Forecasting House Price in the United States: A Time Series Study with Focus on Multiple Structural Breaks

by

Mahua Barari*, Nityananda Sarkar**,
Srikanta Kundu** and Kushal Banik Chowdhury**

---

* Corresponding author – Department of Economics, Missouri State University, Springfield, MO 65897
e mail: mahuabarari@missouristate.edu, phone: 417-836-5171, fax: 417-836-4236

** Economic Research Unit, Indian Statistical Institute, Kolkata 700108, India

Forecasting House Price in the United States: A Time Series Study with Focus on Multiple Structural Breaks

Abstract

Despite significant impact of the housing sector on the real sector of the economy, relatively few studies have conducted house price forecasting exercises using alternative modeling approaches. The main objective of this paper is to forecast house prices in the United States for a very recent time period that encompasses the ongoing slump in the housing market. Given the severity of the crisis and that it has been a nation-wide phenomenon; we investigate the possibility of structural breaks in aggregate house price series. Applying Bai-Perron methodology (See Bai-Perron 1998, 2003) to the 10-city composite S&P/Case-Shiller monthly house price index during 1995-2010, we detect four (endogenously determined) structural breaks occurring in the nonstationary component of the series. Next we compare forecasting performance of the nonstationary model with four breaks in trend to five alternative models – namely ARMA, ARMA-Exponential GARCH, Random Acceleration (RA) and two regime switching models. Comparison across models using root mean square error (RMSE) and mean absolute error (MAE) criteria indicate that nonstationary model with break outperforms the rest in in-sample forecasting exercise making it the best fitting model for the sample. However, it does not perform as well in out-of-sample forecasting using 2009-2010 as the hold-out period. Furthermore, our study does not establish a clear superiority of a linear ARMA type model over competing nonlinear regime switch models in out-of-sample forecasting. For policy purposes, we recommend investigating structural breaks/regime switches in the house price series before formulating forecasting models.

Keywords: structural break, house prices, forecasting, nonlinear models, nonstationarity

1. Introduction

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 slight turnaround in late 2009 and early 2010, house prices reverted back to record lows in the latter half of 2010 (see Figure 1).

[Figure 1 here]

Housing is the business cycle (Ed 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 continues due to 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 has become slow and erratic (Bernanke 2011).

In this backdrop, forecasting house prices has become even more important than ever before. Yet 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 that 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. 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 their sub-samples using a sequence of Chow tests and Ramsey's RESET tests.

By contrast, Crawford and Fratantoni (2003) and subsequently Miles (2008) adopted a univariate time series approach with special focus on nonlinear price dynamics in the housing market. Crawford and Fratantoni used a Markov regime switching model to capture the boom-bust cycle of the housing market. The underlying intuition is that price dynamics may vary between booms and busts resulting in discrete changes in time series properties of house prices over different cycles. Although regime switching model performed better in-sample, simple linear Auto Regressive Integrated Moving Average (ARIMA) models still performed generally better out of sample in their study. In view of poor out of sample performance of Markov regime switching models, Miles (2008) employed several nonlinear modeling techniques including the Generalized Autoregressive (GAR) model that he applied to the same state level data used by Crawford and Fratantoni. Miles found GAR to perform the best in out-of-sample forecasting.

In this study, we also focus on nonlinear price dynamics in the housing market but use different modeling techniques. We explore alternatives to Markov regime switch type models since Crawford and Fratantoni findings in this regard that Markov did not perform as well in out of sample were also corroborated by Bessec and Bouabdallah (2005) in a simulation based study. On the other hand, while Miles' GAR modeling approach performed the best in out of sample forecasting exercise, such a model lacks theoretical underpinning of the Markov model as it is primarily a data fitting technique. In this backdrop, we apply special cases of Threshold Autoregressive (TAR) models such as the STAR and SETAR models to analyze regime shifts in the housing market. Miles is the only study we have come across that looked into TAR effects in the house price data but failed to find evidence of it at the state level (for five sample states) for the period 1979:1-2001:4. However, our sample spanning the period 1995-2010 and thus encompassing the most recent housing crisis supported nonlinearity in the aggregate house price data in the form of TAR effects. Most importantly, in the presence of extremely sharp upturns and downturns in the housing market unprecedented in history, we ask if the house price series has undergone fundamental structural shifts during this period.

Existing literature has not looked into the issue of structural change in house price series. However, 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), 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). So it is worth asking whether the institutional changes that took place in the financial market in early 2000 prior to the onset of the housing crisis may have fundamentally altered the time series properties of house price series.

Based on literature search, ours is perhaps the first attempt in modeling structural break in house price series. In a similar vein to Crawford and Fratantoni as well as Miles, we also use a univariate time series modeling approach in this paper. However, our empirical analysis differs from Crawford& Fratantoni and Miles in the following respects. First, by concentrating on a very recent time period that encompasses the current housing crisis, we incorporate a period of extremely sharp upturns and downturns 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 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 (1998, 2003) that endogenously determines break points.

Using a 10-city composite S&P/Case-Shiller aggregate monthly house price index for the time period 1995-2010, our results indicate that house prices have undergone structural changes and regime shifts 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 five additional models – namely Random Acceleration (RA), simple ARMA, ARMA-Exponential GARCH (EGARCH) as well as Self-Exciting Threshold Auto Regressive (SETAR) and Smooth Transition Threshold Auto Regressive (STAR). 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 and ARMA-GARCH type models are widely used in the literature for the purpose of forecasting and have become standard specifications. As discussed above, we apply the SETAR and STAR models to capture nonlinear price dynamics in the housing market as an alternative to Markov Regime Switch model. A brief theoretical description of these competing models is given below in Section 3.

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 nonstationary model with break in trend does not perform as well in out- of -sample forecasting. This may be due to the fact that the model has undergone another structural change in the hold-out period during 2009-2010 which could not be investigated due to trimming considerations associated with Bai-Perron methodology. 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 suggesting once again that house price series has not remained stable during the sample period.

The paper is organized as follows. In Section 2, we provide a review of the existing literature on house price forecasting with special focus on structural breaks. In Section 3, we provide theoretical descriptions of alternative forecasting models and justification for including them. In Section 4, we describe the data. We provide our empirical findings along with explanations in Section 5. Finally, we provide some concluding remarks in Section 6.

2. Literature Review

Several studies have underscored the prominent role the housing sector has played in modern recessions. In a comprehensive study conducted by IMF, Helbling (2003) examined the empirical regularities of housing price booms and busts in 14 industrial countries during 1970-2001 and showed that in all but one country, housing price bust led to outright recessions. Leamer (2007) demonstrated that of the ten recessions that the US has experienced since World War II, eight has been preceded by substantial problems in housing and consumer durables.[1] In a similar vein, Jannsen (2010) showed that housing slumps in the US, UK, Spain and France during 2006-2007 were all followed by exceptionally strong recessions. Furthermore, Jannsen demonstrated that the housing market downturns in these four countries were sufficient to explain to a considerable degree the global worldwide recession that followed via negative international spillover effects.

Time series techniques have been mostly used to forecast house prices. Using multivariate modeling technique, Zhou (1997) constructed a Vector Auto Regressive (VAR) model with error correction to forecast sales and median prices for existing single family homes in the US between 1991 and 1994 using nation-wide 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) challenged the constant coefficient approach adopted by prior studies and instead carried out tests for parameter stability in sub-samples. The statistical results obtained by running a sequence of Chow and Ramsey's RESET tests confirmed coefficient instability in house price equation. Subsequently they

[1] To explain why housing sector plays such an important role in determining the onset of recession, he argued that housing has a volume cycle more than a price cycle. "The sluggishness of price adjustments is what makes the volume cycle so extreme and what makes housing so important in recessions" (Leamer 2007, pp. 178).

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.

Using univariate time series modeling technique, Crawford and Fratantoni (2003) and Miles (2008) focused on the nonlinear price dynamics in the housing market stemming from asymmetric price adjustment over boom and bust cycles. Crawford and Fratatoni (2003) applied Markov's regime switching model to state level data on repeat transactions home price indices for California, Florida, Massachusetts, Ohio and Texas and compared its forecasting performance with that of ARIMA and GARCH models. While Markov regime switching model performed better in sample, simple linear ARIMA model generally performed better out-of sample.[2] Miles (2008) built upon Crawford and Fratantoni study by using same state level data. In view of poor performance of Markov model in out-of-sample forecasting, he explored alternative nonlinear models including TAR and GAR models. He failed to find any empirical evidence for TAR effects in house price data for the sample. GAR was found to perform 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 those states associated with high home price volatility.

None of the existing studies, however, tested for the presence of structural breaks in house price series. For time series data spanning over a long time period, it is hard to justify the assumption of fixed coefficient estimates as was observed by Guirguis et al. (2005). Time series properties of a series could change drastically from one sub-period to the next due to large scale exogenous shocks. For example, the major institutional changes that took place in the financial markets prior to the onset of the current crisis could be one such shock. These institutional changes facilitated the explosion of exotic high risk mortgage products and in turn, may have fundamentally altered the time series properties of house price series.

Earliest attempt in testing for structural break was made by Chow (1960) who 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

[2]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).

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 Chow test, and subsequent research proceeded by relaxing these assumptions. Quandt (1960) proceeded by assuming that break point is unknown and the error variances change in the presence of break. However, Quandt also considered the possibility of only a single break in the series. The major limitation of Quandt procedure lay in characterizing the distribution of the test statistic under the assumption of an unknown break date. Solution was provided by Andrews three decades later (Andrews, 1993) where he derived the distribution of Quandt test statistic as well as computed the critical values. However, Quandt & Andrews still considered only a single break possibility in the series and their methodology proved inadequate to estimate the break date. Bai (1997a) finally derived the asymptotic distribution of the break point estimator. As regards the issue of testing for the presence of multiple breaks, pioneering research in this area has been carried out by Bai & Perron (1998, 2003). Their methodology is by far the most general procedure available for modeling structural breaks in a time series since it can test for any number of break points and also provides for estimation of such points.

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. On the other hand, using a multivariate framework for US stock returns, Hartmann et al. (2008) argued in favor of using publicly available information on economic and financial crises to detect structural breaks in the link between stock returns and macroeconomic predictor variables in order to improve out-of-sample predictability of stock returns. As regards housing market, Kim et al. (2007) found evidence of a single structural break in a multivariate time series model linking returns on 'real estate investment trusts' (REITs) to returns on equities and other related macroeconomic variables for the period 1971-2004. We may, therefore, end this section by noting that the issue of structural change in house price series has not been investigated in any detail. Yet, as explained above, the recent housing crisis makes a strong case for the possibility of a structural change in the housing market, fundamentally altering the time series properties.

3. Models and Methodology

Forecasting exercise was carried out using six different models. These are: Autoregressive Moving Average (ARMA) Model, Random Acceleration (RA) Model, ARMA – Generalized Autoregressive Conditional Heteroskedastic (GARCH) model, Self-Exciting Threshold Autoregressive (SETAR) model, Smooth Transition Autoregressive (STAR) model, and a model with structural breaks in the trend function. These models were chosen based on consideration of models already used by other researchers in analyzing house prices as well as incorporation of nonlinear dependences and structural breaks in the analysis. It may be stated that the last two aspects, namely, nonlinearity and structural breaks seem to be especially relevant for modeling house prices since the recent global economic crisis has had profound impact on economies all over the world. A brief description of each model follows along with justifications for choosing it.

3.1 ARMA Model

In the univariate stationary time series literature, a series is modeled only in terms of its own past values as well as of present and past values of the noise term. To this end, the most popular model is the autoregressive moving average (ARMA) model whose special cases are the autoregressive (AR) and moving average (MA) models. An ARMA model of orders *p* and *q* for a stationary time series $\{y_t\}$, denoted by ARMA (*p*, *q*), is represented as

$$y_t - \Phi_1 y_{t-1} - \ldots - \Phi_p y_{t-p} = a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q} \tag{3.1}$$

where $\{a_t\}$ is a white noise process with $E(a_t) = 0$ and $V(a_t) = \sigma_a^2$ for all *t* and there are restrictions on the coefficient parameters so that the conditions of stationarity and invertibility of $\{y_t\}$ hold.

Estimation of the parameters of an ARMA (*p*, *q*) model, and also the procedure of choosing appropriate values of the orders, *p* and *q*, are very well-known (see, for instance, Mills (1999), for details).

3.2 Random Acceleration Model

A time series $\{y_t\}$ is said to follow a random acceleration model if Δy_t defined as $\Delta y_t = y_t - y_{t-1}$ follows a random walk model with a drift i.e.,

$$\Delta y_t = c + \Delta y_{t-1} + a_t \tag{3.2}$$

where c is the drift parameter and a_t is white noise with mean zero and variance σ_a^2. While analyzing quarterly aggregate US house price index series spanning 1975-2009, this model was used by Larson (2010) as a naïve baseline model for comparison purposes since it contains no estimated parameters.

3.3 GARCH and EGARCH Models

In his seminal paper, Engle (1982) introduced the class of autoregressive conditional heteroskedastic (ARCH) models to capture the volatility prevalent in a time series. He applied this model to estimate the variance of inflation in the United Kingdom. The original ARCH model was later generalized by Bollerslev (1986) and this is known as the generalized ARCH (GARCH) model. Since its introduction, this class of models has been applied in countless empirical studies, especially those involving financial market data, due to their success in modeling persistence and volatility which are typically present in financial variables.

An AR (k)-GARCH (p, q) model for a stationary time series $\{y_t\}$ is represented as

$$y_t = \Phi_0 + \Phi_1\, y_{t-1} + \Phi_2\, y_{t-2} + \cdots + \Phi_k\, y_{t-k} + \varepsilon_t$$

where $\varepsilon_t \big| \Psi_{t-1} \sim N(0, h_t), \Psi_{t-1} = \{y_{t-1}, y_{t-2}, \cdots\}$ is the information set at $t-1$ and

$$h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \cdots + \alpha_q \varepsilon_{t-q}^2 + \beta_1\, h_{t-1} + \cdots + \beta_p\, h_{t-p}\,. \tag{3.3}$$

All the roots of $\Phi(B) = 1 - \Phi_1 B - \cdots - \Phi_k B^k$ lie outside the unit circle so that the time series $\{y_t\}$ is stationary. To ensure positivity of the conditional variance, h_t, the following conditions on the parameters in (3.3) namely, $\alpha_0 > 0$ and $\alpha_i \geq 0$ for $i = 1, 2, \cdots, q$ and $\beta_i \geq 0$ for $i = 1, 2, \cdots, p$, are required. Weaker sufficient conditions are also available (see Nelson and Cao (1992), for details).

Several alternative specifications of volatility have also been found to be very useful. One such useful specification is known as the exponential GARCH (EGARCH) model. This model was proposed by Nelson (1991) so as to capture what is known as 'leverage effect'. This form of conditional heteroskedasticity is given by

$$\ln\, h_t = \alpha_0 + \sum_{i=1}^{q}\, \alpha_i\, f\,(\nu_{t-i}) + \sum_{i=1}^{p}\, \beta_i \ln(h_{t-i})$$

where $f(\nu_t) = \theta \nu_t + \eta [| \nu_t | - E | \nu_t |]$ and $\nu_t = \frac{\varepsilon_t}{\sqrt{h_t}}$ is independently and identically (normally, as per assumption in (3.3)) distributed with mean zero and variance one. Thus, $ln\ h_t$ can be expressed as

$$\ln h_t = \alpha_0^* + \sum_{i=1}^{q} \eta_i^* |\nu_{t-i}| + \sum_{i=1}^{q} \theta_i^* \nu_{t-i} + \sum_{i=1}^{p} \beta_i\ ln(h_{t-i}) \qquad (3.4)$$

where $\alpha_0^* = \alpha_0 - \sum_{i=1}^{q} \alpha_i\ \eta\ E|\nu_t|$, $\eta_i^* = \alpha_i \eta$, $\theta_i^* = \alpha_i \theta$.

It is easy to note that $f(\nu_t)$ is independent with zero mean and constant finite variance. It may also be noted that unlike GARCH, EGARCH does not require any non-negativity restrictions on the parameters involved in h_t .

Estimation of GARCH and EGARCH models along with a conditional mean specification of the ARMA kind can no longer be done by straightforward application of the ordinary least squares (OLS) method since the volatility model is not of the usual linear form. Estimation of the models is carried out by applying the maximum likelihood (ML) method of estimation (see, for instance, Bera and Higgins (1993), for details). This involves nonlinear optimization. To that end, the BHHH algorithm, as proposed by Berndt et al. (1974), and also Marquardt (1963) algorithm are often used in standard software.

3.4 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]. This is a well-known class of nonlinear models used in describing the dynamic behavior of many economic and financial variables (see, for instance, Tong and Lim (1980), Tong (1978,1983,1990), Chan and Tong (1986) and Tsay (1989)).

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. Following Tong (1990) and Brooks (2002), we specify a general threshold autoregressive model (TAR) for a time series y_t , as

$$y_t = \sum_{j=1}^{J} I_t^{(j)} \left(\phi_0^{(j)} + \sum_{i=1}^{p_i} \phi_i^{(j)}\ y_{t-i} + \sigma_j\ a_t \right), s_{j-1} < z_{t-d} \le s_j, t = 1, 2, \cdots, T \qquad (3.5)$$

where $I_t^{(j)}$ is an indicator function for the $j-th$ regime taking the value one if the underlying variable is in state *j* and zero otherwise, z_{t-d} is an observed variable determining the switching point after some delay $d,\ s'_j\ s$ are the threshold values, a_t is an independently and identically distributed (i.i.d.) error process with zero mean and unit variance and σ_j is the standard deviation of the errors for regime *j*. If the regime changes are driven by own lags of the underlying variable, i.e., $z_{t-d} = y_{t-d}$, then the model is called the self-exciting TAR (SETAR) model. Standard conditions, as stated in Tong (1990), for stationarity of TAR/SETAR models are assumed to hold. An interesting point about these models is that the stationarity of y_t does not necessarily require the model to be stationary at each regime.

Estimation of the parameters $\theta = (\phi_i^{(j)}, s_j, d, p_i, \sigma_j)'$ of the SETAR model can be basically divided into four steps. The first step is the determination of the threshold values and this is done by a method where initially s_j is determined using a grid search procedure that seeks the minimal residual sum of squares over a range of probable values of thresholds of an assumed model.

The second step in the estimation procedure is concerned with the determination of appropriate lag lengths, and this involves the use of an approach which is conditional upon the specified threshold values. In this approach, a modified version of Akaike's information criterion (AIC), as given in (3.6) below for the two-regime case, is used.

$$AIC(p_1, p_2) = \tilde{T} \ln \hat{\sigma}_1^2 + (T - \tilde{T}) \ln \hat{\sigma}_2^2 + 2(p_1 + 1) + 2(p_2 + 1) \qquad (3.6)$$

where $\tilde{T}$ is the number of observations in regime 1, *T* is the total number of observations, p_1 and p_2 are the lag lengths, and $\hat{\sigma}_1^2$ and $\hat{\sigma}_2^2$ the residual variances, of regimes 1 and 2, respectively.

The determination of delay parameter, *d*, is carried out in the third step. This can be done in a variety of ways one of which is the use of an information criterion. However, in many applications it is typically set to 1 on theoretical grounds. Finally, the autoregressive coefficients are estimated by using the nonlinear least squares estimation procedure.

Testing for linearity against 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 y_t. 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. Thus, for instance, in case of a two-regime SETAR model, the null (H_0) and alternative (H_1) hypotheses are specified, assuming, without any loss of generality, that $p_1 = p_2 = p$ (say), as

$$H_0 : \phi_0^{(1)} = \phi_0^{(2)}, \quad \phi_1^{(1)} = \phi_1^{(2)}, \ldots, \phi_p^{(1)} = \phi_p^{(2)}$$

and $H_1 : \phi_i^{(1)} \neq \phi_i^{(2)}$ for at least one $i \in \{0, 1, \ldots, p\}$.

Now, the underlying statistical test for this testing problem suffers from the problem of unidentified nuisance parameters- the threshold parameters- under the null hypothesis (see, Chan (1990, 1993), Chan and Tong (1990), Hansen (1997, 2000) and Franses and Dijk (2000), for details on this problem). The main problem in such cases is that the conventional statistical theory cannot be applied to obtain the (asymptotic) distribution of the test statistic [see Davies (1997) and Hansen (1996)], and consequently the test statistic has nonstandard distribution under H_0, and its critical values are to be obtained by means of simulation and/or bootstrap method(s).

3.5 STAR Model

A variant of the SETAR model introduced by Chan and Tong (1986) and extensively explored by Terasvirta and Anderson (1992), Granger and Terasvirta (1993) and Terasvirta (1994), can be obtained if, unlike the SETAR model where an indicator function is used to incorporate regime-switching of on-off kind, the parameters are allowed to change smoothly over time. The resulting model is called the smooth transition autoregressive (STAR) model (see Terasvirta (1998) for a comprehensive review). The basic smooth transition autoregressive (STAR) model for a univariate time series y_t, originally proposed by Chan and Tong (1986), is given by

$$y_t = (\phi_{1,0} + \phi_{1,1} y_{t-1} + \cdots + \phi_{1,p} y_{t-p})\ (1 - G(s_t; \gamma, c))$$

$$+ (\phi_{2,0} + \phi_{2,1} y_{t-1} + \cdots + \phi_{2,p} y_{t-p})\ G(s_t; \gamma, c) + \varepsilon_t, \quad t = 1, 2, \cdots, T. \tag{3.7}$$

This may be conveniently expressed as

$$y_t = \phi'_1\, \tilde{y}_t\, (1 - G(s_t; \gamma, c)) + \phi'_2\, \tilde{y}_t\, G(s_t; \gamma, c) + \varepsilon_t \tag{3.8}$$

where $\tilde{y}_t = (1, y_{t-1}, \cdots, y_{t-p})'$ and $\phi_i = (\phi_{i,0}, \phi_{i,1}, \cdots, \phi_{i,p})'$, $i = 1, 2$. Assumptions about ε_t are that $E(\varepsilon_t \mid \Psi_{t-1}) = 0$ and $E(\varepsilon_t^2 \mid \Psi_{t-1}) = \sigma_\varepsilon^2$ where $\Psi_{t-1} = \{y_{t-1}, y_{t-2}, \cdots\}$ is the information set up to time $t-1$.

The transition function $G(s_t; \gamma, c)$ in the STAR model is a continuous function that is bounded between 0 and 1. 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 transition between the different regimes. The transition variable s_t 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

$$G(s_t;\gamma,c)=\frac{1}{1+\exp\{-\gamma(s_t-c)\}},\quad \gamma>0 \tag{3.9}$$

and the resultant model is called the logistic STAR (LSTAR) model. The parameter *c* in (3.9) 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 in the sense that the logistic function changes monotonically from 0 to 1 as s_t increases, while $G(c;\gamma,c)$ = 0.5. The parameter γ determines the smoothness of the change in the value of the logistic function, and thus the transition from one regime to the other. As γ becomes very large, the change of $G(s_t;\gamma,c)$ from 0 to 1 becomes almost instantaneous at $s_t=c$ and, consequently, the logistic function $G(s_t;\gamma,c)$ approaches the indicator function $I[s_t>c]$. Hence, the LSTAR model represented by (3.8) and (3.9) nests a two-regime threshold autoregressive (TAR) model as a special case. In fact, if $s_t=y_{t-d}$, the STAR model reduces to the two-regime SETAR model. When $\gamma\to 0$, the logistic function tends to the constant 0.5, and when $\gamma=0$, the STAR model reduces to a linear model.

Hypothesis testing in STAR framework

The first step towards building a STAR model is to test linearity against STAR. In terms of model (3.8), the null hypothesis is then $H_0:\phi_1=\phi_2$ and the alternative hypothesis is $H_1:\phi_{1,j}\neq\phi_{2,j}$ for at least one $j\in\{0,1,\cdots p\}$. As in case of SETAR model, this testing problem is complicated by the presence of unidentified nuisance parameters, γ and *c*, under the null hypothesis. The solution to this problem was suggested by Luukkonen *et al.* (1988). They proposed replacing the transition function $G(s_t;\gamma,c)$ by a suitable Taylor series approximation. In the reparameterized model, the identification problem is no longer present, and the null of linearity can be tested using a Lagrange multiplier (LM) / Rao's score (RS) test which follows a standard χ^2 distribution (asymptotically) under the null hypothesis.

The estimation of the parameters of the STAR model is done using conditional maximum likelihood method, and the log-likelihood is maximized numerically by using efficient iterative algorithms.

3.6 Model with Structural Breaks in Trend

In this modeling approach, we first test for the presence of structural breaks/changes in the level values of the series by applying the method proposed by Bai and Perron (1998, 2003). Based on the findings of this test whereby presence of one or more breaks in the series are detected, we partition the whole sampled period into sub-periods of stable parameter each, and then obtain dynamic model for the sub-period concerned. Of course, before building these models, we carry out standard unit root tests – augmented Dickey-Fuller (ADF) and /or Phillips-Perron tests (see, for instance, Maddala and In-Moo Kim (1998), for details of these tests) – to find the status regarding stationarity or nonstationarity of the series in the sub-periods. In what follows, we describe briefly the methodology suggested by Bai and Perron (1998, 2003) for testing the presence of multiple structural breaks in nonstationary time series. Since we also test for the presence of structural breaks in the stationary time series we first describe very 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.

3.6.1 Quandt-Andrews Test

The first classical test of an exogenously given structural change in the econometric literature is due to Chow (1960). At 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 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. In this test, 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 if the maximum of this sequence of test statistic values exceeds the Andrew's appropriate critical values. If it does, then we conclude that the time series has a structural break.

When the Quandt-Andrews test suggests a structural break, we then need to estimate the break point. Following Bai (1994, 1997a,1997 b), the sample is split at each possible break date, 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.

3.6.2 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. However, it was in 1998 that Bai and Perron considered estimating multiple structural changes in a linear model. 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 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.

(iii) 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* $F(\ell+1\,|\,\ell)$ 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* $F(\ell+1\,|\,\ell)$ are insignificant for $\ell \geq m$). This method leads to the best results and is recommended for empirical applications.

3.7 *Forecasting*

In order to assess forecasting performance of the models discussed in the preceding sections, we have obtained both in-sample and out-of-sample forecasts and then compared these with the actual values by 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 the y_t 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

$$MAE = \frac{1}{T_1} \sum_{t=T+1}^{T+T_1} \left| \, y_{t+s} - f_{t,s} \, \right|.$$

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 and 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.

4.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 use a 10-city composite S&P/Case-Shiller aggregate house price index that is seasonally adjusted at monthly frequency for the time period January, 1995 - December, 2010. 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----

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.[3]

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

[3]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.

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

5. 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 structural stability of the series using relevant testing procedures. Thereafter, we report the estimated versions of the six 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. 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 as needed.

For the purpose of this study, the observations were changed to their logarithmic values i.e., $In P_t = p_t \text{ (say)}$. Therefore, difference between p_t 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.

5.1 Testing for stationarity

As is apparent from the graphical plot of the house price index in Figure 1, the series is not stationary. In order to verify this statistically, we carried out the Augmented Dickey Fuller (ADF) as well as Phillips-Perron (PP) tests on the level values of the series i.e., p_t. The appropriate lag value for the ADF test was chosen using Schwarz's (1970) Bayesian Information Criterion (SBIC[4]) and 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 conclude that the null hypothesis of unit root cannot be rejected for the Case – Shiller house price index,

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

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 the 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.

$$\Delta p_t{*} = \underset{(-0.68)}{0.0002} + \underset{(0.59)}{(-2.02\times 10^{-6})}\, t - \underset{(-11.44)}{0.818}\, p_{t-1}{*} + \hat{a}_t\,. \tag{5.1}$$

where $p_t^* = \Delta^2 p_t$, and the numbers in parentheses underneath denote the *t*-ratios.

As is evident from equation 5.1, the ADF test statistic value stands at -11.44,suggesting that there is no more unit roots in the second differenced values. As for the PP test, the test statistic value was found to be -11.34which reinforced our conclusion based on the ADF test.

Further, the *p*-value associated with testing the significance of the deterministic trend term in the ADF estimating equation was found to be 0.59 which is highly insignificant. The constant (or the drift parameter) was also found to be insignificant. We therefore concluded that Case-Shiller house price series is integrated of order 2, i.e., it becomes stationary after second differencing. The plots of $\log P_t (= p_t)$, Δp_t and $\Delta^2 p_t$ against time was shown in *Figures* 1, 2 and 3, also exhibit graphically that the series has become stationary after second differencing.

[Figure 2 here]

[Figure 3 here]

5.2 Testing for structural stability in the stationary series

We next considered the stationary house price series to determine if there was any structural break in the series. To that end, we applied Quandt (1960) – Andrews (1993). It may be relevant to point out that 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. We have taken the break points to be determined endogenously by Bai’s (1994, 1997a, 1997b) least squares based procedure if the Quandt-Andrews test concludes that there exists a break in the stationary series.

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. 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[5].

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.

[Figure 4 here]

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, the overwhelming conclusion is that the US house prices, in its stationary values, have remained structurally stable during the entire sample period. To put it in the context of the recent global economic crisis, our findings suggest that even during the period of this crisis, the stationary component of the series has not undergone any structural change or adjustment. In other words, it empirically establishes the fact that the underlying autocorrelation pattern in the stationary component of this series has remained unchanged despite the crisis. Therefore, in all likelihood, structural change(s) must have occurred in the nonstationary component of the series, which in this case is the trend, and we report our findings for the nonstationary series in the following section.

5.3 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 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 comprise a constant term, a time trend and the first lagged value of p_t, apart from the noise term, a_t i.e.,

$$p_t = \eta + \delta t + \alpha_1 p_{t-1} + a_t \tag{5.2}$$

While applying this test, we set the value of the trimming parameter, τ, to equal 0.15. As described in Section 3.6.2., we first carried out, as per the suggestion by Bai and Perron, the *UD max* and *WD max* tests and 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

[5] Choice of AR(1) is also supported by Monte-Carlo evidence. See Maddala and In-Moo Kim (1998).

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.

5.4 Estimated models

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

5.4.1 ARMA Model

Since the time series of the Case-Shiller house prices was found to be integrated of order 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

$$\tilde{p}_t = \underset{(3.001)}{0.178^*}\ \tilde{p}_{t-1} - \underset{(-12.529)}{0.717^*}\ \tilde{p}_{t-2} - \underset{(-8.244)}{0.153^*}\ \hat{a}_{t-1} + \underset{(78.670)}{0.983^*}\ \hat{a}_{t-2} + \hat{a}_t \qquad (5.3)$$

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 Ljung-Box Q(.) statistic, based on the residuals of this model suggested that were no significant autocorrelations left in the residuals at 1% level of significance. These test statistics values along with p-values are reported in the second and third columns of Table 5.1 below. The Q (.) test was also applied to the squared residuals so as to determine if volatility is present in the residuals of the estimated ARMA model. These computations are reported in the fourth and fifth columns of Table 1.

[Table 1 here]

A closer look at this table suggested that the squared residuals are mostly highly significant. This implies that the stationary series has significant volatility which needs to be incorporated suitably in the modeling framework. This is what we report next.

5.4.2 ARMA-EGARCH Model

Since the residuals of the fitted ARMA (2,2) model showed that there is at least squared nonlinear dependence amongst them, we first considered the usual GARCH model to depict conditional heteroscedasticity and estimated the model with ARMA (2,2) as conditional mean specification. It was, however, found that the estimated GARCH model violated the condition, $\left(\sum_{i=1}^{q} \alpha_i + \sum_{i=1}^{p} \beta_i\right) < 1,$ required for the existence of unconditional variance. Hence the GARCH specification for volatility was found to be untenable for modeling volatility present in the Case-Shiller house prices.

Therefore, we next considered an alternative specification for h_t *viz.*, the EGARCH model (*cf.* equation (3.4)). The estimated ARMA-EGARCH model is reported below.

$$\tilde{p}_t = \underset{(2.997)}{0.209^*}\ \tilde{p}_{t-1} - \underset{(-13.453)}{0.761^*}\ \tilde{p}_{t-2} - \underset{(-8.327)*}{0.152}\ \hat{a}_{t-1} + \underset{(86.509)}{0.982^*}\ \hat{a}_{t-2} + \hat{\varepsilon}_t \tag{5.4}$$

and

$$\log\ \hat{h}_t = \underset{(-50.249)}{-0.125^*} - \underset{(-4.074)}{0.096^*}\ \hat{v}_{t-1} - \underset{(-30.841)}{0.131^*}\ |\hat{v}_{t-1}| + \underset{(2259101)}{0.982^*}\ \log\ \hat{h}_{t-1} \tag{5.5}$$

where $\widehat{v_t} = \widehat{\varepsilon_t} / \sqrt{\widehat{h_t}}$.

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

It is evident from equation (5.5) that the volatility exhibited in the time series of Case-Shiller house price is 'best' explained by an EGARCH (1, 1) model with all the coefficients being highly significant. Further, the coefficient of $\hat{v}_{t-1}$ in (5.4) is negative implying thereby that the relationship between volatility and rate of change in Case-Shiller house prices is negative. The standardized residuals of the estimated ARMA-EGARCH model i.e., $\hat{v}_t$ were once again checked for the presence of any remaining volatility, and the test could not reject the null of 'no volatility' at 5% level of significance. Our empirical findings in this context, namely that the volatility present in this series is adequately captured by an EGARCH (1,1) model and that the coefficient of $\hat{v}_{t-1}$ has a significant negative sign, suggest that the 'leverage effect' is highly present in the Case-Shiller house prices. This seems quite plausible because, like in equity markets, some sort of asymmetric response of volatility to positive and negative shocks is expected in the housing market as well. For example, Guirgius and Vogel (2006) studied asymmetry in house prices using regional house price data for three cities in California using a dynamic conditional correlation multivariate GARCH model and provided empirical evidence that lagged positive changes in house prices played a far more important role than lagged negative price changes in shaping the current changes in real house prices. This downward stickiness in house-prices was also emphasized by Leamer (2007) where he talked about housing cycle being more of a volume cycle and less of a price cycle and how that prolongs recessions.

5.4.3. Random Acceleration Model

This is a very simple model to estimate whereby the stationary series $\tilde{p}_t$ was regressed on a constant and an error term. 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 last ADF estimating equation given in equation (5.1), was found to be insignificant.

5.4.4. SETAR Model

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. As noted in Section 3.4, the distribution of the underlying test statistic is nonstandard. Its critical values have been obtained with bootstrap-based computations using 1000 replications and also allowing for White 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 : $$\tilde{p}_t = 7.04\times10^{-5} + 0.301^{*}\,\tilde{p}_{t-1} + \hat{a}_t \qquad (5.6)$$

$(\tilde{p}_{t-4} \le 0.001004)$ (0.485) (3.213)

Regime II: $$\tilde{p}_t = -0.0009^{*} - 0.365^{**}\,\tilde{p}_{t-1} + \hat{a}_t \qquad (5.7)$$

($\tilde{p}_{t-4} > 0.001004$)   (2.62)   (2.028)

*[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. 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.

5.4.5. STAR Model

Like in case of SETAR model, we first carried out a test for linearity *versus* nonlinearity, as proposed by the STAR model. This test has already been described in Section 3.5. 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 and results are reported below.

$$\tilde{p}_t = \begin{pmatrix} 9.053 \times 10^{-5} \\ (0.519) \end{pmatrix} + \begin{pmatrix} 0.328^* \tilde{p}_{t-1} \\ (3.258) \end{pmatrix} \begin{pmatrix} 1 - G(\tilde{p}_{t-4}; 18.68^*, 0.0008^*) \\ (12.37)(4.17) \end{pmatrix} + \begin{pmatrix} -0.00076^* \\ (-2.76) \end{pmatrix} - \begin{pmatrix} 0.325^* \tilde{p}_{t-1} \\ (-2.58) \end{pmatrix} \begin{pmatrix} G(\tilde{p}_{t-4}; 18.68^*, 0.0008^*) \\ (12.37)(4.17) \end{pmatrix} + \hat{a}_t \qquad (5.8)$$

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

5.4.6 Model with Breaks in Trend

Unlike the preceding models, this model 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 (spanning January 1995 to December 2010) which were estimated to be February 2001, October 2003, April 2006, August 2008. Therefore we estimated a model using four dummy variables for each of the intercept and slope parameters along with a sufficient number of lag values of p_t. The following best fitted model was thus obtained. It may be noted that none of the break points falls in the hold-out sample.

$$p_t = \underset{(1.146)}{0.027} - \underset{(-2.684)}{0.002}\, D_{1t} + \underset{(2.900)}{0.003}\, D_{2t} - \underset{(-2.050)}{0.002}\, D_{3t} - \underset{(-1.461)}{0.003}\, D_{4t}$$

$$+ \underset{(3.499)}{0.001}\, t + \underset{(0.137)}{5.43\times 10^{-6}}\, K_{1t} - \underset{(-3.583)}{0.0002}\, K_{2t} - \underset{(-2.800)}{0.0004}\, K_{3t}$$

$$+ \underset{(2.025)}{0.002}\, K_{4t} + \underset{(21.708)}{1.780}\, p_{t-1} - \underset{(-3.987)}{0.669}\, p_{t-2} - \underset{(-1.782)}{0.316}\, p_{t-3}$$

$$- \underset{(-0.156)}{0.029}\, p_{t-4} + \underset{(1.726)}{0.331}\, p_{t-5} - \underset{(-1.019)}{0.206}\, p_{t-6}$$

$$+ \underset{(1.627)}{0.339}\, p_{t-7} - \underset{(-2.293)}{0.443}\, p_{t-8} + \underset{(2.308)}{0.206}\, p_{t-9} + \hat{a}_t \,. \qquad (5.9)$$

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 $K_{it} = t\text{-}T_i$, 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.

5.5 Forecast Performance

In order to assess the performance of these six 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 the six models. We report the RMSE and MAE values for out-of-sample forecasts in Table 2 and in-sample forecasts in Table 3.

[Table 2 here]

[Table 3 here]

It is evident from Table 2 that both the RMSE and MAE values for out-of-sample forecasts are, as such, very small for all the models. Upon careful inspection, a few distinct patterns emerge. First, the RMSE and MAE values, while differing from each other, are more or less identical for models M1, M2, M4 and M5 for all forecast horizons. In that sense, the simple ARMA specification performs as well as ARMA – EGARCH, SETAR and STAR. Our results 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. Second, the RMSE and MAE values for model M3 are slightly higher than those for M1,M2, M4 and M5 for near term forecast horizons (1 through 3) and slightly lower than M1,M2,M4 and M5 for longer term forecast horizons (4 thorough 6). In that sense, the RA Model performs slightly better when the forecast horizon is somewhat longer. A plausible explanation for this is that unlike in the RA model, the forecast errors for other models are cumulated as the forecast horizon increases. Third, the RMSE and MAE values for model M6 are slightly higher than those for M1, M2, M4 and M5 for all forecast horizons, indicating that the nonstationary series with break in trend does not necessarily result in superior out-of- sample forecasts. When compared with model M3, M6 performs slightly better only in forecast horizon 1 and 2, but this improvement disappears for forecast horizons 3 through 6. A probable reason for this could be the fact that toward the end of the series covering the hold-out period, the model appears to have changed (see *Figure* 1) which could not be detected by the Bai Perron test because of consideration to trimming associated with the break tests. On the whole we may conclude that insofar as the out-of-sample forecasts are concerned, ARMA (2,2) model still performs as well as EGARCH, SETAR and STAR, outperforms nonstationary model with break in trend throughout and yields slightly better forecasts over the RA model for the short horizon periods (*viz*.,1,2 and 3 step ahead forecast horizon). For the longer horizons (*viz*., 4,5 and 6 step ahead), the RA model on the other hand has a slight edge over the others.

Comparison of models based on in-sample forecasts yield different results. As is evident from the RMSE and MAE values reported in Table 3, the nonstationary model with breaks in trend produces the best in-sample forecast. Its RMSE and MAE values are much lower than those for the ARMA model and the other models in general. A possible explanation for the best forecasting performance of M6 in-sample (January 1995-December 2008) lies in the fact that the apparent change in the house price index series towards the end of the hold-out period is taken into account in this case.

6. Concluding Remarks

The main objectives of this empirical study were to i) detect the possibility of multiple structural breaks in the US house price series data (that are endogenously determined)for a recent time period exhibiting very sharp upturns and downturns and ii) carry out house price forecasting exercises using alternative time series models – both linear and nonlinear. In view of the risky financial innovations that took place in the housing market prior to the recent crisis fueling the speculative housing boom, we investigated the possibility of structural breaks fundamentally altering the time series properties of the house price series. Ours is perhaps the very first attempt in this regard. Using Bai-Perron 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 five other models comprising simple ARMA, ARMA-EGARCH, RA, 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 and ARMA-GARCH may not accurately capture the house price dynamics.

For each of the models, in-sample forecasts, as expected, were found to perform better than the corresponding out of sample forecasts. 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 be due to the fact that the model has undergone yet another structural change in the hold-out period. Because of consideration of trimming associated with the Bai-Perron 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. Forecasting performance may improve as more data become available, and in that case the test for the existence of another break could be carried out. Finally, 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 that of ARMA in our study. In view of our findings, we recommend investigating structural breaks/regime switches in the house price series before formulating forecasting models for house price series.

Acknowledgement: *The corresponding author thanks the College of Humanities and Public Affairs at Missouri State University for granting a sabbatical leave. The research was carried out at the Economic Research Unit of the Indian Statistical Institute.*

References

Andreou, E. and Ghysels, E. (2002). Detecting multiple breaks in financial market volatility dynamics. *Journal of Applied Econometrics*, 17, 579-600.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica , 61,* 821-856.

Andrews, D. W. (2003). Tests for parameter instability and structural change with unknown change point: A corrigendum. *Econometrica, 71* , 395-397.

Andrews, D. W., & Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica, 62* , 1383-1414.

Bai, J. (1994). Least squares estimation of a shift in linear processes. *Journal of Time Series Analysis 15*, 453-472.

Bai, J. (1997a). Estimation of a change point in multiple regression models. *Review of Economics and Statistics 79* , 551-563.

Bai, J. (1997b). Estimating multiple breaks one at a time. *Econometric Theory, 13* , 315-352.

Bai, J., & Perron, P. (1998). Estimating and testing linear models with multiple structural changes. *Econometrica, 66*, 47-78.

Bai, B., & Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics 18*, 1-22.

Bera, A. K., & Higgins, M. L. (1993). A survey of ARCH models: properties, estimation and testing . *Journal of Economic Surveys, 7* , 305-366.

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

Bernanke B.S. (2010).Monetary policy and the housing bubble. Speech delivered at the annual meeting of the American Economic Association Meeting, Atlanta, Georgia, January 2010. Retrieved from http://www.federalreserve.gov/newsevents/speech/bernanke20100103a.htm Accessed 20 September, 2011.

Berndt, E. R., Hall, B. H., Hall, R. E., & Hausman, J. A. (1974). Estimation and inference in nonlinear statistical models. *Annals of Economic and Social Measurement, 3*, 653-665.

Bessec, M., and Bouabdallah, O. (2005). What causes the forecasting failure of Markov-switching models? A Monte Carlo study. *Studies in Nonlinear Dynamics and Econometrics*, 9.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics 31*, 307-327.

Brooks, C. (2002). *Introductory econometrics for finance.* Cambridge University Press.

Case, K., and Shiller, R. (1989). The efficiency of the market for single-family homes. *American Economic Review*, 79, 125–137.

Chan, K. S., & Tong, H. (1986). On estimating thresholds in autoregressive models. *Journal of Time Series Analysis 7*, 179-190.

Chan, K.S. (1990). Testing for Threshold Autoregression, *The Annals of Statistics, 18,* 1886-1894.

Chan, K. S., & Tong, H. (1990). On likelihood ratio tests for threshold autoregression. *Journal of the Royal Statistical Society B 52*, 469-476.

Chan, K. S. (1993), consistency and limiting distribution of least squares estimation of a threshold autoregressive model, *Annals of statistics, 21,* 520-533.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. *Econometrica, 28* , 591-605.

Crawford, G., and Fratantoni, M. (2003). Assessing the forecasting performance of Regime-Switching, ARIMA and GARCH models of house prices. *Real Estate Economics*, 31, 223–243.

Dacco, R. and Satchell, C. (1999) Why do regime–switching forecast so badly? *Journal of Forecasting*, 18, 1-16.

Davies, R.B.(1997), Hypothesis testing when a nuisance parameter is present only under the alternative, *Biometrica, 64,* 247-254.

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

Ducca, J.V., Luttrell, D. and Murphy, A (2011). When will the U.S. housing market stabilize? *Economic Letter*, Federal Reserve Bank of Dallas, 6, 1-4.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica, 50* , 987-1007.

Franses, P. H., and Dijk, D. V. (2000). *Non-linear Time Series Models in Empirical Finance* . Cambridge University Press.

Gerlach, R., Wilson, P. and Zurbruegg, R. (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, 974-91.

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

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

Hansen, B. E. (1996), Inference when a nuisance parameter is not identified under the null hypothesis, *Econometrica, 64,* 413-430.

Hansen .B. E. (1997). Inference in TAR Models, *Studies in Nonlinear Dynamics and Econometrics, 2,* 1-14.

Hansen. B. E. (2000). Sample Splitting and threshold estimation, *Econometrica, 68,* 575-603.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in U.S. labor productivity. *Journal of Economic Perspectives, 15,* 117-128.

Hartmann, D., Kempa, B. and Pierdzioch, C. (2008). Economic and Financial Crises and the Predictability of U.S. Stock Returns. *Journal of Empirical Finance*, 15, 468-80.

Ho, A. K.F. and Wan, A.T. K. (2002). Testing for covariance stationarity of stock returns in the presence of structural breaks: An intervention analysis. *Applied Economics Letters*, 9, 441-47.

Helbling, T.F. (2003). Housing Price Bubbles – A Tale Based on Housing Price Booms and Busts. *Real Estate Indicators and Financial Stability*, Proceedings of a joint conference organized by the Bank for International Settlements (BIS), Switzerland and the IMF in Washington DC, 27–28 October 2003.

Jannsen, N. (2010). National and International Business Cycle Effects of Housing Crises. *Applied Economics Quarterly*, 56, 175-206.

Kim, J.W., Leatham, D. J. and Bessler, D.A. (2007). REITs' dynamics under structural change with unknown Break Points. *Journal of Housing Economics*, 16, 37-58.

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

Larson, W.D. (2010) Evaluating Alternative Methods of Forecasting House Prices: A Post-Crisis Reassessment. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1709647 Accessed 20 March, 2011.

Leamer, E. (2007). Housing is the business cycle. *NBER Working Paper 13428*, Retrieved from http://www.nber.org/papers/w13428 Accessed 20 March, 2011.

Lucey, B. M and Voronkova, S. (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, 1303-24.

Luukkonen, R., Saikkonen, P., & Terasvirta, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika, 75* , 491-499.

Maddala, G. S., & Kim, In-Moo (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, 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. *Econometrica 59*, 347-370.

Nelson, D. B., & Cao, C. Q. (1992). Inequality constraints in the univariate GARCH model. *Journal of Business & Economic Statistics, 10*, 229-235.

Obtsfeld, M and Rogoff K. (2009). "Global Imbalances and the Financial Crisis: Products of Common Causes", Federal Reserve Bank of San Francisco, Asia Economic Policy Conference, October 2009. Retrieved from http://elsa.berkeley.edu/~obstfeld/santabarbara.pdf Accessed 25 March, 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. Retrieved from http://www.kc.frb.org/publicat/sympos/2007/pdf/2007.09.27.shiller.pdf Accessed 2 March, 2011.

Terasvirta, T. (1998). Modeling economic relationships with smooth transition regressions. In A. Ullah, & D. E. Giles, *Handbook of Applied Economic Statistics,* 507-552. New York: Marcel Dekker.

Terasvirta, T. (1994). Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association, 89* , 208-218.

Terasvirta, T., & Anderson, H. M. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics, 7* , 119-136.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach.* Oxford University Press.

Tong, H. (1983). Threshold models in non-Linear time series analysis. In *Lecture Notes in Statistics.* New York: Springer.

Tong, H. (1978). On a threshold model. In C. H. Chen, *Pattern Recognition and Signal Processing*, 101-141. Amsterdam: Sijthoff & Noordhoff.

Tong, H., & Lim, K. S. (1980). Threshold autoregressions, limit cycles, and data . *Journal of the Royal Statistical Society B 42*, 245-292.

Tsay, R. S. (1989). Testing and modeling threshold autoregressive process. *Journal of the American Statistical Association 84* , 231-240.

Tsouma, E. (2007). Stock return dynamics and stock market interdependencies. *Applied Financial Economics,* 17, 805-25.

Zhou, Z. (1997). Forecasting sales and price for existing single-family homes: a VAR model with error correction. *Journal of Real Estate Research*, 14, 155–167.

Figure 1: Plot of Case-Shiller house price index (in log value), p_t

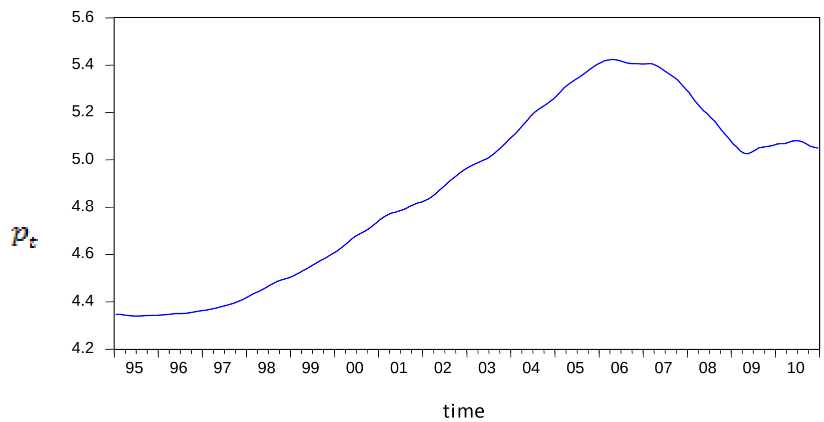

Figure 2: Plot of first differenced values of p_t

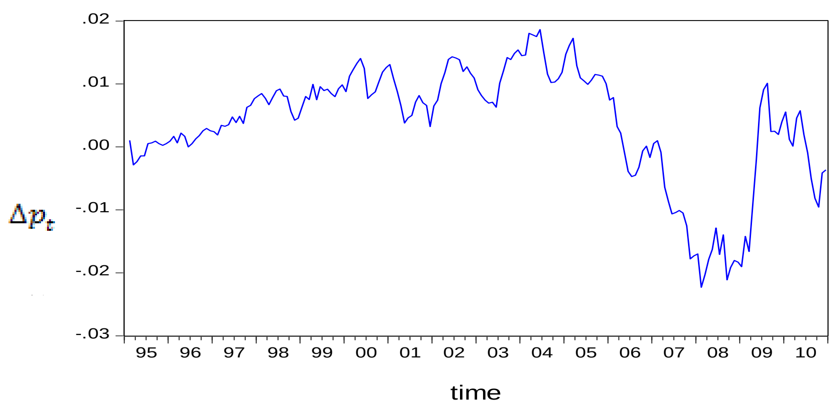

Figure3: Plot of second differenced values of p_t

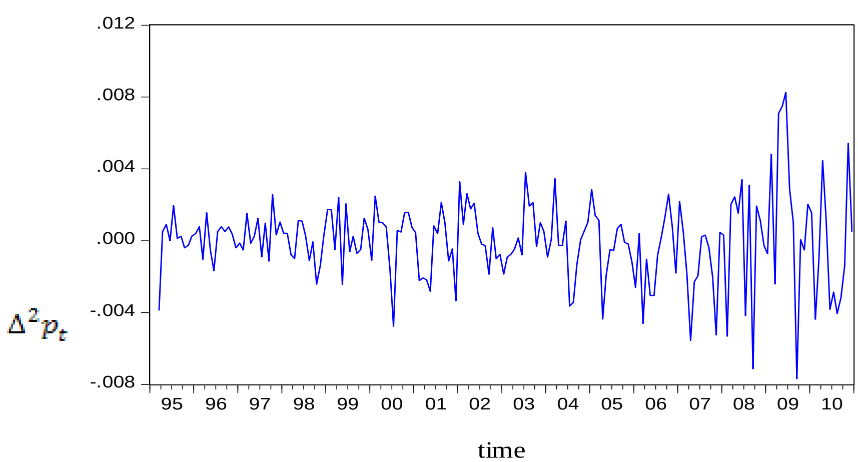

Figure 4: Plot of Quandt-Andrews test statistic values

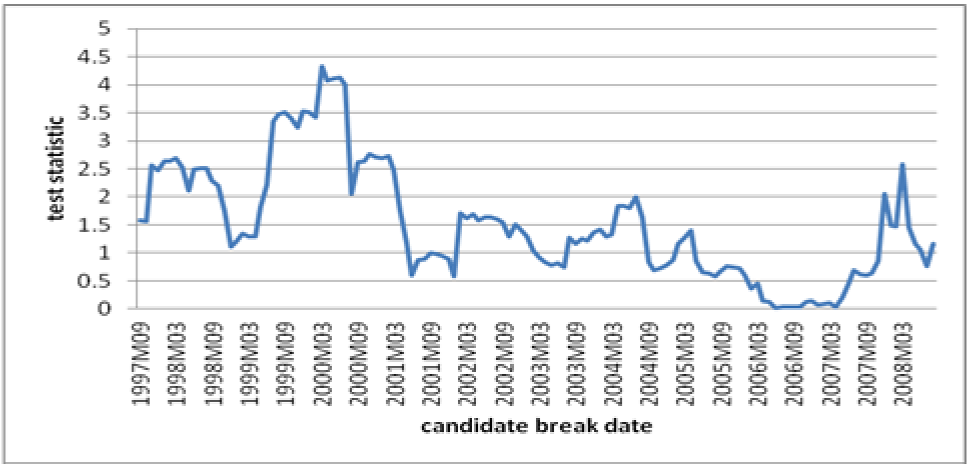

Table 1: Ljung-Box test statistic and p-values for residuals and squared residuals of the ARMA(2,2)model

Lag value	**Residuals**		**Squared residuals**	
	Test statistic	***p-value***	**Test statistic**	***p-value***
1	2.881	0.410	6.533	0.088
2	9.350	0.053	6.548	0.162
3	10.991	0.052	11.355	0.045
6	14.264	0.075	24.445	0.002
12	21.613	0.087	31.169	0.003
18	28.984	0.088	37.361	0.007
24	32.917	0.165	41.995	0.018

Table 2: Out-of-sample forecast performance

Forecast horizon	RMSE						MAE					
	M1	M2	M3	M4	M5	M6	M1	M2	M3	M4	M5	M6
1	0.004	0.004	0.011	0.004	0.004	0.005	0.003	0.003	0.009	0.003	0.003	0.004
2	0.010	0.010	0.014	0.010	0.010	0.011	0.007	0.008	0.011	0.008	0.008	0.009
3	0.017	0.017	0.018	0.017	0.017	0.020	0.013	0.013	0.014	0.014	0.014	0.017
4	0.027	0.027	0.026	0.027	0.027	0.031	0.021	0.021	0.020	0.021	0.021	0.026
5	0.039	0.039	0.036	0.039	0.039	0.041	0.029	0.029	0.028	0.029	0.029	0.034
6	0.051	0.051	0.047	0.051	0.051	0.052	0.038	0.038	0.037	0.038	0.038	0.043

M1: ARMA(2,2) ; M2: ARMA(2,2)-EGARCH(1,1); M3: Random Acceleration; M4: SETAR; M5:STAR; M6: Model with breaks in trend

Table 3: In-sample forecast performance

Model	RMSE	MAE
M1	0.001735	0.001294
M2	0.001740	0.001288
M3	0.001871	0.001401
M4	0.001741	0.001265
M5	0.001744	0.001298
M6	0.001499	0.001141

M1: ARMA(2,2); M2: ARMA(2,2)-EGARCH(1,1); M3: Random Acceleration; M4: SETAR; M5: STAR; M6: Model with breaks in trend