Chapter 14
Time Series Based House Sale Value Market Forecasting Using Genetically Evolved Neural Networks

George S. Atsalakis
Technical University of Crete, Greece

Kimon P. Valavanis
University of Denver, USA

Constantinos D. Zopounidis
Technical University of Crete, Greece

Dimitris Nezis
Technical University of Crete, Greece

ABSTRACT

Accurate forecasting of the house sale value market is important for individual investors, business investors, banks and mortgage companies. This chapter uses fundamentals of Genetic Algorithms (GAs) and Artificial Neural Networks (ANNs) to derive and implement a hybrid, genetically evolved feedforward ANN model that predicts next month house sale prices. Derived model results are compared with results obtained using a linear regression model and an Adaptive Neuro Fuzzy Inference System (ANFIS). The proposed model returned lower Root Mean Square Error (RMSE), Absolute Mean Error (MAE), Mean Square Error (MSE) and Mean Absolute Percent Error (MAPE) results compared with the linear regression and ANFIS models. For case studies real monthly data of USA housing prices from 1963 to 2007 were used.

INTRODUCTION

The housing market has a major impact on the overall economy. As Reichert rightly points out “...housing markets are not entirely isolated and distinct submarkets, for they in turn react to key housing trends in contiguous regions as well as to various national factors” (Meen, 1996, pp.425-446).

Shiller (1993) suggests that developing an efficient housing market is a prerequisite for sustained...
economic development. Therefore, it is essential to predict as accurately as possible housing prices for better asset allocation decisions, including mortgage underwriting.

However, estimating house sales price is a complex procedure coupled with housing sector structural changes, macroeconomic policies, tax changes, relocation, and market variations. As a result, there have been several theoretical and econometric studies attempting to model and forecast volatility in house prices, resulting in valuable insight information related to understanding such pricing dynamics. However, models already developed to analyze the housing sector are limited in reliability because they do not capture the full set of interactions with the rest of the economy.

One of the first attempts to capture housing price dynamics relates to hedonic housing price models that have been used extensively in applied economics since the seminal work of Rosen (1974). Such models examine housing markets based on the Lancastrian concept in which a commodity is treated as an ensemble of characteristics with an estimated sales price. Regardless, a major concern has been whether commonly assumed parametric specifications are adequate as hedonic price functions, that is, whether modelled characteristics through such functions reflect real market prices. Regardless, a major concern has been whether commonly assumed parametric specifications are adequate as hedonic price functions, that is, whether modelled characteristics through such functions reflect real market prices. In principle, the most appropriate hedonic price models are based on: specifications of a house price function that relates observed housing expenditures to selected structural and/ or neighbourhood characteristics influencing house prices; estimated coefficients/indices, behavioural assumptions, random error terms and econometric procedures used to define hedonic function parameters.

A second attempt focuses on Vector Auto Regression (VAR) models pioneered by Sims (1980; 1982) and popularized by Doan (1984) who demonstrated the effect of macroeconomic variables on housing prices and housing construction. VAR time series models use data to specify the dynamic structure of the model, requiring minimum information on the model structure, employing a fixed lag for all model variables.

A third attempt concentrates on repeat-sales sampled data assuming a random walk in housing prices as in Case (1987) used to construct the equivalent hedonic function characteristics, the drawback being not only possible inaccurate representation outside the sampled period, but also lack of recognition of specific house types more frequently sold.

However, existing models do not capture non-linear processes that very common in financial markets including the housing market; thus, they fail to forecast accurately values. Therefore, it is postulated that ANNs represent better solution candidates. The difference between an ANN and other approximation methods is that the ANN uses one or more hidden layers, in which input variables are squashed or transformed by a special function, known as logistic or sigmoid transformation to approximate better dynamic relations of the process model. While this hidden layer approach may seem esoteric, it represents a very efficient way of modeling nonlinear statistical processes.

In particular, in this article the focus is on utilizing hybrid genetically evolved feedforward ANN derived to capture nonlinearities in house sales market trends, where some of the input time series are non-stationary and overfitting is very likely (German, 1981; Verkooijen, 1996).

Obtained results using time series real data over a period of more than forty years demonstrate the superiority of the proposed method compared to a linear regression and ANFIS model in terms of RMSE, MAE, MSE and MAPE.

**BACKGROUND**

There are extensive literatures related to house sales price forecasting. Kim (1993) investigates the speculation possibility that caused land and housing prices to deviate from their given long-run equilibrium levels in Korea and Japan. Markusen