© *Investigaciones Regionales. 21 – Pages 71 to 90*
Section On theory and methods

Investigaciones Regionales

A research agenda on general-to-specific spatial model search

Peter Burridge *

ABSTRACT: The paper sets up a nesting spatial regression model incorporating heteroskedastic shocks, and discusses hypothesis testing in both nested and non-nested cases in a quasi-likelihood framework, suggesting directions for future research effort.

JEL Classification: C21.

Keywords: General-to-Specific, Research Agenda, Nesting Spatial Regression Models, Heteroskedasticity.

Una agenda de investigación sobre la búsqueda de modelos espaciales de lo general a lo particular

RESUMEN: El artículo propone un modelo de regresión espacial anidado en el cual se incorporan también *shocks* heteroscedásticos. Sobre este modelo se analizan contrastes de hipótesis tanto en casos anidados como no anidados, utilizando métodos de cuasi-verosimilitud y proponiendo líneas futuras de investigación.

Clasificación JEL: C21.

Palabras clave: De lo general a lo particular, Agenda de investigación, Modelos de regresión espacial anidados, Heteroscedasticidad.

1. Introduction

The paper is motivated by several recent research strands in which spatial econometric models are studied formally from a statistical perspective. Such models are sometimes criticised for a lack of clear economic foundations, yet there are also examples of models in which the features of interest are developed from first principles, such as the study of spillovers by Ertur and Koch (2007), and of trade flows

* Department of Economics and Related Studies, University of York.

Received: 25 march 2011 / Accepted: 27 july 2011.

by Behrens, Ertur and Koch (2010). Although the economic and social origins of the spatially mediated interactions and structures that enter the formal models are important, the purpose of this paper is to suggest directions in which the more narrowly formal analysis might go. In mainstream (predominantly time-series based) econometrics statistical techniques were developed through the latter half of the last century mostly by elaboration of relatively simple models that failed diagnostic tests - perhaps most notably in response to unfavourable outcomes of the Durbin-Watson test for serial correlation. However, it is now widely accepted that in a contemporary model-building exercise it is inefficient to imitate this historical sequence by starting with a simple model and elaborating it only when diagnostic tests are failed. Rather, a more effective strategy begins with a general model and seeks to reduce this by testing restrictions that lead to simpler models. The latter strategy has come to be associated with the LSE research agenda instigated by Sargan's so-called COMFAC analysis, and carried forward on a wide front in particular by Hendry (for the current state of the art, see Hendry 2011). In the spatial model context, a reconsideration by Mur and Angulo (2009) of the modeling strategies investigated experimentally by Florax, Folmer and Rey (2003) suggests that the so-called general to specific *(Gets)* strategy is superior to the specific-to-general *(Stge)* strategy. This is important, since the prevailing custom of adopting a version of *Stge* will be inefficient in some important cases, in line with the situation prevailing in time series modelling.

Within the model classes over which these searches are conducted, testing between non-nested models may be of interest, either for model selection or for specification checking, and here the improved $J - type$ test of Kelejian and Piras (2011) is a useful advance. Furthermore, there is a general awareness that spatially structured data are likely to be heteroskedastic, and that ignoring this phenomenon may lead at best to inefficient estimation results. Indeed, one of the advantages of the *Gets* strategy identified by Mur and Angulo was that it was much more robust to heteroskedastic, skewed or heavy-tailed disturbances than the competitor *Stge* strategy. Among papers dealing formally with heteroskedasticity, Anselin (1988a) devises a Lagrange Multiplier specification test for a classical linear regression model against a heteroskedastic spatially dependent alternative, and recently a practical algorithm for estimating Anselin's model by maximizing the Normal likelihood has been proposed by Yokoi (2010). IV/GMM-based estimators for Anselin's model with unknown heteroskedasticity have also recently been published by Kelejian and Prucha (2010) and Lin and Lee (2010).

These strands taken together suggest a research agenda:

a) Set up a satisfyingly general spatial model class from which the *Gets* strategy could begin.

b) Investigate identification and estimation algorithms for the general model.

c) Investigate tests of non-nested models for this class.

d) Devise and / or investigate tests for (nested) model reduction where these are unavailable or their properties are not known.

e) Investigate the performance of the *Gets* and *Stge* strategies in this richer setting.

The paper comments on some aspects of a) - d). A significant part of the discussion is speculative. The proposed general model is introduced next, and is seen to nest the SARAR model and the Spatial Durbin Model (SDM), also described as the first order spatial Autoregressive Distributed Lag (ADL) model by Bivand (1984, eq. 4), the Spatial Durbin Error Model, the Spatial Lag Model and the Spatial Error Model, each of which is defined below.

2. A heteroskedastic general nesting model (HGNM)

It would be natural to start a *Gets* - type analysis with a model in which popular simpler ones are nested. In principle, this is achieved by what Elhorst (2010) calls the Manski model, after Manski (1993), also mentioned as a possibility by LeSage and Pace (2009, p. 53), and which could form a starting point for a *Gets* procedure, or a possible endpoint for a *Stge* procedure. This paper prefers to call the model the HGNM because the identification problem discussed by Manski does not arise, in general, for this model, contrary to the impression given by some authors because of its formal similarity to Manski's model. The nesting model is elaborated slightly here by the inclusion of heteroskedastic shocks and by relaxing the restriction that weight matrices are equal [1]:

$$\begin{aligned} \mathbf{Y} &= \lambda_0 \mathbf{W}_0 \mathbf{Y} + 1\delta_0 + \mathbf{X}_0\beta_0 + \mathbf{Q}_0\mathbf{X}_0\gamma_0 + \mathbf{U}_0 \\ \mathbf{U}_0 &= \rho_0 \mathbf{M}_0 \mathbf{U}_0 + \varepsilon_0 \\ & \quad \varepsilon_0 \sim N(0, \Omega_0) \\ \omega_{0,ii} &= h_0(\alpha_0' \mathbf{Z}_{0,i}) > 0, \quad \omega_{0,ij} = 0, \quad i \neq j. \end{aligned} \tag{1}$$

The constant regressor, $\mathbf{1} = [1, 1, 1,..., 1]'$ is separately treated in the notation to allow for weight matrices that are row-normalised, such that, for example, $\mathbf{Q}_0 1 = 1$. In a very simple case, the variance of the i^{th} shock might be proportional to some measure of the «size» of region i. An alternative Bayesian approach to heteroskedasticity that does not depend on a prespecified $h()$ function is described by LeSage and Pace (2009, Section 5.6.1). As is often remarked, a more local spatial averaging of shocks could be achieved by the use of a moving average specification, such as $\mathbf{U}_0 = \varepsilon_0 + \rho_0 \mathbf{M}_0 \varepsilon_0$ but this possibility is not taken up here.

As soon as the model (1) is contemplated, an obvious restriction that might need to be tested is that the weight matrices are the same: $\mathbf{W} = \mathbf{Q} = \mathbf{M}$; indeed, only if $\mathbf{W} = \mathbf{Q}$ does the possible existence of the common factor mentioned below arise. Also, there may be competing models within the same class, just as in the $J - test$ set-up adopted by Kelejian (2008) and Kelejian and Piras (2011).

[1] Anselin (1988b) attributes to Hordijk (1979) the introduction of a SARAR model with weights that are different for the spatial lag and spatial error.

Before settling on (1) as the general model, however, we should consider whether or not a yet more general starting point is required. In the time series context, it is now usual to regard models with serially correlated disturbances as restricted forms of more general models with richer dynamics. Following Hendry and Mizon (1978) who implemented the COMFAC analysis being developed by Sargan in the mid 1970's that was eventually published in Sargan (1980), we might consider (1) as itself a restricted form of the model,

$$\mathbf{Y} = \lambda_1 \mathbf{W}_1 \mathbf{Y} + \lambda_2 \mathbf{W}_2 \mathbf{Y} + 1\delta + \mathbf{X}\beta_0 + \mathbf{W}_3 \mathbf{X}\beta_1 + \mathbf{W}_4 \mathbf{X}\beta_2 + \varepsilon. \tag{2}$$

This possibility has been discussed by Blommestein (1983), and again recently by Mur and Angulo (2006). If we take (2) seriously, then a first model simplification step would seem to be to test the hypothesis,

$$\mathcal{H}_w : \mathbf{W}_2 = \mathbf{W}_1^2, \mathbf{W}_4 = \mathbf{W}_3^2, \mathbf{W}_3 = \mathbf{W}_1; \tag{3}$$

however, the essential difference between the time series and spatial cross-section cases then becomes apparent: while in time series the term, $\mathbf{W}_2\mathbf{Y}$ just represents a two-period lag of $\mathbf{Y}$ which results from applying the lag operator twice, in the spatial setting there is in general no obvious equivalent construction[2]. Thus if the analysis were to start from (2) the specification of the four weight matrices would be problematic from the outset if it were desired to test for the possible simplification. Of course, a more feasible alternative starting point would be to impose $\mathcal{H}_w$ and test the implied common factor restriction that would then reduce (2), with $\mathcal{H}_w$ maintained, to (1).

In time series models, there is an obvious value in representations in which the unobserved shocks may be treated as innovations, that is, as independent of the previous history of the quantities under study, including previous innovations. How far it is appropriate to seek models in which shocks are independent over space has been, I think, much debated. The key may be in the conditioning information brought into the analysis at the outset. For example, as long argued in the literature, and described by LeSage and Pace (2009, pp. 27-28, 67-68) when spatially-patterned explanatory variables are omitted from the model's mean function, they will enter the disturbance term, thus producing a spatially autocorrelated disturbance that could be eliminated by their inclusion in the mean. On the other hand, rather stronger grounds may be found for introducing spatially lagged dependent variables to the right-hand-side, such as when data are observed at a lower frequency (in time) than that at which agents take decisions that can be influenced by those of their neighbours, or in the group interaction models now gaining in popularity (see Lee, Liu and Lin (2010) for a recent example). Although a residual doubt over model specification is unavoidable, to make progress, we have to suppose that the investigator gets something right,

[2] Exceptions are the «two weight matrix» model of Lacombe (2004) discussed by Le Sage and Pace (2009, p. 52), and the model explored by Brandsma and Ketellapper (1979) in which $\mathbf{Y} = \mathbf{X}\beta + \mathbf{U}$ with $(\mathbf{I} - \rho_1 \mathbf{W}_1 - \rho_2 \mathbf{W}_2)\mathbf{U} = \varepsilon$, and a likelihood ratio test of the hypothesis that $\rho_1 = \rho_2 = 0$ is implemented.

and so for this rather pragmatic reason, and because it has not received much attention, this paper treats (1) as the initial general model, supposing it to have passed such diagnostic checks as are available. If, in fact, a test of the hypothesis, $\rho_0 = 0$, failed to reject, our confidence that no major systematic spatially patterned explanatory factor had been omitted would of course increase.

2.1. Nested Models

2.1.1. The SARAR model

Elhorst (2010) designates the model containing a spatially lagged dependent variable and a spatially autoregressive disturbance, the Kelejian-Prucha model - see Elhorst (2010, p. 13). LeSage and Pace (2009 p. 32) on the other hand designate this the SAC model; since Kelejian (2008) calls the model the SARAR model, that name seems a reasonable compromise, the repeated AR a reminder of its essential feature.

In Yokoi (2010) the MLE for the heteroskedastic SARAR model described by Anselin (1988a,b) is studied. The model is:

$$\mathbf{Y} = \lambda_0 \mathbf{W}_0 \mathbf{Y} + 1\delta_0 + \mathbf{X}_0 \beta_0 + \mathbf{U}_0 \quad (4)$$

$$\begin{aligned} U_0 &= \rho_0 \mathbf{M}_0 \mathbf{U}_0 + \varepsilon_0 \\ & \varepsilon_0 - N(0, \Omega_0) \end{aligned} \quad (5)$$

$$\omega_{0,ii} = h_0(\alpha_0' \mathbf{Z}_{0,i}) > 0, \quad \omega_{0,ij} = 0, \quad i \neq j. \quad (6)$$

As can be seen, it arises from the HGNM by the exclusion of the spatially lagged exogenous variables, $\mathbf{Q}_0\ \mathbf{X}_0\ \gamma_0$. However, with a little care over the treatment of any accidental collinearity between **X**, **WX** and **QX**, it is easy to see that the definition of $\mathbf{X}_0\ \beta_0$ in (4) can be expanded to include $\mathbf{Q}_0\ \mathbf{X}_0\ \gamma_0$ from (1). This is useful because it means that estimator properties derived for the SARAR model may, with a little care, apply readily to the more general model. The extra care involved is obvious in the case of IV based estimators that rely on use of instruments such as $\mathbf{W}_0\ \mathbf{X}_0$, and so on, to take care of the correlation between the disturbances and the spatially lagged dependent variable: any spatially lagged exogenous variables that are already present on the right-hand side are not available as additional instruments.

2.1.2. The Spatial Durbin Model

Consider the so-called Spatial Durbin model (SDM), obtained from the HGNM when $\rho_0 = 0$. This model has been widely promoted as a possible starting point because it nests two popular simpler models; LeSage and Pace (2009, pp. 67-68) also argue that the model produces estimates with a degree of robustness to omitted variables not shared for example by the nested models. The SDM is

$$\begin{aligned} \mathbf{Y} &= \lambda_0 \mathbf{W}_0 \mathbf{Y} + 1\delta_0 + \mathbf{X}_0\beta_0 + \mathbf{W}_0\mathbf{X}_0\gamma_0 + \varepsilon_0 \\ & \varepsilon_0 \_ N(0, \Omega_0). \end{aligned} \tag{7}$$

The Spatial Error Model. To see how the SDM may be simplified under certain restrictions, suppose, for convenience, that the rows of $\mathbf{W}_0$ sum to 1 so that $\mathbf{W}_0\ 1 = 1$ and observe that (7) may be written equivalently, by taking out a factor of $(I - \lambda_0\ \mathbf{W}_0)$ on the right-hand side, as

$$(\mathbf{I} - \lambda_0 \mathbf{W}_0)\mathbf{Y} = (\mathbf{I} - \lambda_0 \mathbf{W}_0)1\delta_0 / (1 - \lambda_0) + (\mathbf{I} - \lambda_0 \mathbf{W}_0)\mathbf{X}_0\beta_0 + \mathbf{W}_0\mathbf{X}_0(\lambda_0\beta_0 + \gamma_0) + \varepsilon_0$$

where the remainder, $\mathbf{W}_0\ \mathbf{X}_0\ (\lambda_0\ \beta_0\ +\ \lambda_0)$ is now of interest. If the parameters satisfy the so-called common-factor restriction,

$$\lambda_0 = \lambda_0\beta_0 \tag{8}$$

the remainder vanishes, and the matrix, $(\mathrm{I} - \lambda_0\ \mathbf{W}_0)$, is seen to be a common factor in the model. If this matrix is invertible, as usually assumed, the model simplifies to the spatial error model [3].

$$\mathbf{Y} = 1\delta_0 / (1 - \lambda_0) + \mathbf{X}_0\beta_0 + \mathbf{U}_0 \tag{9}$$

$$= 1\delta_0^* + \mathbf{X}_0\beta_0 + \mathbf{U}_0 \text{ say, with} \tag{10}$$

$$(\mathbf{I} - \lambda_0 \mathbf{W}_0)\mathbf{U}_0 = \varepsilon_0.$$

The Spatial Lag Model. More obviously perhaps, when $\lambda_0 = 0$ the SDM reduces to the spatial lag model, studied in a Normal likelihood framework by Ord (1975). The SLM is:

$$\mathbf{Y} = \lambda_0 \mathbf{W}_0 \mathbf{Y} + 1\delta_0 + \mathbf{X}_0\beta_0 + \varepsilon_0 \tag{11}$$

and is the generic model for spatially interacting responses to changes in conditioning variables and shocks.

2.1.3. The Spatial Durbin Error model

Elhorst (2010) comments that the model that results from (1) when $\lambda = 0$, called the SDEM by LeSage and Pace (2009, p. 42) does not seem to have been used much.

[3] See also Burridge (1981), Bivand (1984), Anselin (1988b), Folmer Florax and Rey (2003), Elhorst (2001, 2010), LeSage and Pace (2009), and Mur and Angulo (2009) for more discussion.

I don't know why it should have been overlooked, however, though nesting it in the more general model being discussed here may lead to its more frequent use.

2.1.4. Preference for the SDM

Of course, neither the restriction that **Q** = **W**, nor the common-factor restriction (8), nor the zero restrictions for λ_0 ρ_0 or γ_0 leading to the simpler models may be plausible; besides the nesting of (9) and (11) the SDM has other merits, delivering unbiased coefficient estimates, according to LeSage and Pace (2009, pp. 56-158) when the other models may fail to do so, a point echoed by Elhorst (2010, pp. 14-15).

3. The General-to-Specific Strategy in Outline

With a single fixed weight matrix, treated as given, the first few steps, which expand the strategy investigated by Mur and Angulo (2009), could be as follows

1. Estimate (1) with **M** = **W** = **Q** (it is assumed that any available diagnostic tests have been passed, see Section 4.1 below for more on this point).
2. Test for simplification to homoskedasticity $\mathcal{H}_\alpha : \alpha_2 = \alpha_3 = ... = \alpha_m = 0$.
3. Test for simplification to SDM/ADL $\mathcal{H}_\rho : \rho = 0$.
4. If $\mathcal{H}_\rho$ is not rejected test for simplification to SLM $\mathcal{H}_\gamma : \gamma = 0$; if $\mathcal{H}_\rho$ is rejected test for simplification to SARAR $\mathcal{H}_\gamma : \gamma = 0$.
5. If $\mathcal{H}_\rho$ is not rejected at Step 3 but $\mathcal{H}_\gamma$ is rejected at Step 4, test for common factor and reduction to SEM; if $\mathcal{H}_\rho$ is rejected at Step 3, and $\mathcal{H}_\gamma$ is rejected at Step 4, test $\mathcal{H}_\lambda : \lambda = 0$ for simplification to the SDEM.

With different, but fixed, weight matrices, the first step could be to seek a simplification via a non-nested test, as described below.

4. Test procedures

4.1. Diagnostics for the HGNM?

A critical ingredient in the *Gets* strategy is the assumption that the general nesting model is itself an adequate description of the data generating process, the DGP. In the time series context, in which the detection and accommodation of serial correlation was the key problem, the leading requirement was for a test for serial correlation in the disturbances of a dynamic model that could be applied after, say, an ARMA(p,q) model had been fitted to the data. As is well known, the Durbin-Watson test could not be used in such a model as the sampling distribution of the statistic is shifted when lagged values of the dependent variable are present resulting in a bias towards acceptance of the null hypothesis. The critical advance here was the development of a

Lagrange multiplier test for serial correlation in dynamic models by Breusch (1978) and Godfrey (1978). In the spatial case the main diagnostic required will play a similar role, and is thus a test for neglected spatial correlation in the disturbance, ε_0 of (1). Such a test has yet to be developed, apparently.

4.2. A general non-nested test procedure $\bar{\beta}_0$

Consider the problem of testing a model of the form (1) against a non-nested alternative, of the same form, Model$_1$, say. Broadly speaking, the *J*-test approach implemented for the SARAR model by Kelejian (2008), as modified by Kelejian and Piras (2011), would entail the construction of a prediction of $(\mathbf{I} - \rho_0\, \mathbf{M}_0)Y$ from Model$_1$ which would be added as an explanatory variable to an equation predicting $(\mathbf{I} - \rho_0\, \mathbf{M}_0)\mathbf{Y}$ using Model$_0$. Suppose the models satisfy relevant sets of sufficient conditions for identification, and that Gaussian quasi-maximum likelihood[4] estimates of the parameters of the two models are available, and write these as $\tilde{\delta}_0$, $\bar{\beta}_0$, ..., $\tilde{\delta}_1$, $\bar{\beta}_1$, ..., and so on. Imitating the Kelejian and Piras approach but implementing QMLE for all but the final test regression leads to the following. Initially, ignoring the heteroskedasticity, using Model$_1$ construct the predictor,

$$\bar{\mathbf{Y}}^1 = \tilde{\lambda}_1 \mathbf{W}_1 \mathbf{Y} + 1\tilde{\delta}_1 + \mathbf{X}_1 \bar{\beta}_1 + \mathbf{Q}_1 \mathbf{X}_1 \bar{\gamma}_1.$$

From Model$_0$ estimate $\tilde{\lambda}_0$, $\tilde{\delta}_0$, $\bar{\beta}_0$, $\bar{\gamma}_0$, $\tilde{\rho}_0$. Using $\tilde{\rho}_0$ construct the «whitened» dependent variable,

$$\mathbf{Y}^*(\tilde{\rho}_0) = (\mathbf{I} - \tilde{\rho}_0 \mathbf{M}_0)\mathbf{Y} \tag{12}$$

together with the transformed RHS variables,

$$\mathbf{Z}_0^*(\tilde{\rho}_0) = (\mathbf{I} - \tilde{\rho}_0 \mathbf{M}_0)\,[\mathbf{X}_0,\ \mathbf{Q}_0\mathbf{X}_0,\ \mathbf{W}_0, \mathbf{Y}] \tag{13}$$

and the transformed predictor,

$$\bar{\mathbf{Y}}^1(\tilde{\rho}_0) = (\mathbf{I} - \tilde{\rho}_0 \mathbf{M}_0)\ \bar{\mathbf{Y}}^1. \tag{14}$$

The idea behind the test is now to add $\bar{\mathbf{Y}}^1\,(\tilde{\rho}_0)$ to the right-hand side of the equation,

$$\mathbf{Y}^*(\tilde{\rho}_0) = \mathbf{Z}_0^*(\tilde{\rho}_0)\hat{\phi}_0^* + \bar{\mathbf{Y}}^1(\tilde{\rho}_0)\hat{\psi}_{01}^* + \mathbf{e}_0^*,\ \text{say}, \tag{15}$$

[4] Kelejian and Piras do not employ quasi-maximum likelihood estimators, but they are preferred here for use in modestly-sized samples because they satisfy the determinantal conditions on λ and ρ.

and test the hypothesis that $\psi^*_{01} = 0$. To extend the procedure to the heteroskedastic case it would seem natural to premultiply (15) by $\tilde{\Omega}_0^{-1/2}$ which is the estimate of the diagonal variance matrix of the disturbance that corresponds to the residual, $\mathbf{e}^*_0$ under the null hypothesis. The specification of the test based on (15) differs from the Kelejian and Piras test for the SARAR model in two respects. Evidently, the model has been expanded by the introduction of the spatially lagged exogenous regressors, $\mathbf{Q}_i\ \mathbf{X}_i$ $(i = 0, 1)$; however, their presence introduces nothing of great significance since the various conditions imposed by Kelejian and Piras should require only a very minor expansion to accommodate this change - conditions on the matrix, **X** must now be applied to the matrix, [**X**, **QX**] and in their approach instruments would need to be chosen with care to avoid rank deficiency. Secondly, except for the final equation which is estimated using instrumental variables, the parameters are estimated by Gaussian QML to guarantee that they satisfy the determinantal conditions, $|(\mathbf{I} - \rho\mathbf{M}_0)| > 0$, $|(\mathbf{I} - \lambda\mathbf{W}_0)| > 0$ and similarly for Model_1 . To the best of the author's knowledge, a J-type test adapted to accommodate heteroskedasticity has not yet been implemented, and so its development along the lines above seems warranted. The tasks involved include establishing the asymptotic sampling distribution of such a statistic, checking its small sample performance and devising any correction that may be necessary to control significance levels.

4.3. Information Criteria and the Likelihood

In a time series modelling exercise it is usual to examine a so-called «information criterion» such as AIC, or BIC, to select model order. For example, when fitting an AR(p) model to a single time series, such as

$$y_t = \sum_{j=1}^{j=p_0} \phi_j y_{t-j} + \varepsilon_t \tag{16}$$

under the maintained assumption that ε_t is while noise, the order of the autoregressive operator could be chosen to minimise the BIC, $\ln \hat{\sigma}^2_p + p \ln n/n$, in which $\hat{\sigma}^2_p$ is the quasi maximum likelihood estimate of the innovation variance from the model with $j = p$, and is proportional to the negative of the log conditional likelihood. That the choice, $\tilde{p}_n$ say, which minimises this criterion, is consistent in the sense that $\lim_{n\to\infty} \Pr[\tilde{p}_n = p_0] = 1$ has been demonstrated under very general conditions, reviewed and extended in a recent contribution by Burridge and Hristova (2008). However, although the parallel with time series modelling is appealing, and model selection via an information criterion was suggested as a simpler alternative to use of a $J - type$ test 25 years ago by Haining (1986), there does not appear to have been a systematic investigation of its properties in the spatial case; it should be noted that a treatment of consistency would require explicit conditions relating to the evolution of the weights matrices as sample size increased, moment conditions on regressors, and the like, similar in nature to those introduced by Lee (2004), but also that a fundamental problem remains to be addressed. The difficulty arises from the fact that the competing models are not nested; because of this, the fact that Model A delivers a higher value of its maximised likelihood than Model B does of *its* likelihood is not sufficient for Model A to be preferred, and introduction of a «penalty» for additional parameters, as in the BIC, has no bearing on this fundamental problem. Nevertheless, as suggested by a referee, a comparison between such model selection and use of the $J - test$ in finite samples could be interesting. Closely related is the Bayesian approach described by Le Sage and Pace (2009, Section 6.3) and applied by Pijnenburg and Kholodilin (2011) who consider 43 different weight matrices in their study of entrepreneurial spillovers, choosing the one that delivers the highest posterior model probability. In this framework, there are three components to the model posterior probability, a prior over the various weight matrices, $\pi(W)$, a prior over the parameters for each W, $\pi(\theta|W)$, and the likelihood of the data given W and θ, $p(D|\theta, W)$. In effect, if $\pi(W)$ is chosen to be uninformative, choosing the model with the highest posterior probability amounts to choosing the model for which the smoothed ([i.e. integrated over $\pi(\theta|W)$] likelihood is highest. The problem of comparing likelihoods from different|probability models remains, therefore, within the Bayesian formalism.

4.4. Tests of nested models

With the rather general starting point, (1), natural hypotheses to test are parametric restrictions that simplify the model. These could be of various kinds, of which several are described below.

4.4.1. Tests on weight matrices

Hypotheses that might be tested include, as an example, $H_{01} : \mathbf{Q} = \mathbf{W} = \mathbf{W}_0$ say, with $\mathbf{M} = \mathbf{M}_0$ maintained vs $H_{11} : \mathbf{Q} = \mathbf{Q}_1$ and $\mathbf{W} = \mathbf{W}_0$ with $\mathbf{M} = \mathbf{M}_0$ maintained. The point here is to seek model simplification, since the common factor reduction only arises as a possibility if $\mathbf{Q} = \mathbf{W}$. At the current state of development of the field (in which results for structures within which $\mathbf{W}$ and so on may be *estimated* from the sample, are not yet available) such hypotheses should probably be approached via methods developed for testing non-nested models. Thus H_{01} would correspond to Model_0 and H_{11} to Model_1 and a test could be based on (15).

4.4.2. A test for heteroskedasticity

To test parametric restrictions that simplify the model, the usual likelihood ratio machinery could be used, or LM-type tests be developed for cases in which the restricted model was significantly simpler to estimate than the unrestricted one. A case in point could be, given a parametric model (or a linear approximation to such a model) for the covariance matrix, Ω, a test of the homoskedastic null hypothesis, $h(\alpha'\mathbf{z}_i) = h([\alpha_1, 0, ..., 0]'\mathbf{z}_i) = \sigma^2$ = a constant, say. To see the form such a test could take[5], consider the first-order condition, (27)

$$\frac{\partial l}{\partial \alpha_p} = -\frac{1}{2}Tr\{\Omega^{-1}\mathbf{H}_p\} + \frac{1}{2}\varepsilon'\Omega^{-2}\mathbf{H}_p\varepsilon$$

where $\mathbf{H}_p = diag\{\partial\omega_{ii}/\partial\alpha_p\}$, and the corresponding block of the information matrix, which has typical element (76)

$$-E\left\{\frac{\partial^2 l}{\partial\alpha_p\partial\alpha_q}\right\} = \frac{1}{2}Tr\{\Omega^{-2}\mathbf{H}_p\mathbf{H}_q\}.$$

Stacking the first derivatives for $\alpha_2, ..., \alpha_m$ into the vector, $\mathbf{d}$, and writing $\mathbf{I}_{\alpha\alpha}$ for the corresponding part of the information matrix, the usual form for the LM statistic (in the block diagonal case) would be

$$LM = \bar{\mathbf{d}}'\bar{\mathbf{I}}_{\alpha\alpha}^{-1}\bar{\mathbf{d}}$$

in which both $\mathbf{d}$ and $\mathbf{I}_{\alpha\alpha}$ are evaluated at the null hypothesis. Now, observe, as in Breusch and Pagan (1979) and Anselin (1988), that

$$\mathbf{H}_p(i, i) = \partial\omega_{ii}/\partial\alpha_p = \frac{\partial}{\partial\alpha_p}\{h(\alpha'\mathbf{z}_i)\} = h\cdot(\alpha'\mathbf{z}_i)z_{ip}$$

where $h\cdot(s_i) = \partial h/\partial s_i$ with $s_i = \alpha'\mathbf{z}_i$. Under the homoskedastic null, the covariance matrix estimator reduces to $\bar{\Omega} = \hat{\sigma}^2\mathbf{I}$ and $h\cdot(\alpha'\mathbf{z}_i) = h\cdot(\alpha_1)$, so that, evaluated at the constrained estimator,

$$\frac{\partial l}{\partial\alpha_p}|_{null} = -\frac{1}{2}\sum_{i=1}^{i=n}\tilde{\sigma}^{-2}h\cdot(\alpha_1)z_{ip} + \frac{1}{2}\sum_{i=1}^{i=n}\tilde{\sigma}^{-4}h\cdot(\alpha_1)z_{ip}\tilde{\varepsilon}_i^2.$$

Writing $g_i = \dfrac{\tilde{\varepsilon}_i^2}{\tilde{\sigma}^2}$ this can be simplified to

$$\frac{\partial l}{\partial\alpha_p}|_{null} = \frac{1}{2}\tilde{\sigma}^{-2}h\cdot(\alpha_1)\sum_{i=1}^{i=n}z_{ip}(g_1 - 1). \qquad (17)$$

[5] Ignoring the off-diagonal blocks involving ρ and λ.

Similarly,

$$-E\left\{\frac{\partial^2 l}{\partial\alpha_p\partial\alpha_q}\right\}\Big|_{null} = \frac{1}{2}\tilde{\sigma}^{-4}\sum_{i=1}^{i=n}[h\cdot(\alpha_1)]^2 z_{ip}z_{ip}. \tag{18}$$

Putting these objects together the test statistic,

$$LM = \bar{\mathbf{d}}'\bar{\mathbf{I}}_{\alpha\alpha}^{-1}\bar{\mathbf{d}} = \frac{1}{2}\mathbf{f}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{f} \tag{19}$$

is obtained, where $\mathbf{Z} = [\mathbf{z}_1, ..., \mathbf{z}_n]'$ and $\mathbf{f} = \mathbf{g} - \mathbf{1}$. As in Breusch and Pagan (1979, p. 1290) it is found that (19) is one half the explained sum of squares from regression of g_i on $\mathbf{z}_i$. Notice that the test for heteroskedasticity in the presence of spatially lagged dependent variables devised by Anselin (1988a) maintains $(\lambda, \rho) = (0, 0)$ which is quite restrictive. Whether or not information about (ρ, λ) can be exploited to improve the test at (19) is a question that should be investigated.

Kelejian and Robinson (1998) present a test they designate, KR-SPHET, that has the absence of both spatial correlation and heteroskedasticity as its null hypothesis, mentioning in a remark (Remark 5, p. 395) a possible modification that could be used to test for heteroskedasticity with spatial correlation maintained. Their test is similar in spirit to the Breusch-Pagan test in that it employs a regression of squares and cross-products of residuals on regressors supposed related to the heteroskedasticity under the alternative.

4.4.3. A better approach to tests on the weights matrices?

While the non-nested testing procedure could be used to test hypotheses about the weights, a more natural and flexible approach would be to have a parametric model for the weights matrices derived from economic theory, and to construct tests in a nesting model. Suppose **W** has elements $w_{ij} = f(d_{ij}, \tau_w)$, **M** has elements $m_{ij} = f(d_{ij}, \tau_m)$ and **Q** has elements $q_{ij} = f(d_{ij}, \tau_q)$ in which the d_{ij} are observed (distances or adjacency measures, or other indices of interactivity) and the τ are parameters to be estimated. In this framework, likelihood ratio tests of restrictions on the τ parameters can easily be formulated. To begin the development of such tests, a simpler homoskedastic nesting model could be studied. Consider the following model, in which $\varepsilon = \eta\cdot\sigma$ with $\eta \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$. The log-likelihood can be written

$$l(\mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{Q}, \mathbf{M}, \delta, \beta, \gamma, \lambda, \rho, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\sigma^2 + \ln\left|\mathbf{I} - \lambda\mathbf{W}\right| + \ln\left|\mathbf{I} - \rho\mathbf{M}\right| - \frac{1}{2}\eta'\eta$$

where as above, the sum of squares term is

$$\begin{aligned}\eta'\eta &= \varepsilon'\varepsilon / \sigma^2 \\ \varepsilon &= (\mathbf{I} - \rho\mathbf{M})([\mathbf{I} - \lambda\mathbf{W}]\mathbf{Y} - 1\delta - \mathbf{X}\beta - \mathbf{QX}\gamma) \\ &= (\mathbf{I} - \rho\mathbf{M})\mathbf{U}.\end{aligned}$$

The matrices **M**, **W** and **Q** are defined by

$$\begin{aligned}m_{ij} &= f(d_{ij}, \tau_m) \\ w_{ij} &= f(d_{ij}, \tau_w) \\ q_{ij} &= f(d_{ij}, \tau_q)\end{aligned}$$

For compactness, as before, write $\mathbf{B} = (\mathbf{I} - \rho\mathbf{M})$ and $\mathbf{A} = (\mathbf{I} - \lambda\mathbf{W})$, with **A** being non-singular for (λ, τ_w) in a neighbourhood of $(\lambda_0, \tau_{w,0})$ and similarly **B** being non-singular for (ρ, τ_m) in a neighbourhood of $(\rho_0, \tau_{m,0})$. Evidently, provided the model is identified, LR tests could be constructed numerically. Whether convenient alternative tests can be devised is another open question. A precedent for estimating the weights does exist, in the work of Bodson and Peeters (1975, p.467), though no systematic treatment appears to be available in the literature.

4.5. Approximate sampling distributions and the bootstrap

In the model class under discussion here, neither least squares regression estimates nor likelihood ratio statistics will have exactly known sampling distributions except possibly in very special cases. There are at least two responses to this. First, it is possible to search for meaningful conditions under which the sampling distributions of estimators and test statistics converge to known standard distributions as the sample size increases. If such conditions turn out to be difficult to obtain, or at odds with the way in which empirical models are usually specified, then due caution needs to be exercised. However, even if the conditions under which the relevant convergence in distribution can be established are empirically reasonable, there remains the problem of controlling significance levels in finite samples. This motivates the second response, namely the use of resampling to obtain approximate sampling distributions. The properties of bootstrap-based approximations to sampling distributions have yet to be investigated in the context of this model.

5. Final comments

The formal statistical analysis of regression models that embody spatial interactions is enjoying a resurgence of interest, and some of the important properties of estimators and test statistics have been established with the help of equipment developed over the past decade and a half in numerous papers by Kelejian and Prucha, and Lee, their collaborators, and others. These authors' work provides a rigorous account of the large sample behaviour of various tests and estimators in which, as the sample size grows, so the elements of the spatial weight matrix, **W**, evolve in a specific way, and in which the regressors obey some quite natural restrictions. These are real advances. However, in spite of all this progress, we are still unable to provide satisfactory answers to some seemingly obvious questions about the structure of the models themselves. These questions are the subject of the present paper, and of other recent contributions that focus on model selection. In writing rather speculatively about a model that nests many of those currently in favour for handling data derived from a single crosssection, my purpose has been to suggest avenues that merit further exploration and formal study.

6. Appendix: The likelihood for the general nesting model

For convenience, write $\varepsilon = \Omega^{1/2}\eta$ where $\eta \sim \mathbf{N}(\mathbf{0}, \mathbf{I})$. The heteroskedastic nesting model with Normal shocks has log-likelihood that can be written

$$l(\mathbf{Y}, \mathbf{X}, \mathbf{W}, \mathbf{Q}, \mathbf{M}, \delta, \beta, \gamma, \lambda, \rho, \Omega) = -\frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Omega| + \ln|\mathbf{I} - \lambda\mathbf{W}| + \ln|\mathbf{I} - \rho\mathbf{M}| - \frac{1}{2}\eta'\eta \tag{21}$$

where the sum of squares term is

$$\begin{aligned} \eta'\eta &= \varepsilon'\Omega^{-1}\varepsilon \\ \varepsilon &= (\mathbf{I} - \rho\mathbf{M})([\mathbf{I} - \lambda\mathbf{W}]\mathbf{Y} - 1\delta_0 - \mathbf{X}\beta - \mathbf{QX}\gamma) \\ &= (\mathbf{I} - \rho\mathbf{M})\mathbf{U}. \end{aligned}$$

For compactness, write $\mathbf{B} = (\mathbf{I} - \rho\mathbf{M})$ and $\mathbf{A} = (\mathbf{I} - \lambda\mathbf{W})$, both matrices being non-singular by assumption. The first partial derivatives of the log-likelihood are (cf. Anselin (1988a) where the roles of λ and ρ are reversed, and our **W**, **M** are his $\mathbf{W}_1$, $\mathbf{W}_2$ but the lagged exogenous variables, **QX** do not appear in his model):

$$\frac{\partial l}{\partial \delta} = \mathbf{1}'\mathbf{B}'\Omega^{-1}\varepsilon \tag{22}$$

$$\frac{\partial l}{\partial \beta} = \mathbf{X}'\mathbf{B}'\Omega^{-1}\varepsilon \tag{23}$$

$$\frac{\partial l}{\partial \gamma} = \mathbf{X'Q'B'}\Omega^{-1}\varepsilon \tag{24}$$

$$\frac{\partial l}{\partial \lambda} = -Tr\{\mathbf{WA}^{-1}\mathbf{W}\} + \varepsilon'\Omega^{-1}\mathbf{BWY} \tag{25}$$

$$\frac{\partial l}{\partial \rho} = -Tr\{\mathbf{B}^{-1}\mathbf{M}\} + \varepsilon'\Omega^{-1}\mathbf{MB}^{-1}\varepsilon \tag{26}$$

$$\frac{\partial l}{\partial \alpha_p} = -\frac{1}{2}Tr\{\Omega^{-1}\mathbf{H}_p\} + \frac{1}{2}\varepsilon'\Omega^{-2}\mathbf{H}_p\varepsilon \tag{27}$$

where $\mathbf{H}_p = diag\{\partial\omega_{ii}/\partial\alpha_p\}$. The second partial derivatives are

$$\frac{\partial^2 l}{\partial \delta^2} = -\mathbf{1'B'}\Omega^{-1}\mathbf{B1} \tag{28}$$

$$\frac{\partial^2 l}{\partial \beta \partial \beta'} = -\mathbf{X'B'}\Omega^{-1}\mathbf{BX} \tag{29}$$

$$\frac{\partial^2 l}{\partial \gamma \partial \gamma'} = -\mathbf{X'Q'B'}\Omega^{-1}\mathbf{BQX} \tag{30}$$

$$\frac{\partial^2 l}{\partial \lambda^2} = -Tr\{[\mathbf{A}^{-1}\mathbf{W}]^2\} \tag{31}$$

$$-\mathbf{Y'W'B'}\Omega^{-1}\mathbf{BWY} \tag{32}$$

$$\frac{\partial^2 l}{\partial \rho^2} = -Tr\{[\mathbf{B}^{-1}\mathbf{M}]^2\} - \mathbf{U'M'}\Omega^{-1}\mathbf{MU} \tag{33}$$

$$\frac{\partial^2 l}{\partial \alpha_p^2} = \frac{1}{2}\{Tr[\Omega^{-2}\mathbf{H}_p^2 - \Omega^{-1}\mathbf{H}_{pp}] + \varepsilon'\Omega^{-2}\mathbf{H}_{pp}\varepsilon\} \tag{34}$$

$$-\varepsilon'\Omega^{-3}\mathbf{H}_p^2\varepsilon \tag{35}$$

with cross-partials

$$\frac{\partial^2 l}{\partial \delta \partial \beta'} = -\mathbf{1'B'}\Omega^{-1}\mathbf{BX} \tag{36}$$

$$\frac{\partial^2 l}{\partial \delta \partial \gamma'} = -\mathbf{1'B'}\Omega^{-1}\mathbf{BQX} \tag{37}$$

$$\frac{\partial^2 l}{\partial \delta \partial \lambda} = -\mathbf{1'B'}\Omega^{-1}\mathbf{BWX} \tag{38}$$

$$\frac{\partial^2 l}{\partial\delta\partial\rho} = -\mathbf{1}'(M'\Omega^{-1}\mathbf{B} + \mathbf{B}'\Omega^{-1}\mathbf{M})\mathbf{U} \tag{39}$$

$$\frac{\partial^2 l}{\partial\delta\partial\alpha_p} = -\mathbf{1}'\mathbf{B}'\Omega^{-2}\mathbf{H}_p\varepsilon \tag{40}$$

$$\frac{\partial^2 l}{\partial\delta\partial\gamma'} = -\mathbf{X}'\mathbf{B}'\Omega^{-1}\mathbf{BQX} \tag{41}$$

$$\frac{\partial^2 l}{\partial\beta\partial\lambda} = -\mathbf{X}'\mathbf{B}'\Omega^{-1}\mathbf{BWY} \tag{42}$$

$$\frac{\partial^2 l}{\partial\beta\partial\rho} = -\mathbf{X}'(\mathbf{M}'\Omega^{-1}\mathbf{B} + \mathbf{B}'\Omega^{-1}\mathbf{M})\mathbf{U} \tag{43}$$

$$\frac{\partial^2 l}{\partial\beta\partial\alpha_p} = -\mathbf{X}'\mathbf{B}'\Omega^{-2}\mathbf{H}_p\varepsilon \tag{44}$$

$$\frac{\partial^2 l}{\partial\gamma\partial\lambda} = -\mathbf{X}'\mathbf{Q}'\mathbf{B}'\Omega^{-1}\mathbf{BWY} \tag{45}$$

$$\frac{\partial^2 l}{\partial\gamma\partial\rho} = -\mathbf{X}'\mathbf{Q}'(\mathbf{B}'\Omega^{-1}\mathbf{M} + \mathbf{M}'\Omega^{-1}\mathbf{B})\mathbf{U} \tag{46}$$

$$\frac{\partial^2 l}{\partial\gamma\partial\alpha_p} = -\mathbf{X}'\mathbf{Q}'\mathbf{B}'\Omega^{-2}\mathbf{H}_p\varepsilon \tag{47}$$

$$\frac{\partial^2 l}{\partial\gamma\partial\rho} = -\mathbf{X}'\mathbf{W}'(\mathbf{B}'\Omega^{-1}\mathbf{M} + \mathbf{M}'\Omega^{-1}\mathbf{B})\mathbf{U} \tag{48}$$

$$\frac{\partial^2 l}{\partial\lambda\partial\alpha_p} = -\varepsilon'\Omega^{-2}\mathbf{H}_p\mathbf{BWY} \tag{49}$$

$$\frac{\partial^2 l}{\partial\rho\partial\alpha_p} = -\varepsilon'\Omega^{-2}\mathbf{H}_p\mathbf{MU} \tag{50}$$

$$\frac{\partial^2 l}{\partial\alpha_\rho\partial\alpha_p} = -\frac{1}{2}\{Tr[\Omega^{-1}\mathbf{H}_{pq} - \Omega^{-2}\mathbf{H}_p\mathbf{H}_q] - \varepsilon'\Omega^{-2}\mathbf{H}_{pq}\varepsilon\} \tag{51}$$

$$-\varepsilon'\Omega^{-3}\mathbf{H}_p\mathbf{H}_q\varepsilon \tag{52}$$

The corresponding elements of the information matrix are thus:

$$-E\left\{\frac{\partial^2 l}{\partial\lambda^2}\right\} = -\mathbf{1}'\mathbf{B}'\Omega^{-1}\mathbf{B1} \tag{53}$$

$$-E\left\{\frac{\partial^2 l}{\partial\beta\partial\beta'}\right\} = \mathbf{X'B'\Omega^{-1}BX} \quad (54)$$

$$-E\left\{\frac{\partial^2 l}{\partial\gamma\partial\gamma'}\right\} = -\mathbf{X'Q'B'\Omega^{-1}BQX} \quad (55)$$

$$-E\left\{\frac{\partial^2 l}{\partial\lambda^2}\right\} = Tr\{[\mathbf{A}^{-1}\mathbf{W}]^2 + \Omega(\mathbf{A'B'})^{-1}\mathbf{W'B'\Omega^{-1}BW(BA)}^{-1} \quad (56)$$

$$+ (1\delta + \mathbf{X}\beta + \mathbf{QX}\gamma)'(\mathbf{A}')^{-1}\mathbf{W'B'\Omega^{-1}BWA}^{-1} \quad (57)$$

$$(1\delta + \mathbf{X}\beta + \mathbf{QX}\gamma)\} \quad (58)$$

$$-E\left\{\frac{\partial^2 l}{\partial\rho^2}\right\} = Tr\{[\mathbf{B}^{-1}\mathbf{M}]^2 + \Omega(\mathbf{B}')^{-1}\mathbf{M'\Omega^{-1}MB}^{-1}\} \quad (59)$$

$$-E\left\{\frac{\partial^2 l}{\partial\alpha_p^2}\right\} = \frac{1}{2}Tr\{\Omega^{-2}\mathbf{H}_p^2\} \quad (60)$$

$$-E\left\{\frac{\partial^2 l}{\partial\delta\partial\beta'}\right\} = -\mathbf{1'B'\Omega^{-1}BX} \quad (61)$$

$$-E\left\{\frac{\partial^2 l}{\partial\delta\partial\gamma'}\right\} = -\mathbf{1'B'\Omega^{-1}BQX} \quad (62)$$

$$-E\left\{\frac{\partial^2 l}{\partial\delta\partial\lambda}\right\} = -\mathbf{1'B'\Omega^{-1}BWA}^{-1}(\mathbf{1}\delta + \mathbf{X}\beta + \mathbf{QX}\gamma) \quad (63)$$

$$-E\left\{\frac{\partial^2 l}{\partial\delta\partial\rho}\right\} = 0 \quad (64)$$

$$-E\left\{\frac{\partial^2 l}{\partial\delta\partial\alpha_\rho}\right\} = 0 \quad (65)$$

$$-E\left\{\frac{\partial^2 l}{\partial\beta\partial\gamma'}\right\} = -\mathbf{X'B'\Omega^{-1}BQX} \quad (66)$$

$$-E\left\{\frac{\partial^2 l}{\partial\beta\partial\lambda}\right\} = \mathbf{X'B'\Omega^{-1}BWA}^{-1}(\mathbf{1}\delta + \mathbf{X}\beta + \mathbf{QX}\gamma) \quad (67)$$

$$-E\left\{\frac{\partial^2 l}{\partial\beta\partial\rho}\right\} = 0 \quad (68)$$

$$-E\left\{\frac{\partial^2 l}{\partial\beta\partial\alpha_p}\right\} = 0 \quad (69)$$

$$-E\left\{\frac{\partial^2 l}{\partial\lambda\partial\gamma'}\right\} = -(\mathbf{1}\delta + \mathbf{X}\beta + \mathbf{QX}\gamma)'(\mathbf{A}')^{-1}\mathbf{W}'\mathbf{B}'\Omega^{-1}\mathbf{B}\,\mathbf{QX} \tag{70}$$

$$-E\left\{\frac{\partial^2 l}{\partial\rho\partial\gamma'}\right\} = -0' \tag{71}$$

$$-E\left\{\frac{\partial^2 l}{\partial\lambda\partial\alpha_\rho}\right\} = 0 \tag{72}$$

$$-E\left\{\frac{\partial^2 l}{\partial\rho\partial\lambda}\right\} = Tr\{\Omega(\mathbf{B}')^{-1}(\mathbf{B}'\Omega^{-1}\mathbf{M} + \mathbf{M}'\Omega^{-1}\mathbf{B})\mathbf{WA}^{-1}\mathbf{B}^{-1}\} \tag{73}$$

$$-E\left\{\frac{\partial^2 l}{\partial\lambda\partial\alpha_p}\right\} = Tr\{\Omega^{-1}\mathbf{H}_p\mathbf{BWA}^{-1}\mathbf{B}^{-1}\} \tag{74}$$

$$-E\left\{\frac{\partial^2 l}{\partial\rho\partial\alpha_p}\right\} = Tr\{\Omega^{-1}\mathbf{H}_p\mathbf{MB}^{-1}\} \tag{75}$$

$$-E\left\{\frac{\partial^2 l}{\partial\alpha_p\partial\alpha_q}\right\} = \frac{1}{2}Tr\{\Omega^{-2}\mathbf{H}_p\mathbf{H}_q\} \tag{76}$$

The information matrix is of the form

$$I(\theta) = \begin{pmatrix} Dim & 1 & k & k & 1 & 1 & m \\ 1 & I_{\delta\delta} & \mathbf{I}_{\delta\beta'} & \mathbf{I}_{\delta\gamma'} & I_{\delta\lambda} & 0 & \mathbf{0}' \\ k & \mathbf{I}_{\beta\delta} & \mathbf{I}_{\beta\beta'} & \mathbf{I}_{\beta\gamma'} & \mathbf{I}_{\beta\lambda} & \mathbf{0} & \mathbf{0}' \\ k & \mathbf{I}_{\gamma\delta} & \mathbf{I}_{\gamma\beta'} & \mathbf{I}_{\gamma\gamma'} & \mathbf{I}_{\gamma\lambda} & \mathbf{0} & \mathbf{0}' \\ 1 & I_{\lambda\delta} & \mathbf{I}_{\lambda\beta'} & \mathbf{I}_{\lambda\gamma'} & I_{\lambda\lambda} & I_{\lambda\rho} & \mathbf{I}_{\lambda\alpha'} \\ 1 & 0 & \mathbf{0}' & \mathbf{0}' & I_{\rho\lambda} & I_{\rho\rho} & \mathbf{I}_{\rho\alpha'} \\ m & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{\alpha\lambda} & \mathbf{I}_{\alpha\rho} & \mathbf{I}_{\alpha\alpha'} \end{pmatrix} \tag{77}$$

$$= \begin{pmatrix} & 2k+1 & m+2 \\ 2k+1 & \mathbf{I}_{11} & \mathbf{I}_{12} \\ m+2 & \mathbf{I}_{21} & \mathbf{I}_{22} \end{pmatrix}, \text{ say.} \tag{78}$$

in which the dimensions of the blocks appear in the margins. As can be seen, this matrix is *not* block-diagonal between the mean and variance-covariance parameters of the model; this is because the spatial lag parameter, λ, enters both mean and covariance structure in this formulation.

References

Anselin, L. (1988a): «Lagrange Multiplier Test Diagnostics for Spatial Dependence and Spatial Heterogeneity», *Geographical Analysis,* vol. 20, n.º 1, pp. 1-17.

— (1988b): *Spatial Econometrics: Methods and Models,* Chapter 6. Kluwer Academic.

Behrens, K.; Ertur, C., and Koch, W. (2010): Dual Gravity: «Using Spatial Econometrics to Control for Multilateral Resistance. Forthcoming» *Journal of Applied Econometrics*.

Bivand, R. S. (1984): «Regression modeling with spatial dependence: an application of some class selection and estimation methods», *Geographical Analysis,* vol. 16, n.º 1, pp. 25-37.

Blommestein, H. (1983): «Specification and Estimation of Spatial Econometric Models: Discussion of Alternative Strategies for Spatial Economic Modelling», *Regional Science and Urban Economics,* vol. 13, pp. 251-270.

Bodson. P., and Peeters, D. (1975): «Estimation of the coefficients of a linear regression in the presence of spatial autocorrelation. An application to a Belgian labour demand function», *Environment and Planning A*, vol. 7, pp. 455-472.

Brandsma, A. S., and Ketellapper, R. H. (1979): «A biparametric approach to spatial autocorrelation», *Environment and Planning A*, vol. 11, pp. 51-58.

Breusch, T. S. (1978): «Testing for Autocorrelation in Dynamic Linear Models», *Australian Economic Papers,* vol. 17, pp. 334-355.

Breusch, T. S., and Pagan, A. R. (1979): «A simple test for heteroskedasticity and random coefficient variation», *Econometrica,* vol. 47, n.º 5, pp. 1287-1294.

Burridge, P. (1981): «Testing for a common factor in a spatial autoregression model», *Environment and Planning A,* vol. 13, pp. 795-800.

Burridge, P., and Hristova, D. (2008): «Consistent estimation and order selection for nonstationary autoregressive processes with stable innovations», *Journal of Time Series Analysis,* vol. 29, n.º 4, pp. 695-718.

Elhorst, J. P. (2001): «Dynamic Models in Space and Time», *Geographical Analysis,* vol. 33, n.º 2, pp. 119-140.

— (2010): «Applied Spatial Econometrics: Raising the Bar», *Spatial Economic Analysis,* vol. 5, n.º 1, pp. 9-28.

Ertur, C., and Koch, W. (2007): «Growth, Technological Interdependence and Spatial Externalities: Theory and Evidence», *Journal of Applied Econometrics,* vol. 22, pp. 1033-1062.

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

Godfrey, L. G. (1978): «Testing Against General Autoregressive and Moving Average Error Models when the Regressors Include Lagged Dependent Variables», *Econometrica*, vol. 46, pp. 1293-1302.

Haining, R. (1986): «Spatial models and regional science: a comment on Anselin's paper and research directions», *Journal of Regional Science,* 26, 4, 793-798.

Hendry, D. F. (2011): «Empirical Economic Model Discovery and Theory Evaluation», University of Oxford, Department of Economics Discussion Paper n.º 529, february 2011.

Hendry, D. F., and Mizon, G. E. (1978): «Serial Correlation as a Convenient Simplification, not a Nuisance: Comment on a Study of the Demand for Money by the Bank of England», *Economic Journal,* vol. 88, pp. 549-563.

Hordijk, L. (1979): «Problems in Estimating Econometric Relations in Space», *RSA Papers,* vol. 42, pp. 99-115.

Kelejian, H. H. (2008): «A spatial J-test for model specification against a single or a set of nonnested alternatives», *Letters in Spatial and Resource Sciences,* vol. 1, pp. 3-11.

Kelejian, H. H., and Piras, G. (2011): «An extension of Kelejian's J-test for non-nested spatial models», *Regional Science and Urban Economics,* vol. 41, pp. 281-292.

Kelejian, H. H., and Prucha, I. R. (2010): «Specification and Estimation of Spatial Autoregressive models with autoregressive and heteroskedastic disturbances», *Journal of Econometrics,* vol. 157, pp. 53-67.

Kelejian, H. H., and Robinson, D. P. (1998): «A suggested test for spatial autocorrelation and/or heteroskedasticity and corresponding Monte Carlo results», *Regional Science and Urban Economics,* vol. 28, pp. 389-417.

Lacombe, D. (2004): «Does Econometric Methodology Matter? An Analysis of Public Policy Using Spatial Econometric Techniques», *Geographical Analysis,* vol. 36, pp. 87-89.

Lee, L. F. (2004): «Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models», *Econometrica,* vol. 72, n.º 6, pp. 1899-1925, nov. 2004.

Lee, L. F.; Liu, X., and Lin, X. (2010): «Specification and Estimation of Social Interaction Models with Network Structures», *Econometrics Journal,* vol. 13, pp. 145-176.

LeSage, J. P., and Pace, R. K. (2009): *Introduction to Spatial Econometrics*, Boca Raton, Taylor and Francis.

Lin, X., and Lee. L. F. (2010): «GMM estimation of spatial autoregressive models with unknown heteroskedasticity», *Journal of Econometrics,* vol. 157, pp. 34-52.

Manski, C. F. (1993): «Identification of endogenous social effects: the reflection problem», *Review of Economic Studies,* vol. 60, pp. 531-542.

Mur, J., and Angulo, A. (2006): «The Spatial Durbin Model and the Common Factor Tests», *Spatial Economic Analysis,* vol. 1, n.º 2, pp. 207-226.

— (2009): «Model selection strategies in a spatial setting: Some additional results», *Regional Science and Urban Economics,* vol. 39, pp. 200-213.

Ord, K. (1975): «Estimation Methods for Models of Spatial Interaction», *Journal of the American Statistical Association,* vol. 70, n.º 349, pp. 120-126.

Pijnenburg, K., and Kholodilin, K. (2011): «Do regions with entrepreneurial neighbours perform better?», presented at the *Spatial Econometric Association World Conference,* Toulouse, july 2011.

Sargan, J. D. (1980): «Some tests of dynamic specification for a single equation», *Econometrica,* vol. 48, pp. 879-897.

Yokoi, T.: «Efficient Maximum Likelihood Estimation of Spatial Autoregressive Models with Normal but Heteroskedastic Disturbances», presented at *9th International Workshop on Spatial Econometrics and Statistics,* june 24/25, Orléans.