Dynamic Generalized Structural Equation Models

Luigi Ippoliti

Department of Economics, University G. d'Annunzio, Chieti-Pescara, Italy

Abstract

This work discusses a number of conceptual issues pertaining to the study of the relationship existing between two groups of variables which are supposed to be spatially and temporally correlated. Since it is assumed that this relationships can be studied in a reduced latent space, we provide an overview of the motivations for including spatial effects in a dynamic factor model, both from a theory-driven as well as from a data-driven perspective. Considerable attention is paid to the inferential framework necessary to carry out estimation and to the different assumptions, constraints and implications embedded in the various model specifications. The discussion combines insights from (spatial) econometrics and geostatistical literature.

Key Words: Dynamic factor models, Structural equation models, Spatio-temporal processes, Bayesian inference, Spatial econometrics, Geostatistics.

1 Introduction

The idea of borrowing information from different but related sources can be very powerful for statistical analysis. It proved to be very useful in the last decades where complex data structures began to be tackled, as they required sophisticated modeling strategies. In this work we consider the problem of modeling high-dimensional multivariate, spatially and temporally referenced data. This problem has enjoyed widespread popularity in the last years and requires the definition of general and flexible statistical models where the temporal and cross-sectional dependencies must be accommodated.

Spurred by recent advances in Geo-spatial data acquisition technologies, it is often desirable for these data to examine the relationships existing between one or more dependent variables and some other linked covariates. This can be achieved in a number of ways, though there might be no single approach which can be considered uniformly as being the most appropriate.

Among the different methodologies proposed in the literature, dynamic factor models (DFMs) have grown significantly in popularity and have been shown to be very useful for exploratory analysis, policy analysis and forecasting in a data-rich environment. DFMs have been widely developed in both methodological and practical issues, and have become a standard tool for increasingly high-dimensional modeling of time series. They have been extensively used in macroeconomics and finance with the core idea of explaining the common dynamic structure of the multivariate time series through a set of common (time series) factors. This is achieved by the introduction of flexible temporal correlation structures for the latent factors, previously assumed to be independent. This renders the DFM capable of assessing the complexity of time series data. Models along these lines include Geweke (1977), Molenaar (1985), Engle and Watson (1981), Peña and Box (1987), Forni et al. (2000).

Special attention will be devoted here to the use of a dynamic factor analytic approach in the framework of spatial statistics. It will be shown that this is not only an important area of application but also that this area can receive several benefits from this modeling approach (see, for example, Ippoliti et al., 2012; Valentini et al., 2013).

A key property of much spatio-temporal data is that observations at nearby sites and times will tend to be similar to one another. Then, factor analysis assumes that the cross dependence can be characterized by a finite number of unobserved common factors, possibly due to common shocks that affect all the spatial sites, albeit with different intensities. Thus, the strong co-movement and the high correlation among the series, amplified by the presence of spatial correlation, suggest that both observable and unobservable factors must be at place.

In this paper, we thus approach the analysis of multivariate spatio-temporal processes from the perspective of recent developments of dynamic factor models. Through a fully Bayesian approach, we contribute to the recent literature by melding together dynamic factor models, spatial regression models and geostatistical techniques, in order to explain the multifactorial nature of many spatio-temporal data. We assume that the relationships existing between the groups of dependent and regressor variables can be studied through a temporally dynamic and spatially descriptive model, hereafter referred to as *dynamic generalized structural equation* (DGSE) model.

The proposed model has an intuitive appeal and enjoys several advantages. First, our model formulation exploits the spatio-temporal nature of the data and explicitly defines a non-separable spatio-temporal covariance structure of the multivariate process. Second, since the data have a multivariate and multidimensional structure, in that several time series can be measured at specific spatial sites, the temporal relationships between dependent and regressor variables is modeled in a latent space. The observed processes are thus described by a potentially small set of common dynamic latent factors with the advantage of overcoming the difficulties of interpreting the relationships under study due to collinearity and low signal-to-noise ratio issues. Temporal forecasts of the variables of interest can also be obtained by only modeling the dynamics of a few common factors. Third, by modeling the spatial variation via spatially structured factor loadings, we entertain the possibility of identifying clusters of spatial sites that share common time series components. Through the spatial modeling of the factor loadings, spatial interpolations of the observed variables are also straightforward. Fourth, several general structures that make use of different covariate information, can be easily accommodated in the different levels of the hierarchy. Fifth, the DGSE model offers a unified approach suitable to deal with variables and indicators measured at different scales and coming from different spatial sources. Hence, the model provides a simple solution to the misalignment problems which, for example, normally occurs in health care research. Lastly, the model specification is not limited to normally distributed variables, but it can be extended to handle more types of variables from an exponential family.

Details on data analysis will be given in an extended version of the present abstract in the framework of environmental and health care research.

References

Engle, R., Watson, M. (1981). A one-factor multivariate time series model of metropolitan wage rates. *Journal of the American Statistical Association*, **76**, 774-781

Forni, M., Hallin, M., Lippi, M., Reichlin, L. (2000). The generalized dynamic factor model: identification and estimation. *The Review of Economics and Statistics*, **82**, 540-554

Geweke, J. (1977). The dynamic factor analysis of economic time series models. In *Latent Variables in Socio-Economic Models (eds. D. J. Aigner and A. S. Goldberger)*, pp. 365-383, North Holland, Amsterdam

Ippoliti, L., Valentini, P., Gamerman, D. (2012). Space-Time Modelling of Coupled Spatio-Temporal Environmental Variables. *Journal of the Royal Statistical Society*, Series C (Applied Statistics), **61**, 175-200

Molenaar, P. C. M. (1985). A dynamic factor model for the analysis of multivariate time series. *Psychometrika*, **50**, 181-202

Pena D., Box, G. (1987). Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, **82**, 836-843

Valentini, P., Ippoliti, L., Fontanella, L. (2013). Modeling US housing prices by spatial dynamic structural equation models. *The Annals of Applied Statistics*, **7**, 763-798