Economic Modelling 45 (2015) 259–267

Contents lists available at ScienceDirect

Economic Modelling

journal homepage: www.elsevier.com/locate/ecmod

Forecasting the U.S. real house price index

Vasilios Plakandaras [a], Rangan Gupta [b], Periklis Gogas [a,*], Theophilos Papadimitriou [a]

[a] Department of Economics, Democritus University of Thrace, Greece
[b] Department of Economics, Pretoria University, South Africa

ARTICLE INFO

Article history:
Accepted 30 October 2014
Available online 24 November 2014

Keywords:
House prices
Forecasting
Machine learning
Support Vector Regression

ABSTRACT

The 2006 sudden and immense downturn in U.S. house prices sparked the 2007 global financial crisis and revived the interest about forecasting such imminent threats for economic stability. In this paper we propose a novel hybrid forecasting methodology that combines the Ensemble Empirical Mode Decomposition (EEMD) from the field of signal processing with the Support Vector Regression (SVR) methodology that originates from machine learning. We test the forecasting ability of the proposed model against a Random Walk (RW), a Bayesian Autoregressive and a Bayesian Vector Autoregressive model. The proposed methodology outperforms all the competing models with half the error of the RW model with and without drift in out-of-sample forecasting. Finally, we argue that this new methodology can be used as an early warning system for forecasting sudden house price drops with direct policy implications.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction and literature review

The last global financial crisis restored a keen interest in both academic and policy circles on the role of asset prices and in particular housing prices in economic activity. As Leamer (2007) notes the housing market predicted eight of the ten post-World War II recessions, acting as a leading indicator for the real sector of the economy. In fact he goes as far as to state that "Housing is the business cycle". Vargas and Silva (2008) argue that housing price adjustments play an important role in the determination of the phase of the business cycle. When the economy booms, construction and employment in the housing sector expand rapidly to respond to excess demand, rapidly pushing nominal house prices upwards. During the contraction phase, the drop in private income reduces aggregate demand and nominal house prices. Nevertheless, nominal house prices typically fall sluggishly since householders are unwilling to lower their prices. Most of the adjustment is achieved through decreases in sales volume resulting in a drop in the construction sector and the housing based employment. Moreover, during contraction and recession real house prices fall rapidly as general inflationary trends reduce real house prices even with sticky nominal prices.

Recently, several authors reach to empirical findings that house prices can be instrumental in forecasting output (Das et al., 2009, 2010, 2011; Forni et al., 2003; Gupta and Das, 2010; Gupta and Hartley, 2013; Stock and Watson, 2003). The housing construction sector represents a large part of total economic activity expressed in the GDP. Consequently, as it reflects a large portion of the overall wealth of the economy, house price fluctuations can be an indicator of the evolution of GDP (Case et al., 2005). As it is the case with other assets, the movement of house prices can be also an indicator of the future direction of inflation (Gupta and Kabundi, 2010). Overall, accurate forecasting of the evolution path of house prices can be a useful tool both to house market participants and monetary policy authorities.

There is a vast literature regarding U.S. house price forecasting. Rapach and Strauss (2007) use an autoregressive distributed lag (ARDL) model framework, containing 25 determinants to forecast real housing price growth for the individual states of the Federal Reserve's Eighth District. They find that ARDL models tend to outperform a benchmark AR model. Rapach and Strauss (2009) extend the same analysis on the 20 largest U.S. states based on ARDL models examining state, regional and national level variables. Once again, the authors reach similar conclusions on the importance of combining forecasts of models with different lag structures. Gogas and Pragidis (2011) use the risk premium calculated as the difference between various long-term interest rates and the agents' expectations about future short-term rates as an input variable in predicting the future direction of house prices. They conclude that investors and analysts can use effectively the information provided by the interest rate risk premium today in order to estimate the probability of obtaining a below-trend S&P CS-10 index three months ahead.

Gupta and Das (2010) also forecast the recent downturn in real house price growth rates for the twenty largest U.S. states. The authors use Spatial Bayesian VARs (BVARs), based only on monthly real house price growth rates, to forecast their downturn over the period 2007:01 to 2008:01. They find that BVAR models are well-equipped in forecasting the future direction of real house prices, though they

* Corresponding author at: Department of Economics, Democritus University of Thrace, University Campus, Komotini, Greece. Tel.: +30 2531039555.
E-mail address: pgkogkas@ierd.duth.gr (P. Gogas).

http://dx.doi.org/10.1016/j.econmod.2014.10.050
0264-9993/© 2014 Elsevier B.V. All rights reserved.

significantly underestimate the decline. They attribute this underprediction of the BVAR models to the lack of any information on fundamentals in the estimation process.

Das et al. (2010) use small-scale BVARs, Bayesian Factor Augmented VARs (BFAVARs) and large-scale BVARs in forecasting house prices of nine census regions. The authors use the standard Minnesota Bayesian prior to estimate the Bayesian models. They indicate that the BFAVARs are best-suited for forecasting the house price growth rates of the nine census divisions. Gupta et al. (2011) examine the explanatory power of small and large sets of economic variables, using atheoretical models such as VAR, BVAR, FAVAR, BVAR, BFAVAR, and forward-looking structural Dynamic Stochastic General Equilibrium (DSGE) models. Based on the average root mean-squared errors for the one-, two-, three-, and four-quarters-ahead forecasts, they find that the small-scale Bayesian-shrinkage model fed with 10 variables outperforms all the other input sets. Gupta (2013) uses dynamic factor and Bayesian shrinkage models in a large number of predictors (145 variables) and forecasts of house prices for four U.S. census regions and for the aggregate economy. The results show that the BFAVAR models exhibit the best forecasting ability. Similar results were also obtained by Gupta and Kabundi (2010) and Aye and Gupta (forthcoming) using Bayesian predictive regressions in forecasting the overall U.S. house price index. More recently, Balcilar et al. (forthcoming) compared the ability of nonlinear AR models in forecasting nominal house price growth rates of four U.S. census regions and the aggregate economy, relative to an AR model. Interestingly, even though they could detect nonlinearity in the in-sample for all the 5 growth rates of house prices, when it came to out-of-sample point, interval and density forecasting, the evidence in favor of the nonlinear model was virtually non-existent.[1] In light of the above discussion, it is clear that, in general, Bayesian models are best-suited for forecasting house prices — regional or for the aggregate U.S.

In this paper we build on previous empirical studies by comparing the econometric Bayesian Vector Autoregressive (BVAR) and Bayesian Autoregressive (BAR) models, instead of Bayesian predictive regressions to avoid issues of endogeneity, with a novel forecasting methodology on one-year-ahead forecasting. We propose a methodology that combines Ensemble Empirical Mode Decomposition from the field of signal processing with the machine learning Support Vector Regression methodology for constructing forecasting models. In comparison to previous studies we use a much longer sample spanning annual observations from 1890 to 2012 and evaluate the use of 11 macroeconomic variables. Finally, we expand our research framework into multi-period ahead forecasting in an effort to evaluate our forecasting methodology as a house price early-warning system, focusing on the 2006–2007 house price downturn.

The rest of the paper is organized as follows: the methodology is discussed in Section 2, the dataset is presented in Section 3 while the empirical results are reported in Section 4. Finally, Section 5 concludes.

2. Methodology

2.1. Overview

A common issue among econometricians when developing forecasting models is that almost all time series are aggregations of more volatile elements. For instance, the annual dataset of real annual housing prices in the U.S. consists of quarterly or even monthly price surveys which are aggregated. As the sampling frequency decreases short-term dynamics diminish and long-term characteristics such as trend or seasonality are to be observed. The use of such input variables in a forecasting system incorporates error, making it difficult to create accurate models. Smoothing techniques can reduce the influence of noise and errors in observed data, offering a less volatile representation of the underlying phenomenon. Nonetheless, essential information may be lost during the smoothing process.

A key issue in all smoothing methodologies is the definition of the optimal point, beyond which smoothing results to distortion rather than noise reduction. In real applications, noise-free time series are unobservable and we can only approximate them through their smoothed counterparts. Many smoothing implementations select *ad hoc* the smoothness degree imposed; see for example the Hodrick and Prescott (1997) filter and the discussion regarding the optimum value of the smoothness parameter λ with respect to data frequency. To mitigate proper model selection dilemmas, our implementation exploits a relatively novel signal decomposition method called Ensemble Empirical Mode Decomposition (EEMD) used as a smoothing function. The main advantages of this smoothing framework are a) the absence of *a priori* assumptions and that b) the tuning of the model's parameters is solely based on data characteristics.

In this paper we propose a three step forecasting methodology as follows:

Step 1 A smoothing of the initial series employing the EEMD
Step 2 A variable selection process using the Elastic Net approach
Step 3 Fitting a Support Vector Regression model for forecasting.

In the first step we decompose all input series with EEMD into a low frequency (smoothed) and a remaining highly volatile (fluctuating) component that approximate the long-term trend and short-term dynamics,[2] respectively. By de-composing the initial time series into two distinct components we are able to use a distinct forecasting model for each of the individual components. In this way each model is better suited to the individual data characteristics. The model regarding the smoothed component of the output time series is fed with the smoothed component of the input time series. The same mechanism is used accordingly for the fluctuating component. The variable selection in every case is conducted using the Elastic Net (Zou and Hastie, 2005) and the selected variable set is forwarded to an SVR model. The proposed system yields one forecasted time series for the smoothed and one for the fluctuating component of the initial time series. Overfitting is avoided by adopting a 4-fold cross-validation scheme in the training step. The sum of the two forecasted components is the final output of our methodology and it is evaluated for its out-of-sample forecasting accuracy. An overview of the method is depicted in Fig. 1.

2.2. VAR and BVAR specifications

Following the work of Sims (1980), we create an unrestricted VAR model as follows:

$$\boldsymbol{y}_t = A_0 + A(L)\boldsymbol{y}_t + e_t, \tag{1}$$

assuming that $\varepsilon_t \sim N(0, \sigma^2 I_n)$ and I_n is an $n \times n$ identity matrix where $\boldsymbol{y}_t$ is an $n \times 1$ vector of variables, $A(L)$ is an $n \times n$ polynomial matrix in the backshift operator L with lag length p $[A(L) = A_0 + A_1 L + A_2 L^2 + \cdots + A_p L^p]$, A_0 is an $n \times n$ vector of constant terms, and e_t is an $n \times n$ vector of error terms.

VAR models use equal lag lengths for all variables in the model, which implies that the researcher must estimate many parameters, including some that may be statistically insignificant. This over-parameterization problem can create multicollinearity, leading to possibly large out-of-sample forecasting errors. Litterman (1981, 1986), Doan et al. (1984) and Spencer (1993) use a Bayesian VAR (BVAR) model to overcome the over-parameterization problem. Rather

[1] For a detailed literature review on forecasting involving the U.S. commercial and residential real-estate markets, refer to Ghysels et al. (2013).

[2] For more information about the representation of short- and long-term data characteristics during EEMD decomposition see Wu et al. (2007).

Download English Version:

https://daneshyari.com/en/article/5053956

Download Persian Version:

https://daneshyari.com/article/5053956

Daneshyari.com