Journal of Econometrics 199 (2017) 12–34

Contents lists available at ScienceDirect

Journal of Econometrics

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

Semiparametric estimation and testing of smooth coefficient spatial autoregressive models☆

Emir Malikov [a,*], Yiguo Sun [b]

[a] Department of Agricultural Economics and Rural Sociology, Auburn University, Auburn, AL, United States
[b] Department of Economics and Finance, University of Guelph, ON, United States

ARTICLE INFO

Article history:
Received 21 June 2015
Received in revised form 16 July 2016
Accepted 27 February 2017
Available online 19 April 2017

JEL classification:
C13
C14
C21

Keywords:
Consistent test
Constrained estimation
Local linear fitting
Nonparametric GMM
Partially linear
Quadratic moments
SAR
Spatial lag

ABSTRACT

This paper considers a flexible semiparametric spatial autoregressive (mixed-regressive) model in which unknown coefficients are permitted to be nonparametric functions of some contextual variables to allow for potential nonlinearities and parameter heterogeneity in the spatial relationship. Unlike other semiparametric spatial dependence models, ours permits the spatial autoregressive parameter to meaningfully vary across units and thus allows the identification of a neighborhood-specific spatial dependence measure conditional on the vector of contextual variables. We propose several (locally) nonparametric GMM estimators for our model. The developed two-stage estimators incorporate both the linear and quadratic orthogonality conditions and are capable of accommodating a variety of data generating processes, including the instance of a pure spatially autoregressive semiparametric model with no relevant regressors as well as multiple partially linear specifications. All proposed estimators are shown to be consistent and asymptotically normal. We also contribute to the literature by putting forward two test statistics to test for parameter constancy in our model. Both tests are consistent.

1. Introduction

While spatial econometric methods have solidly become part of a standard methodological toolkit of applied researchers in many fields of economics that deal with spatial data which include applications such as land use, hedonic pricing or cross-country growth studies, most empirical work has confined its analysis to linear spatial models only. However, to paraphrase Paelinck and Klaassen (1979), econometric relations in space result more often than not in highly nonlinear specifications (as cited in van Gastel and Paelinck, 1995). For instance, allowing for nonlinearities in the hedonic house price function is often argued to be crucial in order to obtain realistic marginal valuations of housing attributes (e.g., see Parmeter et al., 2007). Taking such potential nonlinearities in spatial models for granted is likely to lead to inconsistent parameter estimates and thus misleading conclusions.

This paper proposes a semiparametric method to handling nonlinearity (and parameter heterogeneity) in models of spatial dependence. We extend a particular class of semiparametric models in which parameters of a linear regression are permitted to be unspecified smooth functions of some contextual variables (Hastie and Tibshirani, 1993; Cai et al., 2000; Li et al., 2002) to the case of data with spatial dependence. Specifically, we generalize a popular parametric spatial autoregressive mixed-regressive model by allowing its coefficients, including the spatial autoregressive parameter, to be nonparametric functions of unknown form [for concreteness, see Eq. (2.1)].

While our "smooth coefficient" spatial autoregressive model closely relates to the family of partially linear semiparametric spatial models recently proposed in the literature (Su and Jin, 2010; Su, 2012; Zhang, 2013; Sun et al., 2014), its distinct feature is that it permits the spatial autoregressive parameter to meaningfully vary across units. The latter may be highly desirable from

☆ We would like to thank the editor, the associate editor and two anonymous referees for many insightful comments and suggestions that helped improve this article. We are also thankful to conference participants at the NY Camp Econometrics X, 2015 Midwest Econometrics Group Meeting at the St. Louis Fed, 85th Annual Meeting of the Southern Economic Association and the 2015 Tsinghua International Conference for Econometrics in Beijing for many helpful comments and suggestions. This project was supported by the Alabama Agricultural Experiment Station and the Hatch program of the National Institute of Food and Agriculture, U.S. Department of Agriculture. Yiguo Sun gratefully acknowledges financial support from the Social Sciences and Humanities Research Council of Canada Insightful Grant 435-2016-0340. Any remaining errors are our own.

* Corresponding author.
E-mail addresses: emalikov@auburn.edu (E. Malikov), yisun@uoguelph.ca (Y. Sun).

http://dx.doi.org/10.1016/j.jeconom.2017.02.005
0304-4076/

a practitioner's point of view since it allows the identification of a neighborhood-specific spatial dependence measure conditional on the vector of contextual variables. For instance, when a spatial autoregressive model is game-theoretically rationalized as a "response function", our model empowers a researcher to estimate heterogeneous "reaction" parameters that can vary with some environmental control factors. Some potential applications of our model, for instance, include the estimation of growth models that explicitly account for technological interdependence between countries in the presence of spillover effects. Such technological interdependence is usually formulated in the form of spatial externalities (e.g., see Ertur and Koch, 2007). However, the intensity of knowledge spillovers is naturally expected to greatly depend on institutional and cultural compatibility of neighboring countries (Kelejian et al., 2013). Our smooth coefficient spatial autoregressive model presents a practical, easy-to-implement way to allow for such indirect effects of institutions on the degree of spatial dependence in the cross-country conditional convergence regressions.

Our semiparametric treatment of nonlinearities is also relatively more flexible than pioneer nonlinear modeling approaches in spatial dependence models put forward by van Gastel and Paelinck (1995), Baltagi and Li (2001), Pace et al. (2004) and Yang et al. (2006). At the same time, like in all these studies as well as many others (e.g., Kelejian and Prucha, 1998, 1999, 2010; Lee, 2004, 2007; Su and Jin, 2010; Su, 2012), the consistency of our estimator rests on an admittedly rather restrictive assumption of a correctly pre-specified spatial weighting matrix. Dispensing with this assumption requires either making an alternative assumption of strong spatial mixing along with spatial stationarity or modeling spatial weights as nonparametric functions of the distance between neighbors (see Sun, 2016 and the references therein). The acute disadvantage of both of these alternative approaches is an inability to quantify a spatial autoregressive parameter (a "reaction" parameter) which many empirical studies are specifically interested in. In this paper, we therefore abstract from the issue concerning the correct specification of spatial weights.[1]

We propose several (locally) nonparametric Generalized Method of Moments (GMM) estimators for our model. The developed estimators incorporate both the linear and quadratic orthogonality conditions and are capable of accommodating a variety of data generating processes, including the instance of a pure spatially autoregressive semiparametric model with no relevant regressors as well as multiple partially linear specifications. To this end, we contribute to the literature on four fronts. First, our paper is the first (to the best of our knowledge) attempt in the nonparametric estimation literature to make use of local quadratic orthogonality conditions, which are necessary for the IV identification of spatially autoregressive models in the case when all explanatory covariates are irrelevant in predicting the outcome variable (see Lee, 2007). Second, we propose a two-stage estimation procedure whereby we first obtain an initial estimator of unknown parameter functions using feasible, but likely not so strong, instruments which we then use for the construction of more natural instruments suggested by the model's reduced form. Again, to our knowledge, no prior attempt has been made in the nonparametric econometrics literature to study such a class of estimators which themselves are based on the estimated instruments. Third, we also consider two special cases of our model by allowing some of its parameter functions to be constant thus resulting in a partially linear specification. Our proposed estimators present an alternative to those by Su (2012) and Zhang (2013) who study a similar class of partially linear spatial models. Unlike their estimators, ours however preserves its consistency property if the true model is a pure spatial autoregression. Fourth, we discuss ways of ensuring that the estimated model satisfies the non-singularity condition needed to rule out unstable Nash equilibria. In the instance of a mixed-regressive model, we impose this non-singularity restriction via the "tilting" procedure à la Hall and Huang (2001) whose theoretical results we generalize to the case of GMM estimators in the presence of endogenous regressors. Under fairly mild regularity conditions, we show that all our proposed estimators are consistent and asymptotically normal.

Further, we contribute to the literature by putting forward two test statistics to test for parameter constancy in our model. These model specification tests allow us to discriminate between a standard linear spatial autoregressive and our semiparametric models. The first consistent test utilizes a popular residual-based specification test technique which we extend to spatial data with cross-sectional dependence.[2] Given the well-known poor performance of nonparametric residual-based tests in finite samples, we also suggest a (wild) bootstrap procedure for it which we show to be asymptotically valid in approximating the null distribution of our test statistic regardless of whether the null hypothesis holds true or not. However, our residual-based test is impractical when the spatial model has no regressors. We therefore propose an alternative consistent test statistic à la Henderson et al. (2008) which provides a vehicle for testing for parameter constancy in our model even when the model is a pure spatial autoregression.

We investigate the finite sample performance of the proposed estimators and test statistics in a small set of Monte Carlo experiments. The results are encouraging and show that all estimators and tests perform well in finite samples with considerable improvements as the sample size increases. Overall, simulation experiments lend support to our asymptotic results. To showcase our methodology, we then apply it to estimate a spatial hedonic price function using the well-known Harrison and Rubenfeld's house price data from Gilley and Pace (1996), where we let unknown parameter functions to vary with the NO_x concentration in the air. We find that spatial dependence between house prices is statistically significant only at higher values of the NO_x concentration in the air and that the degree of this spatial dependence, on average, increases as the air quality declines. This finding suggests that locational similarity may matter little for house valuations in pollution-free localities.

The rest of the paper proceeds as follows. Section 2 outlines the model. We present our estimators in Section 3, where we also provide their large-sample statistical properties. In Section 4, we consider two different types of partially linear spatial autoregressive models. Section 5 discusses specification tests for parameter constancy. The results of Monte Carlo simulations are described in Section 6. Section 7 concludes.

Throughout the paper, we use M to denote a generic finite constant that can take different values at different appearances and vec $\{\mathbf{A}\}$ stacks columns of an $n \times m$ matrix $\mathbf{A}$ into an $(nm) \times 1$ vector. Lastly, tr $\{\mathbf{A}\}$ refers to the trace of a square matrix $\mathbf{A}$.

2. Semiparametric spatial autoregressive model

Consider a semiparametric generalization of the conventional (linear) spatial autoregressive mixed-regressive model, where the

[1] Furthermore, theoretical basis for the widely-believed sensitivity of the estimates to the choice of spatial weights remains rather unclear, as recently argued by LeSage and Pace (2014).

[2] Asymptotic properties of such a test have been studied for independent data (e.g., Zheng, 1996; Li and Wang, 1998; Stengos and Sun, 2001), weakly dependent time series data (e.g., Fan and Li, 1999; Li, 1999) and integrated time series data (e.g., Gao et al., 2009; Wang and Phillips, 2012; Sun et al., 2015), to mention few among many contributions.