International Journal of Forecasting 16 (2000) 229–246

international journal of forecasting

www.elsevier.com/locate/ijforecast

A method for spatial–temporal forecasting with an application to real estate prices

R. Kelley Pace[a,*], Ronald Barry[b], Otis W. Gilley[c], C.F. Sirmans[d]

[a]E.J. Ourso College of Business Administration, Louisiana State University, Baton Rouge, LA 70803, USA
[b]Department of Mathematical Sciences, University of Alaska, Fairbanks, AK 99775-6660, USA
[c]Department of Economics and Finance, College of Administration and Business, Louisiana Tech University, Ruston, LA 71272, USA
[d]Center for Real Estate and Urban Studies, 368 Fairfield Rd., Rm 426, U-41RE, Storrs, CT 06269-2041, USA

Abstract

Using 5243 housing price observations during 1984–92 from Baton Rouge, this manuscript demonstrates the substantial benefits obtained by modeling the spatial as well as the temporal dependence of the errors. Specifically, the spatial–temporal autoregression with 14 variables produced 46.9% less SSE than a 12-variable regression using simple indicator variables for time. More impressively, the spatial–temporal regression with 14 variables displayed 8% lower SSE than a regression using 211 variables attempting to control for the housing characteristics, time, and space via continuous and indicator variables. One-step ahead forecasts document the utility of the proposed spatial–temporal model. In addition, the manuscript illustrates techniques for rapidly computing the estimates based upon an interesting decomposition for modeling spatial and temporal effects. The decomposition maximizes the use of sparsity in some of the matrices and consequently accelerates computations. In fact, the model uses the frequent transactions in the housing market to help simplify computations. The techniques employed also have applications to other dimensions and metrics.

Keywords: Real estate; Spatial–temporal forecasting; Modeling

1. Introduction

Data sets have often been organized by units of time such as quarters or years as well as by geographical constructs such as regions, states, or counties. In reality, the data often represent an aggregation of individual observations which have more precise temporal and spatial characteristics. Much of the governmentally collected economic, medical, and social data falls into this category. However, the increasing capabilities of information systems and especially geographic information systems (GIS) have greatly aided work with disaggregated data having precise spatial and temporal references. For example, automated point-of-sale data from individual stores have a precise identification in

*Corresponding author. Tel.: +1-225-388-6238; fax: +1-225-334-1227.

E-mail addresses: kpace@lsu.edu (R.K. Pace), ffrpb@aurora.alaska.edu (R. Barry), gilley@cab.latech.edu (O.W. Gilley), fnceadm5@uconnvm.uconn.edu (C.F. Sirmans)

0169-2070/00/$ – see front matter
PII: S0169-2070(99)00047-3

time and space. Even governmentally collected data, where privacy concerns dictate minimum levels of aggregation, have become more precise in their identification of space or time. For example, the Home Mortgage Disclosure Act (HMDA) data for 1993 contained approximately 15 million individual transactions identified by over 60 000 locations (census tracts).

The trend towards large data sets with substantial spatial and temporal detail raises the issue of how to forecast such data. Moreover, such data raise computational issues as well as conceptual issues of how to plausibly model the spatial and temporal dependence.

Housing prices provide another example of this type of data. First, over 4.5 million houses sold during 1997 alone. Most data sources (multiple listing services or assessor databases) record the day, month, and year of the transaction. Given an address, geographic information systems can provide the corresponding latitude and longitude (or other locational coordinates) for 80% or more of the records (Johnson, 1998).

At least five common applications employ housing transaction data. First, most houses in the US (and many in other countries) have their assessed value for tax purposes determined by the predictions from statistical models calibrated using individual housing transactions (Eckert, 1990, p. 27). Second, the movement of many primary and secondary lenders to some form of automated appraisal places an added premium on prediction accuracy (Gelfand, Ghosh, Knight & Sirmans, 1998). Third, the spatial and product differentiated nature of housing makes it difficult to compare prices across time and space. The desire to make such comparisons has spurred substantial activity in creating constant quality price indices by location (Hill, Knight & Sirmans, 1997). Fourth, hedonic pricing models use housing data to estimate the costs and benefits associated with such items as pollution, growth controls, and tax policies (Case, Rosen & Hines, 1993; Brueckner, 1997). Fifth, as a house comprises a large fraction of an individual's wealth, a number of parties follow local price forecasts. Collectively, these applications involve the goals of accurate prediction, efficient coefficient estimation, valid inference, and the desire to understand both the temporal and spatial dependencies in prices.

For housing data, the benefits from modeling the error dependence over time are well-known and the benefits from modeling the error dependence over space are becoming better known.[1] Such benefits include more efficient asymptotic parameter estimation, less biased inference (positive correlations among errors artificially inflate *t*-statistics), and more precise predictions.

Joint modeling of errors in both time and space offers the potential for further gains.[2] However, these techniques have not been applied extensively to economic data and the best ways of specifying spatial, temporal, and spatial–temporal interactions do not appear obvi-

[1] Hill, Knight and Sirmans (1997) found annual AR(1) estimates on annual data for the autoregressive parameter of 0.54. Pace and Barry (1997a) found estimates of the spatial autoregressive parameter of 0.8536 using row-standardized weight matrices while Dubin (1988) found houses 0.5 miles apart exhibited a correlation among errors of 0.58.

[2] We have not found many applications of joint modeling of the spatial and temporal errors to housing data, although various approaches to this have been taken in different areas. For example, Pfeifer and Bodily (1990) jointly model errors to aid in the prediction of revenue for a small sample of hotels. In fact, Pfeifer and Bodily claim theirs was the only application of space–time autoregressive moving average techniques to any kind of business data (we cannot find other studies involving actual data either). Although, as they point out, these have been applied often in the physical sciences. For example, Szummer and Picard (1996) use a space–time autoregressive (STAR) model to aid in the synthesis of images of phenomenon such as moving water, fire, or other evolving textures. Deutsch and Ramos (1986) use these techniques to examine river flows. Pace, Barry, Clapp, and Rodriguez (1998) did apply a simplified version of the model developed here to a dataset from Fairfax, Virginia with over 70 000 observations. They also found great improvements from using spatial–temporal information.