Journal of Agricultural and Applied Economics, 43,3(August 2011):339–343

Discussion: Applications and Innovations in Spatial Econometrics

James P. LeSage

These article provide a discussion of studies presented in a session on spatial econometrics, focusing on the ability of spatial regression models to quantify the magnitude of spatial spillover impacts. Both articles presented argue that a proper modeling of spatial spillovers is required to truly understand the phenomena under study, in one case the impact of climate change on land values (or crop yields) and in the second the role of regional industry composition on regional business establishment growth.

Key Words: lagged variables, panel data, spatial spillovers

JEL Classifications: C33, C51

Both articles presented in this session make use of spatially lagged dependent variables with the article by Baylis, Paulson, and Piras entitled, "Spatial Approaches to Panel Data in Agricultural Economics" dealing with a panel data model setting and that by Lambert and Xu entitled "Business Establishment Growth and Technology Clusters in Appalachia, 2000–2007: An Exploration with Smooth Transition Spatial Process Models" in a simpler cross-sectional data setting.

A key point about models that rely on spatial lags of the dependent variable is that they allow us to quantify spatial spillovers, an important phenomenon that frequently arises in agricultural economics. I follow LeSage and Pace (2009) in defining spatial spillovers as nonzero cross-partial derivatives that show how changes in the characteristics, decisions, or actions of one economic agent influence outcomes of other agents. In an independent world, these cross-partial derivatives would be zero, indicating that only own-individual characteristics influence outcomes, not those of other individuals.

The article by Baylis, Paulson, and Piras uses the concept of spatial spillovers to assess how temperature and precipitation in one county can influence crop yields (and therefore land values) in neighboring counties. They argue that a proper modeling of spatial spillovers is required to truly understand the impact of climate change on land values (or crop yields) and that past studies assumed independence between land values in neighboring counties, making those studies deficient.

The article by Lambert and Xu uses spatial spillovers when considering the impact of industry composition on regional growth of business establishments. They argue that spatial spillovers are important, but they also posit considerable spatial variation in the magnitude of spillover impacts associated with the relationship between industry composition and business formation. To this end, they extend conventional spatial regression models involving spatial lags of the dependent variable to include a smooth transition autoregressive process (STAR), which allows for regime changes in the regression

James P. LeSage is the fields chair in Urban and Regional Economics, Department of Finance and Economics, Texas State University–SanMarcos, San Marcos, Texas.

relationship with respect to space. For example, metropolitan and nonmetropolitan spillovers might differ in a manner consistent with a two-regime relationship rather than a single global relationship for all observational units.

Discussion of Lambert and Xu Spatial Autoregressive Models–Smooth Transition Autoregressive Process Modeling

My comments regarding the work by Lambert and Xu pertain to the issue of spatial variability in the parameter impacts on the dependent variable, which is inherent in conventional spatial autoregressive models (SAR) without the space time autoregressive (STAR) process. It is generally not recognized that spatial regressions (SAR models) of the type shown in (1) used as the basis for the STAR extension by Lambert and Xu already allow implicitly for inherent variability of spillover impacts over spatial locations.

To see the nature of spatial variation that arises as a result of changes in the explanatory variables of the SAR model in (1), consider the own- and cross-partial derivatives: $\partial y/\partial x_r$ for this model shown in (2), where x_r denotes the r th explanatory variable from the matrix X. These take the form of an $N \times N$ matrix that can be expressed as in (1) (see LeSage and Pace, 2009).

(1) $$y = \rho W y + X\beta + \varepsilon$$

(2) $$\partial y/\partial x_r = (I_N - \rho W)^{-1} I_N \beta_r$$

The $N \times N$ matrix of partial derivatives show how changes in the r th variable for each observation/region i will impact the dependent variable y_j, $j = 1, \ldots, N$ in all other regions in the sample. In a model in which we have dependence among observations, changes taking place in each and every region can (potentially) impact outcomes in all other regions. Changes in the i th region give rise to N possible responses that can be found in one column of the $N \times N$ matrix. Because it is possible to consider changes in all $i = 1, \ldots, N$ regions, we have an $N \times N$ matrix that reflects impacts arising from changes in the r th explanatory variable at each location.

An implication is that responses to changes in each explanatory variable in each region differ across the sample of spatial locations. That is, the SAR model exhibits inherent spatial variation in the relationship between explanatory variables and the dependent variable responses. LeSage and Pace (2009) suggest converting the $N \times N$ matrix to scalar expressions for the own-partial derivatives $\partial y_i / \partial x_{ir}$, $i = 1, \ldots, N$ using an average of the diagonal elements from the $N \times N$ matrix as a scalar summary measure of the own-partial derivatives that they label a direct (own-region) effect. They also propose an average of the (cumulative) off-diagonal elements over all rows (observations) as a scalar summary that corresponds to the cross-partial derivative or indirect (spillover) effect associated with changes in the r th explanatory variable. These scalar summary measures of direct and indirect effects are convenient for reporting estimation results, because we do not need to report $N \times N$ matrices of results. A real attractive feature of spatial regression models that include a spatial lag of the dependent variable is that they allow us to quantify direct and indirect or spatial spillover impacts that frequently arise in agricultural economic applications. It should be noted that free software for estimating these models and reporting direct and indirect effects estimates (along with the usual t - statistics for significance) is available in both LeSage's Spatial Econometrics Toolbox routines for Matlab and Roger Bivand's R Language spdep package for estimating spatial regression models. A new set of Stata procedures for estimation spatial regression models should be available shortly.

The focus of the work by Lambert and Xu that uses a STAR extension of this SAR model is an attempt to allow for spatial variation in the relationship between explanatory variables and the dependent variable responses. From this reasoning, it should be clear that we could consider the diagonal of the matrix $(I_N - \rho W)^{-1} I_N \beta_r$ rather than the scalar summary measures if we were interested in observation-level direct impacts arising from changes in the r th explanatory variable. We could also consider off-diagonal elements of this matrix (summed for each row) to determine observation-level indirect impacts. Both of these sets of observation-level impacts would exhibit spatial variation over the locational observations of the type that interest Lambert and

Xu. It would be of interest to see how the variation in impacts from the basic SAR model corresponds to that found using the STAR approach proposed by Lambert and Xu.

Another suggestion for the model of Lambert and Xu would be use of a spatial Durbin model, which takes the form shown in (3) with the corresponding partial derivatives shown in (4).

$$(3) \quad y = \alpha \iota_n + \rho W y + X\beta + WX\theta + \varepsilon$$

$$(4) \quad \partial y / \partial x_r = (I_N - \rho W)^{-1}(I_N \beta_r + W\theta_r)$$

This model allows for characteristics of neighboring regions (WX) to influence the dependent variable (establishment growth in Lambert and Xu's application). It seems plausible that characteristics of neighboring regions would influence establishment growth, and LeSage and Pace (2009) point out that the partial derivatives for this model are much less restrictive than those from the SAR model and allow for richer variation in the impacts over space.

Discussion of Baylis, Paulson, and Piras Panel Data Modeling

It should be clear from the earlier discussion that interpretation of cross-sectional spatial regression models including spatial lags of the dependent variable requires additional work. This is because the typical situation in which the coefficient estimates of the model can be interpreted as partial derivative impacts on the dependent variable is not valid for spatial lag regression models. This has caused a great deal of confusion in past empirical studies using cross-sectional spatial regression models, and the confusion has spilled over to space–time panel data models.

The most general dynamic space–time panel model is shown in (5), in which we have $N \times 1$ variable vectors of observations y_t,x_t for time t. This model allows for time dependence by including an $N \times 1$ vector of past period values of the dependent variable y_{t-1}, cross-sectional spatial dependence through the spatial lag of the dependent variable vector Wy_t, a cross-product spatial-time lag term Wy_{t-1} that reflects diffusion over space and time as well as characteristics of neighboring regions represented by Wx_t. The $N \times 1$ vector η_t represents random effects for the N regions/observations.

$$(5) \quad y_t = \varphi y_{t-1} + \rho W y_t + \theta W y_{t-1} + x_t\beta + W x_t \gamma + \eta_t$$

$$(6) \quad \eta_t = \mu + \varepsilon_t \quad t=1, \ldots, T,$$

The (random effects) spatial lag variant of the model considered by Baylis, Paulson and Piras shown in (7) represents a special (restricted) case of this general model that excludes the dynamic/time dependence and does not allow for characteristics of neighboring regions (Wx_t) to exert an influence.

$$(7) \quad y_t = \rho W y_t + x_t\beta + \eta_t$$

$$(8) \quad \eta_t = \mu + \varepsilon_t \quad t = 1, \ldots, T,$$

For the issue of temperature and precipitation impacts on agricultural land values, dynamics are likely to play an important role. To see this, consider the space–time dynamic model application from Parent and LeSage (2010), who relate commuting times to highway expenditures. Expenditures for an improvement in a single highway segment at time t (say segment i) will improve commuting times for those traveling on this highway segment (say y_{it}) and also future travel time benefits for segment i (y_{it+T}, $T = 1, \ldots$). Equally important are commuting times on neighboring roadways, which we might denote as: y_{jt} and y_{jt+T} where $j \neq i$. This is because less congestion on one highway segment will spill over to improve traffic flow on neighboring segments.

Dynamic space–time panel data models have the ability to quantify these changes, which should prove extremely useful in development, environmental, production, and land economics as well as finance and risk management suggested by Baylis, Paulson, and Piras as areas for application of panel data models. This more general model would also provide a better fit to the illustrative application in the article involving climate change and agricultural productivity.

Debarsy, Ertur, and LeSage (2011) show that the partial derivatives $\partial y_t / \partial x_{rt}$ for these models take the form of an $N \times N$ matrix for time t, and those for the cumulative effects of a change taking place in time t at future time horizon T take the form of a sum of T different $N \times N$ matrices. Debarsy, Ertur, and LeSage (2011) derive explicit forms for these as a function of the dynamic space–time panel data model parameter estimates. This allows calculation of the dynamic responses over time and space that arise from changes in the explanatory variables.

For the special case of the spatial lag panel model in (7), interpretation of the partial derivative impacts are the same as those discussed earlier for the cross-sectional model. These impacts simply average over all cross-sectional units and time periods. It is important to note that panel data models that consider dependence only in the model disturbances such as those reported in the fourth and fifth column of Table 2 in Baylis, Paulson, and Piras can be interpreted in the same fashion as our standard (nonspatial) regression models. That is, the coefficient estimates reflect partial derivative impacts on the dependent variable that would arise from changes in the explanatory variables. This is not the case for panel data models that incorporate a spatial lag of the dependent variable such as those reported in the sixth and seventh column of Table 2. This makes direct comparison of the coefficient estimates reported in the sixth and seventh columns with those from all other columns in the table impossible. Elhorst (2011) has recently extended his MATLAB (nondynamic) spatial lag panel data functions to calculate (and print out) direct and indirect effects estimates for the models used by Baylis, Paulson, and Piras.

There is a great deal of literature regarding the asymptotic properties of various approaches to estimating simultaneous space–time panel models, and this article also provides a great deal of discussion regarding how the model is estimated. These issues are fairly well understood for maximum likelihood estimation (Elhorst, 2003), generalized methods of moments estimation (Lee and Yu, 2009), and Bayesian Markov Chain Monte Carlo (Parent and LeSage, 2011). Too little attention has been paid to how the model estimates should be interpreted. The motivation for use of space–time panel models is that they can provide us with information not available from cross-sectional spatial regressions. LeSage and Pace (2009) show that cross-sectional simultaneous spatial autoregressive models can be viewed as a limiting outcome of a dynamic space–time autoregressive process. A valuable aspect of dynamic space–time panel data models is that the own- and cross-partial derivatives that relate changes in the explanatory variables to those that arise in the dependent variable are explicit. This allows us to use parameter estimates from these models to quantify dynamic responses over time and space as well as space–time diffusion impacts. Diffusion impacts are those that arise over time as impacts travel to neighbors to neighboring regions, neighbors to those regions, and so on, producing a propagation of effects arising from changes made in one location at a single point in time.

Conclusion

Spatial regression models hold a great deal of promise for empirical applications typically encountered in agricultural economics. The ability to quantify direct and spatial spillover effects should be extremely useful when it comes to policy implications that we typically derive from empirical work. As an example, consider cost–benefit analysis of an agricultural program. The costs of the program are likely to involve direct effects associated with the agricultural entities involved in the program, whereas the benefits could accrue to those entities participating in the program as well as others who receive spillover benefits. Quantifying the costs vs. benefits requires that we take the spillover benefits into account if we do not wish to undervalue the true program benefits. Spatial regression models provide a simple approach to quantifying spillover benefits.

Decisions made by economic/agricultural agents located in space are likely to be influenced by decisions of neighbors. For example, in land use decisions regarding agricultural vs. nonagricultural use, decisions made by neighbors may exert an influence. Probit variants of spatial regression models (LeSage and Pace, 2009, Chapter 10) can quantify the spillover impact of one agent's decision on the probability of neighboring entities making similar decisions.

Spatiotemporal panel data models hold the promise of quantifying future period dynamic responses to changes that take place at one point in space and time. These responses would incorporate changing behavior that arises over time as economic actors adjust their behavior in response to the dynamically changing environment.

References

Debarsy, N., C. Ertur, and J.P. LeSage. "Interpreting Dynamic Space–Time Panel Data Models." *Statistical Methodology* (2011).

Elhorst, J.P. "Specification and Estimation of Spatial Panel Data Models." *International Regional Science Review* 26,3(2003):244-68.

———. "MATLAB Software for Spatial Panels." *International Regional Science Review* Forthcoming. doi:10.1016/j.stamet.2011.02.002

Lee, L., and J. Yu. "Estimation of Spatial Autoregressive Panel Data Models with Fixed Effects." *Journal of Econometrics* 154,2(2009):165–85.

LeSage, J.P., and R.K. Pace. *Introduction to Spatial Econometrics*. New York: Taylor & Francis/CRC Press, 2009.

Parent, O. and J.P. LeSage. "A Spatial Dynamic Panel Model with Random Effects Applied to Commuting Times." Transportation Research Part B: Methodological 44B(2010):633–45.

———. "A Space–Time Filter for Panel Data Models Containing Random Effects." *Computational Statistics & Data Analysis* 55,1(2011): 475–90.