Forecasting House Prices in the United States with Multiple Structural Breaks

Mahua Barari — Missouri State University

Nityananda Sarkar — Indian Statistical Institute

Srikanta Kundu — Indian Statistical Institute

Kushal Banik Chowdhury — Indian Statistical Institute

ABSTRACT

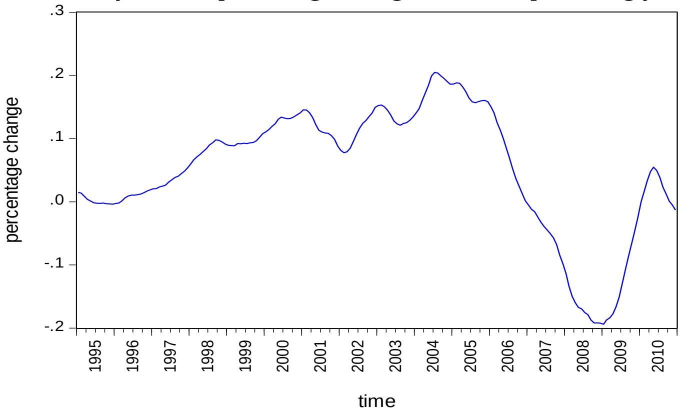
The boom-bust cycle in U.S. house prices has been a fundamental determinant of the recent financial crisis leading up to the Great Recession. The risky financial innovations in the housing market prior to the recent crisis fueled the speculative housing boom. In this backdrop, the main objectives of this empirical study are to i) detect the possibility of multiple structural breaks in the US house price data during 1995-2010, exhibiting very sharp upturns and downturns; ii) endogenously determine the break points and iii) conduct house price forecasting exercises to see how models with structural breaks fare with competing time series models – linear and nonlinear. Using a very general methodology (Bai-Perron, 1998, 2003), we found four break points in the trend in the S&P/Case-Shiller 10 city aggregate house-price index series. Next, we compared the forecasting performance of the model with structural breaks to four competing models – namely, Random Acceleration (RA), Autoregressive Moving Average (ARMA), Self-Exciting Threshold Autoregressive (SETAR), and Smooth Transition Autoregressive (STAR). Our findings suggest that house price series not only has undergone structural changes but also regime shifts. Hence, forecasting models that assume constant coefficients such as ARMA may not accurately capture house price dynamics.

Key words: *Structural Break, House Prices, Forecasting, Non-linear Models, Nonstationarity*
JEL Classifications: C13, C22, C53

1. INTRODUCTION AND LITERATURE REVIEW

Between the bottom in the 4th quarter of 1996 and the peak in the first quarter of 2006, real home prices rose 86% nationally in the United States (Shiller, 2007). However, there was a dramatic fall in house prices beginning mid-2006. While there was a slight turnaround in late 2009 and early 2010, house prices reverted back to record lows in the latter half of 2010 (see Figure 1.1).

Figure 1.1 S&P/Case Shiller 10 City Index - percentage changes from the preceding year



"Housing is the business cycle" (Leamer, 2007). What goes on in the housing sector has a significant impact on the real sector of the economy. Housing prices affect GDP growth both directly via new home construction and indirectly through changes in private household wealth, leading to changes in consumer spending (Ducca et al., 2011). Given the significant share of housing wealth in the overall private household wealth, it is not surprising that the severe downturn in the housing market ushered in the worst recession since the Great Depression of the 1930s. As the slump in the housing market continued due to the overhang of distressed and foreclosed properties, tight credit conditions, and ongoing concerns among potential borrowers and lenders about continued decline in house prices, the economic recovery process became slow and erratic (Bernanke, 2011).

In this backdrop, forecasting house prices has become even more important than ever before. But what types of forecasting models should be used? Our literature search indicates that relatively few studies have conducted house price forecasting exercises using alternative modeling techniques. The pioneering work in this respect was carried out by Case and Shiller (1989), in which they performed tests of market efficiency for the housing market using their weighted repeat sales price index for the first time. Existing studies on house price forecasting have mostly used time series models. For example, Zhou (1997) and Guirguis et al. (2005) utilized multivariate time series modeling approach, which presupposes an underlying theoretical relationship. Zhou used a Vector Error Correction (VEC) model to forecast sales and median prices of existing single family homes in the US between 1991 and 1994 using national data. He found that the predicted values of sales and prices fitted the actual data well and hence would be useful in guiding policy decisions. Guirguis et al. (2005) acknowledged that modeling house price appreciation has been a challenge for theoreticians and econometricians alike due to the strong vulnerability of the housing sector to structural changes, macro policies, regime switching, and market imperfections. They justifiably questioned the validity of constant coefficient approaches of prior studies to forecast house prices and instead first tested for parameter instability in the sub-samples using a sequence of Chow tests and Ramsey's RESET tests. Their findings confirmed coefficient instability in the house price equation. Subsequently, they applied time-varying coefficients approach to estimate GARCH, AR, Kalman filters, and VEC models from 1975-1985 and generated forecasts of house prices from 1985-1998. Based on Mean Square Forecast Error (MSFE) comparisons, a rolling GARCH model as well as a Kalman filter model with autoregressive representation outperformed the rest.

By contrast, Crawford and Fratantoni (2003), and subsequently Miles (2008), adopted a univariate time series approach with a special focus on nonlinear price dynamics in the housing market. Crawford and Fratantoni (2003) used a Markov regime switching model to capture the boom-bust cycle of the housing market. The underlying intuition addressed the price dynamics that may vary between booms and busts, resulting in discrete changes in time series properties of house prices over different cycles. They estimated the Markov model using state-level data on repeat transactions home price indices for California, Florida, Massachusetts, Ohio, and Texas. They compared the model's forecasting performance with that of ARIMA and GARCH models, and while Markov regime switching model performed better in-sample, simple linear ARIMA model generally performed better out-of-sample. Miles (2008) built upon the study of Crawford and Fratantoni (2003) by using the same state-level data. In view of the poor out-of-sample performance of Markov regime switching models, Miles (2008) employed a few other nonlinear modeling techniques, including the Threshold Autoregressive (TAR) and Generalized Autoregressive (GAR) models. He failed to find any empirical evidence for TAR effects in house price data for the sample. GAR performed generally better than ARIMA and GARCH models in out-of-sample forecasting. His general conclusion was that GAR performs substantially better than Markov switching models at forecasting house prices, particularly in states associated with high home price volatility.

In this study, we also focus on nonlinear price dynamics in the housing market but use a very different modeling technique. We explore alternatives to Markov regime switch type models since Crawford and Fratantoni (2003) findings in this regard (that Markov regime switching did not perform as well in out-of-sample compared to linear ARMA type model) were also corroborated by Bessec and Bouabdallah (2005) in a simulation based study[1]. On the other hand, while Miles' GAR modeling approach performed the best in out-of-sample forecasting exercise, such a model lacks the theoretical underpinning of the Markov model as it is primarily a data fitting technique. Instead, in the presence of extremely sharp and unprecedented upturns and downturns in the housing market, we ask if the house price series has undergone fundamental structural shifts during this period.

Our literature search indicates that the issue of structural break in time series data has been studied to some extent in the financial market literature, especially in the aftermath of the Asian or Russian financial crisis in 1997 and 1998, to analyze dynamic market linkages before and after the crisis. For example, see Andreou and Ghysels (2002), Ho and Wan (2002), Gerlach et al. (2006), Tsouma (2007), Lucey and Voronkova (2008) in this regard. However, existing literature has not looked into the issue of structural change in house price series. And yet some of the explanations offered for the recent housing crisis have made it imperative that we examine the possibility of structural breaks in house price series before conducting forecasting experiments. For example, Shiller (2007) characterized the housing boom that lasted till 2006 as a classic speculative bubble driven largely by expectations of unusually high future price increases. This speculative psychology, in turn, brought forth institutional changes in the form of proliferation of new mortgage credit institutions, deterioration of lending standards, growth of subprime loans among others. Similar views were also expressed by Bernanke (2010), Kohn (2007), and Dokko et al. (2009). In the end, the market dynamics were such that they created a vicious cycle in which the expectation of rapidly rising house prices fed mortgage credit expansion, which in turn pushed housing prices up even further until it became unsustainable (Obtsfeld and Rogoff, 2009). Hence it is worth asking whether the institutional changes that took place in the financial market in the first half of the last decade prior to the onset of the housing crisis may have fundamentally altered the time series properties of house price series.

To the best of our knowledge, our paper is the first attempt in endogenously modeling structural break in house price series. In a similar vein to Crawford and Fratantoni (2003) as well as Miles (2008), we also use a univariate time series modeling approach in this paper. However, our empirical analysis differs from Crawford and Fratantoni (2003) and Miles (2008) in the following respects: First, by concentrating on a very recent time period that encompasses the current housing crisis, we incorporate not only a period of prolonged sharp upturn but also a period of sharp downturn in house prices. Second, as Shiller (2007) observed, the last boom in the housing market differed from prior booms in that it was more of a nationwide event rather than a regional event. Therefore we use an aggregate composite

[1] Using Monte Carlo study on a wide range of specifications, Besecand Bouabdallah (2005) found Markov regime model to perform poorly in general in out-of-sample forecasting due to its failure to predict future regimes. Their findings lend empirical support to the theoretical results obtained by Dacco and Satchell (1999).

house price index instead of state-level data. Third, we perform tests for multiple structural breaks in the house price series using the recent Bai-Perron methodology (Bai and Perron, 1998, 2003) that endogenously determines break points.

Using a 10-City Composite S&P/Case-Shiller aggregate monthly seasonally adjusted house price index for the time period 1995-2010, our results indicate that the nonstationary house price series has undergone important structural changes during the sample period. Fundamental structural shifts in the series have occurred at February 2001, October 2003, April 2006, and August 2008, with the last shift coinciding with the recent housing market collapse. Hence any time series forecasting exercise that ignores the structural break possibility may run into model misspecification.

Next we compare the forecasting performance of nonstationary models (with break related information incorporated) to four competing models – namely, Random Acceleration (RA), simple ARMA as well as Self-Exciting Threshold Autoregressive (SETAR) and Smooth Transition Autoregressive (STAR) models. In view of the fact that the S&P/Case-Shiller house price index series is found to be I(2), we model the first difference in house price series to follow a random walk; i.e., the RA model. ARMA type model is most widely used in the literature for the purpose of forecasting and has become the standard specification. We apply the SETAR and STAR models as alternatives to Markov Regime Switch model to capture nonlinear price dynamics in the housing market.[2]

Comparison across alternative models using Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) criteria indicates that the nonstationary model with break in trend outperforms all other models in terms of in-sample forecasting. In that sense, it is the best fitted model for the given time series. However, the structural breaks model does not yield the best results in out-of-sample forecasting. Further, in terms of performance by the modified Diebold-Mariano test (see Harvey et al., 1997), none of the models performs significantly better than the rest. This may have been due to the fact that the house price series has undergone yet another structural change in the hold-out period during 2009-2010 which could not be investigated due to trimming considerations associated with the Bai-Perron methodology. Furthermore, while ARMA has typically outperformed the Markov in out-of-sample forecasting in the literature, we do not find empirical evidence of ARMA outperforming SETAR and STAR in our empirical study. Our overall findings clearly demonstrate that models of house prices have not remained stable during the sample period. We strongly recommend taking such nonlinearities into account when conducting forecasting exercises and formulating housing policy.

The paper is organized as follows. In Section 2, we provide a brief theoretical description of models that investigate structural breaks in time series data. In Section 3, we describe the data. We provide our empirical findings along with explanations in Section 4. Finally, we provide some concluding remarks in Section 5.

2. MODELS AND METHODOLOGY

Our primary focus in this paper is the model with structural breaks. A brief description of this model follows

[2] Miles (2008) also looked into TAR effects in the house price data but failed to find evidence of it at the state-level (for five sample state states) for the period 1979:1-2001:4.

2.1. Model with Structural Breaks

In this modeling approach, we first test for the presence of structural breaks in house price series – both stationary and nonstationary. For stationary series, we apply the Quandt (1960)-Andrews (1993) test whereas for nonstationary series, we apply the Bai-Perron (Bai and Perron, 1998, 2003) methodology. Before building these models, we carry out standard unit root tests to comment on the stationarity property of the series – namely, the augmented Dickey-Fuller (ADF) and /or Phillips-Perron tests (see, for instance, Maddala and In-Moo Kim, 1998, for details of these tests). In this context, it is worth noting that Maasoummi et al. (2010) have recently suggested another test for detecting structural breaks in a time series. This test is fairly general and robust even in the presence of multiple breaks. We have, however, carried out the Quandt (1960)-Andrews (1993) test along with the Bai-Perron test for studying multiple breaks in the house price series in this paper. This is because the latter tests have been applied widely in the literature and found to be very useful in detecting structural breaks in time series.

In what follows, we first describe briefly the Quandt-Andrews test for detecting a single endogenous structural break in a stationary time series and the subsequent works by Bai (1994, 1997a, 1997b) for estimating the break point. Next, we describe the methodology suggested by Bai and Perron (1998, 2003) for testing the presence of multiple structural breaks in a nonstationary time series.

2.1.1. The Quandt-Andrews Test for Single Structural Break

The first classical test of an exogenously given structural change in econometric literature is due to Chow (1960). He modeled it under the assumptions of i) a single break point, ii) *a priori* knowledge of exact break date, and iii) equality of error variance over the two sub-periods even when there is a break in terms of coefficient parameters. Clearly, these assumptions proved to be major limitations of the Chow test. At around the same time, Quandt (1960) also discussed the problem of testing the null hypothesis of constant coefficients against a more general alternative, where the break point is unknown and the error variance is also allowed to change. However, because of the lack of a proper distribution theory, this test could not be applied. It was only after three decades that Andrews (1993) and Andrews and Ploberger (1994) derived the asymptotic distributions of the likelihood-ratio test statistic, as well as the analogous Wald and Lagrange multiplier/Rao's score test statistics for a one-time unknown structural change. These distributions are valid for models with no deterministic or stochastic trends as well as for nonlinear models. Andrews (1993, 2003) also provided the asymptotic critical values of these distributions under the null hypothesis of no structural break. The test statistics are obtained as a function of all possible break dates. However, as noted by Hansen (2001), the break dates cannot be considered to be too close to the beginning or end of sample, because otherwise there are not enough observations to identify the sub-sample parameters. This is called trimming, and conventionally, the trimming parameter, τ, is taken to be 0.15, and thus the search is confined to the range between 15% and 85% of the observations. It is then checked to see if the maximum of this sequence of test statistic values exceeds Andrew's appropriate critical values. If it does, then we conclude that the time series has a structural break.

The Quandt-Andrews method, however, does not estimate the break point. It is the subsequent works by Bai that provide the methodology for estimating the break points. Following Bai (1994, 1997a,1997b), the sample is split at each possible break date and the parameters of the model are then estimated by ordinary least squares method and the sum of squared errors calculated. The least squares break date estimate is the date that minimizes the full-sample sum of squared errors.

2.1.2. The Bai-Perron's Test for Multiple Structural Breaks

Precursor to Bai and Perron's papers in 1998 and 2003 on testing the presence of multiple structural breaks in time series, Bai (1997b) and Bai and Perron (1998) discussed how to estimate multiple break dates sequentially. The results were obtained under very general conditions of the data and the errors, and the framework also allowed a subset of the parameters not to change. They proposed a number of test statistics for identifying multiple break points, and these are stated below.

(i) The sup $F_T(k)$ test i.e., a *supF*-type test of the null hypothesis of no structural break versus the alternative of a fixed number of breaks, T representing the sample size.

(ii) Two tests, designated by them as *UD max* test and *WD max* test, from consideration of having equal

(iii) weighting scheme and unequal weighting scheme where weights depend on the number of regressors and the significance level of the tests. For these two max tests, the alternative hypothesis is somewhat different from that in (i) *viz*., the number of breaks/changes is arbitrary/unknown, but up to some specified maximum.

(iv) The *sup* $F_T(l + 1| l)$ test i.e., a sequential test of the null hypothesis of l breaks versus the alternative of $(l + 1)$ breaks.

It should be quite obvious that size and power of these tests are important issues for final testing conclusions. Based on extensive simulation exercise, they have suggested the following useful strategy: "First look at the *UD max* or *WD max* tests to see if at least one break is present. If these indicate the presence of at least one break, then the number of breaks can be decided based upon a sequential examination of the *sup* $F(l + 1| l)$ statistics constructed using global minimizers for the break dates (i.e., ignore the test $F(1| 0)$ and select m such that the tests *sup* $F(l + 1| l)$ are insignificant for $l \geq m$). This method leads to the best results and is recommended for empirical applications" (Bai and Perron, 2003, pp. 16).

2.2. Forecasting

In order to assess the forecasting performance of the models discussed above, we compared both in-sample and out-of-sample forecasts with the actual values using standard forecast evaluation criteria. One such well known criterion is the root mean squared error (RMSE) of the forecasts defined as:

$$RMSE = \sqrt{\left(\frac{1}{T_1}\sum_{t=T+1}^{T+T_1}(y_{t+s} - f_{t,s})^2\right)}$$

where $f_{t,s}$ is the s-step ahead forecast from time t and y_{t+s} is the actual value of y_t the at time $t + s$, $T + T_1$ is the total sample size (in-sample plus out-of-sample), and $(T + 1)$-*th* observation is the first out-of-sample forecast observation so that the total hold-out sample size is T_1. Another standard criterion for evaluating forecasting performance is the mean absolute forecast error which is given by:

$$MAE = \frac{1}{T_1} \sum_{t=T+1}^{T+T_1} \mid y_{t+s} - f_{t,s} \mid$$

Insofar as generation of s-step ahead forecasts are concerned, we have used a recursive window where the series of forecasts is generated with the initial estimation date fixed. Additional observations are added one at a time to the estimation period. By these criteria, a model is said to be better than another if the RMSE/MAE value of the former is smaller than the latter.

In order to compare between any two models in terms of their forecast performances, a number of tests are available in recent literature. To this end, we performed the modified Diebold-Mariano test (*cf.* Harvey et al., 1997) involving all the models considered in our study. The test we used is a modified version of the original Diebold-Mariano (1995) test, proposed by Harvey et al. (1997). The original test specifies the null hypothesis as of equal predictive ability of two models against the alternative hypothesis that one model has a smaller RMSE than the other one. The modification is based on a correction factor designed to account for potential finite sample size distortion of the original test. We computed this modified test statistic for all possible pairs involving the five models considered in this paper.

3. DATA

Given that the housing crisis that erupted in 2006 morphed into a full blown nation-wide phenomenon, an aggregate house price index seemed appropriate. We used a 10-City Composite S&P/Case-Shiller aggregate house price index that was seasonally adjusted at monthly frequency for the time period January, 1995 - December, 2010.[3] S&P/Case-Shiller indices have become one of the most consistent benchmarks of housing prices in the US. Their purpose is to track average change in single family house prices in different geographical regions. The indices are calculated using the repeat-sales methodology, first developed by Case and Shiller (1989), which uses data on properties that have sold at least twice in order to capture the true appreciated value of constant-quality homes.[4]

There are three aggregate house price indexes that are routinely published by S&P/Case-Shiller - i) National U.S. Home Price Index is a quarterly composite of single-family home price indices for the nine U.S. Census divisions dating back to 1987, ii) The S&P/Case-Shiller monthly 10-City Composite is a value-weighted average of 10 metro area indices dating back to 1987 and iii) the S&P/Case-Shiller monthly 20-City Composite is a value-weighted average of 20 metro area indices dating back to 2000. Given our focus on the dramatic rise in house prices over the decade spanning 1996-2006 and the subsequent meltdown, we decided to use the monthly 10-City Composite index with January 1995 as the starting point.

---

[3] Data were obtained from the following site: http://www.standardandpoors.com/indices/sp-case-shiller-home-price-indices/en/us/?indexId=spusa-cashpidff--p-us----

[4] Several existing studies on house prices have used Federal Housing Finance Agency (FHFA) house price indices. While both indices use repeat sales methodology, S&P/Case-Shiller indices represent an improvement over FHFA house price indices in several aspects. For example, unlike FHFA index which is a quarterly index, S&P/Case-Shiller is a monthly index. Besides, S&P/Case-Shiller indices include foreclosed properties while FHFA indices do not. Given the significant increase in number of foreclosed properties in the wake of the current crisis, S&P/Case-Shiller indices are expected to more accurately tract the decline in house prices. Furthermore, by restricting to Fannie May and Freddie Mac conforming mortgages, FHFA indices concentrate more on the lower end of the housing markets.

Furthermore, it is worth noting that despite the difference in coverage, all three aggregate indices track each other fairly closely.

4. EMPIRICAL FINDINGS AND ANALYSIS

In this section, we first report and discuss the results of our time series analysis characterizing the data generation process of Case-Shiller house price index (P_t) , covering the entire sample period from January 1995 to December 2010. We comment on the stationarity property of (P_t) as well as the structural stability of the series using relevant testing procedures. Thereafter, we report estimated versions of the five different models discussed earlier and the respective in-sample and out-of-sample forecasts. We also provide a comparison of forecasts using the RMSE and MAE values based on these forecasts as well as the modified Diebold-Mariano test statistic. For purposes of forecasting, we take the time-period covering January 2009 to December 2010 to be the out-of-sample period; thus all the models are estimated based on in-sample observations from January 1995 till December 2008. The estimation was carried out using Eviews, and programs were written in Gauss, based on codes available on the official homepage of Dick van Dijk (http://people.few.eur.nl/djvandijk/).[5]

For the purpose of this study, the observations were changed to their logarithmic values i.e., ln $P_t = p_t$ (say). Therefore, the difference between pt and p_{t-1} constitutes house price inflation. Furthermore, by compressing the scale, the logarithmic transformation makes the distribution more akin to a symmetric distribution.

4.1. Testing for Stationarity

As is apparent from the graphical plot of the percentage in house price index (from a year earlier) in Figure 1.1, the series is not stationary. In order to verify this statistically, we carried out the Augmented Dickey-Fuller (ADF) as well as the Phillips-Perron (PP) tests on the level values of the series i.e., p_t using estimated equations with intercept and a linear trend term. The appropriate lag value for the ADF test was chosen using Schwarz's (1970) Bayesian Information Criterion (SBIC[6]). Subsequently the Ljung Box Q (.) test was carried out to make sure that the residuals have indeed become white noise.

The ADF test statistic value was obtained as –1.301. Since the critical value at 5% level of significance is - 3.438, we concluded that the null hypothesis of unit root could not be rejected for the Case – Shiller house price index, and hence the series has a unit root. Our findings based on the PP test for unit root was also the same. The PP test statistic value of 0.279 was found to be significant only at a p-value of 0.99.

The first difference of the series was considered next. The unit root tests yielded test statistic values of -2.499 and -2.283 for the ADF and PP test respectively. Since both test statistics suggested the presence of unit root in the Δp_t series as well, yet another round of differencing was done, and unit root tests were carried out once again. The estimating equation for the second differenced series is given below:

$$\Delta p_t^* = \underset{(-0.68)}{0.0002} + \underset{(0.59)}{(-2.02\times10^{-6})}\, t - \underset{(-11.44)}{0.818}\, p_{t-1}^* + \hat{a} \qquad (4.1)$$

[5] The algorithms as well as the codes are available, on request, from the third author (kundu.srikanta@gmail.com).

[6] We choose SBIC since it imposes a stiffer penalty term than Akaike's Information Criterion (AIC).

where $p_t^* = \Delta^2 p_t$, $\hat{a}_t$ is the residual of the ADF estimating equation, and the numbers in parentheses underneath denote the t-ratios.

As is evident from equation (4.1), the intercept and the coefficient of the linear trend term were not significant. The ADF test statistic value of –11.44 indicated that there was no unit root present in the second differenced values. As for the PP test, the test statistic value was found to be -11.34, which reinforced our conclusion based on the ADF test.

Figure 4.2 Plot of Case-Shiller house price index (in log value), p_t

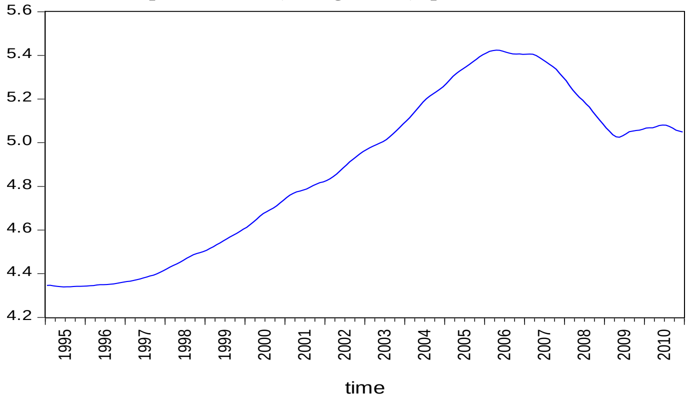


Figure 4.3 Plot of first differenced values of p_t

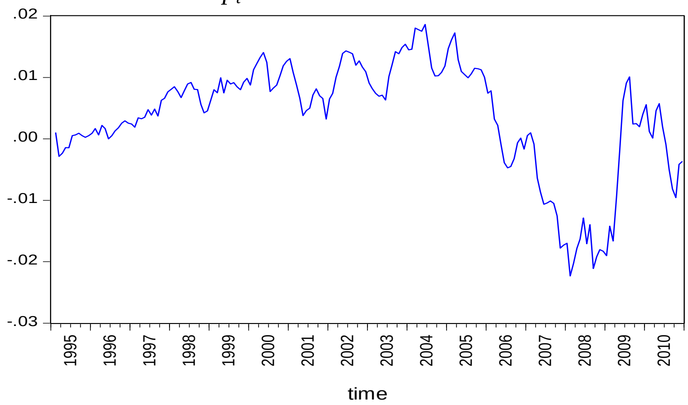
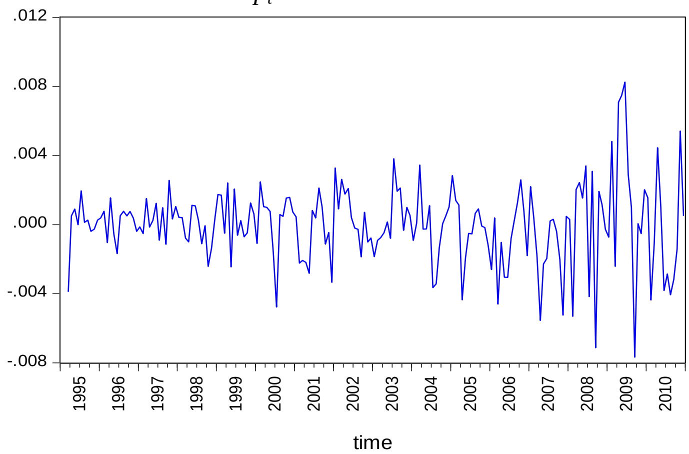


Figure 4.4 Plot of second differenced values of p_t



We, therefore, concluded, following the procedure of Box-Jenkins of achieving stationarity, that Case-Shiller house price series is integrated of order 2, i.e., it becomes stationary after second differencing. This was also evident in the graphical plots of ln P_t (= p_t), Δp_t and $\Delta^2 p_t$

against time as shown in Figures 4.2, 4.3, and 4.4 respectively. From Figure 4.1, it appeared that the house price series at level values had a rising trend until 2006, followed by a declining trend suggesting a possible break in the series sometime in 2006. The statistical testing of the presence of this break, as well as any other break, and their role in the modeling of house prices are discussed in the following sections.

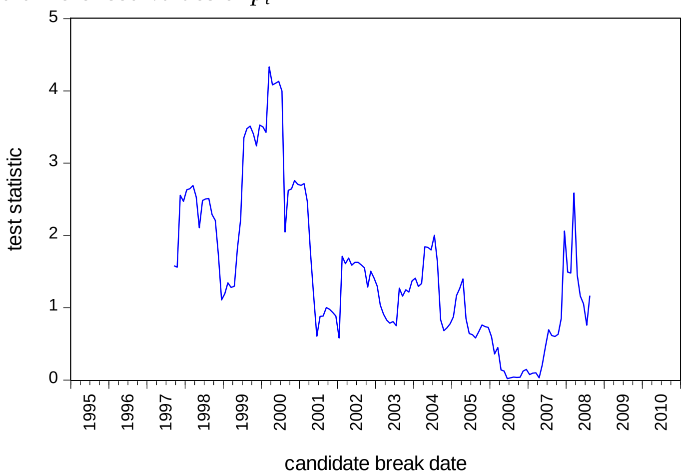
4.2. Testing for Structural Stability in the Stationary Series

Next we formally investigated if there was any structural break in the stationary house price series. To that end, we applied Quandt (1960) – Andrews (1993). As mentioned in Section 2.1.1, in this testing procedure, the null hypothesis of no structural break is tested against the alternative where a single structural break has occurred at some unknown time point, and the error variance is allowed to change from pre-break to post-break period. If the Quandt-Andrews test concluded that there existed a break in the stationary series, we determined the break points endogenously by applying Bai's (1994, 1997a, 1997b) least squares based procedure.

For the purpose of carrying out the Quandt-Andrews test for parameter stability, we first considered an AR(1) model. Thereafter, we considered some higher order AR models as well viz., AR(p), $p = 2, 3, 4$ and 5. However, results of the tests were found to have hardly changed with higher lags, and hence we are reporting here the computational figures for AR(1) only[7].

Next we calculated Andrews' Wald (W) statistic to test for stability of the stationary series. For the purpose of computing a sequence of Wald statistics as a function of candidate break dates, we eliminated the first and the last 15% of the data points. A plot where values of the Wald statistic are plotted on the *Y*-axis against the candidate break dates on the *X*-axis is given in Figure 4.5.

Figure 4.5 Plot of second differenced values of p_t



It is evident from this plot that the maximum value of the sequence of Wald statistics, 4.331, lies below the Andrews' critical value of 11.72 at 5% level of significance, and hence, the null hypothesis of 'no structural break' could not be rejected. Thus, based on the Quandt-Andrews test, the US house price series, in its stationary values, was found to remain stable during the entire sample period. In other words, our findings suggest that even during the period of this

[7] Choice of AR (1) is also supported by Monte-Carlo evidence (see, for instance, Maddala and In-Moo Kim, 1998).

crisis, the stationary component of the series did not undergo any structural change or adjustment. However, it appears from Figure 4.2 that structural changes might have occurred in the trend component of the series, especially in 2006. We investigated this possibility next by applying the Bai-Perron test (1998, 2003), the modeling specification which allows for a trend component.

4.3. Testing for Structural Stability in the Nonstationary Series

As stated earlier, we applied the testing procedure proposed by Bai and Perron (1998, 2003) for finding the presence of structural breaks, including multiple ones, if that be the case, in the nonstationary time series.

The model considered by Bai and Perron (1998) is fairly general allowing for, *inter alia*, trending regressor so that the test can be carried out with nonstationary data having a deterministic trend as well. For the purpose of our study, we considered a similar model, where the regressors for p_t comprised a constant term, a time trend and the first lagged value of p_t, apart from the noise term, a_t i.e.,

$$p_t = \eta + \delta t + \alpha_1 p_{t-1} + a_t \tag{4.2}$$

While applying this test, we set the value of the trimming parameter, τ, to equal 0.15. As described in Section 2.1.2, we first carried out, as per suggestion by Bai and Perron (1998), the *UD max* and *WD max* tests. Their test statistic values were found to be 405.502 and 518.169, respectively. These were compared with their respective critical values of 11.70 and 12.81 at 5% level of significance, leading us to conclude that at least one structural break is present in the nonstationary i.e., trended series. We then performed the sequential $sup F_T(l + 1|\ l)$, test and the test statistic values were obtained as 80.85, 31.45, 15.06, and 0.0001 for $F_T(2|1)$, F_T (3|2), F_T (4|3), and F_T (5|4), respectively. The comparison with the critical values of 12.95, 14.03, 14.85, and 15.29 at 5% level of significance suggested the presence of four breaks in the nonstationary series. Finally, the four break points were estimated following the procedure proposed by Bai and Perron (1998, 2003), and these were found to be February 2001, October 2003, April 2006, and August 2008, respectively. In search of an intuitive explanation for the break dates, we note that the first break occurred around the time the US economy dipped into a mild recession in 200,1 which was preceded by the dot com bust. The subsequent recovery was fairly anemic, which prompted the Fed to engage in an aggressive rate cut from 6.5% in late 2000 to 1% in June 2003 to accelerate the pace of recovery. The housing market grew vigorously during this period till the peak was reached in summer 2006. Finally, the last break date effectively coincided with the collapse of the housing market in 2008.

4.4. Estimated Models

In this section, we present the five estimated models. As stated earlier, the estimation was carried out based on data covering the period January 1995 to December 2008.

4.4.1. Model with Breaks in Trend

Our model with structural breaks is estimated with the level values, p_t. As discussed before, the Bai-Perron test, in the framework of a model where break is captured through the deterministic trend function, produced four break points in the entire series. Therefore, we estimated a model using four dummy variables for each of the intercept and slope parameters along with a sufficient number of lagged values of[8] p_t. The following best fitted model was thus obtained:

$$\begin{aligned} p_t &= \underset{(1.146)}{0.027} - \underset{(-2.684)}{0.002}\, D_{1t} + \underset{(2.900)}{0.003}\, D_{2t} - \underset{(-2.050)}{0.002}\, D_{3t} - \underset{(-1.461)}{0.003}\, D_{4t} + \underset{(3.499)}{0.001}\, t \\ &+ \underset{(0.137)}{5.43\times 10^{-6}}\, K_{1t} - \underset{(-3.583)}{0.0002}\, K_{2t} - \underset{(-2.800)}{0.0004}\, K_{3t} + \underset{(2.025)}{0.002}\, K_{4t} + \underset{(21.708)}{1.780}\, p_{t-1} - \underset{(-3.987)}{0.669}\, p_{t-2} \\ &- \underset{(-1.782)}{0.316}\, p_{t-3} - \underset{(-0.156)}{0.029}\, p_{t-4} + \underset{(1.726)}{0.331}\, p_{t-5} - \underset{(-1.019)}{0.206}\, p_{t-6} + \underset{(1.627)}{0.339}\, p_{t-7} - \underset{(-2.293)}{0.443}\, p_{t-8} + \underset{(2.308)}{0.206}\, p_{t-9} + \hat{a}_t \end{aligned} \tag{4.3}$$

[The values in parentheses indicate corresponding values of t-ratios.]

where D_{it}, $i = 1, 2, 3, 4$ stands for the i-th break dummy for the intercept i.e., D_{it} takes the value 1 if the observation falls in the sub-group of observations as characterized by the i-th dummy. Similarly, K_{it} stands for the corresponding slope dummy variables defined as $K_{it} = t - T_i$, where T_i is the i-th break point. From the diagnostic tests of the residuals of this estimated model, it was found that the residuals have become white noise. A few p-values of the $Q(k)$ test statistic are given here to lend support to the conclusion. For instance, for $k = 1, 7, 13, 19, 25$ and 31, the p-values were obtained as 0.98, 0.95, 0.09, 0.34, 0.27 and 0.43.

4.4.2. Random Acceleration Model

This is a very simple model to estimate whereby the stationary series $\tilde{p}_t$ (i.e. the second difference in p_t) was regressed on a constant and an error term. In other words, this is a random walk model in first difference and was particularly chosen to characterize a house price series that is found to be $I(2)$. The constant, estimated as -0.001, was found to be statistically insignificant with a p-value of 0.425, which should indeed be so since the constant in the ADF estimating equation given in equation (4.1), was found to be insignificant.

4.4.3. ARMA Model

Since the time series of the Case-Shiller house prices was found to be $I(2)$, the series was differenced twice so as to obtain the stationary series, and then the ARMA model was fitted on the stationary series, $\tilde{p}_t$ say. The best fitted ARMA model for the stationary series was found to be an ARMA (2,2) model, as given by:

$$\tilde{p}_t = -8.28\times 10^{-5} + \underset{(3.001)}{0.178^*}\, \tilde{p}_{t-1} - \underset{(-12.529)}{0.717^*}\, \tilde{p}_{t-2} - \underset{(-8.244)}{0.153^*}\, \hat{a}_{t-1} + \underset{(78.670)}{0.983^*}\, \hat{a}_{t-2} + \hat{a}_t \tag{4.4}$$

where $\tilde{p}_t = \Delta p_t - \Delta p_{t-1}$. [The values in parentheses indicate corresponding values of t-ratios. * indicates significance at 1% level of significance.]

The orders of the ARMA model were obtained by following the Schwarz's (1978) BIC criterion. The highest orders of the ARMA model were taken to be the usual (2,2). The values of the BIC criterion for all possible combinations of the orders are given in Table 4.1 below. The minimum BIC value was attained at orders (2,2). Further, the Ljung-Box Q(.) statistic, based on the residuals of this model, suggested that there were no significant autocorrelations left in the residuals at 1% level of significance[9]. These test statistic values along with p-values are reported in the second and third columns of Table 4.2 below.

[8] Since the data are at monthly frequency, we started with 12 lags of p_t and checked for the white noise property of the residuals thereof. It was at lag value 9 that the residuals were found to be white noise, and the estimated model is accordingly reported in equation (4.3).

[9] Since residuals of the ARMA (2,2) model in equation (4.4) turned out to be white noise, any other higher orders for ARMA model were not considered.

	MA(q)		
AR(p)	0	1	2
0	-9.694	-9.668	-9.688
1	-9.688	-9.658	-9.678
2	-9.680	-9.652	**-9.722**

Table 4.1 Values of the BIC for ARMA(p,q) model.

	Residual	
Lag value	**Test statistic**	***p*-value**
1	2.881	0.410
2	9.350	0.053
3	10.991	0.052
6	14.264	0.075
12	21.613	0.087
18	28.984	0.088
24	32.917	0.165

Table 4.2 The Ljung-Box test statistic and p-values based on residuals of the ARMA (2,2) model.

4.4.4. SETAR Model

Based on our literature search, we have found that threshold autoregressive (TAR) models have not been applied to forecast house prices [except for Miles (2008) who failed to find any threshold effect]. Threshold autoregressive model is an entirely different class of nonlinear models in the sense that this is a simple relaxation of the class of linear autoregressive models, which allow a locally linear approximation over a number of states (regimes) so that globally the model is nonlinear. Tong and Lim (1980) proposed a special case of TAR model, where the state-determining variable is the variable under study itself, and in that case the model is called the self-exciting TAR or SETAR model.

While considering the SETAR model instead of a single regime linear AR model, an important question that naturally arises is whether the additional regimes add significantly to explaining the dynamic behavior of the variable under question, the house price series in our case. A natural approach to answering this question empirically is to take the single regime linear model as the null hypothesis and the regime switching SETAR model as the alternative. Therefore, we first tested for the null hypothesis of 'no threshold' against the alternative of 'threshold' before actually fitting a threshold model of the SETAR kind. To this end, the Rao's score/LM test was computed and test statistic value was found to be 10.80. Given that the distribution of the underlying test statistic is nonstandard, its critical values were obtained with bootstrap-based computations using 1000 replications and also allowing for White (1980) corrected heteroskedastic errors. The test rejected the null hypothesis with *p*-value 0.017. A two-regime SETAR model was then considered, and the estimated model was obtained as follows.

Regime I: $$\tilde{p}_t = -\underset{(0.485)}{8.28\times 10^{-5}} + \underset{(3.213)}{0.301^{*}}\ \tilde{p}_{t-1} + \hat{a}_t \qquad (4.5)$$

($\tilde{p}_{t-4} \le 0.001004$)

Regime II: $$\tilde{p}_t = -\underset{(2.62)}{0.0009^{*}} + \underset{(2.028)}{0.365^{**}}\ \tilde{p}_{t-1} + \hat{a}_t \qquad (4.6)$$

($\tilde{p}_{t-4} > 0.001004$)

[The values in parentheses indicate corresponding values of t-ratios. * and ** indicate significance at 1% and 5% levels of significance, respectively.]

The estimation technique first searches for the appropriate threshold variable, which is an appropriate lag value of the variable concerned, along with the estimate of the threshold value. These were found to be $\tilde{p}_{t-4}$ and 0.001004, respectively[10].

4.4.5. STAR Model

This is a variant of the SETAR model that was introduced by Chan and Tong (1986) and extensively explored by Terasvirta and Anderson (1992), Granger and Terasvirta (1993), and Terasvirta (1994). It is worth noting that unlike the SETAR model, where it is assumed that the border between the two regimes is given by a specific value of the threshold variable y_{t-d}, the STAR model allows for a gradual smooth transition between the different regimes. The resulting model is called the Smooth Transition Autoregressive (STAR) model (see Terasvirta, 1998, for a comprehensive review). The transition variable is often assumed to be a lagged endogenous variable i.e., $s_t = y_{t-d}$ for certain integer $d > 0$.

Following the 'two regime' interpretation, which is very common in the STAR literature, we may state that different choices for the transition function $G(s_t; \gamma, c)$ give rise to different types of regime-switching behavior. The most popular choice for $G(s_t; \gamma, c)$ is the first-order logistic function:

$$G(s_t;\gamma,c) = \frac{1}{1+\exp\{-\gamma(s_t - c)\}}, \quad \gamma > 0 \tag{4.7}$$

and the resultant model is called the logistic STAR (LSTAR) model. The parameter c in (4.7) can be interpreted as the threshold between the two regimes corresponding to $G(s_t; \gamma, c) = 0$ and $G(s_t; \gamma, c) = 1$. The parameter γ determines the smoothness of change in the value of the logistic function, and thus the transition from one regime to the other.

As in case of SETAR model, we first carried out a test for linearity versus nonlinearity in the STAR model. By comparing all possible combination of different lag values with different threshold, we observe that the lowest p-values for both the F-type test and chi-square version of the test statistic occur at $d = 4$ and lag value 1. The *p-values* for F-type test and chi-square type test are reported as 0.020 and 0.021, respectively, thus rejecting linearity against LSTAR. Thereafter the LSTAR model was estimated using conditional maximum likelihood method. The results are reported below.

$$\tilde{p}_t = \left(\underset{(0.519)}{9.053}\times 10^{-5} + \underset{(3.258)}{0.328^*}\,\tilde{p}_{t-1}\right)\left(1 - G(\tilde{p}_{t-4}; \underset{(12.37)}{18.68^*}, \underset{(4.17)}{0.0008^*})\right) + \left(-\underset{(-2.76)}{0.00076^*} - \underset{(-2.58)}{0.325^*}\,\tilde{p}_{t-1}\right)\left(G(\tilde{p}_{t-4}; \underset{(12.37)}{18.68^*}, \underset{(4.17)}{0.0008^*})\right) + \hat{a}_t \tag{4.8}$$

[The values in parentheses indicate corresponding values of t-ratios. *indicates significance at 1% level of significance.]

After all of the five models were estimated, we first plotted the fitted values of the four stationary models along with the stationary (log) price series. We also plotted the fitted values of the nonstationary model and the actual (log) prices in level. These plots, shown in Figures 4.6a and 4.6b, present a visual comparison amongst the fitted models as well as the closeness of their fits in comparison with the actual prices. These plots suggest that there was not much of a difference in the performance of the five models at their level values. Moreover, since all these plots are also very close to the plot of actual values, it can be concluded that all the fitted models performed quite well for the given sample period.

---

[10] It is further to be noted that only the first lag was found to be significant for both the regimes, and that the intercept of the model for Regime II was found to be statistically significant unlike in the case of 'no threshold' i.e., single ARMA model for the entire series.

Figure 4.6a Fitted (in-sample) values of the four stationary models and the stationary data

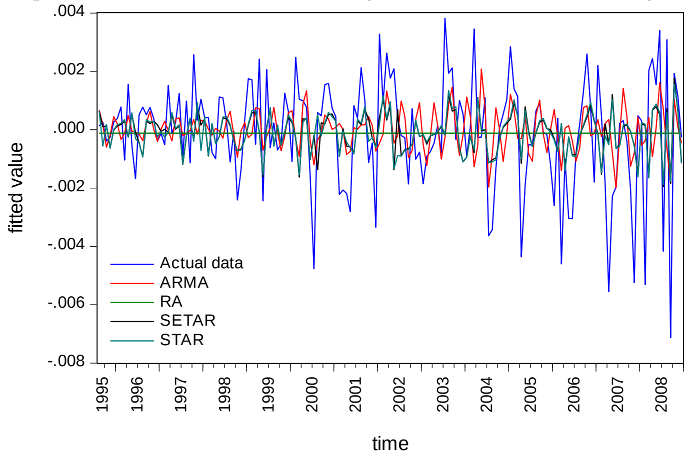
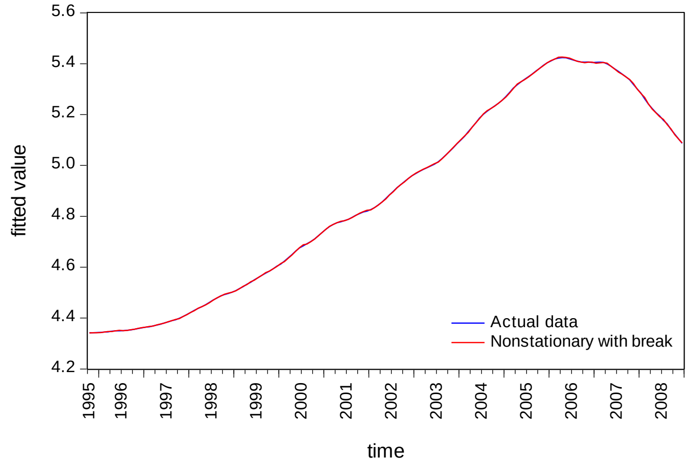


Figure 4.6b Fitted (in-sample) values of the four stationary models and the stationary data



4.5. Forecast Performance

In order to assess the performance of these five estimated models in terms of forecasts, we obtained both out-of-sample and in-sample forecasts. We calculated out-of-sample forecasts for 1-, 2-, 3-, 4-, 5-, and 6- step ahead horizons for the hold-out period ranging from January 2009 to December 2010. In the forecast computations, we applied the recursive window method where the initial estimation date is fixed, but additional observations are added one at a time to the estimation period. RMSE and MAE values were then computed based on the 24 forecasts thus obtained for all six models. The out-of-sample forecast values were subsequently plotted along with the actual house prices separately for each forecast horizon. These plots are given in Figures 4.7 through 4.12.

We report the RMSE and MAE values for in-sample forecasts in Table 4.3 and out-of-sample forecasts in Table 4.4. We also report the values of the modified Diebold-Mariano (MDM) test statistic in Table 4.5, where the comparison in terms of forecast performance of each of the four models namely, RA, SETAR, STAR, and the model with trend break, were made with reference to the ARMA (2,2) model[11].

[11] For brevity of space, the test statistic values for the other pairs of models are not reported. But the conclusion is the same.

Figure 4.7 One-step-ahead out-of-sample forecasts by the five models.

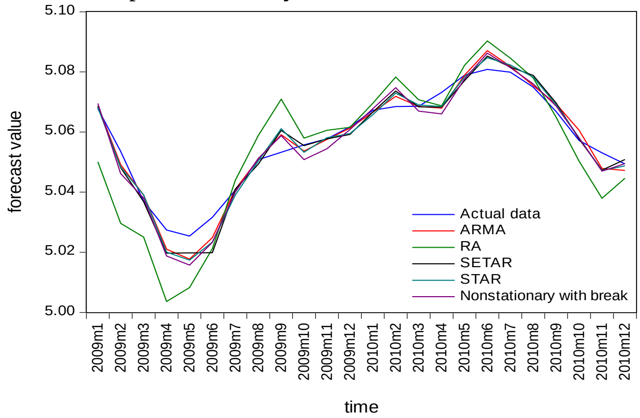


Figure 4.8 Two-step-ahead out-of-sample forecasts by the five models

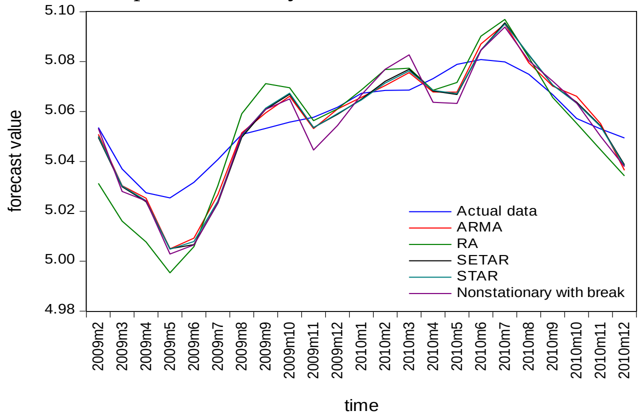


Figure 4.9 Three-step-ahead out-of-sample forecasts by the five models

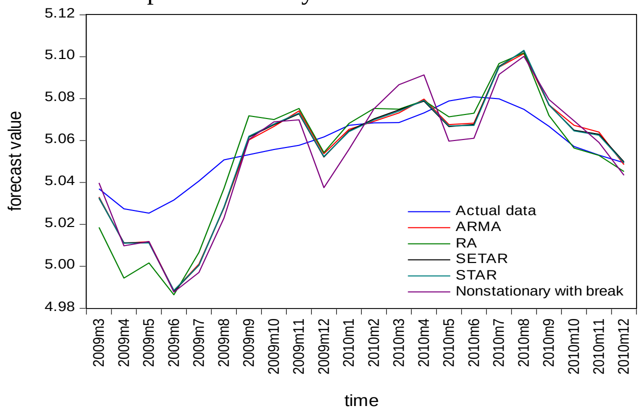


Figure 4.10 Four-step-ahead out-of-sample forecasts by the five models

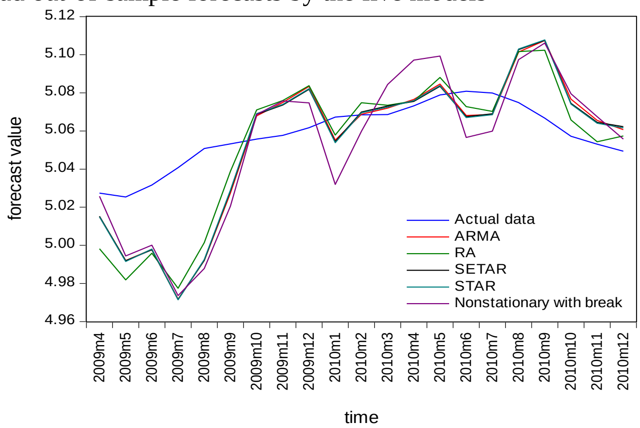


Figure 4.11 Five-step-ahead out-of-sample forecasts by the five models

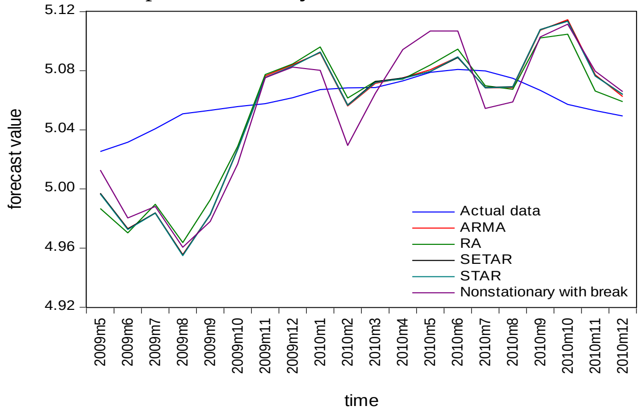
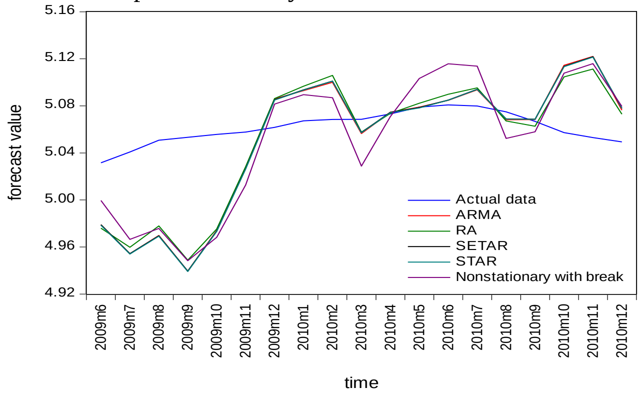


Figure 4.12 Six-step-ahead out-of-sample forecasts by the five models



Model	RMSE	MAE
M1	0.001499	0.001141
M2	0.001871	0.001401
M3	0.001735	0.001294
M4	0.001741	0.001265
M5	0.001744	0.001298

Table 4.3 In-sample forecast performance
Notes: M1: Model with breaks in trend; M2: Random Acceleration; M3: ARMA (2,2) ; M4: SETAR; M5:STAR

Comparisons of RMSE and MAE values presented in Table 4.3 and Table 4.4 suggest that in-sample forecasts perform very well for all the five models, while the out-of-sample forecast performances of all the models, in terms of these two criteria, are also remarkably good. In addition, a careful inspection of Table 4.3 values indicates that the model with structural breaks outperforms all the competing models in terms of in-sample forecasting. In this sense, the structural breaks model is the best fitting model for the sample. As regards out-of-sample forecasting, Table 4.4 values indicate that SETAR and STAR exhibit identical performance with ARMA model using RMSE and near identical performance using MAE for all forecast horizons. It is also noted from the values of the MDM test statistic that for all possible pairs based on the five models, none of the models performs significantly better than any other model. Our results in this regard stand in contrast with Crawford and Fratantoni (2003) who found simple ARIMA model to always outperform GARCH and regime switching models in out-of-sample forecasts. Based on the RMSE and MAE values in Table 4.4, the structural breaks model does not yield the best out-of-sample forecasts. A probable reason for this could be the fact that toward the end of the series covering the hold-out period (January 2009-December 2010), the model appeared to have undergone yet another structural change (see Figure 1.1). Because of consideration of trimming associated with the Bai-Perron (1998) methodology, the entire hold-out period could not be considered as candidate breakpoints.

Hence, the possibility of finding another structural break in the hold-out period was simply ruled out by this test. As it turns out, the RA model provides the best out-of-sample forecasts for longer term horizons (namely 4, 5, and 6 step ahead). A plausible explanation could be that the forecast errors of the RA model are not cumulated as forecast horizon increases, unlike the other models.

Forecast			**RMSE**					**MAE**		
horizon	M1	M2	M3	M4	M5	M1	M2	M3	M4	M5
1	0.005	0.011	0.004	0.004	0.004	0.004	0.009	0.003	0.003	0.003
2	0.011	0.014	0.010	0.010	0.010	0.009	0.011	0.007	0.008	0.008
3	0.020	0.018	0.017	0.017	0.017	0.017	0.014	0.013	0.014	0.014
4	0.031	0.026	0.027	0.027	0.027	0.026	0.020	0.021	0.021	0.021
5	0.041	0.036	0.039	0.039	0.039	0.034	0.028	0.029	0.029	0.029
6	0.052	0.047	0.051	0.051	0.051	0.043	0.037	0.038	0.038	0.038

Table 4.4 Out-of-sample forecast performance
Notes: M1: Model with breaks in trend; M2: Random Acceleration; M3: ARMA (2,2) ; M4: SETAR; M5:STAR

Forecast horizon	**ARMA *versus* RA**	**ARMA *versus* SETAR**	**ARMA *versus* STAR**	**ARMA *versus* nonstationary model with break**
1	1.293 (0.104)	0.592 (0.280)	1.012 (0.161)	2.522 (0.009)
2	0.640 (0.264)	0.656 (0.259)	0.504 (0.310)	0.957 (0.174)
3	0.386 (0.352)	0.507 (0.309)	0.555 (0.292)	0.547 (0.295)
4	0.186 (0.427)	0.437 (0.333)	0.599 (0.278)	0.179 (0.430)
5	-0.088 (0.535)	0.122 (0.452)	0.944 (0.178)	-0.099 (0.539)
6	-0.298 (0.615)	-0.547 (0.704)	0.649 (0.262)	-0.167 (0.565)

Table 4.5 Values of the modified Diebold-Mariano test statistic
Notes: *p*-values are reported in parentheses

5. CONCLUDING REMARKS

The boom-bust cycle in U.S. house prices has been a fundamental determinant of the recent financial crisis leading up to the Great Recession. The risky financial innovations that took place in the housing market prior to the recent crisis fueled the speculative housing boom. In this backdrop, the main objectives of this empirical study were to i) detect the possibility of multiple structural breaks in the US house price series data for this recent time period exhibiting very sharp upturns and downturns; ii) endogenously determine the break points and iii) carry out house price forecasting exercises to see how the structural breaks model fares with competing time series models – both linear and nonlinear.

Our study is perhaps the very first attempt in investigating the possibility of structural breaks in house price series fundamentally altering the time series properties of the series. Using the Bai-Perron (1998) methodology that allows for multiple break points and determines them endogenously, we found four break points during the sample period in the Case-Shiller 10 city aggregate house-price index series. As noted earlier, the last break point coincided with the time period when the housing market effectively collapsed.

For the purpose of forecasting, we used the house price series that was found to be nonstationary and explicitly incorporated break related information in it. We then compared the performance of this model with four other models comprising RA, simple ARMA, SETAR, and STAR. Our findings suggest that house price series not only has undergone structural changes but also regime shifts during the sample period. Hence, models that assume constant coefficients such as ARMA may not accurately capture the house price dynamics.

Comparison of forecasts across alternative models using RMSE and MAE criteria indicated that the nonstationary model with break in trend outperformed all other models in terms of in-sample forecasting. In that sense, it was found to be the best fitted model for the given time series. The superior performance of the nonstationary model with break in trend, however, did not extend to out-of-sample forecasting. This may have been due to the fact that the model has undergone yet another structural change in the hold-out period, which could not be detected because of trimming consideration associated with the Bai-Perron (1998) methodology. We expect out-of-sample forecasting performance of the structural breaks model to improve as more data become available for the post crisis period. Finally, contrary to many existing findings, our empirical findings did not clearly establish superiority of simple ARMA model over others in out-of-sample. In fact, the TAR models performed almost as well as ARMA in our study.

Stabilizing house prices is going to be the key to a healthy economic recovery. Where are house prices heading in the US? Answering such questions will be critical in formulating a path out of the current lackluster recovery. Therefore forecasting house prices accurately has never been more important. What type of forecasting model should be used to that end? While the Random Acceleration model provided the best out-of-sample forecasts in our sample, it was only for longer term forecast horizons, and furthermore, such a model lacks theoretical underpinning. Instead, based on our findings, we strongly recommend investigating the presence of structural breaks/regime switches in house price series before formulating forecasting models. Although such models will add to complexity compared to rival linear models, they will also be able to capture the unique nonlinear properties of house price series as we found in this study. One way of extending our study to gain further insight will be to forecast house prices using a multivariate VAR framework with structural breaks.

REFERENCES

Andreou, E. and E. Ghysels (2002). Detecting Multiple Breaks in Financial Market Volatility Dynamics. *Journal of Applied Econometrics*, 17 (5), 579–600.

Andrews, D. W. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica*, 61 (4), 821–856.

Andrews, D.W. (2003). Tests for Parameter Instability and Structural Change with Unknown Change Point: A Corrigendum. *Econometrica*, 71 (1), 395–397.

Andrews, D. W. and W. Ploberger (1994). Optimal Tests when A Nuisance Parameter is Present Only Under the Alternative. *Econometrica*, 62 (6), 1383–1414.

Bai, J. (1994). Least Squares Estimation of A Shift in Linear Processes. *Journal of Time Series Analysis*, 15 (5), 453-472.

Bai, J. (1997a). Estimation of A Change Point in Multiple Regression Models. *Review of Economics and Statistics*, 79 (4), 551–563.

Bai, J. (1997b). Estimating Multiple Breaks One at A Time. *Econometric Theory*, 13 (3), 315–352.

Bai, J. and P. Perron (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66 (1), 47–78.

Bai, J. and P. Perron (2003). Computation and Analysis of Multiple Structural Change Models. *Journal of Applied Econometrics*, 18 (1), 1–22.

Bera, A. K. and M. L. Higgins (1993). A Survey of ARCH Models: Properties, Estimation and Testing. *Journal of Economic Surveys*, 7 (4), 305–366.

Bernanke, B. S. (2011). The Near- and Longer-Term Prospects for the U.S. Economy. In *Federal Reserve Bank of Kansas City Economic Symposium.* Wyoming: Jackson Hole. http://www.federalreserve.gov/newsevents/speech/bernanke20110826a.htm (accessed October 28, 2011).

Bernanke, B. S. (2010). Monetary Policy and the Housing Bubble. In *Annual Meeting of the American Economic Association Meeting*. Atlanta, Georgia, http://www.federalreserve.gov/newsevents/speech/bernanke20100103a.htm (accessed September 20, 2011).

Berndt, E. R., B. H. Hall, R. E. Hall and J. A. Hausman (1974). Estimation and Inference in Nonlinear Statistical Models. *Annals of Economic and Social Measurement*, 3 (4), 653–665.

Bessec, M and O. Bouabdallah (2005). What Causes the Forecasting Failure of Markov-Switching Models? A Monte Carlo Study. *Studies in Nonlinear Dynamics and Econometrics*, 9 (2), DOI: 10.2202/1558-3708.1171.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31 (3), 307–327.

Case, K. and R. Shiller (1989). The Efficiency of the Market for Single-Family Homes. *American Economic Review*, 79 (1), 125–137.

Chan, K. S and H. Tong (1990). On Likelihood Ratio Tests for Threshold Autoregression. *Journal of the Royal Statistical Society B*, 52 (3), 469–476.

Chow, G. C. (1960). Tests of Equality Between Sets of Coefficients in Two Linear Regressions. *Econometrica*, 28 (3), 591–605.

Crawford, G. and M. Fratantoni (2003). Assessing the Forecasting Performance of Regime-Switching, ARIMA and GARCH Models of House Prices. *Real Estate Economics*, 31 (2), 223–243.

Dacco, R. and S. Satchell (1999). Why Do Regime–Switching Forecast So Badly? *Journal of Forecasting*, 18 (1), 1–16.

Diebold, F. X. and R. S. Mariano (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics*, 13 (3), 253–263.

Dokko, J., B. Doyle, M. T. Kiley, J. Kim, S. Sherlund, J. Sim and S. Van Den Heuvel (2009). Monetary Policy and the Housing Bubble. In *Finance and Economics Discussion Series*. Washington DC: Division of Research & Statistics and Monetary Affairs, Fed. Reserve Board. http://www.federalreserve.gov/pubs/feds/2009/200949/200949pap.pdf (accessed March 15, 2011).

Ducca, J. V., D. Luttrell, and A. Murphy (2011). When Will the U.S. Housing Market Stabilize? *Economic Letter*, 6 (8), 1–4.

Engle, R. F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50 (4), 987–1007.

Gerlach, R., P. Wilson and R. Zurbruegg (2006). Structural Breaks and Diversification: The Impact of the 1997 Asian Financial Crisis on the Integration of Asia-Pacific Real Estate Markets. *Journal of International Money and Finance*, 25 (6), 974-91.

Granger, C. J. and T. Terasvirta (1993). *Modeling Nonlinear Economic Relationships*. Oxford University Press.

Guirguis, H. S., C. Giannikos and R. I. Anderson (2005). The US Housing Market: Asset Pricing Forecasts Using Time Varying Coefficients. *The Journal of Real Estate Finance and Economics*, 30 (1), 33–53.

Guirguis, H. S. and R. Vogel (2006). Asymmetry in Regional Real House Prices. *Journal of Real Estate Portfolio Management*, 12 (3), 293–298.

Hansen, B. E. (2001). The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity. *Journal of Economic Perspectives*, 15 (4), 117–128.

Harvey, D., S. Leybourne and P. Newbold (1997). Testing the Equality of Prediction Mean Squared Errors. *International Journal of Forecasting*, 13 (2), 281-291.

Ho, A. K. F. and A. T. K. Wan (2002). Testing for Covariance Stationarity of Stock Returns In the Presence of Structural Breaks: An Intervention Analysis. *Applied Economics Letters*, 9 (7), 441–47.

Kohn, D. L. (2007). Economic Outlook. http://www.federalreserve.gov/newsevents/speech/kohn20071004a.htm (accessed March 15, 2011).

Leamer, E. (2007). Housing is the Business Cycle. *NBER Working Paper* No.13428. http://www.nber.org/papers/w13428 (accessed March 20, 2011).

Lucey, B. M. and S. Voronkova (2008). Russian Equity Market Linkages Before and After the 1998 Crisis: Evidence from Stochastic and Regime-Switching Cointegration Tests. *Journal of International Money and Finance*, 27 (8), 1303–24.

Luukkonen, R., P. Saikkonen and T. Terasvirta (1988). Testing Linearity Against Smooth Transition Autoregressive Models. *Biometrika*, 75 (3), 491–499.

Maasoumi, E., A. Zaman, and M. Ahmed (2010). Tests for Structural Change, Homogeneity, and Aggregation. *Economic Modeling*, 27 (6), 1382–1391.

Maddala, G. S. and In-Moo. Kim (1998). *Unit Roots, Cointegration, and Structural Change*. Cambridge University Press.

Marquardt, D. (1963). An Algorithm for Least-Squares Estimation of Nonlinear Parameters. *SIAM Journal on Applied Mathematics*, 11 (2), 431–441.

Miles, W. (2008). Boom-Bust Cycles and Forecasting Performance of Linear and Nonlinear Models of House Prices. *Journal of Real Estate Finance and Economics*, 36 (3), 249–264.

Mills, T. C. (1999*). The Econometric Modeling of Financial Time Series*. Cambridge University Press.

Nelson, D. B. (1991). Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica*, 59 (2), 347–370.

Nelson D. B. and C. Q. Cao (1992). Inequality Constraints in the Univariate GARCH Model. *Journal of Business & Economic Statistics*, 10 (2), 229–235.

Obtsfeld, M. and K. Rogoff (2009). Global Imbalances and the Financial Crisis: Products of Common Causes. In Asia Economic Policy Conference. San Francisco: Federal Reserve Bank. http://elsa.berkeley.edu/~obstfeld/santabarbara.pdf (accessed March 25, 2011).

Quandt, R. E. (1960). Tests of the Hypothesis that a Linear Regression System Obeys Two Separate Regimes. *Journal of the American Statistical Association*, 55, 324–330.

Shiller, R. J. (2007). Understanding Recent Trends in House Prices and Home Ownership. http://www.kc.frb.org/publicat/sympos/2007/pdf/2007.09.27.shiller.pdf (accessed March 2, 2011).

Terasvirta, T. (1998). *Modeling Economic Relationships with Smooth Transition Regressions*. In *Handbook of Applied Economic Statistics*, New York: Marcel Dekker Publisher, 507-552, ED. A. Ullah and D. E. Giles.

Terasvirta, T. (1994). Specification, Estimation, and Evaluation of Smooth Transition Autoregressive Models. *Journal of the American Statistical Association*, 89 (425), 208–218.

Terasvirta, T. and H. M. Anderson (1992). Characterizing Nonlinearities in Business Cycles Using Smooth Transition Autoregressive Models. *Journal of Applied Econometrics*, 7 (S), 119–136.

Tong, H. and K. S. Lim (1980). Threshold Autoregressions, Limit Cycles, and Data . *Journal of the Royal Statistical Society*, B 42 (3), 245–292.

Tsouma, E. (2007). Stock Return Dynamics and Stock Market Interdependencies. *Applied Financial Economics*, 17 (10), 805–25.

White, H. (1980). A Heteroskedasticity- Consistent Covariance Matrix Estimator and A Direct Test for Heteroskedasticity. *Econometrica,* 48 (4), 817–838.

Zhou, Z. (1997). Forecasting Sales and Price for Existing Single-Family Homes: a VAR Model with Error Correction. *Journal of Real Estate Research*, 14 (2), 155–167.