmented equation our test is based on. That equation involves predictions of the dependent variable under the alternative hypothesis. These predictions are based on information sets which are also given in this section. We also develop notation for the instruments which we suggest for estimating the augmented equation. In the next section we give our modeling assumptions which use the notation developed in this section.

**Some preliminaries**

Let

$$Q_0 = (I_T - \frac{1}{T}J_T) \otimes I_n; \;\; J_T = e_T e_T' \tag{7}$$

and note that

$$Q_0(I_T \otimes A) = (I_T \otimes A)Q_0 \qquad (8)$$
$$Q_0(e_T \otimes I_N) = 0$$

where $A$ is any $N \times N$ matrix.[5] It then follows from (2) that pre-multiplying the null model by $Q_0$[6] eliminates the vector $(e_T \otimes I_N)\mu$ and therefore

$$\begin{aligned} H_0 &: \qquad (9) \\ Q_0 y &= Q_0 X\beta_1 + Q_0 P\beta_2 + \lambda_0 Q_0(I_T \otimes W)y + \alpha_0 Q_0 y_{-1} + Q_0 u \\ &= Q_0 Z\gamma_0 + Q_0 R\varepsilon \\ Z &= (X, P, (I_T \otimes W)y, y_{-1}); \gamma_0' = (\beta_1', \beta_2', \lambda_0, \alpha_0) \end{aligned}$$

For future reference note that the pre-multiplication of the alternative models in (6) by $Q_0$ yields

$$\begin{aligned} H_1 &: \qquad (10) \\ Q_0 y &= Q_0 M_J\phi_{1,J} + Q_0 P\phi_{2,J} + \lambda_J Q_0(I_T \otimes W_J)y + \alpha_J Q_0 y_{-1} + Q_0\nu_J;\ J = 1, ..., G \\ &= Q_0 Z_J\gamma_J + Q_0\nu_J \\ Z_J &= [M_J, P, (I_T \otimes W_J)y, y_{-1}];\ \gamma_J' = (\phi_{1,J}', \phi_{2,J}', \lambda_J, \alpha_J) \end{aligned}$$

**The augmented equation**

[5] See, e.g., Kapoor et al. (2007).

[6] There is more than one way to eliminate fixed effects. We chose to eliminate them by premultiplying the model by $Q_0$ because it was convenient. We note that some researchers may eliminate the fixed effects from a dynamic panel model by taking a time difference. This is done because, under certain assumptions, it facilitates the use of specific time lagged dependent variables as part of the instrument set. We did not take this approach for at least two reasons. First, our error specification in (3) implies that we are allowing for a general pattern of time series correlation in the errors and, therefore, time lagged values of the dependent variable can not be used as instruments. Second, even if they could be used as instruments, there is, at present, no central limit theorem that could then be applied to obtain the large sample distribution of the model parameter estimators because of our general model specifications.

The $J$-test is based on augmenting the null model in (9) by the predicted values of $Q_0y$ based on each of the $G$ alternatives under $H_1$ and then testing for the significance of the augmenting variables.[7] Denote the $J^{th}$ alternative under $H_1$ as $H_{1,J}$, and let $INFO_J$ be the information set underlying the prediction of $Q_0y$ under $H_{1,J}$. Possible information sets are described below. At this point let

$$\begin{aligned} Y_J^+ &= E[Q_0y|H_{1,J}, INFO_J] \\ &= Q_0E[y|H_{1,J}, INFO_J] \\ &= Q_0Y_J^E,\ J = 1, ..., G \end{aligned} \tag{11}$$

where $Y_J^E = E[y|H_{1,J}, INFO_J]$. The information sets we consider are described below. At this point we note that these sets contain variables of the model which are correlated with $Q_0\nu_J$ and so $E[y|H_{1,J}, INFO_J]$ involves $E[\nu_J|H_{1,J}, INFO_J] \neq 0$ - see, e.g., (10).

**A result relating to the case G=1**

Results for the general case in which $G > 1$ are given below. At this point we note a result relating to the typical case in which $G = 1$. Specifically, under reasonable conditions, we demonstrate in the appendix that, asymptotically, if $G = 1$ and the true model under the alternative is $H_{1,J}$, the term $E[\nu_J|H_{1,J}, INFO_J]$ can be ignored even if it is observed.[8] We also note that $Q_0Y_J^E$ can be taken as one of two forms, given below, which are both efficient as well as asymptotically equivalent in terms of the power of the $J$-test. In finite samples results relating to these two forms need not be identical.

**The general case: The two forms**

Assuming the inverse exists,[9] let

$$\Pi_J = (I_T \otimes (I_N - \lambda_J W_J)^{-1}), J = 1, ..., G \tag{12}$$

[7] Essentially, the rational of the $J$-test is that if the null model is correct, these augmenting variables should not add to the explanation of the dependent variable in (9).

[8] Assuming a linear conditional mean, $E[\nu_J|H_{1,J}, INFO_J]$ can be estimated given reasonable data assumptions since it only involves variances and covariances of the elements of $\nu_J$ and $INFO_J$.

[9] Further specifications are given below.

Then, the two forms are $Q_0 Y_J^{E,A}$ and $Q_0 Y_J^{E,B}$ where

$$\begin{aligned} Q_0 Y_J^{E,A} &= Q_0 M_J \phi_{1,J} + Q_0 P \phi_{2,J} + \lambda_J Q_0 (I_T \otimes W_J) y + \alpha_J Q_0 y_{-1} \\ Q_0 Y_J^{E,B} &= Q_0 \Pi_J M_J \phi_{1,J} + Q_0 \Pi_J P \phi_{2,J} + \alpha_J Q_0 \Pi_J y_{-1}, J = 1, ..., G \end{aligned} \tag{13}$$

Note that $Q_0 Y_J^{E,A}$ corresponds to the right hand side of the null model, while $Q_0 Y_J^{E,B}$ corresponds to its reduced form.

Let $\hat{\phi}_{1,J}$, $\hat{\phi}_{2,J}, \hat{\lambda}_J$, and $\hat{\alpha}_J$ be the estimated values of $\phi_{1,J}, \phi_{2,J}, \lambda_J$, and $\alpha_J$ based on the specifications of $H_{1,J},\ J = 1, ..., G$. Let

$$\begin{aligned} \overbrace{Q_0 Y_J^{E,A}} &= Q_0 M_J \hat{\phi}_{1,J} + Q_0 P \hat{\phi}_{2,J} + \hat{\lambda}_J Q_0 (I_T \otimes W_J) y + \hat{\alpha}_J Q_0 y_{-1} \\ \overbrace{Q_0 Y_J^{E,B}} &= Q_0 \Pi_J M_J \hat{\phi}_{1,J} + Q_0 \Pi_J P \hat{\phi}_{2,J} + \hat{\alpha}_J Q_0 \Pi_J y_{-1} \end{aligned} \tag{14}$$

Finally, let

$$\hat{Y}_{1,G}^i = [\overbrace{Q_0 Y_1^{E,i}}, ..., \overbrace{Q_0 Y_G^{E,i}}], i = A, B \tag{15}$$

and note that $\hat{Y}_{1,G}^i$ is an $NT \times G$ matrix.

Given this notation, and for a preselected value of $i = A, B$,[10] our augmented equation for the $J$-test is

$$\begin{aligned} Q_0 y &= Q_0 Z \gamma_0 + Q_0 \hat{Y}_{1,G}^i \psi_i + Q_0 R \varepsilon \\ &= Q_0 \hat{\Gamma}_i \xi_i + Q_0 R \varepsilon, i = A, B \end{aligned} \tag{16}$$

where $\psi_i$ is a $G \times 1$ vector of parameters, $\hat{\Gamma}_i = [Z, \hat{Y}_{1,G}^i]$ and $\xi_i' = (\gamma_0', \psi_i'), i = A, B$. The $J$-test relates to the significance of the estimator of $\psi_i$ based on (16). Details concerning this are given below.

**The information sets**

We consider two information sets for the $J^{th}$ model, for each time $t = 1, ..., T$, under $H_1$.

[10] We indicate that the value of $i$ should be preselected in the test in order to avoid a selection which is based on the results obtained.

The first is a minimum set in the sense that it only relates to the variables that enter the right hand side of the model in (5) with the exception of $\theta_J$ because this term does not appear in the augmented model (16), nor is it relevant for estimating that model. The second is a full information set in the sense that it includes all possible variables that could be relevant in the prediction described in (13).

Our minimum and maximum information sets are

$$\begin{aligned} INFO_{J,t}^{\min} &= (\Upsilon_J, W_J y_t, y_{t-1}) \\ INFO_{J,t}^{\max} &= (INFO_{J,t}^{\min}, y_{-i,t}, y_0, ..., y_{t-1});\ J = 1, ..., G \end{aligned}$$

where $y_{-i,t}$ is identical to $y_t$ except it does not contain its $i^{th}$ element. Note that, since the diagonal elements of the weighting matrices are zero,[11] $W_J y_t$ does not involve the $i^{th}$ element of $y_t$. As a consequence, $W_J y_t$ is given if both $W_J$ and $y_{-i,t}$ are given. Clearly, $INFO_{J,t}^{\min}$ relates to the variables that enter the transformed model (10). On the other hand, $INFO_{J,t}^{\max}$ contains all of the elements of $INFO_{J,t}^{\min}$ as well as all of the lagged dependent variable vectors which would be available at time $t$. These additional vectors could be of use in predicting if the error term is time autocorrelated.

**The instruments**

Under our assumptions given in the next section the spatially lagged dependent variable as well as the time lag which appears in the augmented equation in (16) are endogenous. Therefore, we suggest an instrumental variable procedure for estimating that model.

Let $X_{-1}$ and $M_{J,-1}$ be identical, correspondingly, to $X$ and $M_J$ except that all of their elements are lagged one time period, $J = 1, ..., G$. Let

$$F_J = [M_J, (I_T \otimes W_J)M_J, M_{J,-1}, (I_T \otimes W_J)M_{J,-1}],\ J = 1, ..., G$$

[11]The assumption that the diagonal elements of the weighting matrices are zero is standard, and is given in our formal list of specifications in the next section.

Our suggested instruments are $H$ where:[12]

$$H = Q_0[X, P, (I_T \otimes W)X, X_{-1}, (I_T \otimes W)X_{-1}, F_1, ..., F_G]_{LI} \tag{17}$$

where $LI$ in (17) denotes the linearly independent columns of the matrix in brackets. For future reference, and without loss of generality, we will refer to $H$ as an $NT \times k_h$ matrix, where $k_h > k_x + T + G + 1$, which is the number of parameters in the augmented model.

## 5 Model specifications

In this section we first list our assumptions and then give their interpretation.

**The assumptions**

**Assumption 1** *(a) $T$ is a fixed positive integer. (b) For all $1 \leq t \leq T$ and $1 \leq i \leq N$, $N \geq 1$ the elements of $\varepsilon$, namely $\varepsilon_{it}$ are identically distributed with mean and variance $(0, 1)$, and have a finite fourth moments. In addition for each $N \geq 1$ and $1 \leq t \leq T$, $1 \leq i \leq N$ the error terms $\varepsilon_{it}$ are independently distributed. (c) The row and column sums of the $NT \times NT$ matrix $R$ is uniformly bounded in absolute value.*

**Assumption 2** *(a) For all $J = 1, ..., G$, $E(\nu_J | H_{1,J}, \Upsilon_J) = 0$ and $E(\nu_J \nu'_J | H_{1,J}, \Upsilon_J) = V_{v_J}$, where the row and column sums of $V_{v_J}$ are uniformly bounded in absolute value. (b) $(I_N - aW_J)$ is nonsingular for all $|a| < 1.0$, $J = 1, ..., G$.*

**Assumption 3** *(a) The diagonal elements of $W$ and $W_J$, $J = 1, ..., G$ are zero. (b) $|\lambda_0| <$ 1.0; (c) $(I_N - aW)$ is nonsingular for all $|a| < 1.0$. (d) The row and column sums of $W$, $(I_N - aW)^{-1}$, $R$, and $W_J, J = 1, ..., G$ and are uniformly bounded in absolute value.*

**Assumption 4** *(a) The elements of the instrument matrix $H$, and therefore of $X$ and $M_J, J = 1, ..., G$ are uniformly bounded in absolute value. (b) $H, X$, and $M_J, J = 1, ..., G$ have full column rank for $N$ large enough.*

[12] Recall from footnote 5 that the general specifications of our model preclude the use of lagged values of the dependent variable as instruments.

**Assumption 5** *The data corresponding to the alternative models, and the estimation procedures are such that, under* $H_{1,J}$ :

$$(\hat{\phi}_{1,J}, \hat{\phi}_{2,J}, \hat{\lambda}_J, \hat{\alpha}_J) \xrightarrow{P} (c_{1,J}, c_{2,J}, L_J, a_J), J = 1, ..., G.$$

*where* $c_{1,J}, c_{2,J}, L_J, a_J$ *are finite constants which need not equal, respectively,* $\phi_{1,J}, \phi_{2,J}, \lambda_J$, *and* $\alpha_J$.

**Assumption 6** *Let* $\Gamma_i$ *be identical to* $\hat{\Gamma}_i$ *except that* $\hat{\phi}_{1,J}, \hat{\phi}_{2,J}, \hat{\lambda}_J$, *and* $\hat{\alpha}_J$ *are replaced, respectively, by* $c_{1,J}, c_{2,J}, L_J$, *and* $a_J$, $J = 1, ..., G$. *Then, we assume*

$$\begin{aligned}
(a) &: \lim_{N\to\infty} (NT)^{-1} H'H = \Omega_{HH} \\
(b) &: p\lim_{N\to\infty} (NT)^{-1} H'Q_0\Gamma_i = \Omega_{HQ_0\Gamma_i}, i = A, B \\
(c) &: \lim_{N\to\infty} (NT)^{-1} H'Q_0RR'Q_0H = \Omega_{H'Q_0RR'Q_0H}
\end{aligned}$$

*where* $\Omega_{HH}, \Omega_{HQ_0\Gamma_i}$, *and* $\Omega_{H'Q_0RR'Q_0H}$ *are finite full column matrices, and therefore* $\Omega_{HH}$ *and* $\Omega_{H'Q_0RR'Q_0H}$ *are positive definite.*

**Interpretations**

In the great majority of spatial panel models the time dimension is short relative to the number of cross sectional units. Assumption 1 (a) is consistent with this and with large sample theory based on $N \to \infty$. Part (b) of this assumption accounts for triangular arrays in specifying the elements of $\varepsilon$. Since the product of matrices whose row and column sums are uniformly bounded in absolute value also have this property, it follows from part (c) that the row and column sums of the variance-covariance matrix of $u$, namely, $RR'$, are also uniformly bounded in absolute value.

As we demonstrate in the appendix, it turns out that specifications beyond those given in Assumption 2 relating to $\nu_J$ are of no consequence asymptotically concerning either the size or power of our suggested $J$-test. Obviously, part (b) of this assumption is needed in reference to $Q_0Y_J^{E,B}$ in (13). Assumptions 3 and 4 are standard. The force of Assumption 5 is that, asymptotically, the size of our test does not depend upon the consistent estimation of the models under $H_1$; on the other hand, it should be evident that the power of our test will be higher if the alternative models are consistently estimated! Assumption 6 is somewhat standard.

# 6 The $J$-Test

Our test is based on the $2SLS$ estimator of the parameter vector $\xi_i$ in the augmented equation (16). Let $P_H = H(H'H)^{-1}H'$, $\hat{\Phi}_i = Q_0\hat{\Gamma}_i$ and $\tilde{\Phi}_i = P_H\hat{\Phi}_i$. Then the $2SLS$ estimator of $\xi_i$ based on (16) is

$$\tilde{\xi}_i = (\tilde{\Phi}_i'\tilde{\Phi}_i)^{-1}\tilde{\Phi}_i'Q_0y, i = A, B \tag{18}$$

**Theorem 1** *Given the model in* (16) *and the assumptions in Section 5*

$$(NT)^{1/2}[\tilde{\xi}_i - \xi_i] \xrightarrow{D} N(0, p\lim_{N\to\infty} A[\Omega_{H'Q_0RR'Q_0H}]A') \tag{19}$$

$$\Omega_{HQ_0RRQ_0H} = \lim_{N\to\infty}(NT)^{-1}(R'Q_0H)'(R'Q_0H)$$

$$A = (NT)(\tilde{\Phi}_i'\tilde{\Phi}_i)^{-1}[\hat{\Gamma}_i'Q_0H](H'H)^{-1}$$

*The proof on Theorem 1 is given in the appendix.*

The suggested small sample approximation to the distribution of $\tilde{\xi}_i$ based on (19) is

$$\begin{aligned} \tilde{\xi}_i &\simeq N(\xi_i, \tilde{V}_{\tilde{\xi}_i}) \\ \tilde{V}_{\tilde{\xi}_i} &= (NT)^{-1}A[\tilde{\Omega}_{H'Q_0RR'Q_0H}]A' \end{aligned} \tag{20}$$

where $\tilde{\Omega}_{H'Q_0RR'Q_0H}$ is a HAC estimator of $\Omega_{H'Q_0RR'Q_0H}$ - see, e.g. Kelejian and Prucha (2007) and Kim and Sun (2011). Let $\tilde{\xi}_i' = (\tilde{\gamma}_0', \tilde{\psi}_i')$ and let $\tilde{V}_{\tilde{\psi}_i}$the lower right $G \times G$ block diagonal submatrix of $\tilde{V}_{\hat{\xi}_i}$. Then, at the 5% level, our suggested $J$-test will reject the null model if

$$\tilde{\psi}_i' \tilde{V}_{\tilde{\psi}_i}^{-1} \tilde{\psi}_i' > \chi^2_G(.95). \tag{21}$$

# 7 Empirical Application

In our empirical application, we use a dynamic demand model for cigarettes (Baltagi and Levin, 1986, 1992). Our data set is based on a panel from 46 US states over the period 1963-1992, and it has been used for illustrative purposes in a number of spatial econometric studies (see e.g. Elhorst, 2005; Debarsy et al., 2012, among others). This data set (over a limited period) was originally used in Baltagi and Levin (1986), who estimated a dynamic demand for cigarettes to address several important policy issues. They found that cigarette sales in a given state are negatively (and significantly) affected by the average retail price of cigarettes in that state with a price elasticity of -0.2. They also found that the income effect was not significant. A distinctive characteristic of their model is that cigarette sales in each state is assumed to depend upon, among other things, the lowest cigarette price in neighboring states. This price variable is meant to capture cross state shopping by cigarette consumers, as well as a "bootlegging" effect, where cigarette consumers purchase cigarettes from "agents" who obtain their supplies from states which have lower prices. This bootlegging effect is found to be positive and statistically significant. Baltagi and Levin (1992) updated the results of their previous analysis (on an extended time frame), and considered various ways of modeling the bootlegging effect. In particular, they analyzed the sensitivity of their results by replacing the minimum price of cigarettes in neighboring states, by the maximum neighboring price. However, they did not consider replacing the minimum price with an average price of the neighboring states, which would seem to be the most intuitive thing to do in a spatial context.

With our $J$-test, we wish to test two competing non-nested alternatives. The null model we consider is the one estimated in Baltagi and Levin (1992) and includes, along with time dummies and spatial fixed effects, the minimum price variable in neighboring states. The alternative model is one that is specified in terms of the average price of the neighboring states, as well as a spatially lagged dependent variable.

More formally, the model under the null, $H_0$, is the following:

$$\ln C_{it} = \beta_1 \ln C_{it-1} + \beta_2 \ln p_{it} + \beta_3 \ln I_{it} + \beta_4 \ln \bar{p}_{it} + \mu_i + \delta_t + u_{it} \tag{22}$$

where $i = 1, \ldots, N$ denotes states, $t = 1, \ldots, T$ denotes time periods. In our sample $N = 46$, and $T = 29$. In (22) $C_{it}$ is cigarette sales per capita in constant dollars to persons of smoking age in state $i$ at time $t$; $p_{it}$ is the real price of cigarettes in state $i$ at time $t$; $I_{it}$ is real per capita disposable income in state $i$ at time $t$; $\bar{p}_{it}$ denotes the minimum real price of cigarettes in a neighboring state; $\mu_i$ is the fixed effect for state $i$, and $\delta_t$ is the fixed time effect for period $t$. The error term $u_{it}$ is assumed to have the non-parametric specification in (3). We expect $\beta_1 > 0, \beta_2 < 0$, and $\beta_3 > 0$. The expectations concerning $\beta_1, \beta_2$, and $\beta_3$ relate, respectively, to habit effects, the usual negative price effect on demand, and the positive effect of income. We also expect $\beta_4$ to be positive. The reason for this is that higher prices in neighboring states should lead to greater sales in state $i$.

Stacking the data over $i$ and $t$ (22) can be expressed as

$$y = XB_0 + P\Delta_0 + \alpha_0 y_{-1} + (e_T \otimes I_N)\mu + u_0 \tag{23}$$

where $y$ is the $NT \times 1$ vector of observations on $\ln C_{it}$; $X$ is the matrix of observations on the price variable $\ln p_{it}$, the income variable $\ln I_{it}$, and on the minimum price variable $\ln \bar{p}_{it}$; $P$ is an $N \times (T-1)$ matrix of observations on $T-1$ dummies ($\delta_t$); $y_{-1}$ is identical to $y$ except all of its elements are lagged one time period. $\mu$ is the $N \times 1$ vector of fixed effects; $e_T$ is the $T \times 1$ vector of unit elements; and $u_0$ is the corresponding vector of disturbance terms.[13]

[13]To estimate the null model we use the following matrix of instruments:

$$H_0 = Q_0[X, X_{-1}, P]$$

Results from the estimation of the null, alternative and augmented model are reported in the Appendix.

We assume the alternative model, $H_1$, to be:

$$\ln C_{it} = \beta_5 \ln C_{it-1} + \beta_6 \ln p_{it} + \beta_7 \ln I_{it} + \beta_8 \sum_{j=1}^{N} w_{ij} \ln p_{jt} + \lambda \sum_{j=1}^{N} w_{ij} \ln C_{jt} + \pi_i + \delta_t + v_{it} \quad (24)$$

where the spatial weight $w_{ij}$ is a measure of the distance between states,[14] $\sum_{j=1}^{n} w_{ij} \ln C_{jt}$ is the spatial lag of consumption, $\sum_{j=1}^{N} w_{ij} \ln p_{jt}$ is the average price of cigarettes in neighboring states, and $\pi_i$ is the fixed effect corresponding to the $i^{th}$ unit. We will again assume that the error term is specified non-parametrically. The remaining notation should be evident. Note that the $\sum_{j=1}^{N} w_{ij} \ln p_{jt}$ is different from the minimum price in Baltagi and Levin (1992) so that the models under $H_0$ and $H_1$ are non-nested. In this formulation we would assume that $\beta_8$ is positive for reasons similar to those relating to $\beta_4$ in $H_0$, namely cross state purchases. Finally, $\lambda$ is the coefficient that measures spatial spillover effects.

Stacking the data, the alternative model can be written in the usual spatial form as

$$y = MB_1 + (I_T \otimes W)M^{\star}B_2 + \lambda(I_T \otimes W)y + P\Delta_1 + \alpha_1 y_{-1} + (e_T \otimes I_N)\pi + u_1 \quad (25)$$

where $M$ is the $NT \times 2$ matrix of observations on the price variable $\ln p_{it}$, and the income variable $\ln I_{it}$; $M^{\star}$ is an $NT \times 1$ vector of observations on the spatial lag of the price variable, and the remaining notation should be evident.

As indicated, the $J$-test is based on augmenting the null model by the predicted values of $Q_0 y$ based on the alternative model. The procedure then is to test for the significance of the augmenting variable.

Given the specifications of the model, the first step in the procedure is to estimate the alternative model by 2SLS. The matrix of instruments used to estimate the alternative model is [15]

$$H_1^* = Q_0[M, (I_T \otimes W)M, (I_T \otimes W^2)M, M_{-1}, (I_T \otimes W)M_{-1}, (I_T \otimes W^2)M_{-1}, P]$$

[14] The weighting matrix employed in this paper is based on the six nearest neighbors.

[15] Note that the spatial lag in the price variable is one of the columns in the spatial lag of $M$.

Given the estimated parameter values, we can obtain the two predictors corresponding to (14).[16] Let $\bar{M}$ be the $NT \times 3$ matrix whose columns are observations on $\ln p_{it}$, $\ln I_{it}$, and the minimum price variable in (22). Then, the augmented equation is also estimated by 2SLS using the matrix of instruments which is identical to $H_1^*$ except that $M$ is replaced by $\bar{M}$.

One final point relates to statistical inference. Standard errors are produced using the spatial HAC estimator of Kelejian and Prucha (2007) with a Parzen kernel. Following previous literature (e.g. Anselin and Lozano-Gracia, 2008), we specify a variable bandwidth based on the distance to the six nearest neighbors.[17]

At the 5% level the $J$-test rejects the null model since the Chi-squared variable $= 19.063 > \chi_1^2 = 3.841$. We conclude then that cross state purchases are better captured by the average price in neighboring states.

## 8 Monte Carlo Design

The design for the Monte Carlo simulation is based on the typical format used in studies on spatial panel models (e.g. Kapoor et al., 2007; Lee and Yu, 2010c; Piras, 2013) and in studies on non-nested tests (Pesaran, 1982; Davidson and MacKinnon, 1981; Godfrey and Pesaran, 1983; Delgado and Stengos, 1994).

Following previous literature (see e.g. Florax et al., 2003; Baltagi et al., 2003, 2007b, among others), our experimental design relates to data generated on regular grids. Specifically, we generate the data from two regular grids of dimensions $10 \times 10$ and $20 \times 20$, corresponding to sample sizes of 100 and 400 observations. For each sample size we construct three row normalized weighting matrices. Following Kelejian and Prucha (1999), the first of this matrix is defined in a circular world and it is generally referred to as the $k$ ahead and $k$ behind spatial weighting matrix. Specifically, in our first matrix ($W_0$), $k$ is set to five. Our second

[16] In the paper we only present the test based on the predictor corresponding to the minimum information set. The results for the other predictor are qualitatively similar and, therefore, are only available upon request from the authors.

[17] In the empirical application as well as in all of the Monte Carlo experiments in Section 8, the denominator of $\tilde{\Omega}_{H'Q_0RR'Q_0H}$ is $N(T-1)-k$ where $k$ is the number of regressors in the model. This degrees of freedom correction in typically done in fixed effects studies (see e.g. Baltagi, 2008, equation 2.24). Also, in order to estimate the residuals more efficiently, $Q_0u$ is estimated from the null model.

matrix ($W_1$) is a distance matrix based on the three nearest neighbors. The distance measure employed to calculate the three nearest neighbors matrix is the Euclidean distance

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{26}$$

where $(x_i, y_i)$ and $(x_j, y_j)$, are the coordinates of the two units. The last spatial weighting matrix considered ($W_3$) is a contiguity matrix based on the queen criterion (i.e. common borders and vertex).

For each sample size we design four sets of experiments. In the first set of experiments the null and alternative models only differ in terms of the weighting matrix employed. In the second set the null and alternative models differ both in terms of the spatial weighting matrix and the regressor matrix. In the third and fourth set of experiments we consider more than one model under the alternative.

In the first set of experiments, the null model is generated from:

$$\begin{aligned} y &= X_0\beta_0 + \lambda_0(I_T \otimes W_0)y + \alpha_0 y_{-1} + u \\ u &= N(0, \sigma^2 I_{NT}) \end{aligned} \tag{27}$$

where the regressors matrix $X_0$ is taken as $X_0 = (x_1, x_2)$, where the $NT \times 1$ values of $x_1$ are generated from a uniform distribution over (0,4); the $NT \times 1$ values of $x_2$ are generated from a chi-squared with three degrees of freedom.[18] The elements of the parameter vector $\beta_0$ were set to 0.5, and $\sigma^2 = 1.0$; this value of $\sigma^2$ lead to $R^2$ values of, approximately 0.6. The alternative model is also generated from (27) except that it is specified in terms of the spatial weights matrix $W_1$. In this experiment, and in the experiments below, once generated, the values of the regressors are held fixed in the Monte Carlo trials. The parameter values, and the specification of the innovation term are given below.

[18]We generate the spatial panel data with $100 + T$ periods and then take the last $T$ as our sample and we set $T$ equal to 5 in all experiments. The initial values are generated as

$$y_0 = (I_T \otimes (I_N - \lambda_0 W_0)^{-1})X_0\beta_0$$

In the second set of experiments, the null model is identical to (27), while the alternative model differs in terms of both the weighting matrix and the regressor matrix $X_1 = (z_1, z_2)$. Specifically, the weighting matrix employed in this set of experiments is $W_1$ (the three nearest neighbors). Additionally, the first column of the regressors matrix ($z_1$) is generated from an uniform distribution over (0,3), and the values of $z_2$ are generated as

$$z_2 = ax_1 + \xi \tag{28}$$

where $a = 0.5$ and $\xi \sim N(0, I_{NT})$. This value of $a$ leads to a correlation between $z_2$ and $x_1$ of, approximately, 0.5.

The third set of experiments is different from the previous two in that it accounts for two models under the alternative. Again, the null model is specified as in (27), and the two alternative models are specified in terms of $X_0$ and $W_1$, and $X_1$ and $W_0$, respectively.

Finally, the fourth set of experiments accounts for three models under the alternative. While the null model is again specified as in (27), the models under the alternative are specified in terms of $X_0$ and $W_1$, $X_1$ and $W_0$, and $X_1$ and $W_2$, respectively. The four sets of experiments are summarized in Table 1.

In all four sets of experiments, six values are considered for $\lambda$, namely -0.6, -0.4, -0.2, 0.2, 0.4 and 0.6; and two values for $\alpha$, namely -0.2, 0.2. Our parameter combinations are consistent with the stability conditions given in Elhorst (2001) and Parent and LeSage (2011).

The total number of combinations relating to $\lambda$, $\alpha$, $N$ and the four sets of experiment combining different definitions of $W$ leads to a total of $6 \times 2 \times 2 \times 4 = 96$ experiments for each of which the two predictors are calculated. For all experiments, 2,000 replications are performed. This is roughly the number of replications needed to obtain a 95% confidence interval of length .02 on the size of a test statistic.

# 9 Monte Carlo Results

Our Monte Carlo results corresponding to each of the four experiments are summarized in Tables 2-9. For each experiment and sample size, these tables give the frequency of rejection of the null hypothesis at the 5% level. Reading across, the results in the first two sections give the estimated size of the test, while power estimates are given in the second two sections. The results in the tables relate to the use of our two predictors, $Y^{(A)}$ and $Y^{(B)}$. Finally, the first four tables refers to the smaller sample size ($n = 100$), while the other four to the larger sample size ($n = 400$).

Consider the results in Table 2, which are based on the first set of experiments for $n = 100$. Looking first at column averages, the empirical size of the test, based on both predictors in (14) is reasonably close to the theoretical 5% level. There is only one exception when $\alpha = 0.2$ and $\lambda = 0.6$. In this case the average sizes of the test corresponding to the predictor $y^{(A)}$ is, in percentage points, 6.15. The theoretical level, in percentage terms, is 5.00. A glance at the corresponding results in Table 6, based on $n = 400$, suggests that as the sample size increases the empirical size gets closer to the theoretical level.

Let us turn now to the reported power calculations of the $J$-test in Table 2. Again, in terms of column averages, it should be clear that the $J$-test exhibits "very" high power in all cases relating to the first set of experiments in which the null and alternative models differ only in the spatial weight matrix. In these cases, the power results seem to be degraded for low values of $\lambda$.[19] Given our Monte Carlo design, the power is very high and, in general, differences relating to the use of the two predictor are small. Therefore, based on computational simplicity we suggest the use of our test in this paper based on the predictor $y^{(A)}$.[20] When the sample size increases (see Table 6) the power calculation are virtually all equal to 1.0.

[19] In a sense this result is not very surprising given that the power of any test will depend on the extent to which the null and alternative hypotheses differ. A further discussion of this is given by LeSage and Pace (2009). Using Bayesian posterior model comparisons, they illustrate for alternative weight matrices that as the spatial dependence approaches low levels, the posterior probabilities approach the prior probabilities. In the limiting case, if the spatial dependence were zero an empirical test would not be able to distinguish between two different weighting matrices.

[20] However, as pointed out by Jin and Lee (2013), there might be situations where the gap in the power for the two versions of the spatial $J$-test may be large. This is an issue that could be explored in a larger Monte Carlo study, and we leave it for further research.

Consider now the results in Table 3, where the difference between the null and alternative model pertains to both the weighting matrix and the regressors matrix. The empirical size of the test, based on the predictor $y^{(B)}$, is reasonably close to the theoretical 5% level. There are very few exceptions mostly related to low values of $\lambda$. Specifically, these exceptions are, in percentage points, 6.3, 6.25 and 6.45. On the other hand, concerning the predictor $y^{(A)}$, none of the estimates of the empirical size fall into the acceptance interval (.041, .060) and the test seems to systematically over-reject the null hypothesis.[21] Fortunately, we note that this "size of test problem" diminishes as the value of $n$ increases. In fact, looking at column averages in Table 7, we note that the empirical size of the test, based on both predictors, is reasonably close to the theoretical 5% level when the sample size is $n = 400$. There are a few exceptions that do not fall into the "acceptance interval". Moving to the power of the test we observe from Table 3 that, for all combinations of model parameters, the power of the test corresponding to the use of both predictors is equal to one. The suggestion is that, in terms of power, if the null and alternative models differ in both the spatial weighting matrix and the regressor matrices, the two tests we suggest in this paper are equally good.

The last two sets of experiments are specified in such a way that the null model is tested against two (or more) possible alternatives. In particular, Table 4 relates to a situations where there are two models under the alternatives, while the results in Table 5 are obtained when there are three models under the alternative.

Results in Table 4 and 5 suggest that, when there is more than one model under the alternative hypothesis the empirical size of the test, based on both predictors, is reasonably close to the theoretical 5% level. Interestingly, there are no exceptions to this in Table 4. However, Table 5 presents a few cases in which the empirical sizes do not fall into the "acceptance interval". These cases are mostly related to the use of the first predictor, and only one case relates to the second predictor. Fortunately, we note again that this "size of test problem" diminishes as the value of $n$ increases. In fact, looking at Table 9, we note that there are only two individual size estimates significantly different from 5% when the sample

[21] Some studies have suggested to implement bootstrap testing procedure to improve the small sample performance of the test (see, e.g., Burridge and Fingleton, 2010, for an example in a spatial context). We decided to leave this for future research.

size is $n = 400$. Finally, the reported powers in Table 4 and 5 are very high and suggest that in reasonably large samples, the two tests considered are quite “powerful”.

## 10 Conclusion

In this paper we extended the $J$-test to a spatial panel model containing fixed effects, a spatially lagged dependent variable, and a time lagged dependent variable. The disturbance term in our model was specified non-parametrically and allows for general patterns of spatial and time correlation, as well as heteroskedastity. The alternative models were specified in such a way that both spatial and time correlation of various sorts, as well as general patterns of heteroskedasticity are special cases. These alternative models can have either fixed or random effects. Given reasonable assumptions, our test is computationally simple.

We gave formal large sample results, as well as small sample Monte Carlo results that suggested, among other things, that our proposed $J$-test has good power, and proper size for small to moderately sized samples.

Finally, we implemented our procedure to test the specifications of a demand for cigarette model. Our empirical results suggested that the most appropriate specification was the one involving a spatial lag of cigarette consumption in neighboring states.

One suggestion for future research would be an extension of our results to the case in which both $N \to \infty$ and $T \to \infty$. In doing this one should, among other things, account for the possible limits of $N/T$ - e.g., to $0, \infty$, or a finite constant. Another extension would be to non-linear spatial models in a panel framework. Among others, such a framework would arise in qualitative, or limited dependent variable models. Still another suggestion for future research relates to small sample issues which would arise in a Monte Carlo study in which the true value of a parameter relating to the spatial lag of the dependent variable is “close” to a limiting value of the parameter space- e.g., if 1.0 is a limiting value then the true value might be .9. Assuming the stability conditions described in Parent and LeSage (2011), in this framework estimates of this parameter would, in some trials, exceed that upper limit. There are various ways of handling such cases but guidance on this issue would be relevant.

## References


Anselin, L., Le Gallo, J., and Jayet, H. (2008). Spatial panel econometrics. In Matyas, L. and Sevestre, P., editors, *The econometrics of Panel Data, Fundamentals and Recent Developments in Theory and Practice (3rd Edition)*, pages 624 – 660. Springer-Verlag, Berlin Heidelberg.

Anselin, L. and Lozano-Gracia, N. (2008). Errors in variables and spatial effects in hedonic house price models of ambient air quality. *Empirical Economics*, 34(1):5–34.

Baltagi, B. (2008). *Econometric Analysis of Panel Data, 4th edition.* Wiley, New York.

Baltagi, B., Egger, P., and Pfafermayr, M. (2007a). Estimating models of complex fdi: are there third-country effects? *Journal of Econometrics*, 140:260–281.

Baltagi, B., Egger, P., and Pfaffermayr, M. (2013). A generalized spatial panel data model with random effects. *Econometric Reviews*, 32(5 –6):650 – 685.

Baltagi, B., Kelejian, H., and Prucha, I. (2007b). Analysis of spatially dependent data. *Journal of Econometrics*, 140:1–4.

Baltagi, B. and Liu, L. (2008). Testing for random effects and spatial lag dependence in panel data models. *Statistics and Probability Letters*, 78:3304–3306.

Baltagi, B., Song, S., Jung, B., and Koh, W. (2007c). Testing for serial correlation, spatial autocorrelation and random effects using panel data. *Journal of Econometrics*, 140(1):5–51.

Baltagi, B., Song, S., and Koh, W. (2003). Testing panel data regression models with spatial error correlation. *Journal of Econometrics*, 117:123–150.

Baltagi, B. H. and Levin, D. (1986). Estimating dynamic demand for cigarettes using panel data: The effects of bootlegging, taxation and advertising reconsidered. *The Review of Economics and Statistics*, 68:148–155.

Baltagi, B. H. and Levin, D. (1992). Cigarette taxation: raising revenues and reducing consumption. *Structural Change and Economic Dynamics*, 3:321–335.

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

Dastoor, N. (1983). Some Aspects of Testing Non-Nested Hypothesis. *Journal of Econometrics*, 21:213–228.

Davidson, R. and MacKinnon, J. (1981). Several tests for model specification in the presence of alternative hypotheses. *Econometrica*, 49:781–794.

Debarsy, N. and Ertur, C. (2010). Testing for spatial autocorrelation in a fixed effects panel data model. *Regional Science and Urban Economics*, 40:453–470.

Debarsy, N., Ertur, C., and LeSage, J. P. (2012). Interpreting dynamic space-time panel data models. *Statistical Methodology*, 9:158 – 171.

Delgado, M. and Stengos, T. (1994). Semiparametric specification testing of non-nested econometric models. *Review of Economic Studies*, 61:291–303.

Elhorst, J. (2001). Dynamic models in space and time. *Geographical Analysis*, 33(2):119–140.

Elhorst, J. (2003). Specification and estimation of spatial panel data models. *International Regional Sciences Review*, 26(3):244–268.

Elhorst, J. (2008). Serial and spatial error correlation. *Economics Letters*, 100:422–424.

Elhorst, J. (2009). Spatial panel data models. In Fischer, M. M. and Getis, A., editors, *Handbook of Applied Spatial Analysis*. Springer, Berlin, Heidelberg, New York.

Elhorst, J. (2010). Dynamic panels with endogenous interactions effects when T is small. *Regional Science and Urban Economics*, 40:272–282.

Elhorst, J. and 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, J., Piras, G., and Arbia, G. (2010). Growth and convergence in a multi-regional model with space-time dynamics. *Geographical Analysis*, 42:338–355.

Elhorst, J. P. (2005). Unconditional maximum likelihood estimation of linear and log-linear dynamic models for spatial panels. *Geographical Analysis*, 37:62–83.

Florax, R., Folmer, H., and Rey, S. (2003). Specification searches in spatial econometrics: the relevance of Hendry's methodology. *Regional Science and Urban Economics*, 33:557–579.

Godfrey, L. (1983). Testing non-nested Models After Estimation by Instrumental Variables or Least Squares. *Econometrica*, 51:355–366.

Godfrey, L. and Pesaran, M. (1983). Test of non-nested regression models. *Journal of Econometrics*, 21:133–154.

Greene, W. (2003). *Econometric Analysis (fifth edition)*. Prentice Hall, Upper Saddle River.

Jin, F. and Lee, L. (2013). Cox-type tests for competing spatial autoregressive models with spatial autoregressive disturbances. *Regional Science and Urban Economics*, 43:590 – 616.

Kapoor, M., Kelejian, H., and Prucha, I. (2007). Panel Data Model with Spatially Correlated Error Components. *Journal of Econometrics*, 140(1):97–130.

Kelejian, H., Murrell, P., and Shepotylo, O. (2013). Spatial spillovers in the development of institutions. *Journal of Development Economics*, pages 297–315.

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

Kelejian, H. and Prucha, I. (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review*, 40(2):509–533.

Kelejian, H. and Prucha, I. (2007). HAC estimation in a spatial framework. *Journal of Econometrics*, 140:131–154.

Kelejian, H. H. (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.

Kim, M. S. and Sun, Y. (2011). Spatial heteroskedasticity and autocorrelation consistent estimation of covariance matrix. *Journal of Econometrics*, 160(2):349 – 371.

Kmenta, J. (1986). *Elements of Econometrics (second edition)*. Macmillan, New York.

Lee, L. and Yu, J. (2010a). Estimation of spatial autoregressive panel data models with fixed effects. *Journal of Econometrics*, 154:165–185.

Lee, L. and Yu, J. (2010b). Some recent development in spatial panel data models. *Regional Science and Urban Economics*, 40:255–271.

Lee, L. and Yu, J. (2010c). A spatial dynamic panel data model with both time and individual fixed effects. *Econometric Theory*, 26:564–597.

Lee, L. and Yu, J. (2010d). A unified transformation approach to the estimation of spatial dynamic panel data models: stability, spatial cointegration and explosive roots. In A., U. and Giles, D., editors, *Handbook of Empirical Economics and Finance*, pages 397–434. Chapman and Hall - CRC.

LeSage, J. and Pace, K. (2009). *Introduction to Spatial Econometrics*. CRC Press, Boca Raton, FL.

MacKinnon, J., White, H., and 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. and Pfaffermayr, M. (2011). The Hausman test in a Cliff and Ord panel model. *Econometrics Journal*, 14:48–76.

Parent, O. and LeSage, J. (2010). A spatial dynamic panel model with random effects applied to commuting times. *Transportation Research Part B*, 44:633–645.

Parent, O. and LeSage, J. (2011). A space-time filter for panel data models containing random effects. *Computational Statistics and Data Analysis*, 55(1):475–490.

Pesaran, H. M. and Tosetti, E. (2011). Large panels with common factors and spatial correlations. *Journal of Econometrics*, 161(2):182 –202.

Pesaran, M. (1974). On the general problem of model selection. *Review of Economic Studies*, 41:153–171.

Pesaran, M. (1982). Comparison of local power of alternative tests of non-nested regression models. *Econometrica*, 50:1287–1305.

Pesaran, M. and Deaton, A. (1978). Testing non-nested non-linear regression models. *Econometrica*, 46:677–694.

Pesaran, M. and Weeks, M. (2001). Non-nested hypothesis testing: an overview. In Baltagi, B., editor, *A companion to theoretical econometrics*. Blackwell, Malden.

Piras, G. (2013). Efficient gm estimation of a cliff and ord panel data model with random effects. *Spatial Economic Analysis*, forthcoming.

Pötscher, B. M. and Prucha, I. R. (2000). Basic elements of asymptotic theory. In Baltagi, B. H., editor, *A Companion to Theoretical Econometrics*, pages 201 – 229. Basil Blackwell.

Yu, J., de Jong, R., and Lee, L. (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.

Yu, J. and Lee, L. (2010). Estimation of unit root spatial dynamic panel data models. *Econometric Theory*, 26:1332–1362.

Table 1: Experimental Designs Relating to the Regressor and Weighting Matrices

| **Experiments** | | **Regressors** | **Weights** |
|---|---|---|---|
| Set 1 | $H_0$ | $X_0$ | $W_0$ |
| | $H_1$ | $X_0$ | $W_1$ |
| Set 2 | $H_0$ | $X_0$ | $W_0$ |
| | $H_1$ | $X_1$ | $W_1$ |
| Set 3 | $H_0$ | $X_0$ | $W_0$ |
| | $H_1$ | $X_1$ | $W_0$ |
| | | $X_0$ | $W_1$ |
| Set 4 | $H_0$ | $X_0$ | $W_0$ |
| | $H_1$ | $X_0$ | $W_1$ |
| | | $X_1$ | $W_0$ |
| | | $X_1$ | $W_2$ |

Table 2: Frequency of rejection of the null hypothesis: two predictors (2000 replications), $n = 100$, $t = 5$.

First set of experiments.
$H_0 : X_0, W_0$ and $H_1 : X_0 W_1$

| | | size | | power | |
|---|---|---|---|---|---|
| | | $Y^{(A)}$ | $Y^{(B)}$ | $Y^{(A)}$ | $Y^{(B)}$ |
| $\alpha = -0.2$ | $\lambda = -0.6$ | 0.0505 | 0.0525 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0500 | 0.0510 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0540 | 0.0545 | 0.8885 | 0.8985 |
| | $\lambda = 0.2$ | 0.0495 | 0.0450 | 0.9155 | 0.8655 |
| | $\lambda = 0.4$ | 0.0560 | 0.0565 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0555 | 0.0590 | 1.0000 | 1.0000 |
| $\alpha = 0.2$ | $\lambda = -0.6$ | 0.0575 | 0.0525 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0515 | 0.0435 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0475 | 0.0525 | 0.8405 | 0.8510 |
| | $\lambda = 0.2$ | 0.0580 | 0.0595 | 0.8475 | 0.7670 |
| | $\lambda = 0.4$ | 0.0445 | 0.0450 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0615 | 0.0565 | 1.0000 | 1.0000 |
| Average | | 0.0530 | 0.0523 | 0.9577 | 0.9485 |

Table 3: Frequency of rejection of the null hypothesis: two predictors (2000 replications), $n = 100$, $t = 5$.

Second set of experiments.
$H_0 : X_0, W_0$ and $H_1 : X_1 W_1$

| | | size | | power | |
|---|---|---|---|---|---|
| | | $Y^{(A)}$ | $Y^{(B)}$ | $Y^{(A)}$ | $Y^{(B)}$ |
| $\alpha = -0.2$ | $\lambda = -0.6$ | 0.0650 | 0.0535 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0645 | 0.0565 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0665 | 0.0520 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0685 | 0.0630 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0690 | 0.0520 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0685 | 0.0530 | 1.0000 | 1.0000 |
| $\alpha = 0.2$ | $\lambda = -0.6$ | 0.0760 | 0.0625 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0670 | 0.0530 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0640 | 0.0645 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0670 | 0.0575 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0680 | 0.0575 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0675 | 0.0515 | 1.0000 | 1.0000 |
| Average | | 0.0676 | 0.0564 | 1.0000 | 1.0000 |

Table 4: Frequency of rejection of the null hypothesis: two predictors (2000 replications), $n = 100$, $t = 5$.

Third set of experiments.
$H_0 : X_0, W_0$ and $H_1 : X_0 W_1; X_1 W_0$

| | | size | | power | |
|---|---|---|---|---|---|
| | | $Y^{(A)}$ | $Y^{(B)}$ | $Y^{(A)}$ | $Y^{(B)}$ |
| $\alpha = -0.2$ | $\lambda = -0.6$ | 0.0575 | 0.0580 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0510 | 0.0485 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0430 | 0.0460 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0540 | 0.0445 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0490 | 0.0410 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0510 | 0.0460 | 1.0000 | 1.0000 |
| $\alpha = 0.2$ | $\lambda = -0.6$ | 0.0475 | 0.0495 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0545 | 0.0490 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0540 | 0.0470 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0485 | 0.0415 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0570 | 0.0445 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0440 | 0.0465 | 1.0000 | 1.0000 |
| Average | | 0.0509 | 0.0468 | 1.0000 | 1.0000 |

Table 5: Frequency of rejection of the null hypothesis: two predictors (2000 replications), $n = 100$, $t = 5$.

Fourth set of experiments.
$H_0 : X_0, W_0$ and $H_1 : X_0W_1; X_1W_0; X_1W_2$

| | | size | | power | |
|---|---|---|---|---|---|
| | | $Y^{(A)}$ | $Y^{(B)}$ | $Y^{(A)}$ | $Y^{(B)}$ |
| $\alpha = -0.2$ | $\lambda = -0.6$ | 0.0645 | 0.0465 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0580 | 0.0550 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0525 | 0.0495 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0545 | 0.0460 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0555 | 0.0505 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0480 | 0.0450 | 1.0000 | 1.0000 |
| $\alpha = 0.2$ | $\lambda = -0.6$ | 0.0625 | 0.0510 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0690 | 0.0555 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0680 | 0.0505 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0580 | 0.0505 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0530 | 0.0390 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0575 | 0.0440 | 1.0000 | 1.0000 |
| Average | | 0.0584 | 0.0486 | 1.0000 | 1.0000 |

Table 6: Frequency of rejection of the null hypothesis: two predictors (2000 replications), $n = 400$, $t = 5$.

First set of experiments.
$H_0 : X_0, W_0$ and $H_1 : X_0W_1$

| | | size | | power | |
|---|---|---|---|---|---|
| | | $Y^{(A)}$ | $Y^{(B)}$ | $Y^{(A)}$ | $Y^{(B)}$ |
| $\alpha = -0.2$ | $\lambda = -0.6$ | 0.0515 | 0.0490 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0520 | 0.0535 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0480 | 0.0460 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0430 | 0.0445 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0520 | 0.0505 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0555 | 0.0510 | 1.0000 | 1.0000 |
| $\alpha = 0.2$ | $\lambda = -0.6$ | 0.0465 | 0.0445 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0530 | 0.0535 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0500 | 0.0500 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0455 | 0.0480 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0465 | 0.0445 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0490 | 0.0510 | 1.0000 | 1.0000 |
| Average | | 0.0494 | 0.0488 | 1.0000 | 1.0000 |

Table 7: Frequency of rejection of the null hypothesis: two predictors (2000 replications), $n = 400$, $t = 5$.

Second set of experiments.
$H_0 : X_0, W_0$ and $H_1 : X_1 W_1$

| | | size | | power | |
|---|---|---|---|---|---|
| | | $Y^{(A)}$ | $Y^{(B)}$ | $Y^{(A)}$ | $Y^{(B)}$ |
| $\alpha = -0.2$ | $\lambda = -0.6$ | 0.0560 | 0.0545 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0510 | 0.0510 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0670 | 0.0635 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0505 | 0.0540 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0645 | 0.0625 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0565 | 0.0580 | 1.0000 | 1.0000 |
| $\alpha = 0.2$ | $\lambda = -0.6$ | 0.0610 | 0.0585 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0510 | 0.0540 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0500 | 0.0520 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0495 | 0.0500 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0580 | 0.0570 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0610 | 0.0615 | 1.0000 | 1.0000 |
| Average | | 0.0563 | 0.0564 | 1.0000 | 1.0000 |

Table 8: Frequency of rejection of the null hypothesis: two predictors (2000 replications), $n = 400$, $t = 5$.

Third set of experiments.
$H_0 : X_0, W_0$ and $H_1 : X_0 W_1; X_1 W_0$

| | | size | | power | |
|---|---|---|---|---|---|
| | | $Y^{(A)}$ | $Y^{(B)}$ | $Y^{(A)}$ | $Y^{(B)}$ |
| $\alpha = -0.2$ | $\lambda = -0.6$ | 0.0455 | 0.0470 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0475 | 0.0505 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0530 | 0.0610 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0505 | 0.0490 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0450 | 0.0550 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0545 | 0.0540 | 1.0000 | 1.0000 |
| $\alpha = 0.2$ | $\lambda = -0.6$ | 0.0460 | 0.0480 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0540 | 0.0590 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0560 | 0.0580 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0470 | 0.0470 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0490 | 0.0475 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0500 | 0.0535 | 1.0000 | 1.0000 |
| Average | | 0.0498 | 0.0525 | 1.0000 | 1.0000 |

Table 9: Frequency of rejection of the null hypothesis: two predictors (2000 replications), $n = 400$, $t = 5$.

Fourth set of experiments.
$H_0 : X_0, W_0$ and $H_1 : X_0W_1; X_1W_0; X_1W_2$

| | | size | | power | |
|---|---|---|---|---|---|
| | | $Y^{(A)}$ | $Y^{(B)}$ | $Y^{(A)}$ | $Y^{(B)}$ |
| $\alpha = -0.2$ | $\lambda = -0.6$ | 0.0460 | 0.0485 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0525 | 0.0485 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0530 | 0.0515 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0495 | 0.0550 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0535 | 0.0530 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0615 | 0.0625 | 1.0000 | 1.0000 |
| $\alpha = 0.2$ | $\lambda = -0.6$ | 0.0475 | 0.0545 | 1.0000 | 1.0000 |
| | $\lambda = -0.4$ | 0.0495 | 0.0435 | 1.0000 | 1.0000 |
| | $\lambda = -0.2$ | 0.0495 | 0.0450 | 1.0000 | 1.0000 |
| | $\lambda = 0.2$ | 0.0565 | 0.0455 | 1.0000 | 1.0000 |
| | $\lambda = 0.4$ | 0.0585 | 0.0495 | 1.0000 | 1.0000 |
| | $\lambda = 0.6$ | 0.0510 | 0.0500 | 1.0000 | 1.0000 |
| Average | | 0.0524 | 0.0506 | 1.0000 | 1.0000 |

Table 10: Estimation of the alternative model.

| | Coefficients | se | t-stat | p-val |
|---|---|---|---|---|
| $\ln p_{i,t}$ | -0.4366 | 0.0454 | -9.6174 | 0.0000 |
| $\ln I_{i,t}$ | 0.1661 | 0.0316 | 5.2492 | 0.0000 |
| $\sum_{j=1}^{N} w_{ij} \ln p_{j,t}$ | 0.1776 | 0.0974 | 1.8237 | 0.0682 |
| $\sum_{j=1}^{N} w_{ij} \ln C_{j,t}$ | 0.2169 | 0.0564 | 3.8440 | 0.0001 |
| $\ln C_{i,t-1}$ | 0.6433 | 0.0371 | 17.3191 | 0.0000 |

Table 11: Estimation of the null model.

| | Coefficients | se | t-stat | p-val |
|---|---|---|---|---|
| $\ln p_{i,t}$ | -0.4957 | 0.0492 | -10.0726 | 0.0000 |
| $\ln I_{i,t}$ | 0.1894 | 0.0361 | 5.2467 | 0.0000 |
| $\ln p_{\min}$ | -0.0159 | 0.0359 | -0.4441 | 0.6569 |
| $\ln C_{i,t-1}$ | 0.6016 | 0.0410 | 14.6690 | 0.0000 |

Table 12: Estimation of the augmented model using the first predictor $y^{(A)}$ based on the minimum information set.

| | Coefficients | se | t-stat | p-val |
|---|---|---|---|---|
| $\ln p_{i,t}$ | -0.0749 | 0.1206 | -0.6212 | 0.5345 |
| $\ln I_{i,t}$ | 0.0203 | 0.0551 | 0.3687 | 0.7123 |
| $\ln p_{\min}$ | -0.3479 | 0.2121 | -1.6401 | 0.1010 |
| $\ln C_{i,t-1}$ | 0.0747 | 0.1868 | 0.3997 | 0.6894 |
| $y^{(A)}$ | 0.8288 | 0.2737 | 3.0281 | 0.0025 |

# Appendix

**A: Asymptotic equivalence of** $Q_0 Y_J^{E,A}$ and $Q_0 Y_J^{E,B}$ **when** $G = 1$

**A.1: Results relating to** $\overbrace{Q_0 Y_J^{E,A}}$

Let

$$Q_0 = [Q'_{0,1}, ..., Q'_{0,T}]'$$

where $Q_{0,t}$ is the $N \times NT$ matrix which consists of the $t^{th}$ block of $N$ consecutive rows of $Q_0$ - e.g., $Q_{0,1}$ is the first $N$ consecutive rows of $Q_0$, etc. Let $Q_{0,t,i.}$ be the $i^{th}$ row of $Q_{0,t}$ and note from (10) and (11) that $E[Q_{0,t,i.}y|H_{1,J}, INFO_{J,t}^{\max}] = Q_{0,t,i.}E[y|H_{1,J}, INFO_{J,t}^{\max}]$ where

$$\begin{aligned} Q_{0,t,i.}E[y|H_{1,J}, INFO_{J,t}^{\max}] &= Q_{0,t,i.}M_J\phi_{1,J} + Q_{0,t,i.}P\phi_{2,J} \\ &+\lambda_J Q_{0,t,i.0}(I_T \otimes W_J)y + \alpha_J Q_{0,t,i.}y_{-1} \\ &+Q_{0,t,i.}E[\nu_J|H_{1,J}, INFO_{J,t}^{\max}] \end{aligned} \quad (29)$$

Let

$$r_{J|t,i} = Q_{0,t,i.}E[\nu_J|H_{1,J}, INFO_{J,t}^{\max}] \quad (30)$$

and note that $r_{J|t,i} \neq 0$ since the elements of the $NT \times 1$ vector $\nu_J$ are both spatially and time correlated, and $INFO_{J,t}^{\max}$ contains the vectors $[y_0, ..., y_{t-1}, y_{-i,t}]$. However, using the iterated expectations principle, recalling that $INFO_{J,t}^{\max}$ contains $\Upsilon_J = (M_J, P, W_J)$, and Assumption 2 it follows that

$$\begin{aligned} E[r_{J|t,i}|H_{1,J}, \Upsilon_J] &= Q_{0,t,i.}E[E(\nu_J|H_{1,J}, INFO_{J,t}^{\max})|H_{1,J}, \Upsilon_J] \\ &= Q_{0,t,i.}E(\nu_J|H_{1,J}, \Upsilon_J) = 0 \end{aligned} \quad (31)$$

The results in (30) and (31) imply

$$Q_{0,t,i.}\nu_J = r_{J|t,i} + \Theta_{J|t,i} \quad (32)$$

where $E[\Theta_{J|t,i}|H_{1,J},\Upsilon_J] = 0$.[22]

Let $r_{J|t} = [r_{J|t,1},...,r_{J|t,N}]'$ and $r_J = [r_{J|1},...,r_{J|T}]'$. It follows from (30) - (32) that

$$Q_0\nu_J = r_J + \Theta_J \tag{33}$$

where

$$E[Q_0\nu_J|H_{1,J},\Upsilon_J] = E[r_J|H_{1,J},\Upsilon_J] = E[\Theta_J|H_{1,J},\Upsilon_J] = 0 \tag{34}$$

and the VC matrix of $Q_0\nu_J$ is

$$Q_0V_{v_J}Q_0 = V_{r_J} + V_{\Theta_J} + C_{r_J,\Theta_J} + C'_{r_J,\Theta_J} \tag{35}$$

where $V_{r_J}$ and $V_{\Theta_J}$ are, respectively, the variance covariance matrices of $r_J$ and $\Theta_J$, and $C_{r_J,\Theta_J}$ is the covariance matrix between $r_J$ and $\Theta_J$. Clearly the row and column sums of $Q_0$ are uniformly bounded in absolute value by $\frac{T-1}{T} < 2.0$; By Assumption 2, the row and column sums of $V_{v_J}$ are also uniformly bounded in absolute value. Since the product of matrices whose row and column sums are uniformly bounded in absolute value also have rows and column sums which are so uniformly bounded, the row and column sums of $Q_0V_{v_J}Q_0$ are also uniformly bounded in absolute value. It then follows from (35) that the row and column sums of $V_{r_J}$ are uniformly bounded in absolute value.

If $r_J$ were observed and used, the predictor $\overbrace{Q_0Y_J^{E,A}}$ in (14) would be replaced by $\overbrace{Q_0Y_J^{E,A}}+r_J$ in the augmented regression (16). We now show that the large sample distribution of $\tilde{\xi}_i$ in (18) does not involve $r_J$.

The estimator $\tilde{\xi}_i$ can be expressed

$$\begin{aligned}(NT)^{1/2}[\tilde{\xi}_i - \xi_i] &= NT(\tilde{\Phi}_i'\tilde{\Phi}_i)^{-1}(NT^{-1/2})\tilde{\Phi}_i'Q_0R\varepsilon \\ &= \left\{[(NT)^{-1}\hat{\Phi}_i'H]\,[(NT)(H'H)^{-1}]\,[(NT)^{-1}H'\hat{\Phi}_i]\right\}^{-1} * \\ &\quad [(NT)^{-1}\hat{\Phi}_i'H]\,[(NT)(H'H)^{-1}]\,[(NT)^{-1/2}H'Q_0R\varepsilon]\end{aligned} \tag{36}$$

[22] To see this, take expectations across in (32) conditional on $H_{1,J}, INFO_{J,t}^{\max}$ and use the result in (31).

Since $\hat{\Phi}_A = Q_0\hat{\Gamma}_A$ and, in this case $\hat{\Gamma}_A = [Z, \overbrace{Q_0 Y_J^{E,A}} + r_J]$, the term $r_J$ only arises in (36) in terms of the product $(NT)^{-1}H'Q_0[Z, \overbrace{Q_0 Y_J^{E,A}} + r_J]$. Let $\eta_J = (NT)^{-1}H'Q_0 r_J$. Then, in light of (36) the term $r_J$ enters into the large sample distribution of $\tilde{\xi}_i$ only via $\eta_J$. It follows from (33) - (35) that the mean and variance covariance matrix of $\eta_J$ are

$$\begin{aligned} E[\eta_J | H_{1,J}, \Upsilon_J] &= 0 \\ E[\eta\eta_J' | H_{1,J}, \Upsilon_J] &= (NT)^{-2}H'[Q_0 V_{r_J} Q_0]H \end{aligned} \tag{37}$$

In light of (35) the row and column sums of $Q_0 V_{r_J} Q_0$ are uniformly bounded in absolute value. By Assumption 4, the elements of $H$ are uniformly bounded in absolute value and so the elements of $H'[Q_0 V_{r_J} Q_0]H$ are $0(N)$; it follows from (37) that

$$E[\eta\eta_J' | H_{1,J}, \Upsilon_J] \to 0 \tag{38}$$

The results in (37), (38), and Chebyshev's inequality imply

$$\eta_J \xrightarrow{P} 0 \tag{39}$$

It then follows that, asymptotically, $r_J$ is of no consequence in our $J$-test.

**A.2: Equivalence of $Q_0 Y_J^{E,A}$ and $Q_0 Y_J^{E,B}$, asymptotically, when $G = 1$.**

Let

$$K_J = [I_T \otimes (I_N - \lambda_J W_J)^{-1}] \tag{40}$$

and note from (8) and (10) that under $H_{1,J}$

$$Q_0 y = K_J Q_0 M_J \phi_{1,J} + K_J Q_0 P \phi_{2,J} + \alpha_J K_J Q_0 y_{-1} + K_J Q_0 \nu_J \tag{41}$$

Consider $Q_0 Y_J^{EA}$ and, recalling (8) and (13), we have

$$
\begin{aligned}
Q_0 Y_J^{E,A} &= Q_0 M_J \phi_{1,J} + Q_0 P \phi_{2,J} + \lambda_J (I_T \otimes W_J) Q_0 y + \alpha_J Q_0 y_{-1} \\
&= Q_0 M_J \phi_{1,J} + Q_0 P \phi_{2,J} + \alpha_J Q_0 y_{-1} + (I_T \otimes \lambda_J W_J) * \\
&\quad [K_J Q_0 M_J \phi_{1,J} + K_J Q_0 P \phi_{2,J} + \alpha_J K_J Q_0 y_{-1} + K_J Q_0 \nu_J]
\end{aligned} \tag{42}
$$

Combining terms in (42) we have

$$
Q_0 Y_J^{E,A} = S_J Q_0 M_J \phi_{1,J} + S_J Q_0 P \phi_{2,J} + S_J Q_0 y_{-1} \lambda_J + (I_T \otimes \lambda_J W_J) K_J Q_0 v_J \tag{43}
$$

where

$$
\begin{aligned}
S_J &= I_{NT} + (I_T \otimes \lambda_J W_J)\ (I_T \otimes (I_N - \lambda_J W_J)^{-1}) \\
&= [I_T \otimes (I_N - \lambda_J W_J) + (I_T \otimes \lambda_J W_J)]\ [I_T \otimes (I_N - \lambda_J W_J)^{-1}] \\
&= I_T \otimes (I_N - \lambda_J W_J)^{-1} \\
&\equiv \Pi_J,\ \ J = 1, ..., G
\end{aligned} \tag{44}
$$

where $\Pi_J$ is defined in (12). Let $h_J = (I_T \otimes \lambda_J W_J) K_J Q_0 v_J$. It then follows from (42) - (44) that

$$
\begin{aligned}
Q_0 Y_J^{E,A} &= \Pi_J Q_0 M_J \phi_{1,J} + \Pi_J Q_0 P \phi_{2,J} + \Pi_J Q_0 y_{-1} \lambda_J + h_J \\
&\equiv Q_0 Y_J^{E,B} + h_J
\end{aligned} \tag{45}
$$

since $Q_0 \Pi_J = \Pi_J Q_0$. Thus, in finite samples the only difference between the use of $Y_J^{E,A}$ and $Y_J^{E,B}$ is the term $h_J$. However, asymptotically, $h_J$ is of no consequence for the same reason that $r_J$ is of no consequence. For example, it follows from (36) that $h_J$ can **only** effect the asymptotic distribution of $\tilde{\xi}_i$ if $plim_{N \to \infty} (NT)^{-1} H' h_J \neq 0$. However, $(NT)^{-1} H' h_J \xrightarrow{P} 0$ and so $h_J$ is of no consequence. To see this, let $\Psi_J = (NT)^{-1}H'h_J$. Now note from (34) that

$$\begin{aligned} E[\Psi_J|H_{1,J},\Upsilon_J] &= (NT)^{-1}H'(I_T \otimes \lambda_J W_J)K_J Q_0 E[v_J|H_{1,J},\Upsilon_J] \\ &= 0 \end{aligned} \tag{46}$$

and

$$E[\Psi_J\Psi_J'|H_{1,J},\Upsilon_J] = (NT)^{-2}H'[(I_T \otimes \lambda_J W_J)K_J][Q_0 V_{v_J} Q_0][K_J'(I_T \otimes \lambda_J W_J')]H \tag{47}$$

It follows from Assumption 3 and (35) that the row and column sums of the matrices in brackets in (47) are uniformly bounded in absolute value; in addition, from Assumption 4 the elements of $H$ are uniformly bounded in absolute value. Therefore, the elements of the VC matrix $E[\Psi_J\Psi_J'|H_{1,J},\Upsilon_J]$ are $0(N^{-1})$ and so $E[\Psi_J\Psi_J'|H_{1,J},\Upsilon_J] \to 0$. Chebyshev's inequality then implies that $\Psi_J \overset{P}{\to} 0$.

Of course, in practice $\overbrace{Q_0Y_J^{E,A}}$ and $\overbrace{Q_0Y_J^{E,B}}$ would be used instead of $Q_0Y_J^{E,A}$ and $Q_0Y_J^{E,B}$. Therefore the use of $\overbrace{Q_0Y_J^{E,A}}$ and $\overbrace{Q_0Y_J^{E,B}}$ would be asymptotically equivalent if, in light of (36), the data and estimation procedure associated with $H_{1,J}$ are such that the parameters are consistently estimated and

$$\begin{aligned} p\lim_{N\to\infty}(NT)^{-1}H'\overbrace{Q_0Y_J^{E,A}} &= p\lim_{N\to\infty}(NT)^{-1}H'Q_0Y_J^{E,A} \\ p\lim_{N\to\infty}(NT)^{-1}H'\overbrace{Q_0Y_J^{E,B}} &= p\lim_{N\to\infty}(NT)^{-1}H'Q_0Y_J^{E,B} \end{aligned} \tag{48}$$

**Proof of Theorem 1**

To simplify notation, we prove Theorem 1 for the case in which $\overbrace{Q_0Y_J^{E,A}}$ is used, $J = 1,...,G$. The proof for the case in which $\overbrace{Q_0Y_J^{E,B}}$ is used is virtually identical.

First note from (10), (14) and (15) that

$$\begin{aligned} \hat{Y}^A_{1,G} &= [Z_1, ..., Z_G]\ [\hat{\gamma}_1', ..., \hat{\gamma}_G']' \\ &= Z_{1,G}\ \hat{\gamma}_{1,G} \end{aligned} \tag{49}$$

where $Z_{1,G} = [Z_1, ..., Z_G]$ and $\hat{\gamma}_{1,G} = [\hat{\gamma}_1', ..., \hat{\gamma}_G']'$ where $\hat{\gamma}_J' = (\hat{\phi}_{1,J}', \hat{\phi}_{2,J}', \hat{\lambda}_J, \hat{\alpha}_J),\ J = 1, ..., G$. For this case $\hat{\Gamma}_A$ in (16) is

$$\hat{\Gamma}_A = [Z, Z_{1,G}\ \hat{\gamma}_{1,G}] \tag{50}$$

and $\Gamma_A = [Z, Z_{1,G}\ \gamma_{1,G}]$. Therefore, recalling that $\tilde{\Phi}_A = P_H \hat{\Phi}_A$ and $\hat{\Phi}_A = Q_0 \hat{\Gamma}_A$ it follows from (16) and (18) that

$$\begin{aligned} (NT)^{1/2}[\tilde{\xi}_i - \xi_i] &= \left\{ (NT)^{-1} \hat{\Gamma}_A' Q_0 H\ [(NT)^{-1} H'H]^{-1} (NT)^{-1} H' Q_0 \hat{\Gamma}_A \right\}^{-1} * \\ & (NT)^{-1} \hat{\Gamma}_A' Q_0 H\ (NT)(H'H)^{-1}\ (NT)^{-1/2} H' Q_0 R \varepsilon \end{aligned} \tag{51}$$

Consider the term $(NT)^{-1} H' Q_0 \hat{\Gamma}_A$ in the inverse on the first line of (51) and note that by Assumption 5, $\hat{\gamma}_{1,G} \xrightarrow{P} C_{1,G} = (c_1', ..., c_G')'$, where $c_J' = (c_{0,J}, c_{2,J}, L_J, a_J), J = 1, ..., G$. Part (b) of Assumption 6, and (49), imply

$$\begin{aligned} (NT)^{-1} H' Q_0 \hat{\Gamma}_A] &= p \lim_{N\to\infty} (NT)^{-1} H' Q_0 [Z, Z_{1,G}\ \hat{\gamma}_{1,G}] \\ &= [p \lim_{N\to\infty} (NT)^{-1} H' Q_0 Z,\ (p \lim_{N\to\infty} (NT)^{-1} H' Q_0 Z_{1,G})\ p \lim_{N\to\infty} \hat{\gamma}_{1,G}] \\ &= [p \lim_{N\to\infty} (NT)^{-1} H' Q_0 Z,\ (p \lim_{N\to\infty} (NT)^{-1} H' Q_0 Z_{1,G})\ \gamma_{1,G}] \\ &= p \lim_{N\to\infty} (NT)^{-1} H' Q_0 \Gamma_A \\ &= \Omega_{H Q_0 \Gamma_A} \end{aligned} \tag{52}$$

Let $\hat{F}$ be the inverse term on the first line of (51) and let

$$\tilde{F} = \hat{F} * [(NT)^{-1}\hat{\Gamma}'_A Q_0 H]\; [(NT)(H'H)^{-1}] \tag{53}$$

Then from (52), and parts (a) and (b) of Assumption 6, first note that

$$\hat{F} \xrightarrow{P} \left\{\Omega'_{HQ_0\Gamma_A}\Omega_{HH}^{-1}\Omega_{HQ_0\Gamma_A}\right\}^{-1} \tag{54}$$

where $\Omega'_{HQ_0\Gamma_A}\Omega_{HH}^{-1}\Omega_{HQ_0\Gamma_A}$ is positive definite and, therefore, nonsingular since $\Omega_{HH}$ is positive definite and so, therefore, is $\Omega_{HH}^{-1}$, and $\Omega_{HQ_0\Gamma_A}$ has full column rank. It then follows from (52) - (54), and and parts (a) and (b) of Assumption 6 that

$$\begin{aligned} \tilde{F} &\xrightarrow{P} F^* \\ F^* &= \left\{\Omega'_{HQ_0\Gamma_A}\Omega_{HH}^{-1}\Omega_{HQ_0\Gamma_A}\right\}^{-1}\Omega'_{HQ_0\Gamma_A}\Omega_{HH}^{-1} \end{aligned} \tag{55}$$

Finally, consider the last term on the second line in (51), namely $(NT)^{-1/2}H'Q_0R\varepsilon$. Since the row and column sums of both $Q_0$ and $R$ are uniformly bounded in absolute value, the row and column sums of $Q_0R$ are also uniformly bounded in absolute value. Since by Assumption 4 (a) the elements of $H$ are uniformly bounded in absolute value it follows that the elements of $H'Q_0R$ are uniformly bounded in absolute value. Given this, and Assumptions 1, 6 part (c), and the central limit theorem (30) in Pötscher and Prucha (2000) it follows that

$$(NT)^{-1/2}H'Q_0R\varepsilon \xrightarrow{D} N(0, \Omega_{H'Q_0RR'Q_0H}) \tag{56}$$

Therefore, by the continuous mapping theorem and (51) - (55)

$$(NT)^{1/2}[\tilde{\xi}_i - \xi_i] \xrightarrow{D} N(0, F^*\Omega_{H'Q_0RR'Q_0H}F^{*\prime}) \tag{57}$$