Journal of Economic Dynamics & Control 48 (2014) 229–245

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Modeling regional economic dynamics: Spatial dependence, spatial heterogeneity and nonlinearities

Roberto Basile [a,*], María Durbán [b], Román Mínguez [c], Jose María Montero [d], Jesús Mur [e]

[a] Department of Economics, Second University of Naples, Corso Gran Priorato di Malta, 1 – 81043 Capua, CE, Italy
[b] Department of Statistics, Carlos III University, Madrid, Spain
[c] Statistics Department, University of Castilla-La Mancha, Cuenca, Spain
[d] Statistics Department, University of Castilla-La Mancha, Toledo, Spain
[e] Department of Economic Analysis, University of Zaragoza, Zaragoza, Spain

A R T I C L E I N F O

Article history:
Received 27 September 2013
Received in revised form
11 May 2014
Accepted 13 June 2014
Available online 24 June 2014

JEL classification:
R11
R12
C14

Keywords:
Spatial econometrics
Nonlinearities
Semiparametric models

A B S T R A C T

Spatial modeling of economic phenomena requires the adoption of complex econometric tools, which allow us to deal with important methodological issues, such as spatial dependence, spatial unobserved heterogeneity and nonlinearities. In this paper we describe some recently developed econometric approaches (i.e. Spatial Autoregressive Semiparametric Geoadditive Models), which address the three issues simultaneously. We also illustrate the relative performance of these methods with an application to the case of house prices in the Lucas County.

1. Introduction

Spatial modeling of economic phenomena (growth, unemployment, wages, location, house prices, crime rates and so on) requires the adoption of complex econometric tools which permit us to control for spatial dependence, unknown functional form and unobserved heterogeneity. The dominant paradigm in spatial econometrics is not well equipped to deal simultaneously with the three topics, which instead have been approached separately.

Spatial dependence reflects a situation where values observed at one location depend on the values of neighboring observations. That is, there are externalities known as *global and local spatial spillovers* (Anselin, 2003).[1] Contrary to what one would expect, in only a few cases spatial externalities have been formally predicted by well-defined theoretical models. Ertur and Koch (2007), for example, propose an extension of the multi-region neoclassical growth model that includes

* Corresponding author.
E-mail addresses: roberto.basile@unina2.it (R. Basile), mdurban@est-econ.uc3m.es (M. Durbán), roman.minguez@uclm.es (R. Mínguez), jose.mlorenzo@uclm.es (J. María Montero), jmur@unizar.es (J. Mur).

[1] There is a large list of models devoted to this topic (LeSage and Pace, 2009): the Spatial Lag or Spatial Autoregressive Model (SAR), the Spatial Error Model (SEM), the Spatial Durbin Model (SDM), the Spatial in X-variables Model (SLX) and a mix of the SAR and SEM (SARSAR) are the most popular.

http://dx.doi.org/10.1016/j.jedc.2014.06.011
0165-1889/

technological interdependence across regions. The reduced form of the growth equation predicted in this case is a linear SDM. Brueckner (2006) also presents several theoretical models of spatial interaction among local governments that lead directly to the SAR model for empirical implementation.

In most of the cases, instead, economic theory suggests the existence of network dependence and spatial spillovers, but it does not predict a well structured model. An example is the literature on the regional knowledge production function and on the diffusion of innovation, where spatial (knowledge) spillovers may occur through collaborative networks or other forms of spatial interactions (Autant-Bernard, 2012). These cases are characterized by uncertainty about the functional form of the model. The premise in applied literature is that a linear structure, coupled with some previous transformation of the data, offers enough flexibility to account for the problem.

However, there is growing evidence showing that this is a quite optimistic view. Strong nonlinearities have been detected in studies on regional growth (Arbia and Paelinck, 2003; Azomahou et al., 2011; Basile and Gress, 2005; Basile, 2008, 2009; Basile et al., 2012; Ertur and Gallo, 2009; Fotopoulos, 2012), urban agglomeration economies (Basile et al., 2013), urban environment (Chasco and Le Gallo, 2011), land prices (McMillen, 1996), urban sprawl (Brueckner, 2000; Brueckner et al., 2001; Irwin and Bockstael, 2007), social interaction (Lee et al., 2010) or house prices (Bourassa et al., 2010; Kim and Bhattacharya, 2009; Goodman and Thibodeau, 2003). Thus, in a typical empirical application, the *functional form is unknown* and the linear form, imposed sometimes rather arbitrarily, represents another source of mis-specification bias.

Controlling for *unobserved heterogeneity* is another fundamental challenge in empirical research, as failing to do so can introduce omitted-variable biases and preclude causal inference. To complicate the analysis, spatial dependence may simply be the consequence of (spatially correlated) omitted variables rather than being the result of spillovers. If this is the case, there are no compelling reasons for using traditional parametric models, like the SAR or SEM. As McMillen (2012) shows, a simple semiparametric model, with a smooth interaction between latitude and longitude (the so-called *Geoadditive Model*), can remove unobserved heterogeneity.

However, as mentioned above, in many cases the aim of the empirical study is to assess the impact of spillover effects (for example the global effect of a localized shock in R&D investment) rather than simply compensate for unobserved heterogeneity. In these cases we need to capture the effect of spatial spillover through the inclusion of spatial interaction terms, besides controlling for unobserved heterogeneity and functional form mis-specifications. This is a complex objective that the parametric paradigm, dominant nowadays, can hardly attain. It must be recognized that there have been attempts to develop a more general framework. This is the case of the parametric model proposed by Lambert et al. (2014), which combines spatial dependence and nonlinearity, or the case of Lotka–Volterra prey-predator model proposed by Griffith and Paelinck (2011). The literature on spatial regimes introduces heterogeneity in models with spatial dependence (Fischer and Stumpner, 2010), from which the SALE (Spatial Association Local Estimation) (Pace and LeSage, 2004) and Zoom algorithms (Mur et al., 2010) can be considered as limiting cases. According to our knowledge, few more references can be added. In fact, the history is very short.

Our impression is that there is a genuine need for more general and powerful approaches to model spatial data, and we are not alone in this position. In fact, several prominent scholars have recently called for a review of the methodological basis of the traditional spatial econometrics. McMillen (2010, 2012) points that there is a fundamental contradiction between the severity of the unknowns in the specification (functional form and spatially correlated omitted variables) and the overwhelming use of maximum likelihood methods (which heavily depend on the assumption of a correct specification). Pinkse and Slade (2010) recognize the intrinsic complexity of spatial data which suffer from so many problems (irregular spatial arrangement, varying density, aggregation, and so on) that precludes the use of naively parsimonious specifications, like the family of SAR models. The comparison with time series literature is deceiving because stationarity is a strange concept over space. Their advice can be summarized in avoiding overparameterized specifications and letting the application guide the theory; this has a clear parallel with the position of McMillen.

According to Gibbons and Overman (2012), the dominant approach in spatial econometrics is not convincing because of the many, sometimes unjustified, hypotheses made about the functional form, the presence of omitted factors, the spatial weights, and so on. These authors confer special relevance to the notions of identification and causality. Spatial models that include spatial lags of the endogenous variable, together with a set of contextual variables in the right hand side of the equation, are not identified because of the essential collinearity between the contextual variables and the output variable. This is the *'reflection problem'* posed by Manski (1993) in relation to 'peer effects' models. However, this problem was solved by Pinkse and Slade (2010) with the distinction between 'expected reaction of the individual', a relevant concept in the peer effects literature, and spillover effects, which is the adequate notion in spatial analysis (that is, the spatial lag is no longer a mere sample analogue of the expected reactions of the neighbors). Causality is a 'gold standard' in economics except in the field of spatial econometrics where, surprisingly, the concept of causality is mixed with that of correlation. However, fit well the data may mean nothing but spurious correlation or common factors. Gibbons and Overman (2012) confer some merit to the description of spatial data, but this cannot be the ultimate goal of the analysis. A further critical issue raised by Gibbons and Overman (2012) concerns the use of lagged values of the regressors as instrument variables (IV) for the spatial lag of the endogenous variable in the SAR-type models. The arguments are somewhat familiar with Pinkse and Slade (2010): the first is the unconvincing exclusion of these terms (spatial lagged regressors) from the structural equation; the second is the unjustified claim of exogeneity for the X's variables in a typical spatial model (contrary, they are expected to be endogenous and correlated with the unobserved determinants to the endogenous variable). The last deficiency can be treated more efficiently by using, once again, less structured models.