Chapter LV
Genetic Programming for Spatiotemporal Forecasting of Housing Prices

Mak Kaboudan
University of Redlands, USA

ABSTRACT

This chapter compares forecasts of the median neighborhood prices of residential single-family homes in Cambridge, Massachusetts, using parametric and nonparametric techniques. Prices are measured over time (annually) and over space (by neighborhood). Modeling variables characterized by space and time dynamics is challenging. Multi-dimensional complexities—due to specification, aggregation, and measurement errors—thwart use of parametric modeling, and nonparametric computational techniques (specifically genetic programming and neural networks) may have the advantage. To demonstrate their efficacy, forecasts of the median prices are first obtained using a standard statistical method: weighted least squares. Genetic programming and neural networks are then used to produce two other forecasts. Variables used in modeling neighborhood median home prices include economic variables such as neighborhood median income and mortgage rate, as well as spatial variables that quantify location. Two years' out-of-sample forecasts comparisons of median prices suggest that genetic programming may have the edge.

INTRODUCTION

Techniques to analyze, model, and forecast spatiotemporal series are far from being established. Although statistical methods that analyze, model, and forecast time series are established, applying them to geographic or spatial data may be problematic. Analysis of spatial data using traditional econometric techniques (such as regression or maximum likelihood) may face spatial correlation, model misspecification, and spatial heterogeneity problems that jeopardize the accuracy of results. Further, advancements in spatial statistics do not offer modeling solutions. They offer measures of global spatial correlation like the Moran I and Geary's c, and of local spatial autocorrelation like G and G* (Haining, 2003).

If complex statistical modeling problems hinder analyzing spatiotemporal data, it seems logical and hopefully helpful to use techniques that circumvent statistical estimation of model parameters. This chapter examines whether use of genetic programming (GP) or artificial neural networks (ANNs) can produce applicable and capable forecasting models. The spatiotemporal variable to model and forecast is annual residential single-family home median neighborhood prices. Values of this variable are discrete time series (collected over a number of years) representing 12 neighborhoods. Model-dependent and independent variables' values associated with space (neighborhoods) are identified by i, where i = 1, …, n locations. Values collected at equally spaced time intervals are identified by t, where t = 1, …, T periods. The objective is to model and forecast the spatial univariate time series price variable P_{it}, where P_{it} is a *K* x 1 vector with *K* = n*T. A general specification to adopt is:

$$P_{it} = f(S_i, X_t, Z_{it}) \qquad (1)$$

where S_i is a set of spatial variables that vary across regions (i) but not over time (t), X_t is a set of time series variables that vary over time but remain constant across regions, and Z_{it} are spatiotemporal variables that vary over both.

It is well known that forecasting residential housing prices is important. Decisions made by lending institutions, tax assessors, and homeowners or buyers are affected by the real estate market dynamics and price predictions. Yet, accurate price-predictions remain a challenge, perhaps because models that explain variations in prices over time as well as between neighborhoods and among houses can be rather complex. Most studies that forecast prices of residential homes rely on property-address-level detailed data. Details at property level furnish housing attributes upon which hedonic pricing models have been based for decades. Applications of hedonic methods to the housing markets are plenty; see Goodman and Thibodeau (2003) for a recent application. Advances based on hedonic methods include work on geographically weighted regression models (Fotheringham, Brunsdon, & Charlton, 2002) and work on local regression (or semi-parametric) models (Clapp, Kim, & Gelfand, 2002). Bin (2004) compares parametric vs. semi-parametric hedonic regression models. Hedonic models can be viewed as representations of localized nano-detailed pricing algorithms since they focus on small-scale variations. They are designed to capture the effects of differences in housing attributes on prices. Housing attributes include quantitative and qualitative characteristics of individual properties such as age, square footage, with garage or not, with air conditioning or not, and so forth, as well as location factors that impact the price of a house (Bin, 2004; Mason & Quigley, 1996). These, however, do not explain temporal or dynamical price changes.

Dynamical changes in housing prices are determined by supply and demand forces. Determinants of demand such as income and mortgage rate change over time. Such temporal changes in determinants cause prices of all residential houses to change over time, *ceteris paribus* attributes. It is dynamical changes and not attributes that cause appraisers to adjust the price of the same exact house over time. They adjust prices to reflect a known "current" neighborhood median price while allowing for differences among properties. If this is the case, it is reasonable to model local patterns of dependency that capture the impact of large-scale variations of inter-neighborhood temporal changes in economic conditions on neighborhood median prices first. What follows are hedonic models that capture the impact of intra-neighborhood variations to account for differences in age, square footage, and so on. A neighborhood median price becomes one of the

Related Content

Unsupervised Methods to Identify Cellular Signaling Networks from Perturbation Data
Madhusudan Natarajan (2011). *Handbook of Research on Computational and Systems Biology: Interdisciplinary Applications (pp. 337-354).*
www.irma-international.org/chapter/unsupervised-methods-identify-cellular-signaling/52323

Symmetric and Asymmetric Encryption Algorithm Modeling on CPU Execution Time as Employed Over a Mobile Environment
Ambili Thomasand V. Lakshmi Narasimhan (2021). *International Journal of Natural Computing Research (pp. 21-41).*
www.irma-international.org/article/symmetric-and-asymmetric-encryption-algorithm-modeling-on-cpu-execution-time-as-employed-over-a-mobile-environment/285450

Energy Resources and Their Consumption
Harpreet Kaur Channi (2022). *Applications of Nature-Inspired Computing in Renewable Energy Systems (pp. 267-276).*
www.irma-international.org/chapter/energy-resources-and-their-consumption/294396

The Grand Challenges in Natural Computing Research: The Quest for a New Science
Leandro Nunes de Castro, Rafael Silveira Xavier, Rodrigo Pasti, Renato Dourado Maia, Alexandre Szaboand Daniel Gomes Ferrari (2011). *International Journal of Natural Computing Research (pp. 17-30).*
www.irma-international.org/article/grand-challenges-natural-computing-research/72692

Preinvexity and Semi-Continuity of Fuzzy Mappings
Yu-Ru Syauand E. Stanley Lee (2010). *International Journal of Artificial Life Research (pp. 53-61).*
www.irma-international.org/article/preinvexity-semi-continuity-fuzzy-mappings/46029