VU Research Portal

Theory and Application of Dynamic Spatial Time Series Models

Andree, B.P.J.

2020

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Andree, B. P. J. (2020). *Theory and Application of Dynamic Spatial Time Series Models*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:
vuresearchportal.ub@vu.nl

Download date: 20. May. 2025

Chapter 1

Introduction

This thesis sets out to develop econometric theory and methods to analyze dynamic interactions between observations that are interrelated across space and time. This type of modeling is becoming increasingly important as sensors and institutions continue to gather rich subnational spatial time series of remotely sensed or surveyed economic variables. Going from finance, to macro-economics or the environment, nearly all policy relevant phenomena in the socio-economic domain involve multivariate interactions across both spatial and temporal dimensions. Analyzing these problems raises a number of inquiries about the econometric methods used that are both practically and theoretically interesting. In particular, cross-sectional data is often spatially dependent. From a data generating perspective, this implies that we may be concerned with models that exhibit instantaneous forms of feedback in space. Together with possible endogenous interactions between the observations of the different variables that are collected sequentially over the time dimension, this produces complex feedback properties that may violate various assumptions made by standard econometric models. Second, as the dimensions of datasets grow, it becomes increasingly unlikely that linear relationships provide a realistic description of these phenomena. The tendency of nonlinearities and the complex feedback properties that characterize spatial time series, render many related estimation problems non-standard.

In many cases, deriving the properties of estimators for multivariate models that have complex nonlinearities over both temporal as well as spatial dimensions, can be achieved by extending the theories used to analyze the estimators of dynamic time series models. In particular, spatial feedback renders the standard Least Squares Estimator (LSE) inconsistent or inefficient depending on the situation, but estimating models that explicitly factor in the dependence and feedback between neighbors can be done within the framework of Maximum Likelihood. Other interesting problems, such as exogenous or non-contemporaneous endogenous nonlinearities, can be estimated in the Least Squares framework. In both cases, this requires modifications to the standard criterion functions used. In particular, nonlinear parametric models of spatial time series introduce new components to the likelihood function that correct for the fact that the conditional densities are derived from a nonlinear transformation of the residuals. This requires new proofs that the well-known theoretical results associated with the standard Maximum Likelihood Estimator (MLE) nonetheless apply. Non-parametric Least Squares estimation of nonlinearities over the levels of cross-sectional observations can be solved as a locally linear problem, but requires penalization techniques to ensure that convergences essentially operate within simple spaces. This may change the interpretation of the limiting result all together. We will further investigate these issues in this thesis.

Many of the ideas produced in this thesis build heavily on the theory that underlies the analysis of time series data. This is a natural angle to view many problems. Early spatial models have been developed primarily to analyze cross-sectional data. As such, the underlying theory relied on taking the number of cross-sectional observations to infinity. While this may be sufficient to establish consistency and normality theoretically, in most real world applications it occurs seldom that new cross-sectional observations are made. Often, new observations are only collected over time while the number of spatial units remains fixed. In addition, when new cross-sectional observations are in fact made, it is difficult to perceive that this change does not somehow involve also an extension in the time dimension.

The analysis of spatial data over time is a concept that is gaining in popularity, but it is still relatively new. It is only since recent that a significant part of our cross-sectional datasets have grown substantially enough in the time dimension to exhibit interesting temporal dynamics. For example, with modern compute it is still not possible for everyone to analyze remotely sensed data at high temporal resolution. Many publicly available datasets are therefore summarized as annual statistics that span only a modest number of years. Economic surveys that are consistently gathered across regions are often expensive. As an effect, surveyed data usually have a similar low temporal frequency. Finance data can be available at higher frequency, but many time series only start after the digital infrastructures that support modern systems matured. When one wishes to analyze a problem that involves multiple sources of data, then the data on which the analysis rests will often be constrained in both frequency and dimension. However, we are now at a point that sufficient data can in many cases be found, resulting in interesting problems that one can analyze with basic theory. In particular, with existing time series theory it is possible to analyze the properties of complex nonlinear dynamic time series models and understand the behavior of general estimators in these settings. However, this theory was not developed with spatial dependence and possible multivariate cross-sectional nonlinearities in mind. Many of the existing spatial analysis techniques have on the other hand not been developed with non-linear, possibly observation-driven, dynamics in mind. Moreover, panel techniques often focus on a single dependent variable, and are less concerned with describing the state transitions and dynamics between multiple spatial variables over time, which is needed for multivariate spatial time series forecasting, stochastic simulation, and impulse response analysis.

Before exploring spatial relationships explicitly, we will first review several important standard theoretical results for the estimation of dependencies in cross-sectional time series. We will use this as a basis to discuss what is further needed to analyze dynamic spatial time series problems. This background theory will be confined to what is needed to read the remainder of this thesis in a relatively self-contained manner. The remainder of this thesis then touches upon five key topics:

i Spatial heterogeneity

ii Parametric spatial nonlinearities

iii Non-parametric cross-sectional nonlinearities

iv Vector spatial time series

v Probability and causality in spatial time series

Chapter 3 analyzes spatial heterogeneity. Specifically, it uses simple linear relationships and spatial explicit data to simulate economic outcomes at high spatial resolution. The analysis highlights how economic outcomes can cluster in space due to the natural clustering of independent geophysical variables that may be of economic importance. Moreover, it reveals that simple relationships at a high spatial resolution can produce nonlinear patterns at aggregated levels.

The concepts of spatial heterogeneity, dependence, and nonlinearity form the basis of Chapter 4 that looks into parametric spatial nonlinearities. This chapter covers the econometric application of spatial autoregressive time series models and extends the theory to cover nonlinear spatial dependence. The model that is introduced allows dependence to vary smoothly across levels in the data in an idiosyncratic manner. It will be shown that this type of spatial modeling captures both spatial and temporal dynamics and performs better than the standard linear spatial autoregressive model on a number of widely used diagnostics. Moreover, the chapter will show that this type of modeling can produce interesting results when both T is large and N is small, or when N is large and T is still relatively modest.

Chapter 5 drops the parametric assumption, and looks at the case of non-parametric panel relationships. In this case, the focus is on nonlinear dependence of spatial time series variables on independent data in a manner that is appropriate when a researcher wishes to impose only mild assumptions about the shape of the functional relationships. This allows for a wide range of functional relationships in the data, but, as we shall see, it is necessary to add additional structure to the criterion function to estimate these type of models. The chapter discusses how this impacts the interpretation of basic estimated quantities, and discusses how an appropriate functional form can be estimated while jointly addressing the need for possible fixed effects. It will then be shown how the resulting models can be used to produce alternative future scenarios that take into account historical nonlinear patterns.

In Chapter 6, the discussion moves away from nonlinearities, and shifts the focus toward inter-temporal dynamics between multiple variables within a spatial system. Estimation of interdependence among multiple time series is often at the center of time series analysis, but many panel methods have traditionally been developed with inferential questions about a single dependent variable in mind. The model introduced in this chapter extends the standard spatial time series model to the multiple variable setting and introduces methods to analyze how finite impulse responses flow through a spatial system in the presence of both spatial and temporal forms of feedback. This is useful to address questions about the order in which effects occur over time when variables are not only temporally, but also spatially dependent. While the chapter introduces the analytical framework in a linear way, focusing on a relatively homogeneous subset of locations, the nonlinear concepts introduced in Chapter 4 and 5 can naturally be applied in similar settings to study nonlinear impulse response behavior in heterogeneous systems.

Finally, Chapter 7 circles back to some of the fundamental concepts that are introduced in Chapter 2 that covers background theory. Only this time, the discussion stays at a more general level and focuses on the concepts of probability and causal inference in dynamical systems. The discussion highlights why using flexible models, such as the ones introduced in this thesis, are desirable in the first place when one is interested in answering basic questions about cause and effect in a multivariate setting. An argument will be provided for flexible specification of the possible time dynamics in a spatial system together with estimation strategies that minimize distance to the true probability measure that underlies the observed data. In practice, this implies a general to specific approach to exclude irrelevant dependencies. The particular case of maximizing penalized Maximum Likelihood will be discussed further, which provides additional support for the estimation strategies used throughout this thesis.