JOURNAL OF REGIONAL SCIENCE

JOURNAL OF REGIONAL SCIENCE, VOL. 52, NO. 2, 2012, pp. 167–171

INTRODUCTION: WHITHER SPATIAL ECONOMETRICS?

Mark D. Partridge

Department of Agricultural, Environment, and Development Economics, The Ohio State University, 2120 Fyffe Road, Columbus, OH 43210. E-mail: partridge.27@osu.edu

Marlon Boarnet

Sol Price School of Public Policy, University of Southern California, Lewis Hall 312, Los Angeles, CA 90089–0626. E-mail: boarnet@usc.edu

Steven Brakman

Faculty of Economics and Business, University of Groningen, P.O. Box 800, NL-9700 AV Groningen, The Netherlands. E-mail: s.brakman@rug.nl

Gianmarco Ottaviano

Department of Economics, The London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom. E-mail: g.i.ottaviano@lse.ac.uk

ABSTRACT. Spatial econometrics has become a mainstay for regional scientists aiming to estimate geographic spillovers in regional outcomes. Yet, many remain skeptical, especially urban economists who prefer natural experimental approaches. Their concerns revolve around identification and a general lack of a theoretical foundation in the estimation of spatial econometric models. This theme issue includes three papers from leading regional scientists to appraise the status of spatial econometrics. The outcome is sweeping proposals from (1) abandoning standard spatial econometrics because it cannot identify causality, (2) using nonparametric approaches, and (3) implementing more nuanced changes revolving around better theoretical and empirical modeling.

1. INTRODUCTION

A primary aim of regional science is understanding and explaining spatial spillovers in which factors or outcomes in one location are related to outcomes in another nearby location. In the early years of regional science, a weakness faced by regional scientists was that standard econometric approaches did not directly account for space or spillovers. This began to change with developments in spatial statistics (primarily) from geography in the 1970s and 1980s. Further refinements helped form what we today call spatial econometrics, though spatial econometrics itself did not gain much favor in regional science until Anselin's (1988) influential book *Spatial Econometrics: Methods and Models*. This book became an impetus that set off a movement of regional scientists aiming to broaden this approach. Some of the early leaders included Arthur Getis, Harry Kelejian, and James LeSage to name a few. In our work surrounding the *Journal of Regional Science's* 50th Anniversary issue, we identify the rise of spatial econometrics along with the advent of the New Economic Geography as the key reasons for the expanding global interest in regional science in the first decade of the 21st century.

The *JRS* has been a proud leader in promoting the development of spatial econometrics. Groundbreaking papers published in the *JRS* include Anselin (1990), Ord and Getis

 DOI: 10.1111/j.1467-9787.2012.00767.x

(2001), LeSage and Pace (2008), and Arraiz et al. (2010). Spatial econometrics is now commonly applied in regional science research. For example, just in the December 2011 issue of the *JRS*, Monchuk et al. (2011) examined rural income growth using a spatial error model, Rickman and Rickman (2011) examined population growth with a spatial lag approach, and Park and von Rabenau (2011) examine agglomeration spillovers with a simultaneous spatial approach.

Spatial econometrics seemed to be the perfect solution to addressing spillovers. While traditional econometrics would ignore (for example) how neighboring home prices affect the sale price of a given house, spatial econometrics directly modeled this spatial mechanism. Likewise, spatial econometrics addresses problems with error terms that are not randomly distributed. For example, labor markets cover wide areas but data are often defined by jurisdictional boundaries. Economists today are much more aware about geographically clustered errors. It is now a very mature field and clearly ripe for an appraisal to assess how it could be improved. One area of concern is that urban economists have been much slower to adopt standard spatial econometric approaches. One common complaint is that standard spatial econometrics do not do a good job of differentiating when outcomes in nearby areas are spatially correlated (perhaps due to common explanatory factors) versus spatial causality when outcomes in area A affect outcomes in area B—i.e., the notion of "spatial dependence" is too often applied in sloppy fashion. Both of these imply very different causal mechanisms. One reason is that the theoretical rationale for a spatial error model and (especially) a spatial lag model is too often underdeveloped and identification of casual mechanisms is questionable. In the *JRS*'s 50th anniversary issue, this was a key theme of Pinkse and Slade (2010), in which they argue that identification in the spatial lag specification is virtually impossible. McMillen (2010) also questions whether the key assumptions of standard spatial econometrics are met and called for other approaches. Indeed, there was a strong preference to use experimental and structural approaches in identifying urban and regional causality (e.g., see Holmes, 2010).

2. PROBLEMS OF IDENTIFICATION IN SPATIAL ECONOMETRIC MODELS

It is useful to discuss some of the identification concerns with standard spatial econometrics, especially with the spatial lag model. The spatial lag model can simply be represented as:

$$Y = \beta \mathbf{X} + \rho \mathbf{W}\mathrm{Y} + e_i, \tag{1}$$

where i is observation, Y is the dependent variable, $\mathbf{X}$ is a vector of explanatory variables, $\mathbf{W}$ is a weight matrix to reflect spatial connectivity among neighbors, and e is an error term (see also LeSage and Pace, 2009). The concern with Equation (1) is that spatial spillovers reflected by **W**Y are quite restrictive and may not generally apply, and in cases where such spillovers may apply, better theory is needed for justification. Tobler's (1970) first law of geography that "everything is related to everything else, but near things are more related than distant things" is obviously true, but such reasoning is more of cliché than something that can form a theoretical or conceptual model to assess real-world spillovers. However, too often little more than statements such as the first law are used to justify spatial econometric specifications such as the spatial lag model.

To illustrate problems with identification, consider the rural Great Plains of the United States in which there are many cases in which there is not even a paved highway connecting contiguous counties, suggesting no meaningful economic spillovers. This would strongly imply that Equation (1) is an inappropriate specification if the goal was to (say) estimate per-capita income growth because it would imply that growth in one rural Great Plains county causes growth in adjacent rural Great Plains counties. Even so, there typically would be high correlation in economic outcomes between two such adjacent counties because they both would be highly dependent on agricultural products and their incomes would move in tandem with world commodity prices. Indeed, Moran I tests and other spatial dependence tests would show that there is high spatial correlation in this region between county income growth. A less than thoughtful econometrician may then conclude that Equation (1) is the proper model. However, this would lead to the wrong interpretation that spillovers from nearby rural counties are a strong determinant of growth when the actual reason is that the **X** variables are highly correlated. The point is that a good conceptual/theoretical understanding would have led the econometrician to appreciate the difference between spatial correlation being caused by a high correlation of the **X** variables versus the case where there is spatial dependence or spatial causation through spillovers. Proper identification would lead to results that could inform policy.

Another common concern regarding identification is when there are spillovers, but not the type envisioned in Equation (1). For example, suppose we are estimating a poverty model using U.S. county data. At the scale of the county, the primary way that nearby counties affect a county's poverty rate is through commuting and labor market linkages. Strong economic growth in County B will reduce poverty in neighboring County A, while it would be a major stretch to argue that through some sort of neighborhood or peer effects cause neighboring county poverty rates to directly affect one another (i.e., the geographical scale is too large). Yet, just as above, a Moran I test or other tests of spatial dependence may imply that Equation (1) is the proper specification, when in fact a specification such as Equation (2) is called for:

$$Y = \beta\mathbf{X} + \delta\mathbf{WX} + e_{\mathrm{i}}, \tag{2}$$

where **WX** reflects that nearby county employment growth, wage structure, industry structure, etc., may influence poverty rates the dependent variable. However, as Corrado and Fingleton (2012) show in this issue, estimation of a spatial lag model in Equation (1) may produce a statistically significant coefficient for the ρ term even though the spatial lag does not belong in the model, leading to exactly the wrong inference for how spillovers affect poverty. Gibbons and Overman (2012) in this issue show that under reasonable circumstances, it is nearly impossible to distinguish the model implied in Equation (1) from the model depicted in Equation (2). Yet, too often econometricians will argue there are omitted variables and estimate Equation (1) without any effort to discover the relevant missing variables and/or even worry that inference is entirely different between the two variables (and policy conclusions would greatly differ).[1]

3. A LOOK AHEAD TO THE THEME ISSUE

Given that standard spatial econometrics is based on a set of restrictive assumptions regarding the particular types of spillovers and there is often a wide gap between theory and empirical practice, we commissioned three peer-reviewed papers to assess how spatial econometrics should be used. In sum, their prescriptions range from sweeping changes to a proposal for more nuanced changes of standard practice. The most sweeping is articulated in the paper by Gibbons and Overman (2012). They contend that identification

[1]There is another related problem when estimating the model in Equation (1) when the model depicted in Equation (2) is correct. Specifically, some researchers will try to estimate indirect effects through the spatial lag term for the variable x_{j}; $\delta\rho\,\mathbf{w}_{\mathrm{i}}y/\delta x_{\mathrm{j}}$ (see LeSage and Pace, 2009). Such an indirect effect implies that there is a causal relationship between $\mathbf{w}y$ and y, in which this is not the case in Equation (2). The upshot is that supporting theory and more careful specification tests are needed before calculating such indirect effects.

is almost always impossible with standard spatial econometric practice. They argue that a close analogy for standard spatial econometrics is Manski's (1993) reflection problem of identifying peer effects from individual effects. They also worry that the **W** weighting matrix itself may be endogenous, which would be difficult to address. Instead they argue that modelers interested in spatial spillovers should incorporate a reduced form specification such as in Equation (2), which better identifies causality in most cases (though it may lack a structural interpretation). Moreover, Gibbons and Overman argue that their preferred starting point is natural experiments that use geographical, institutional, or historic factors to identify causality. While natural experimental approaches raise their own problems of only assessing "cute" experiments or searching for valid instruments, Gibbons and Overman argue that identification is more transparent and less prone to the errors we described above.

The second paper is McMillen (2012). Like Gibbons and Overman (2012), McMillen's starting point is that standard spatial econometrics rests on very restrictive assumptions and that identification is very difficult. Specifically regarding spatial lag and spatial error models, McMillen states "Though they play a useful role in detecting various forms of model misspecification, they are apt to be viewed as the *correct* parametric form for a model when, in fact, they are simply a convenient way to control for unknown sources of spatial clustering among model residuals or the dependent variable." He further notes that the spatial error and lag models are a restrictive form of spatial smoothing that could much more accurately be accomplished with nonparametric approaches including locally weighted regressions or geographically weighted regressions. Using an empirical housing market model from Chicago, McMillen shows that the nonparametric approaches produce more reliable fits than standard spatial econometric practice.

The final paper is by Corrado and Fingleton (2012). They agree with the two other papers that standard spatial econometrics has too often been misapplied. Namely, there is not enough emphasis on both theory or in simply forming a conceptual framework to understand spatial spillovers. Thus, empirical models are misspecified, making accurate inference challenging. Yet, rather than throwing out standard spatial econometrics, Corrado and Fingleton contend that conventional practice needs to become more rigorous in thinking about causality. Besides arguing that theory should better reflect actual spillovers, they argue that economic theory can be better incorporated in developing the **W** matrix. Better specifying **W** leads to a more structural approach in specifying spillovers versus what they contend are *ad hoc* approaches to measuring spillovers such as including distance variables. There are antecedents in early work, including examples that based the **W** matrix on commuting flows to model labor market interactions (e.g., Boarnet, 1994a, 1994b) and efforts to test different **W** matrices based on similarities that likely reflect communication paths or similarities in preferences (e.g., Case, Hines, and Rosen, 1993). Yet, as spatial econometrics has become more popular, such theory-based representations of **W** have remained more an exception than the norm. Looking forward, **W** matrices based on, for example, network analysis or even input-output theory could be used to more rigorously develop **W**. Besides constructing better weight matrices, Corrado and Fingleton's other main point is that standard spatial econometrics could greatly benefit by incorporating multilevel or hierarchical modeling approaches to better establish identification.

Spatial econometrics has clearly helped to reinvigorate regional science, but it is too often applied in a sloppy fashion without consideration of theory. It is also based on very restrictive assumptions regarding spillovers. Thus, identification and accurate inference can be compromised and the implications for policy could be very misleading. The papers in this theme issue propose three very different approaches to assessing spillovers in light of the criticisms of spatial econometrics: (1) abandon spatial econometrics and adopt the experimental route currently popular among urban and labor economists (Gibbons and Overman); (2) use standard spatial econometrics as a diagnostic tool to support more flexible nonparametric approaches (McMillen); and (3) refine standard spatial econometrics with more careful theoretical treatment, constructing better **W** matrices, and using hierarchical approaches to achieve better identification (Corrado and Fingleton). It is hard to predict where the profession will go in adopting these approaches, but we believe the *JRS* will remain the leader in disseminating the new best practice.

REFERENCES

Anselin, Luc. 1988. *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

———. 1990. "Spatial Dependence and Spatial Structural Instability in Applied Regression-Analysis," *Journal of Regional Science*, 30, 185–207.

Arraiz, Irani, David M. Drukker, Harry H. Kelejian, and Ingmar R. Prucha. 2010. "A Spatial Cliff-Ord-Type Model with Heteroskedastic Innovations: Small and Large Sample Results," *Journal of Regional Science*, 50(2), 592–614.

Boarnet, Marlon G. 1994a. "An Empirical Model of Intrametropolitan Population and Employment Growth," *Papers in Regional Science*, 73, 135–152.

———. 1994b. "The Monocentric Model and Employment Location," *Journal of Urban Economics*, 36, 79–97.

Case, Anne C., James R. Hines, Jr., and Harvey S. Rosen. 1993. "Budget Spillovers and Fiscal Policy Interdependence—Evidence from the States," *Journal of Public Economics*, 52, 285–307.

Corrado, Luisa and Bernie Fingleton. 2012. "Where Is the Economics in Spatial Econometrics?" *Journal of Regional Science*, 52(2).

Gibbons, Stephen and Henry G. Overman. 2012. "Mostly Pointless Spatial Econometrics?" *Journal of Regional Science*, 52(2).

Holmes, Thomas J. 2010. "Structural Experimentalist, and Descriptive Approaches to Empirical Work in Regional Economics," *Journal of Regional Science*, 50(1), 5–22.

LeSage, James P. and Kelly R. Pace. 2008. "Spatial Econometric Modeling of Origin-Destination Flows," *Journal of Regional Science*, 48(5), 941–967.

LeSage, James P. and Kelly R. Pace. 2009. *Introduction to Spatial Econometrics*. Boca Raton, FL: Taylor and Francis.

Manski, Charles. 1993. "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531–542.

McMillen, Daniel P. 2010. "Issues in Spatial Data Analysis," *Journal of Regional Science*, 50(1), 119–141.

———. 2012. "Perspectives on Spatial Econometrics: Linear Smooth with Structured Models," *Journal of Regional Science*, 52(2).

Monchuk, Daniel C., Dermot J. Hayes, John A. Miranowski, and Dayton M. Lambert. 2011. "Inference Based on Alternative Bootstrapping Methods in Spatial Models with an Application to County Income Growth in the United States," *Journal of Regional Science*, 51(5), 880–896.

Ord, J. Keith and Arthur Getis. 2001. "Testing for Local Spatial Autocorrelation in the Presence of Global Autocorrelation," *Journal of Regional Science*, 41(3), 411–432.

Park, In Kwon and Burkhard von Rabenau. 2011. "Disentangling Agglomeration Economies: Agents, Sources, and Spatial Dependence," *Journal of Regional Science*, 51(5), 897–930.

Pinkse, Joris and Margaret E. Slade. 2010. "The Future of Spatial Econometrics," *Journal of Regional Science*, 50(1), 103–117.

Rickman, Dan. S. and Shane D. Rickman. 2011. "Population Growth in High-Amenity Nonmetropolitan Areas: What's the Prognosis?" *Journal of Regional Science*, 51(5), 863–879.

Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region," *Economic Geography*, 46(2), 234–240.