doi:10.1111/j.1435-5957.2008.00208.x

Introduction

New spatial econometric techniques and applications in regional science

Giuseppe Arbia[1], **Bernard Fingleton**[2]

[1] DASTA, Faculty of Economics, University-G. d'annunzio- of Pescara, Viale della Pineta, 4, 65127 Pescara, Italy (e-mail: arbia@unich.it)
[2] Department of Economics, Strathclyde University, 130 Rottenrow, Glasgow G4 0GE, United Kingdom (e-mail: bf100@cam.ac.uk)

The papers appearing in this special issue of *Papers in Regional Science*, which is devoted to spatial econometrics, come from the First International Conference of the Spatial Econometrics Association held in Cambridge (UK) 12–14 July 2008. This conference was the first official meeting of the new association, which was established in May 2006 in Rome and which has already attracted more than 150 members from around the world. At the Cambridge conference there were close to 120 delegates and more than 100 papers were presented. With regard to the eight papers appearing in this special issue, we would particularly like to thank the authors and the referees for their contribution to what we believe is an interesting and lively selection.

Recent years have seen a real explosion in the application of spatial statistical models in all branches of social sciences and in particular in economics. Spatial econometrics models have been used to analyse different topics (see for example Anselin et al. 2004 for a review) and as a matter of fact spatial regression techniques are now becoming an established component in the applied econometrics toolbox, as witnessed by the increasing attention given to this topic in standard econometrics textbooks (Maddala 2001; Woolridge 2002; Gujarati 2003; Kennedy 2003; Baltagi 2008).

Of the eight papers published in this issue, three have a major methodological emphasis and the remaining five have a predominantly applied perspective. The methodological papers are concerned with cross-sectional spatial model estimation techniques (Fingleton and Le Gallo), space-time model comparisons (Griffith) and spatial concentration (Bavaud). The other papers touch very diverse applied topics and provide a good insight into the wide variety of subjects that can benefit from a spatial econometrics approach. They range from regional convergence of the per-capita income (Olejnik), land price modelling (Tsutsumi and Seya), and spatial income and educational inequalities (Tselios), to modelling political trends (Santolini) and industrial specialization (Ciriaci and Palma). In the remainder of this introduction we will provide a short summary of these various contributions.

At the heart of spatial econometrics modelling, it is certainly the case that the various spatial regression models originating from the theory of random fields hold a prominent position (see Arbia 2006, for an account of these statistical foundations). By this we mean the spatial error model and the spatial lag model, which exploit autoregression in the error or in the dependent variable respectively; these two models are by far the most widely used in the spatial econometrics literature. One important issue regarding the estimation of these models is the fact that maximum likelihood estimators of the models' parameters are not easy to obtain. This is due to difficulties associated with maximizing the loglikelihood and with computing the determinant of the variance-covariance matrix. An alternative that is becoming rather common in the recent literature is the use of instrumental variables and generalized method of moments estimators, which have been established and popularized in the spatial literature by Kelejian and Prucha (1998). In a previous recent paper Fingleton and Le Gallo (2008) examined the case of regression models including endogenous variables and a spatial moving average error process, extending Kelejian and Prucha's feasible generalized spatial two-stage least squares (FGS2SLS) to this modelling framework. In the paper included in this special issue, entitled "Estimating spatial models with endogenous variables, a spatial lag and spatially dependent disturbances: finite sample properties", the two authors again revisit this problem, developing their previous work in various directions. First they consider the effect of introducing a spatial lag. Second they analyse the properties of the SHAC estimator (spatial heteroscedasticity and autocorrelation consistent estimator) suggested by Kelejian and Prucha (2007), in the context of endogeneity. Finally they tackle the important issue of the scarcity of instruments and consider the properties of suboptimal instruments derived from the 3-group method for measurement errors (Kennedy 2003).

Daniel Griffith's paper, "A comparison of four model specifications for describing small heterogeneous space-time datasets: sugar cane production in Puerto Rico, 1958/59–1973/74", considers modelling approaches given a dependent variable comprising space-time percentages, namely the percentage of the land surface area of each of 73 Puerto Rican municipios[1] that is given over to sugar cane in 16 annual production periods. This presents some challenges, since rather than a normally distributed dependent variable, the modelling context is the less familiar generalized linear modelling approach (which encompasses normal responses) as is appropriate for a binomial random variable with extra-binomial variation. Issues that are considered include how one might model temporal autocorrelation, control for spatial autocorrelation, select appropriate covariates, account for binomial overdispersion and how one might minimize bias due to omitted variables. One of the interesting aspects of the paper is Griffith's introduction of spatial filter eigenvectors as a way of freeing a variable of spatial dependence. Among the modelling strategies considered, random effects are introduced to account for municipal heterogeneity, and there is evidently a need for the random effects to possess spatial structure. Griffith concludes by calling for further research on the issue of residual spatial autocorrelation in generalized linear models.

The paper by François Bavaud entitled "Local concentrations" tackles in an original way an important topic in spatial economic analysis, namely spatial concentration, which has been of interest ever since the seminal papers by Gini (1912) and Lorentz (1905). Most of the literature on objective concentration measurement refers to global indices of spatial concentration (such as mean absolute deviation, Gini, Theil, variance of logarithms and many others) that satisfy axioms such as, for example, anonimity, Pigou-Dalton transfer principle, sensitivity and decomposibility (Sen 1973). Spatial concentration indices typically measure the discrepancy between two distributions, the actual distribution and a benchmarking distribution. Common examples include the concentration of industries, typically measured by comparing one sector's distribution with that of industry as a whole (Krugman 1991; Ellison and Glaeser 1997), and the concentration of incomes, in which case indices compare the regional income and regional population shares.

[1] Municipios are the primary divisions in Puerto Rico. For statistical purposes, the U.S. Census Bureau treats municipios as the statistical equivalents of U.S. counties.

A global index has the advantage of reducing to a single value the phenomenon under study. Its synthetic value is fundamental, but it does not readily allow one to distinguish situations that are spatially very heterogeneous. For this reason we find that, alongside global measures, local indices play a prominent role in spatial econometrics, such as for example the local indicators of spatial association (or LISA) introduced by Anselin (1995).

The paper by Bavaud has several novel elements. First of all it provides an interesting generalization by introducing a family of concentration measures that covers most of the usual concentration indices. Second, it introduces the idea to local concentration measures, as measured using some indicator of the connectivity of the regions, in contrast to the usual global indices, where typically each region contributes in proportion to its relative population regardless of the distribution of its population in geographic space.

The local concentration analysis follows the lines of local variance analysis, which plays an important role in spatial autocorrelation theory. The paper examines the properties of various local concentration indices and in particular, those that are analogous to spatial autocorrelation indices. The new indices are applied to the distribution of wealth across Swiss cantons using a connectivity structure based on inter-regional migration.

Regional growth and convergence of per-capita incomes have been among the most widely studied topics in spatial econometrics for many years and a huge number of papers have been written using spatial methodologies both to analyse the purely cross-sectional case (often spawned by the celebrated Barro and Sala-i-Martin 1992, paper) and the dynamic case based on panel data (see Islam 1995). Many of these papers applied such models to explain growth and convergence dynamics at the EU level (Fingleton 2003), a topic that achieved wider scope with the fifth enlargement to 25 countries that occurred in 2004 and with the political debate that anticipated and followed it. The paper by Alicja Olejnik entitled "Using the spatial autoregressively distributed lag model in assessing the regional convergence of per-capita income in the EU – 25" adds new insight by considering the modelling framework related to the spatial autoregressive distributed lag (SADL) models in order to explain the enlarged EU-25 regions' convergence dynamics. These models are derived from the autoregressive distributed lag (ADL) models familiar to the time series analysts and have been introduced into the spatial literature by Lauridsen (2006), as an alternative to the more classical spatial econometrics modelling approach.

In particular Olejnik's paper originates from the idea of spatial unit roots and spatial cointegration, both relating to the so-called spatial error correction model (Fingleton 1999). It starts from the research hypothesis of a conditional convergence process of per-capita income in Europe and uses the dynamic spatial setting provided by the SADL model as a basis for empirical testing. This shows that, in line with many other studies, human capital is a potentially relevant growth factor. Moreover the paper stresses the importance of the concept of spatial dynamics as a way of emphasizing the forces that are present in the economic system at one single moment in time.

The link between accessibility and house price is rather intuitive and has a long tradition in urban and regional analysis. As a consequence of this link, the study of the impact of new large scale transportation routes (typically railways, motorways or involving air transportation) has an important place in regional planning and regional political economy. Of particular relevance is the case of new transportation routes connecting large metropolitan areas with their surroundings.

The paper by Morito Tsutsumi and Hajime Seya, entitled "Measuring the impact of large-scale transportation projects on land price using spatial statistical models", tackles this problem with explicit reference to models that explain land price variation. The paper presents two alternative approaches that are available in the literature to deal with land price modelling. The first is based on spatial econometrics regression and the second comes from the geostatistics literature.

In the paper, the authors compare four different hedonic price models stemming from these two contrasting approaches: the basic regression model with no consideration of spatial dependence, the spatial error model (together with its Bayesian counterpart) and the direct representation of the spatial field generating the data as obtained via the geostatistical approach. The various methods are employed to estimate hedonic regression parameters with reference to an interesting dataset collected by the Japanese Ministry of Land, Infrastructure and Transport, and describing the effects of the TX railway introduced in 2005 to improve the accessibility to Central Tokyo.

The paper by Vassilis Tselios, entitled “Income and educational inequalities in the regions of the European Union: geographical spillovers under welfare state restrictions”, analyses data for 94 EU regions over six years using a variety of panel estimation approaches. Following an extensive theoretical preamble, which forms the basis of the model specifications, we are guided to various econometric estimators.

In the simplest approach adopted, with fixed region and time effects, Tselios explains income inequalities as a function of educational inequalities, the level of income per capita, and the level of educational attainment. Similarly, educational inequality is regressed on income inequality plus income per capita and educational attainment. However the fixed effects approach assumes exogeneity for these variables, ignores spatial interaction effects and disallows time-constant covariates and consequently an important question the author wishes to address, which is the impact of welfare regime. Welfare regime is spatially differentiated but constant over the period analysed, with regime dummies identifying conservative (Luxembourg, Belgium, France, Germany, Austria), residual (Portugal, Spain, Italy, Greece) and social democratic (Denmark) welfare states, with the liberal (UK, Ireland) omitted to avoid perfect collinearity.

Estimation using random effects to capture inter-regional heterogeneity as an error component has the important advantage of permitting time-constant covariates, although there is a danger of a loss of consistency due to any correlation induced between the error and the regressors. This is tested by the Hausman test, on the assumption that the fixed effect coefficient estimates are themselves consistent. A dynamic specification is introduced with lagged income inequality and lagged educational inequality as additional regressors in the respective income and educational inequality models. Although this inevitably leads to a loss of degrees of freedom, this allows endogenous regressors although it does introduce the question of what is an appropriate set of instruments. Tselios uses two different approaches to instrumentation.

However none of the models mentioned thus far allow for spatial dependence. The final models introduce spatial effects, either as a spatial error process, or as an endogenous spatial lag, although the maximum likelihood approach implies that the (other) regressors are (once again) exogenous. The approach adopted also entails both region and time fixed effects, and therefore once again precludes the introduction of time-constant covariates, particularly the welfare state dummies that could be introduced under random error components. The estimation approach does however introduce welfare regime differentiated spatial effects as part of the W matrix that defines the spatial interaction between regions. The final maximum likelihood models introduce some extra covariates, namely industry and service shares of total value added.

The author concludes that on the whole there exists a positive relationship between income and educational inequality which is robust to model specification, and that welfare regimes are relevant to our understanding of these inequalities. For instance the maximum likelihood estimates show the spillover of income and educational inequalities according to both distance and welfare regime, reflecting the effects of both geography and institutions on the distribution of inequality. He states that “this shows that both spatial autocorrelation (geography) and spatial heterogeneity (institutions) matter in the estimation of inequality models and thus in accounting for the economic performance of the European regions” and that “the results show that spillovers are more easily captured within welfare state boundaries because social, cultural, and institutional linkages across actors and regions are more homogeneous within a specific regime than across regimes". The paper contains many other theoretical and policy related observations that are too numerous to include in this brief description.

The paper by Raffaella Santolini entitled "A spatial cross-sectional analysis of political trends in Italian municipalities" uses the term 'political trend' to describe the mimicking behaviour by politicians due to common party ideology, as opposed to other causes involving strategic interaction. Of the several alternative hypotheses, the paper focuses on yardstick competition, rather than fiscal competition or budget spillover effects. In the spatial econometric context, fiscal interaction emerges as a significant endogenous spatial lag of the dependent variable, which is a local tax rate in 246 municipalities of the Marche region. The W matrix used in this case is a standardized contiguity matrix, based on the hypothesis that mimicking only extends directly to neighbouring municipalities' boundaries, but the paper also considers W matrices that are conditioned by both spatial and political 'contiguity'. The finding that there is indeed interaction across space emerges after controlling for endogeneity and for several covariates, including demographic, social and location variables, the local political structure, the year of the election, electoral 'distance' and various interactions. The results show that the tax rate mainly depends on neighbouring tax rates, the proportion of elderly people, ideology affiliation, and on the coastal location of municipalities. The methodology is able to bring some light to bear on difficult and complex issues such as whether local authorities with large majorities are less likely to mimic the fiscal policies in neighbouring authorities, and whether copy-cat behaviour is more prevalent in an election year. The analysis of tax rate interactions across space is supplemented by fitting spatial lag models in which the dependent variable is current public expenditure.

The paper by Daria Ciriaci and Daniela Palma on "The role of knowledge-based supply specialisation for competitiveness: a spatial econometric approach" is grounded in Kaldorian dynamic increasing returns, which was a precursor, alongside the work of Myrdal, to the contemporary theory of new economic geography, since it involves circular and cumulative causation. This approach is non-orthodox since it does not make explicit micro-economic assumptions about individual utility or firm profit maximisation, and there is no explicit market structure deriving from Dixit-Stiglitz theory, as assumed in much contemporary analysis. However Kaldorian dynamics, initially formalized as a system of equations by Dixon and Thirwall in 1975, remains a popular approach by which one can introduce increasing returns to scale into regional economic analysis. The basic model assumes that growth of output depends on export growth and export growth is partly dependent on productivity growth which is stimulated in turn by the growth of output.

The paper focuses on one of the limitations of the initial specification, which is the rather simple way in which export growth is determined. Productivity growth together with the growth of nominal wages and the rate of change of the mark-up on labour costs determines the growth of domestic export prices, and domestic export prices plus the growth of real income in export markets and the growth of competitor prices determine the rate of growth of exports. In this set-up there is no scope for non-price competitiveness, which is the focus for the analysis by Ciriaci and Palma.

They introduce specialization in high-tech goods as an indicator of knowledge-based competitiveness, and their empirical analysis shows that this has a spatially differentiated effect that changes over time. They show that knowledge-based competitiveness declines in the centre and North of Italy, whereas export growth in the South depends essentially on World demand rather than non-price competitiveness. This conclusion is reached by studying 103 regions at four time snapshots, using geographically weighted regression at each time point to detect parameter heterogeneity across space and to highlight parameter dynamics.

The papers we have described provide a fair representation of much of the current work in spatial econometrics and highlight some of the outstanding issues and problems that the applied spatial analyst encounters. One serious problem is the issue of endogenous variables. Spatial econometricians are familiar with the concept of the endogenous spatial lag in the cross-sectional spatial regression model and well established methods are available to ensure consistent estimation, notably maximum likelihood, two stage least squares or bootstrap methods. Maximum likelihood has the advantage of providing, usually via an iterative bisection search routine, a spatial parameter estimate that always falls within the 'stable' parameter space defined by the reciprocal of the maximum and minimum eigenvalues of the weights matrix, the reason being a penalty function going to infinity as the parameter approaches the bounds of that space. This is not the case with unconstrained two stage least squares or its variants, and this could lead us into the territory of spatial units roots alluded to above. However, one disadvantage of the maximum likelihood approach is that for the single equation format it assumes exogeneity for the other right hand side variables, and yet in spatial systems what we invariably encounter is endogeneity due to reverse causation or simultaneity, omitted variables or measurement error. In order to nullify these effects, the first best option would be to eliminate the causes of endogeneity, for example by modelling the system as a system of simultaneous equations. However this is often not possible and the mis-specification of the system may itself induce problems that are as serious as the endogeneity we are attempting to eliminate. Likewise we should endeavour to avoid measurement error, and there should be no omitted variables. These requirements are difficult if not impossible to satisfy in most if not all cases. The second best option is to accept that we have to work with a single equation with multiple endogenous variables, and strive to eliminate the effects of this via two stage least squares. This leads then to the problem of the selection instrumental variables, which are supposed to correlate with the endogenous variables and be uncorrelated with the error term. Practical experience leads us to believe that this is not a trivial problem.

A perhaps more fundamental question for spatial econometrics is what is the meaning of the spatial lag? Some economists reject it as a substantive and meaningful concept, and would like to treat it more as a nuisance variable, which together with exogenous lags or a spatial error process helps to sweep away the problem of spatially dependent data and allow analysis to focus on more meaningful and substantive concepts. One problem for spatial econometrics therefore is to justify the spatial lag in particular, perhaps in a more substantive way than simply a demonstration of the econometric necessity of spatial effects to avoid biases and inconsistency in estimation and inference. In the papers above there is ample demonstration of the ongoing attempts to provide a more substantive rationale for the spatial lag.

This problem of what the spatial lag actually represents is bound up with the problem of definition of the spatial weights matrix, which is assumed to be a nonstochastic matrix capturing our hypothesis about the nature of the spatial interactions we are modelling. The problem is that, unlike the simple notion of a time series lag, the spatial lag is a very fluid and complex entity open to multiple definitions within a single study. Critics of spatial econometrics almost always in our experience home in on the arbitrary nature of the weights matrix, asking "how is it defined and why is it precisely like that when it could easily have been like this, what does it mean, and are not the results obtained conditional on somewhat arbitrary decisions taken about its structure?". Some future research on the robustness of outcomes to variations in assumptions about the weight matrix structure would be helpful in allaying such criticisms, although ideally carefully structured arguments coming from theory and leading precisely to the typical reduced form spatial econometric model, with a spatial lag and exogenous lags also, are the preferred option.

So far we have mentioned problems emerging as paramount at the very heart of spatial econometrics. Most of the time these problems emerge in the current literature when dealing with cross sectional synchronic spatial series of data. It should be mentioned at the end of this editorial note the growing interest and the exploding literature relating to the spatial econometric analysis of diachronic spatial series in the form of panel data or space-time series. The reason behind this increase is the proliferation of data sets that are both spatially and temporally indexed and the impelling need to understand and interpret them. From a methodological point of view the introduction of time into spatial econometric modelling brings a substantial increase in the scope of the analysis and allows us to make separate interpretation regarding the spatial correlations, temporal correlations and how space interacts with time. In such a framework it is possible to tackle issues like diffusion phenomena, local evolution, dynamic spill-overs, firm demography, and dynamic allocation, to name just a few. The methods to analyse such data are well consolidated in the spatial statistical literature (see e.g. Borovkova et al. 2008; Banerjee et al. 2004) and in the econometric literature on panel data (Baltagi 2008), but a full integration with the existing spatial econometric literature is still well behind and represents a fruitful field of research for future years. A very good example of this on-going integration may be found in the paper by Elhorst and Zeilstra (2007) that was recently awarded the Martin Beckmann Prize for the best paper published in this journal in 2007.

References

Anselin L (1995) Local indicators of spatial association – LISA. *Geographical Analysis* 27: 93–115

Anselin L, Florax RJGM, Rey S (2004) *Advances in spatial econometrics methodology tools and applications*. Springer Verlag, New York

Arbia G (2006) *Spatial econometrics*. Springer, Berlin

Baltagi BH (2008) *Econometric analysis of panel data*. John Wiley and Sons, Chichester

Banerjee S, Carlin BD, Gelfand AE (2004) *Hierarchical modeling and analysis for spatial data*. Chapman & Hall, New York

Barro RJ, Sala-i-Martin X (1992) Convergence. *Journal of Political Economy*, 100: 223–251

Borovkova SA, Lopuhaä HP, Nurani B (2008) Consistency and asymptotic normality of least squares estimators in generalized space-time models. *Statistica Neerlandica*, forthcoming

Dixon RJ, Thirwall AP (1975) *Regional growth and unemployment in the United Kingdom*. Macmillan, London

Elhorst JP, Zeilstra AS (2007) Labour force participation rates at the regional and national levels of the European Union: An integrated analysis. *Papers in Regional Science* 86: 525–549

Ellison G, Glaeser EL (1997) Geographic concentration in U.S. manufacturing industries: A dartboard approach. *Journal of Political Economy* 105: 889–927

Fingleton B (1999) Spurious spatial regression: Some Monte-Carlo results with a spatial unit root and spatial cointegration. *Journal of Regional Science* 39: 1–19

Fingleton B (2003) *European regional growth*. Springer Verlag, Berlin

Fingleton B, Le Gallo J (2008) Finite sample properties of estimators of spatial models with autoregressive, or moving average disturbances and system feedback. *Annals of Economics and Statistics*, forthcoming

Gini C (1912) Variabilità e mutabilità. In: *Studi economico-giuridici della Facoltà di Giurisprudenza*, Università di Cagliari

Gujarati D (2003) *Basic Econometrics*, (fourth edition). McGraw-Hill, New York

Islam N (1995) Growth empirics: A panel data approach. *Quarterly Journal of Economics* 4: 1127–1170

Kelejian HH, Prucha IR (1998) A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *Journal of Real Estate Finance and Economics* 17: 99–121

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

Kennedy P (2003) *A guide to econometrics*, (fifth edition). Blackwell Publishers, Oxford

Krugman P (1991) *Geography and trade*. Leuven University Press/MIT Press, Leuven/Cambridge, MA

Lauridsen J (2006) Spatial autoregressively distributed lag models: Equivalent forms, estimation and an illustrative commuting model. *The Annals of Regional Science* 40: 297–311

Lorentz MO (1905) Methods for measuring concentration of wealth. *Journal of the American Statistical Association* 9: 859–872

Maddala G (2001) *Econometrics*. McGraw-Hill, New York.

Sen A (1973) *On economic inequality*. Clarendon Press, Oxford

Woolridge G (2002) *Econometric analysis of cross section and panel data*. MIT Press, Cambridge, MA