Eastern Mediterranean University

Department of Economics
Discussion Paper Series

The Out-of-Sample Forecasting Performance of Non-Linear Models of Regional Housing Prices in the US

Mehmet Balcilar, Rangan Gupta, Stephen M. Miller

Discussion Paper 15-27

August 2015

Department of Economics
Eastern Mediterranean University
Famagusta, North Cyprus

The Out-of-Sample Forecasting Performance of Non-Linear Models of Regional Housing Prices in the US

Mehmet Balcilar
Department of Economics
Eastern Mediterranean University
Famagusta, NORTHERN CYPRUS, via Mersin 10, TURKEY

Rangan Gupta
Department of Economics
University of Pretoria
Pretoria, 0002, SOUTH AFRICA

Stephen M. Miller*
Department of Economics,
University of Nevada, Las Vegas
Las Vegas, Nevada, 89154-6005 USA

Abstract

This paper provides out-of-sample forecasts of linear and non-linear models of US and Census regions housing prices. The forecasts include the traditional point forecasts, but also include interval and density forecasts of the housing price distributions. The non-linear smooth-transition autoregressive model outperforms the linear autoregressive model in point forecasts at longer horizons, but the linear autoregressive model dominates the non-linear smooth-transition autoregressive model at short horizons. In addition, we generally do not find major differences in performance for the interval and density forecasts between the linear and non-linear models. Finally, in a dynamic 25-step ex-ante and interval forecasting design, we, once again, do not find major differences between the linear and nonlinear models.

Keywords: Forecasting, Linear and non-linear models, US and Census housing price indexes

JEL classification: C32, R31

* *Corresponding author*

1. Introduction

This paper considers the out-of-sample forecasting performance of linear and non-linear models of real house price indexes for the US and its four Census regions – Northeast, South, Midwest, and West. The analysis compares autoregressive (AR) and smooth-transition autoregressive (STAR) models, estimates the models using monthly data over the 1968:1 to 2000:12 in-sample period, and forecasts over the 2001:1 to 2010:5 out-of-sample period. Finally, we also design an ex-ante dynamic 25-step forecasting experiment over the period 2010:6-2012:6 to examine the real world success of the forecasts generated from the linear AR and non-linear STAR models.

Forecasters typically use linear models, albeit the variables forecast probably undergo some transformation (e.g., natural logarithms). A limited number of studies consider non-linear models for forecasting purposes. For example, Rapach and Wohar (2006) employ non-linear methods to perform out-of-sample forecasting of real exchange rates. In that paper, transactions costs provide a band of inactivity around the current real exchange rate that prevents arbitrage from moving the real exchange rate back to equilibrium. Outside the band of inactivity, arbitrage occurs tending to drive the real exchange rate back toward its equilibrium value.

Housing prices in the US rise more quickly and fully to market events that increase the equilibrium price than they do to market events that lower the equilibrium price. The fall of housing prices during the Great Recession and beyond did not fall quickly enough to clear housing markets around the country, significantly slowing the recovery process. Also, Kim and Bhattachya (2009) show that housing prices in the US and three of the four Census regions exhibit non-linearity. The Midwest, the exception, exhibits linear movements. They conclude that the behavior of the housing market differs across phases of expansion and contraction of the residential real estate sector. Seslen (2004) argues that households exhibit forward looking behavior and a higher probability of trading up, during expansions, since equity constraints prove less binding. During the downswing of the housing market cycle, households less likely trade, implying downward rigidity of house prices. Loss aversion during the downswing more likely reduces the mobility of households as well as trading activity. Further, Muellbauer and Murphy (1997) note that the presence of lumpy transaction costs in the housing market can also cause non-linearity. Given these issues, it makes sense to test for non-linear housing price movements.

The housing market traditionally leads business cycle movements. The typical end of an expansion sees the central bank raising interest rates to subdue inflation. Higher interest rates cause a hiccup in the housing market and it turns down prior to the overall decline in economic activity. Also, housing markets generally follow different patterns in different regions, contributing to the regional differences in business cycle movements.

Recently, Leamer (2007) strongly argues that housing *is* the business cycle, indicating "any attempt to control the business cycle needs to focus especially on residential investment." (p. 150). His conclusion comes out of the dynamics of the home construction industry. That is, a building boom over one time interval pushes the stock of new homes above trend and that necessitates with some lag another time interval with a building slump. Thus, monetary policy should focus on preventing booms from occurring to head off eventual slumps. Quoting Leamer (2007), "***The Fed can stimulate now, or later, but not both***." (p. 151, bold, italics in original). Smets (2007) comments on Leamer's paper and argues that interest rates (and monetary policy) crucially determine the linkages between the housing cycle and the business cycle. In the general discussion of his paper, Leamer (2007) responds to Smets (2007) that "in the context of my paper, ... the interest rate spread has its impact through housing, though it surely operates through other channels." (p. 249).

Residential housing enters directly into the calculation of GDP through investment demand. Other researchers consider the effect of housing demand on consumption demand, the largest component of GDP. Case, *et al.* (2005) provide a good recent review. While the original simple life-cycle model of consumption does not distinguish between different types of wealth, implicitly assuming that the marginal propensities to consume out of wealth remains the same across different wealth types, reasons exist to suggest that this implicit assumption is, in fact, invalid. Case, *et al.* (2005) offer five different possible rationalizations for different marginal propensities to consume out of different types of wealth – differing perceptions about the effects of permanent and transitory components, differing bequest motives, differing motives for wealth accumulation, differing abilities to measure wealth accumulation, and differing psychological "framing" effects. Another possible rationalization, not mentioned by Case, *et al.* (2005), involves whether the wealth holder receives consumption services from the holding of wealth. For example, owner occupied housing and consumer durable goods provide consumption services to holders of these components of wealth. Thus, households may adjust their consumption of nondurables and services, the usual measure of consumption for wealth-effect studies, differently to changes in the market values of owner occupied housing and consumer durables than to changes in other forms of wealth that do not possess such services..

The empirical evidence for the effect of changes in real estate and housing values on consumption also provides somewhat mixed findings, with the bulk of the results supporting a significant positive effect. For time-series evidence, Elliot (1980) does not find any significant effect. Peek (1983), Bhatia (1987), Case (1992), and Case, *et al.* (2005) do.[1] The cross-section evidence provides similar findings. Levin (1998) finds no evidence of a real estate and housing value effect on consumption. Skinner (1989) and Engelhardt (1996) discover significant effects. In addition, Engelhardt's findings exhibit asymmetry, where negative "news" affects consumption but positive "news" does not.

A large number of papers (e.g., Green 1997, Iacoviello 2005, Case *et al*. 2005, Rapach and Strauss 2006, Leamer 2007, Pariès and Notarpietro 2008, Vargas-Silva 2008, Bao *et al.* 2009, Christensen *et al.* 2009, Ghent 2009, Ghent and Owyang 2009, Pavlidis *et al.* 2009, and Iacoviello and Neri 2010) show a strong link between the housing market and economic activity. Since housing contributes to a large percentage of private sector wealth (Cook and Speight 2007), house price changes affect household consumption and saving patterns (Englund and Ioannides 1997). Campbell and Cocco (2005) argue that since households view housing expenditure as a consumption good, house prices correlate positively with consumption spending (see also Pavlidis *et al.* 2009). In addition, Forni *et al.* (2003), Stock and Watson (2003), and Gupta and Das (2010) argue that house-price movements lead economics activity, indicating the future direction of economic movements. Moreover, the recent boom-bust cycles in house prices cause much concern and interest amongst policy makers (Borio *et al*. 1994, and Bernanke and Gertler 1995, 1999), since the bust of house price bubbles typically leads to significant contractions in real economic activity, as seen in the current economic downturn.

---

[1] Case, *et al.* (2005) consider the wealth effects of stock-market and real estate and housing values simultaneously for the U.S. states and 14 developed countries, including the U.S. They find strong evidence of a wealth effect on consumption due to real estate and housing value and a weaker effect due to stock-market value. Stock-market value achieves a stronger effect for the U.S. states, which may reflect the larger relative holding of stocks in the U.S. relative to other developed countries.

Models that forecast house price movements can provide policy makers with early information on future movements in economic activity, leading to better policy control. Hence, deducing the underlying nature of the data-generating process for house price (i.e., linear or non-linear) will improve the forecast performance. For example, if house price movements reflect non-linear adjustment, then forecasts from a linear model will generate inaccurate forecasts for not only house prices, but also the economy, given that house prices lead real economic activity.

Recent studies (Genesove and Mayer 2001, Engelhardt 2001, Seslen 2004, Kim and Bhattacharya 2009, Balcilar *et al*. 2011) document evidence of nonlinearity in housing prices. Several reasons exist that support further research on such nonlinearities and specifications of nonlinear models that successfully capture the nonlinearities. Growing evidence exists on the nonlinearity of macroeconomic variables. Neftci (1984), Falk (1986), and Bradley and Jansen (1997) present evidence that many macroeconomic variables behave asymmetrically over the business cycle and, therefore, show nonlinear dynamics. Skalin and Teräsvirta (1999, 2002) conclude that macroeconomic variables such as unemployment and GDP conform to the non-linear framework of a smooth-transition autoregressive (STAR) model. Given that the housing market leads the business cycle and that significant links exist between the housing market and economic activity, nonlinearities in housing prices can indeed explain the existence of nonlinearities in macroeconomic variables. On the contrary, housing prices may exhibit nonlinearities, particularly of threshold variety captured by STAR models, because determinants of house prices such as the interest rate and GDP display asymmetric adjustment (Neftci 1984, Enders and Siklos 2001). Two possible explanations for intrinsic nonlinearity in house prices exist. First, households respond asymmetrically over the business cycle. Abelson *et al*. (2005) argue that households more likely buy when prices rise, because they expect further rises and try to avoid higher payments. Households will less likely buy or sell, however, due to loss eversion with falling house prices. Seslen (2004) and Muellbauer and Murphy (1997) argue for non-linear adjustment in housing prices due to equity constraints and transactions costs, respectively.

In sum, the housing market and its movement prove important for explaining business cycle movements through their effect on investment and consumption spending. In addition, some differences in regional business cycle movements depend on the local nature of the housing market.

To examine the extent of the nonlinearity in housing price adjustments, we conduct an extensive out-of-sample forecast comparison of nonlinear and linear AR models for four regional (Northeast, Midwest, South, West) housing price indexes as well as for the aggregate US housing price index. If the out-of-sample forecasts generated by the nonlinear AR models outperform the forecasts generated by the linear AR models, then evidence exists against the linear models.

The out-of-sample forecast comparisons do not rely on a single criterion, as usually done, such as the root mean square error (RMSE). We compare linear AR and a class of nonlinear AR models in their out-of-sample point, interval, and density forecasts. First, we compare linear and nonlinear AR models in their out-of-sample point forecast performance using the root mean squared error (RMSE) criterion and test for the superiority of the forecasts using the Diebold and Marino (1995) test. Moreover, nonlinear models may exhibit only superior forecasting performance in certain regimes (e.g., recessions) and not in others (e.g., expansions). To examine this possibility, we focus on the forecasting performance for the observations in the tails of the distribution, using weighted version of the Diebold and Mariano test proposed by van Dijk and Franses (2003).

Second, we also compare the superiority of the forecasts in their out-of-sample interval and density forecasting performance, using the approach suggested by Christoffersen (1998) and Debold *et al*. (1998). We compare interval forecasts using the Pearson χ^2 statistics, while we compare the density forecast using the Kolmogorov-Smirnow, Doornik-Hansen (1994), and Ljung-Box tests. To consider the extent of the nonlinearity, we also evaluate the nonlinear AR models using the informal testing approach proposed by Pagan (2002) and Breunig *et al*. (2003). We more formerly compare linear and nonlinear AR models using the Corradi and Swanson (2003) Z_T statistic that is based on the distributional analogue of the mean square error metric of models. This statistic can compare two models, both of which are possibly misspecified. Finally, we use an ex-ante forecast design and compare 25-step dynamic forecasts of the linear and nonlinear AR models over 2010:6 to 2012:6.

The rest of the paper adopts the following structure. Section 2 outlines the methodology of non-linear estimation. Section 3 provides a description of point, interval, and density forecasts. Section 4 discusses the data. Section 5 evaluates the empirical findings. Section 6 concludes.

2. Methodology

We adopt the STAR framework, developed by Luukonnen *et al.* (1988) as extended by Escribano and Jordá (1999), to model house price growth rates as non-linear and statedependent.[2] The STAR framework connects different regimes with a smooth transition function to describe the long-run dynamics of house price growth rates. The STAR framework dominates threshold autoregressive (TAR) (Tsay 1989) and the Markov switching (MS) (Hamilton 1989) models, since the latter two frameworks specify discrete jumps between regimes. In fact, the TAR model emerges as a limiting case of the STAR model. In addition, the low speeds of transition, which we find in the estimation of the non-linear model, support our choice. In housing markets with large number of buyers and sellers with heterogeneous beliefs and unsynchronized responses to news, the STAR framework seems appropriate.

The STAR model of order p, for variable r_t, is specified as follows:[3]

$$\begin{aligned} r_t &= [\phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i}] + [\rho_0 + \sum_{i=1}^{p} \rho_i r_{t-i}].F(r_{t-d}) + u_t \\ &= [\phi_0 + \phi(L) r_t] + [\rho_0 + \rho(L) r_t].F(r_{t-d}) + u_t \end{aligned} \qquad (1)$$

where r_t denotes the housing price growth rate, and $F(r_{t-d})$ denotes the smooth and continuous transition function of past realized housing price growth rates controlling the regime shift mechanism. Thus, house price growth rates evolve with a smooth transition between regimes that depends on the sign and magnitude of past realization of house price growth rates. We generate non-linearities by conditioning the autoregressive coefficients, $\rho(L)$, to change smoothly with past house price growth rates. That is, the past realized home price growth rate

[2] Non-linear estimation, just like linear estimation, requires stationary variables to avoid spurious estimates. Hence, we convert house prices in the US and the four Census regions into annual growth rates. We confirm stationarity of the series, in turn, by the Augmented–Dickey–Fuller (ADF), the Dickey-Fuller with GLS Detrending (DF-GLS), the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS), and the Phillips-Perron (PP) tests. The results are available from the authors.

[3] This discussion relies heavily on the presentation in Kim and Bhattacharya (2009) and Balcilar *et al*. (2011). We retain their symbolic representation of the equations.

r_{t-d} becomes the transition variable with delay parameter d, which indicates the number of periods that r_{t-d} leads the regime switch.

Teräsvirta and Anderson (1992) consider two alternative transition functions that produce the logistic smooth transition autoregressive (LSTAR) model and the exponential smooth transition autoregressive (ESTAR) model. In the LSTAR model, the transition function equals a logistic model as follows:

$$F(r_{t-d}) = [1 + \exp\{-\gamma(r_{t-d} - c\}]^{-1}, \quad \gamma > 0, \tag{2}$$

while in the ESTAR model, the transition function equals an exponential model as follow:

$$F(r_{t-d}) = 1 - \exp\{-\gamma(r_{t-d} - c)^2\}, \quad \gamma > 0. \tag{3}$$

In equations (2) and (3), γ denotes the speed of transition between regimes and c measures the halfway point or threshold between the two regimes. Equations (1) and (2) yield the LSTAR(p) model and equations (1) and (3) yield the ESTAR(p) model. In STAR models, two different economic phases characterize expansions and contractions, but a smooth transition occurs between the two regimes, controlled by r_{t-d} (Sarantis 2001). The LSTAR and ESTAR models describe different dynamic behaviors. The LSTAR model allows the expansion and contraction regimes to exhibit different dynamics whereas the ESTAR model suggests that the two regimes exhibit similar dynamics (Sarantis 2001). When $\gamma \to \infty$, the model degenerates into the conventional *TAR(p)*, while when $\gamma \to 0$, the model degenerates to the linear *AR(p)* model (Teräsvirta and Anderson 1992).

The construction of an appropriate STAR model for a specific variable encompasses three stages. First, we specify a linear *AR(p)* model with p chosen by the unanimity of at least two of the popular lag-length tests – the LR test statistic, Akaike information criterion (AIC), Schwarz information criterion (SIC), the final prediction error (FPE) criterion, and the

Hannan-Quinn (HQ) information criterion. Second, we test for linearity against a non-linear STAR model, for different values of the delay parameter *d* using the linear *AR(p)* model as the null, based on a Lagrange multiplier smooth transition (LM-STR) test for linearity. This requires estimating the following auxiliary regression proposed by Teräsvirta and Anderson (1992):

$$\begin{aligned} r_t = \phi_0 + \sum_{i=1}^{p} \varphi_{1,i} . r_{t-i} + \sum_{i=1}^{p} \varphi_{2,i} r_{t-i} r_{t-d} + \sum_{i=1}^{p} \varphi_{3,i} r_{t-i} r_{t-d}^{2} + \cdots \\ + \sum_{i=1}^{p} \varphi_{k+1,i} r_{t-i} r_{t-d}^{k} + u_t \end{aligned} , \qquad (4)$$

where the null hypothesis of linearity states that H01: $\phi_{2i} = \phi_{2i} = ... = \phi_{k+1i} = 0$ for all *i*. We estimate equation (4) over a range of values, $1 \leq d \leq D$ to identify the appropriate delay parameter *d*. When we reject linearity for more than one value of *d*, we choose such that $d = arg\ min\ p(d)$ for $1 \leq d \leq D$, where $p(d)$ equals the test's *p*-value.

Escribano and Jordá (1999) argue that a first-order Taylor approximation does not sufficiently capture the characteristics of the exponential function and recommend second-order Taylor approximation. Escribano and Jordá propose four Lagrange multiplier (LM) tests of linearity against STAR alternatives. The LM_2 and LM_4 tests exhibit power against ESTAR alternatives while the LM_1 and LM_3 tests exhibit power against LSTAR alternatives.

According to Teräsvirta and Anderson (1992), we choose between the LSTAR and ESTAR models, once we reject linearity, by applying the following sequence of nested tests:

$$H_{04} : \phi_{4i} = 0, \qquad i = 1, \ldots , p; \qquad (5)$$

$$H_{03} : \phi_{3i} = 0 \big| \phi_{4i} = 0, \qquad i = 1, \ldots , p; \text{ and} \qquad (6)$$

$$H_{02} : \phi_{2i} = 0 \big| \phi_{3i} = \phi_{4i} = 0, \qquad i = 1, \ldots , p. \qquad (7)$$

Three possible sequential outcomes exist, given *d*. One, we reject $H_{04}:\phi_{4i}=0$, which implies that we select the LSTAR model. Two, if we do not reject H_{04}, then we test if $H_{03}:\phi_{3i}=0 \mid \phi_{4i}=0$. Rejecting H_{03} implies the selection of the ESTAR model. Three, if we do not reject H_{03}, then we test $H_{02}:\phi_{2i}=0 \mid \phi_{3i}=\phi_{4i}=0$. Rejecting H_{02} implies selection of the LSTAR model.

Escribano and Jordá (1999) propose alternative tests for the selection of LSTAR versus ESTAR models as follows:

$$H_{0E}:\phi_{3i}=0, \phi_{5i}=0, \qquad i = 1, \dots, p; \text{ and} \tag{8}$$

$$H_{0L}:\phi_{2i}=0, \phi_{4i}=0, \qquad i = 1, \dots, p. \tag{9}$$

The tests H_{0E} and H_{0L} come from equation (4) with $k=4$. Rejecting H_{0E} selects the ESTAR model while rejecting H_{0L} selects the LSTAR model.

Various authors (Granger and Teräsvirta 1993, Teräsvirta 1994, Eitrheim and Teräsvirta 1996, and Sarantis 2001) argue that if researchers strictly apply this sequence of tests, then they may reach false conclusions, since the tests ignore higher-order terms of the Taylor expansion used in its derivation. These authors recommend that researchers compute *p*-values for all the *F*-tests of the hypotheses in equations (1) through (3). Then, researchers choose the appropriate STAR model based on the lowest *p*-value or highest *F*-statistic.

3. Point, Interval, and Density Forecasts: Method and Analysis[4]

Our analysis expands beyond the traditional point forecasts to include both interval and density forecasts. Recent studies report that non-linear models produce superior interval and density forecasts to linear models, although inferior point forecasts (e.g., Clements and Smith

[4] This section relies heavily on Rapach and Wohar (2006).

2000, Siliverstovs and van Dijk 2003, and Rapach and Wohar 2006). We develop interval and density forecasts using Christoffersen (1998) and Diebold, *et al.* (1998).

Point, Interval, and Density Forecasts: Method

We use the fitted non-linear AR models reported in Section 2 to calculate out-of-sample point, interval, and density forecasts. We consider whether out-of-sample point, interval, and density forecasts generated by the non-linear models outperform those generated by simple linear AR models. We assume that the non-linear and linear AR models exhibit Gaussian errors.

Generating point, interval, and density forecasts for linear AR models with Gaussian errors proves straightforward.[5] Analytical point, interval, and density forecasts do not generally exist for non-linear AR models with Gaussian errors. We follow Rapach and Wohar (2006) and use their simulation-based procedure to generate forecasts for the non-linear AR models.

Analyzing Point Forecasts

We use the mean-square-forecast-error (MSFE) criterion and adopt the Diebold and Mariano (1995) procedure to test the null hypothesis of equal predictive ability against the one-sided alternative hypothesis that the non-linear AR model exhibits a smaller MSFE than the linear AR model. Following Siliverstovs and van Dijk (2003) and Rapach and Wohar (2006), we use the modified Diebold and Mariano statistic (M-DM) of Harvey, *et al.* (1997), correcting for potential finite-sample size distortions. We use the Student-*t* distribution to determine

[5] We follow the existing literature in treating the parameters of the linear and nonlinear AR models as known in forming forecasts. Hansen (2006) describes how to include parameter estimation uncertainty into interval forecasts for linear models.

significance.[6] We also follow Rapach and Wohar (2006) and consider a weighted Diebold and Mariano (1995) statistic (W-DM) recently developed by van Dijk and Franses (2003), where the observations of different regions receive different weights. Given that our non-linear models include asymmetric adjustment to long-run equilibrium, we adopt the first weight function suggested by van Dijk and Franses (2003), which attaches greater weight to observations in both tails of the distribution. We again follow Siliverstovs and van Dijk (2003) and Rapach and Wohar (2006) and adjust the weighted statistic using the Harvey *et al.* (1997) correction factor to obtain the modified W-DM statistic (MW-DM). We again use the Student-*t* distribution to determine significance.

Analyzing Interval Forecasts

We follow Wallis (2003) and Rapach and Wohar (2006) in analyzing interval forecasts and use the likelihood ratio (LR) tests developed by Christoffersen (1998), who argues that good interval forecasts include good coverage and independently distributed observations over time falling inside or outside of the forecast intervals to prevent clustering. Christoffersen (1998) develops likelihood ratio tests of unconditional coverage, independence, and conditional coverage. We use the Pearson χ^2 versions of these tests, as Wallis (2003) advocates. These statistics include indicator variables that equal one if the actual observation appears in the interval forecast; zero otherwise. We analyze these statistics with contingency tables or matrices, where we compare the observed number of outcomes to the expected number under the appropriate null hypothesis. We follow Wallis (2003) and calculate exact *p*-values based on the observed and expected outcomes using Mehta and Patel (1998). This allows sharper inference, especially for a small number of out-of-sample forecasts. We

[6] We use the Newey and West (1987) procedure with the Bartlett kernel to compute the M-DM statistic.

modify the above procedure to accommodate autocorrelation in the optimal forecasts at horizon h, since the indicator variables used to construct the Pearson χ^2 statistics will also exhibit autocorrelation for the optimal forecasts. We follow Siliverstovs and van Dijk (2003) and Rapach and Wohar (2006), who use the procedure based on Bonferroni bounds, as Diebold *et al.* (1998) suggest. This procedure divides the indicator variable series into h independent sub-groups under the null hypothesis. We then apply the χ^2 tests to each of the h subgroups and reject the relevant null hypothesis for a given test at an overall significance level of α, if we reject the null hypothesis for any of the sub-groups at the α/h significance level. Proceeding in this way can severely restrict the number of indicator variables in each sub-group as h increases, so that we place practical limits on the maximum h we can consider. The declining number of indicator variables available in each sub-group as h increases also helps to motivate our use of exact p-values for inference.

Analyzing Density Forecasts

Diebold *et al.* (1998) develop a method for analyzing density forecasts, using the probability integral transform (PIT). Under the null hypothesis that the density forecast generated by a given forecasting model is true, Diebold *et al.* (1998) demonstrate that the PIT series is distributed *i.i.d. U(0, 1)*. Following Clements and Smith (2000), Siliverstovs and van Dijk (2003), and Rapach and Wohar (2006), we use the Kolmogorov–Smirnov statistic (KS) to test for uniformity. Berkowitz (2001) suggests transforming the PIT series using the inverse of the standard normal cumulative density function. Then, under the null hypothesis that the density forecast is true, the transformed PIT series is distributed *iid N(0, 1)*. Following Clements and Smith (2000), Siliverstovs and van Dijk (2003), and Rapach and Wohar (2006), we test for standard normality using the Doornik and Hansen (1994) statistic (DH).

The KS and DH statistics assume independence. To test explicitly for independence in the PITs, Diebold *et al.* (1998) recommend looking for autocorrelation in the power-transformed PIT series. Following Siliverstovs and van Dijk (2003) and Rapach and Wohar (2006), we use the Ljung–Box (LB) statistic to test for first-order autocorrelation in the power transformed PIT series. Finally, when $h > 2$, we proceed as described above in analyzing interval forecasts and divide the PITs into h independent sub-groups under the null hypothesis. We then apply the KS, DH, and LB tests to each of the h subgroups and reject the null hypothesis at an overall significance level of α, if we reject the null hypothesis for any sub-group at the α/h significance level.

4. Data

The National Association of Realtors (NAR) calculates median and mean (average) prices for the nation and four census regions on a monthly basis. Due to the nature of the distribution of home sales prices, the mean sales price usually exceeds the median price. Although slight seasonal patterns exist in the sales price data, the NAR does not seasonally adjust the data, because the seasonal patterns prove difficult to model. Since home price data is nonstationary, we compute annual natural logarithmic differences in the house price indexes to approximate growth rates to induce stationarity. That is, $r_t = \Delta_{12} \ln P_t = \ln P_t - \ln P_{t-12}$, where P_t is the median home price. We seasonally adjust the data in levels using the Census X-12 method. Figure 1 plots the seasonally adjusted level of the median home sale prices and Figure 2 plots the annual growth rates r_t for the four Census Regions and the US. The analysis uses monthly data over the 1968:1 to 2000:12 in-sample period, and forecasts over the 2001:1 to 2010:5 out-of-sample period. We also compare ex-ante forecasts from 2010:6 to 2012:6.[7]

5. Empirical Findings

This section first considers the LM-STR test for linearity of housing price growth rates and then conducts hypothesis tests to select between the LSTAR and ESTAR models. Once we select the appropriate STAR model, we then estimate the specific STAR model and the linear AR model and compare the in-sample performance over 1968:1 to 2000:12. When conducting the (LM-STR) test for linearity, as discussed above, we choose the optimal lag, *p*, based on the unanimity of at least two of popular lag-length selection tests. We allow the delay lag, *d*, to vary between $1 \leq d \leq 8$. We estimate the optimal delay lag *d* based on the lowest *p*-value or highest *F*-statistic associated with the null hypothesis: H_{01}'' $\phi_{2i} = \phi_{3i} = \phi_{4i} = 0$ for all *i*.

Table 1 indicates delay lags of 3, 8, 1, 1, and 5 for the US, the Northeast, the Midwest, the South, and the West, respectively. Moreover, we reject the null hypothesis of linearity for the US, the Northeast, and the South at the 1-, 5-, and 1-percent levels, respectively. We can only reject the null hypothesis of linearity for the Midwest and the West at the 20-percent level by the LM_3 test. Since Escribano and Jordá (1999) propose four LM tests of linearity, we also report the LM_1, LM_2, and LM_4 test with the following null hypothesis: H_{01} $\phi_{2i} = 0$, H_{01}^{m} $\phi_{2i} = \phi_{3i} = 0$, and H_{01}^{m} $\phi_{2i} = \phi_{3i} = \phi_{4i} = \phi_{5i} = 0$ for all *i,*

---

[7] The four Census regions and the included states are described as follows: Northeast: Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode Island, and Vermont; Midwest: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin; South: Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and West Virginia; and West: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming.

respectively. In this case, we add the West to the Census regions where we can reject the null hypothesis of linearity at the 5-percent level using the LM_4 test and Midwest using the LM_1.[8]

We now need to specify the appropriate STAR model to capture accurately the non-linear dynamics. As proposed by Teräsvirta and Anderson (1992), we need to test for the sequence of nested hypothesis tests H_{04}, H_{03}, and H_{02} for the choice between LSTAR and ESTAR alternatives. Then we implement the H_{0E} and H_{0L} tests proposed by Escribano and Jordá (1999). Table 2 reports the findings. The Teräsvirta and Anderson (1992) method selects the LSTAR model for all the US and the four Census regions. Applying the Escribano and Jordá (1999) test, we also select the LSTAR model in each case, except for the Northeast, where we select the ESTAR model. Comparing the two methods, however, we see that the p-value for the Teräsvirta and Anderson (1992) method proves better than the p-value for the Escribano and Jordá (1999) test. Thus, we choose to adopt the LSTAR model, which implies that house price growth rates exhibit asymmetric dynamics during the phases of contraction and expansion.[9]

Next, we provide further evidence of nonlinearity by providing in-sample comparison based on the estimation of the linear AR model, given in equation (10), and the nonlinear LSTAR model described in equation. (11):

$$r_t = [\phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i}] + u_t \text{, and} \tag{10}$$

$$r_t = [\phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i}] + [\rho_0 + \sum_{i=1}^{p} \rho_i r_{t-i}][1 + \exp\{-\frac{\gamma}{\sigma(r_t)}(r_{t-d} - c)\}]^{-1} + u_t \text{,} \tag{11}$$

[8] In this case, however, the delay lag changes for the Midwest to 8. We can still only reject linearity for the Midwest at the 20-percent level.

[9] Although we can reject linearity with lowest p-value by the LM_4 test for West, which has power against an ESTAR alternative, both Teräsvirta and Anderson (1992) and Escribano and Jordá (1999) methods select an ESTAR model. Given the consistency of both tests, we prefer to specify an ESTAR model for the West Census region.

Teräsvirta (1994) argues that the joint estimation of (γ, c, ϕ_0, ρ_0, ϕ_i, and ρ_i) in the LSTAR model proves difficult, since difficulties arise in the estimation of c and γ. For large γ, we require a large number of observations in the neighbourhood of c to precisely estimate γ. That is, relatively large changes in γ produce only minor effects on shape of $F(.)$. Thus, the estimate for γ may converge slowly. Note that if γ is statistically insignificant, then equation (11) becomes the linear AR model. Following Teräsvirta (1994), we standardize the exponent of the function $F(.)$ of the LSTAR model by multiplying it by the term $1/\sigma(r_t)$, where $\sigma(r_t)$ is the standard deviation of the corresponding yearly housing price growth rate r_t.

Tables 3 and 4 report the results of estimating the LSTAR and AR models, respectively, where we estimate the LSTAR model using nonlinear least-squares (NLS).[10] We use a general-to-specific method to drop insignificant (worse than 10-opercent level) coefficients, but imposing the condition that the adjusted R-squared does not fall. Thus, some insignificant coefficients at the 10-percent level remain in the final models. The logistic function conditions the autoregressive parameters to change smoothly with lagged realized changes in the growth rates of home prices in the LSTAR model, which generates the endogenous nonlinearity. When we compare the estimation results over the period of 1968:1 to 2000:12 of the AR and the LSTAR models, the following features confirm the dominance of the non-linear estimation: (a) The standard errors and the log likelihood values of the nonlinear regression show improvements over those corresponding from the linear regression; (b) The adjusted R^2 values in the nonlinear regression exceed the corresponding

[10] Following the suggestions of van Dijk *et al.* (2002), we examine a battery of misspecifications tests -- no residual autocorrelation parameter constancy, no remaining non-linearity, no autoregressive conditional heteroskedasticity (ARCH), besides the test of normality -- for the LSTAR model. The estimated LSTAR models for the US and its four Census regions do not exhibit any type of misspecification. These results are available on request.

values under the linear regression, implying that a portion of variance in the housing price growth rates in the long-run associates with nonlinear dynamics; (c) Many estimates of the coefficients of the nonlinear portion of equation (11) (i.e., ρ_i's), prove statistically significant; and (d) The speed of adjustment between regimes, γ, proves statistically positive at the 10-percent level or better only for the US and the Northeast Census region. The statistical significance of γ confirms the presence of nonlinearity outlined by the LSTAR model. The estimate of γ, however, does not generally prove precise. Thus, its insignificance does not invalidate the nonlinearity, which we support by the formal tests in Table 2.

These results together provide evidence that the LSTAR model appropriately captures the inherent non-linearity in the long horizon housing price growth rates in the US and the four Census regions housing markets. Thus, a linear model would introduce misspecification, since it does not allow the dynamics of home price growth rates to evolve smoothly between regimes depending on the sign and magnitude of past realization of home price growth rates.[11]

Note that we estimate a relatively small γ for all the categories of housing price growth rates. Relatively small estimates of γ, given that the estimate varies from zero to infinity, suggest a slower transition from one regime to another, which, in turn, contrasts with the TAR or Markov-switching models that witness sudden switches between regimes. The

[11] The Ramsey model specification test provides further evidence of nonlinearity in the housing price growth rates of the US and the four Census regions. We reject the null hypothesis for a linear AR model specification, against a nonlinear LSTAR model, at the 1-percent level of significance for all cases. Note the appropriate F-statistic for the test is: $\frac{(R^2_{nonlinear} - R^2_{linear})/m}{(1-R^2_{nonlinear})/(n-k)}$, where $R^2_{nonlinear}$ (R^2_{linear}) is the R^2 of the LSTAR (AR) model, m denotes the number of restrictions in the linear AR model, and k measures the number of parameters in the LSTAR model. The values of the F-statistic for the US and the four Census regions (Northeast, Midwest, South, and West) equal, respectively, 4.19, 7.30,3.51, 3.47, and 2.77..

parameter *c,* which equals the half-way point between regimes,[12] is positive for the US and all four Census regions, although insignificantly so for the South, indicating that similar values of the housing price growth rate shock trigger a shift in regimes.

5. Forecast Accuracy

Given that we estimate the house price growth rates in the US and its four Census regions using the LSTAR model, this section compares the point, interval, and density forecast performances of the non-linear model with those of the classical linear AR models.

Point Forecasts

Table 5 reports the out-of-sample point forecast evaluation results for the LSTAR and linear AR models for the US and the four Census regions. Columns 2 and 6 report the root mean squared forecast error (RMSFE) for the linear AR model, and columns 3 and 7 report the ratio of the RMSFEs of the LSTAR model to the linear AR model, the relative RMSFE. The relative RMSFE exceeds one for short horizons for the US and each Census region, indicating that the point forecasting performance of the linear AR dominates that of the LSTAR model at short horizons. At longer horizons, the LSTAR models' performance improves, reaching a minimum between 18 and 36 months out. More specifically, the US reaches a minimum of the relative RMSFE at 30 months; the Northeast and the Midwest, at 12 months; the South, at 18 months; and the West, at 30 and 36 months. Then the relative RMSFEs generally increase to the end of the forecast horizon at 48 months, except for the US, which reaches a peak at 30 months.

[12] The parameter c denotes the value for which $G(s_t; \gamma, c)$=.5 at $s_t = c$. Therefore, the process switches monotonically towards Regime 1 as s_t increases. Thus, two regimes exhibit equal weights at the threshold value c and switching occurs exactly at c.

Columns 4 and 8 report the modified Diebold and Mariano (M-DM, 1995) test statistic for the null hypothesis that the linear AR model MSFE equals the LSTAR model MSFE against the alternative hypothesis that the linear AR model MSFE exceeds the LSTAR model MSFE. The numbers in brackets to the right of the M-DM statistic are *p*-values based on the Student's *t* distribution at any horizon. Reasons exist for caution in the use of inferences based on the Student's *t* distribution for the M-DM statistics in Table 5. When h = 1, McCracken (2004) shows that the Diebold and Mariano (1995) statistic exhibits a non-standard limiting distribution when comparing forecasts from two nested linear models. When $h \geq 2$, Clark and McCracken (2004) also show that the Diebold and Mariano (1995) statistic exhibits a nonstandard limiting distribution that depends on nuisance parameters. Thus, we cannot calculate critical values, and they recommend using a bootstrap procedure to generate critical values. The bootstrapped *p*-values appear in braces under the M-DM statistics and reflect 2000 bootstrap simulations.

Columns 5 and 9 report the modified weighted Diebold and Mariano (MW-DM, 1995) test statistic for the null hypothesis that the linear AR model weighted MSFE equals the LSTAR model weighted MSFE against the alternative hypothesis that the linear AR model weighted MSFE exceeds than the LSTAR model weighted MSFE. The MW-DM statistics place greater weight on forecasting house price growth rates farther out in the tails of the unconditional distribution. Columns 5 and 9 also report *p*-values for the MW-DM statistics based on the Student's *t* distribution in square brackets to the right of the MW-DM statistics and bootstrapped *p*-values in braces below the MW-DM statistics.

The findings for the M-DM and MW-DM statistics parallel each other nicely. The LSTAR model provides significantly better point forecasts at the 10-percent level, generally at longer horizons. The *p*-values based on the Student's *t* provide more evidence of LSTAR superiority at more lag lengths than the bootstrapped *p*-values, where significantly better performance by the LSTAR models occurs at the 36, 42, and 48 month horizons. Overall, robust evidence exists in Table 5 that the LSTAR model offers forecasting gains at long horizons relative to simple linear AR models for the US and the four Census regions – the Northeast, Midwest, South and West. No robust evidence exists that the LSTAR models offer forecasting gains at short horizons for the US or the four Census regions.

Interval Forecasts

Table 6 enumerates the Pearson χ^2 statistics used to evaluate interval forecasts for the LSTAR and linear AR models for h =1, 2, 3, and 4. Following Wallis (2003) and Rapach and Wohar (2006), we consider the inter-quartile interval forecasts (i.e., the 0.25 and 0.75 quantiles). For both the LSTAR and linear AR models for the US, we reject correct unconditional coverage at all four reported horizons (see columns 4) and we only reject correct conditional coverage at all four horizons for the linear AR model but only for the 1-, 2-, and 3-month horizons for the LSTAR model. In addition, we can reject independence only at the 1- and 3-month horizons for the linear AR model and at the 3-month horizon for the LSTAR model.

The four Census tracts tell different stories. The best performance occurs for the Northeast. Here, we cannot reject independence at any horizon except for the 3-moth horizon for the LSTAR and linear AR models. Further, we reject correct unconditional coverage at the 1-, 2-, and 4- month horizons for the linear AR model and at the 1- and 2-month horizons for the LSTAR model. Finally, we can reject the correct conditional coverage at the 1- and 3-month horizons for the linear AR and LSTAR models.

The worst performance occurs for the Midwest and the South. For the Midwest, we reject the correct unconditional and conditional coverage at all horizons for both the linear AR and LSTAR models. But, we cannot reject the independence at any horizon for the linear AR and LSTAR models, except for the LSTAR model at the 3-month horizon. The findings for the South match those for the Midwest, except that we cannot reject correct conditional coverage for the 4-month horizon for the LSTAR model and we cannot reject independence at any horizon.

The West provides the most disparate set of findings from the rest. We can reject independence for the 1-. 3-, and 4-month horizons for the LSTAR model and only at the 1- and 3-month horizons for the linear AR model. We also cannot reject the correct unconditional coverage at any horizon for the AR and LSTAR models and the correct conditional coverage at the 4-month horizon for the linear AR model and at the 2-month horizon for the LSTAR model.

In sum, we do not find strong evidence to support the LSTAR model specification over the linear AR specifications. In general, both models produce similar findings with regard to interval forecasts.

Density Forecasts

Table 7 lists the density forecast evaluation findings for the linear AR and LSTAR models for the US and the four Census regions across h =1, 2, 3, and 4. Several observations emerge. First, the results differ across the US and its Census regions. The Ljung-Box tests for no first-order autocorrelation generally reject the null hypothesis of independent PITs for k =2 and 4 across the four Census regions, whereas this test rejects independent PITs for k = 1, 2, 3, and

4 for the US (see columns 6, 7, 8 and 9). These rejections suggest deficiencies in the density forecasts for both the linear AR and LSTAR models.

Second, the KS statistics in column 4 of Table 7 provide support for the LSTAR model over the linear AR model for the US, since the KS statistic proves significant at all reported horizons for the linear AR model but insignificant at all horizons for the LSTAR models. For four Census regions, the KS statistic is not significant for either model at any reported horizon.

Third, the DH statistic proves significant at the 1-month horizon for the linear AR model for the US, whereas the DH statistic proves significant for the linear AR model for the Midwest and West Census regions at 4- and 2-month horizons, respectively. For the Northeast and the South, the DH statistics prove significant for both the linear AR and LSTAR models at varying horizons.

In sum, Table 7 provides limited evidence that the LSTAR model dominates the linear AR model in density forecasting, but only for the US as a whole. Almost no evidence exists supporting this conclusion at the Census region level. That is, the linear AR and LSTAR models produce similar forecasting performance.

6. Comparing In-Sample Conditional Densities and Ex-ante Forecasts

The forecast comparisons in the previous section show that nonlinear AR models only generate slightly better forecasts for some series in terms of interval and density forecasts and only generate better point forecasts at forecast horizons greater than 36 months. Diebold and Nason (1990) list several explanations for failure of nonlinear models to generate better forecasts than their linear counterparts, even though they fit the data better and formal statistical tests strongly reject linearity. They note that slight conditional mean nonlinearities may not produce differences until one uses a large number of observations. We examine why the LSTAR models do not produce notable superior forecasts, following the suggestion made by Pagan (2002) and Breunig *et al*. (2003), and evaluate the conditional expectations functions of fitted LSTAR and AR models for r_t, given the regime switching variable r_{t-d}. We will see how close the nonlinear and linear AR models are in terms of their conditional means, given r_{t-d}. We can evaluate the conditional mean given any lagged value of r_t. In our case, r_{t-d} is a natural choice, since this delay best captures the nonlinearity.

We can evaluate the conditional mean functions of the linear AR models straight from the fitted models. Pagan (2002) suggests for nonlinear models using a large number of simulations from the fitted model to evaluate the conditional mean function. (2002) suggests that a useful informal evaluation fits a nonlinear model, defines its forecasting performance, on the conditional mean function, given a conditioning variable. In our case, this translates into evaluating $E(r_t|r_{t-d})$ against r_{t-d}. Ordering the data according to the magnitude of the conditioning variable r_{t-d} rather than time makes the comparison more sensible. To evaluate the conditional mean function of fitted LSTAR models, we generate 63,000 simulations from each model and discard first 3,000 to remove the burn-in effect. We draw the errors from the actual residuals of the fitted models rather than an assumed distribution.

Figure 3 displays the conditional mean functions of linear (dashed line) and nonlinear (solid line) models given r_{t-d} sorted according to the magnitude of r_{t-d}. We superimpose a scatterplot of annual growth rate of house price r_t against the switch variable r_{t-d} of the estimated LSTAR model in the plots. We generate conditional expectation functions of the fitted LSTAR models by 60,000 bootstrap simulations of the fitted model and estimated using Nadaraya-Watson kernel regression. We choose the kernel regression bandwidth using the least-squares cross validation and a second order Gaussian kernel. Figure 3 gives a good idea on why the LSTAR models do not generate superior forecasts. For the Midwest Census region, linear and nonlinear conditional mean functions are almost the same except for low values where a slight nonlinearity exists. This probably explains, indeed, the non-rejection of linearity in Kim and Bhattacharya (2009) for this series. Only some slight deviation exists from the linearity for the US series and significant deviations in the negative growth rate region. The conditional mean function of the LSTAR model deviates noticeably form the linear conditional mean function in the center of the data only for the South and West Census regions. Interestingly, highly noticeable nonlinearity exists for the Northeast Census region for growth rates higher than 10-percent.

Although Figure 3 usefully compares the conditional mean functions, it does not give any information on the density (or strength) of the various regions in the plots. We gain more insight by considering the density of the conditional mean function and the switch variable. Figure 4 plots the kernel density estimate of the conditional mean function of the fitted LSTAR and the switch variable r_{t-d}. We estimate the kernel densities from 60,000 bootstrap simulations of the fitted models using the Nadaraya-Watson kernel estimator. We choose the kernel regression bandwidth using the least-squares cross validation and a second order Gaussian kernel. The density plots in Figure 4 reveal that a highly dense region exists at the low growth rates for the Northeast, Midwest, and South Census regions, as well as for the US series. The density at low values, where deviation from linearity is particularly prevalent, is high for South and Midwest. A strong peak exists, but dense in a narrow range, at the negative growth rate for Midwest . Actually, this dense range causes the rejection of linearity, otherwise the series behaves close to a linear process. For the West Census region, we see high density at extreme positive growth rates, which radically differs from the other series. For the US series, peaks exist in all regions where there are deviation from linearity, naturally expected as the US series aggregates all Census regions.

Combining the information from Figure 3 and 4, we clearly see why nonlinear AR models do not strongly dominate linear ones. Except for the South and West Census regions, we observe the nonlinearity more in those periods where extremes house price changes occur. Also for the South and West Census regions, nonlinearity exists around the center of the data as well, but these associate with less density than the extremes. Given that nonlinearity dominates usually on the extremes and forecasts even from nonlinear but stationary models return to mean, nonlinear and linear models will produce similar forecasts. This will hold even though the nonlinear models fit and describe the data better.

To compare the fitted AR and STAR models more formally, we follow Rapach and Wohar (2006) and employ the analysis of Corradi and Swanson (2003), who recently developed a formal test of nonlinear (STAR) and linear AR models. Their test provides a distributional analog of the mean squared error metric. This test permits the comparison of the conditional densities for r_t given x_t, where x_t is the vector of lagged r_t values reported in Table 3 and 4 for each model, corresponding to two different fitted models (i.e., LSTAR and linear AR models), each of which may contain some misspecification. More specifically, we use the Corradi and Swanson (2003) Z_T statistic to test the null hypothesis that the conditional densities corresponding to the fitted LSTAR and linear AR models generate equal accuracy relative to the true conditional density against the alternative hypothesis that the conditional density corresponding to the LSTAR model proves more accurate than the conditional density corresponding to the linear AR benchmark model. We compute the Z_T

statistic by integrating over a fine grid running from the minimum to the maximum values of the in-sample r_t observations. A second test statistic, $R–Z_T$, integrates over two grids of values comprising the first and fourth quartiles of the in-sample observations. Thus, in this latter case, we focus our comparison of the conditional distributions corresponding to the fitted LSTAR and linear AR models in the tails of the distributions of in-sample r_t observations. For both tests, we generate bootstrapped critical values using 2,000 replicates with the block bootstrapping method.

Table 8 reports the Corradi and Swanson (2003) test results for the fitted LSTAR and its linear AR counterpart. Following Corradi and Swanson (2003), our inferences rely on block bootstrapped critical values. The Z_T statistics reported in column 2 do not reject the null hypothesis of equal conditional density accuracy for the LSTAR models in the US or its four Census regions relative to the AR benchmark models. This indicates that the conditional densities for r_t given x_t corresponding to the LSTAR models do not significantly differ in accuracy from the conditional densities corresponding to linear AR benchmark models. In addition, limiting our focus to the first and fourth quartiles, the $R–Z_T$ statistic rejects the null hypothesis for none of the series. In sum, the findings in Table 8 imply that fitted LSTAR models generally conform closely to fitted linear AR models. This conclusion matches nicely the fact that the typical point and density forecasts generated by the LSTAR models do not improve much on forecasts generated by linear AR models at short horizons (see Section 5 above).

As a last exercise, we compare the forecasting performance of linear and nonlinear models in an ex-ante dynamic forecasting design. Although the data actually exist for the period that we consider, we use a dynamic forecasting design and do not utilize the actual data for forecasting. Figure 5 plots the 25-step dynamic point forecasts (dashed line) for r_t from the estimated linear AR models for the period 2010:6 to 2012:6 and fan charts formed from 50- to 95-percent interval forecasts. We also plot (solid line) the actual data over 2009:5 to 2012:6. Similarly, Figure 6 plots the forecasts from the LSTAR models. For the LSTAR models, we generate each point forecast by 2,000 parametric bootstrap and we use an additional 2,000 bootstrap simulations to obtain interval forecast for each time point. We calculate the interval forecasts using the highest density region estimator of Hyndman (1996). For the point forecasts, the LSTAR models do better than the linear AR models for the West and Northeast Census regions. Indeed, forecasts for these two regions are exceptionally good. The linear AR model generates poorer forecasts for the West region. For the US, Midwest, and South regions, the AR and LSTAR models generate forecasts that probably do not dominate each other. The LSTAR model certainly performs well for the US series until 2011:12, where an upward trend starts in house prices. For the interval forecasts, both linear AR and LSTAR models do offer good coverage of the actual data. The 95-percent confidence bands almost always cover the actual values. The LSTAR models, however, do in general show narrower interval forecasts, particularly for the Northeast and West regions. Notably, the linear and nonlinear AR models produce the worst forecasts for the Midwest.

7. Conclusion

A large number of recent papers show that a strong link exists between the housing market and economic activity. In addition, these papers also highlight that house-price movements lead real activity, inflation, or both. Given this, models that forecast house price movements can give policy makers insight as to the direction the economy might head and, hence, can improve the design of appropriate policies. Hence, good policy requires that one first deduce the underlying nature of the data-generating process for house prices (i.e., whether linear or non-linear), since presuming that house prices follow a linear process can lead to incorrect forecasts for not only house prices, but the economy, in general.

This paper considers several issues. First, we test housing prices in the US and its four Census regions to see if they conform to nonlinear or linear AR models. We estimate the models using monthly data over the 1968:1 to 2000:12 in-sample period, and forecasts over the 2001:1 to 2010:5 out-of-sample period. That analysis chooses the LSTAR model as the best non-linear specification. In other words, the LSTAR model dominates the ESTAR model.

Second, we compare the one- to 48-month-ahead out-of-sample forecasting performances of the LSTAR model with the linear AR model for point forecasts in the out-of-sample period. We find that the linear model performs the best at short horizons, but the non-linear model dominates at longer horizon.

Third, we do not find strong evidence to support the LSTAR model specification over the linear AR specifications. Both models produce similar findings with regard to interval forecasts, where the South region proves the major exception whereby we usually cannot reject conditional coverage for the LSTAR model, but do usually reject conditional coverage for the linear AR model.

Fourth, we find limited evidence that the LSTAR model dominates the linear AR model in density forecasting, but only for the US as a whole. Almost no evidence exists supporting this conclusion at the Census region level. That is, the linear AR and LSTAR models produce similar forecasting performance.

Fifth, our paper finds that the LSTAR model dominates the linear AR model at point forecasting at longer horizons, but that the linear AR model dominates at shorter horizons. Moreover, little evidence exists suggesting that either model dominates in interval and density forecasting.

Finally, in an ex-ante dynamic 25-step dynamic forecasting design over 2010:6 to 2012:6, we find that the LSTAR model dominates the linear AR model for the Northeast and West regions, as well as for the US. Although both the LSTAR and linear AR models generate interval forecasts with good coverage, the LSTAR models, in general, experience narrower confidence bands.

References

Abelson, P., Joyeux, R., Milunovich. G., & Chung, D., 2005. Explaining house prices in Australia: 1970-2003. *Economic Record* 81, 96-103.

Balcilar, M., Gupta, R., & Shah, Z. B., 2011. An in-sample and out-of-sample empirical investigation of the nonlinearity in house prices of South Africa. *Economic Modelling* 28, 891-899.

Bao, Y. C., Guay, C. L., & Li, S. M., 2009. A small open economy DSGE model with a housing sector. *Preliminary Draft paper prepared for the Conference of Economists*, Department of Economics, University of Melbourne.

Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics* 19, 465-474.

Bernanke, B. S., & Gertler, M., 1995. Inside the black box: The credit channel of monetary policy transmission. *The Journal of Economic Perspectives* 9, 27–48.

Bernanke, B. S., & Gertler, M., 1999. Monetary policy and asset volatility. Federal Reserve Bank of Kansas City*, Economic Review* 84, 17-52.

Bhatia, K., 1987. Real estate assets and consumer spending. *Quarterly Journal of Economics* 102, 437-443.

Borio, C. E. V., Kennedy, N., & Prowse, S. D., 1994. Exploring aggregate asset price fluctuations across countries: Measurement, determinants, and monetary policy implications. *BIS Economics Paper* No. 40.

Bradley, M. D., & Jansen, D. W., 1997. Nonlinear business cycle dynamics: Cross-country evidence on the persistence of aggregate shocks. *Economic Inquiry* 35, 495-509.

Breunig, R., Najarian, S., & Pagan, A., 2003. Specification testing of Markov switching models. *Oxford Bulletin of Economics and Statistics* 65, 703–725.

Campbell, J. Y., & Coco, J. F., 2006. How do house prices affect consumption? Evidence from microdata. *NBER Working Paper* No. 11534.

Case, K. E., 1992. The real estate cycle and the economy: Consequences of the Massachusetts boom of 1984-1987. *Urban Studies* 29, 171-183.

Case, K., Shiller, R., & Quigley, J. 2005. Comparing wealth effects: The stock market versus the housing market. *Advances in Macroeconomics* 5, 1-32.

Christensen, I., Corrigan, P., Mendicino, C. & Nishiyama, S-I., 2009. Consumption, housing collateral, and the Canadian business cycle. Bank of Canada *Working paper 2009-26*, Bank of Canada.

Christoffersen, P., 1998. Evaluating interval forecasts. *International Economic Review* 39, 841-862.

Clark, T.E., & McCracken, M. W., 2004. Evaluating long-horizon forecasts. University of Missouri at Columbia manuscript.

Clements, M. P., & Smith, J., 2000. Evaluating the forecast densities of linear and non-linear models: Applications to output growth and unemployment. *Journal of Forecasting* 19, 255-276.

Cook, S & Speight, A., 2007. Temporal dependencies in UK regional house prices. *Quantitative and Qualitative Analysis in Social Sciences*, 1(3):63-80.

Corradi, V., & Swanson, N. R., 2003. A test for comparing multiple misspecified conditional distributions. Rutgers University, manuscript.

Diebold, F. X., Gunther, T., & Tay, A., 1998. Evaluating density forecasts with applications to financial risk management. *International Economic Review* 39, 863-883.

Diebold, F. X., & Mariano, R. S., 1995. Comparing predictive accuracy. *Journal of Business and Economics Statistics* 13, 253-263.

Diebold, F. X., & Nason, J. A., 1990. Nonparametric exchange rate prediction? *Journal of International Economics* 28, 315–332.

Doornik, J. A., & Hansen, H., 1994. An omnibus test for univariate and multivariate normality. Nuffield College, manuscript.

Eitrheim, O., & Teräsvirta, T., 1996. Testing the adequacy of smooth transition autoregressive models. *Journal of Econometrics* 74, 59–75.

Elliott, J. W., 1980. Wealth and wealth proxies in a permanent income model. *Quarterly Journal of Economics* 95, 509-535.

Enders, W. & Siklos, P., 2001, Cointegration and threshold adjustment. *Journal of Business & Economic Statistics* 19, 166-176.

Engelhardt, G. V., 1996. House prices and home owner saving behavior. *Regional Science and Urban Economics* 26, 313-336.

Engelhardt, G. V., 2001. Nominal loss aversion, housing equity constraints, and household mobility: Evidence from the United States. *Center for Policy Research Working Paper*, No. 42, 1–61.

Englund, P., & Ioannides, Y. M., 1997. House price dynamics: an international empirical perspective. *Journal of Housing Economics* 6, 119–136.

Escribano, A. & Jordá, O., 1999. Improved testing and specification of smooth transition regression models. In P. Rothman, ed., *Nonlinear Time Series Analysis of Economic and Financial Data*, Dordrecht: Kluwer Academic Publishers, 289–320.

Falk, B., 1986. Further evidence on the asymmetric behavior of economic time series over the business cycle. *Journal of Political Economy* 94, 1069-1109.

Forni M, Hallin, M, Lippi, M, & Reichlin, L., 2003. Do financial variables help forecasting inflation and real activity in the euro area*? Journal of Monetary Economics* 50, 1243-1255.

Genesove, D., & Mayer, C. J., 2001. Nominal loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics* 116, 1233–1260.

Ghent, A., 2009. Sticky housing and the real effects of monetary policy. *Mimeo*, Department of Real Estate**,** Zicklin School of Business, Baruch College, CUNY.

Ghent, A., & Owyang, M., 2009**.** Is housing the business cycle? Evidence from US cities, *Mimeo*, Department of Real Estate**,** Zicklin School of Business, Baruch College, CUNY.

Granger, C. W. J., & Teräsvirta, T., 1993. *Modelling nonlinear economic relationships*. Oxford University Press, Oxford.

Green, R., 1997. Follow the leader: How changes in residential and non-residential investment predict changes in GDP. *Real Estate Economics* 25, 253–270.

Gupta, R., & Das, S., 2010. Predicting downturns in the US housing market: A Bayesian approach. *The Journal of Real Estate Finance and Economics* 41, 294-319.

Hamilton, J. D., 1989. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica* 57, 357–384.

Hansen, B. E. 2006. Interval forecasts and parameter uncertainty. *Journal of Econometrics* 135, 377-398.

Harvey, D., Leybourne, S., & Newbold, P., (1997). Testing the equality of prediction mean squared errors. *International Journal of Forecasting* 13, 281-291.

Hyndman, R. J., 1996. Computing and graphing highest density regions. *American Statistician* 50, 120-126

Iacoviello, M., 2005. House prices, borrowing constraints, and monetary policy in the business cycle. *American Economic Review* 95, 739–764.

Iacoviello, M., & Neri S., 2010. Housing market spillovers: Evidence from an estimated DSGE model. *American Economic Journal: Macroeconomics* 2, 125-164.

Kim, S-W., & Bhattacharya. R., (2009). Regional housing prices in the USA: An empirical investigation of nonlinearity. *Journal of Real Estate Finance and Economics* 38, 443-460.

Leamer, E. E., 2007. Housing is the business cycle. In *Housing, Housing Finance, and Monetary Policy*, Economic Symposium Conference Proceedings, Kansas City Federal Reserve Bank, 149-233.

Levin, L., 1998. Are assets fungible? Testing the behavioral theory of life-cycle savings. *Journal of Economic Organization and Behavior* 36, 59-83.

Luukkonen, R., Saikkonen, P., & Teräsvirta, T., 1988. Testing linearity against smooth transition autoregressive models. *Biometrika* 75, 491–499.

McCracken, M. W., (2004). Asymptotics for out-of-sample tests of Granger causality. University of Missouri, Columbia, manuscript.

Mehta, C. R., & Patel, N. R., (1998). Exact inference for categorical data. In P. Armitage, & T. Colton (Eds.), *Encyclopedia of biostatistics* (pp. 1411 – 1422). Chichester: John Wiley.

Muellbauer, J., & Murphy, A., 1997. Booms and busts in the UK housing market. *Economic Journal* 107, 701–727.

Neftci, S., 1984. Are economic time series asymmetric over the business cycle? *Journal of Political Economy* 92, 307-328.

Newey, W. K., & West, K. D., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* 55, 703-708.

Pagan, A., 2002. Learning about models and their fit to data. *International Economic Journal* 16, 1–18.

Pariès, M. D., & Notarpietro, A., 2008.Monetary policy and housing prices in an estimated DSGE model for the US and Euro area. *Working paper series no 972,* European Central Bank.

Pavlidis, E., Paya, I., Peel, D., & Spiru, A., 2009. Bubbles in house prices and their impact on consumption: Evidence for the US. Department of Economics *Working paper 2009/025*. Lancaster University Management School.

Peek, J., 1983. Capital gains and personal saving behavior. *Journal of Money, Credit, and Banking* 15, 1-23.

Rapach, D. E., & Strauss, J. K., 2006. The long-run relationship between consumption and housing wealth in the Eighth District States. Federal Reserve Bank of St. Louis *Regional Economic Development* 2(2), 140-147.

Rapach, D. E., & Wohar, M. E., 2006. The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior. *International Journal of Forecasting* 22, 341-361.

Sarantis, N., 2001. Nonlinearities, cyclical behaviour and predictability in stock markets: International evidence. *International Journal of Forecasting* 17, 459–482.

Seslen, T., 2004. Housing price dynamics and household mobility decisions. Mimeo, University of Southern California.

Siliverstovs, B., & van Dijk, D., 2003. Forecasting industrial production with linear, nonlinear, and structural change models. Econometric Institute Report EI 2003-16.

Skalin, J., & Teräsvirta, T., 1999. Another look at Swedish business cycles: 1861–1988. *Journal of Applied Econometrics* 14, 359–378.

Skalin, J., & Teräsvirta, T., 2002. Modelling asymmetries and moving equilibria in unemployment rates. *Macroeconomic Dynamics* 6, 202—241.

Skinner, J., 1989. Housing wealth and aggregate saving. *Regional Science and Urban Economics* 19, 305-324.

Smets, F., 2007. Commentary: Housing is the business cycle. In *Housing, Housing Finance, and Monetary Policy*, Economic Symposium Conference Proceedings, Kansas City Federal Reserve Bank, 235-243.

Stock, J. H., & Watson, M. W., 2003. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* 41, 788-829.

Teräsvirta, T., & Anderson, H. M., 1992. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics* 7:S119–S136.

Teräsvirta, T., 1994. Specification, estimation and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89:208–218.

Tsay, R., 1989. Testing and modeling threshold autoregressive processes. *Journal of the American Statistical Association* 84:231–240.

van Dijk, D., & Franses, P. H., 2003. Selecting a nonlinear time series model using weighted tests of equal forecast accuracy. *Oxford Bulletin of Economics and Statistics* 65, 727–744.

van Dijk, D., Teräsvirta, T., & Franses, P. H., 2002. Smooth transition autoregressive models-a survey of recent developments. *Econometric Reviews*, 21(1):1-47.

Vargas-Silva, C., 2008. Monetary policy and the US housing market: A VAR analysis imposing sign restrictions. *Journal of Macroeconomics* 30, 977–990.

Wallis, K., 2003. Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts. *International Journal of Forecasting* 19, 165– 175.

Table 1 LM-STR test for linearity

Delay (*d*)	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	LM$_1$: LM Test of H_{01} : $\phi_{2i} = 0$ in equation (4) with *k*=1							
US	**1.745**	1.169	1.579	1.306	1.332	1.377	0.848	0.943
(p^*=15)	**(0.041)**	(0.295)	(0.077)	(0.196)	(0.180)	(0.156)	(0.623)	(0.516)
Northeast	1.413	0.898	1.475	1.356	**1.758**	1.607	1.240	**1.924**
(p^*=14)	(0.144)	(0.561)	(0.118)	(0.173)	**(0.044)**	(0.075)	(0.244)	**(0.023)**
Midwest	1.484	1.538	**1.658**	1.509	1.379	1.119	1.426	1.280
(p^*=17)	(0.098)	(0.080)	**(0.049)**	(0.089)	(0.144)	(0.333)	(0.122)	(0.203)
South	1.422	1.523	1.085	0.965	0.987	0.873	0.828	0.748
(p^*=13)	(0.147)	(0.107)	(0.371)	(0.486)	(0.464)	(0.582)	(0.630)	(0.715)
West	0.834	1.050	1.453	1.101	0.884	0.978	0.788	0.623
(p^*=14)	(0.632)	(0.403)	(0.127)	(0.355)	(0.577)	(0.475)	(0.682)	(0.846)
	LM$_2$: LM Test of H'_{01} : $\phi_{2i} = \phi_{3i} = 0$ in equation (4) with *k*=2							
US	**1.748**	**1.568**	**2.233**	**2.500**	**1.809**	**1.570**	0.932	1.176
(p^*=15)	**(0.011)**	**(0.033)**	**(0.000)**	**(0.000)**	**(0.007)**	**(0.032)**	(0.572)	(0.246)
Northeast	1.378	0.997	0.962	1.100	**1.770**	**1.759**	1.439	**1.768**
(p^*=14)	(0.100)	(0.473)	(0.525)	(0.336)	**(0.011)**	**(0.012)**	(0.074)	**(0.011)**
Midwest	1.265	1.131	1.117	1.034	1.089	0.964	1.174	1.355
(p^*=17)	(0.155)	(0.289)	(0.306)	(0.421)	(0.342)	(0.530)	(0.239)	(0.096)
South	1.471	1.362	1.020	1.050	0.924	0.857	0.951	0.974
(p^*=13)	(0.068)	(0.115)	(0.441)	(0.400)	(0.575)	(0.670)	(0.535)	(0.503)
West	1.356	1.245	1.355	1.283	1.296	1.084	0.717	0.854
(p^*=14)	(0.112)	(0.187)	(0.113)	(0.158)	(0.149)	(0.355)	(0.855)	(0.683)
	LM$_3$: LM Test of H''_{01} : $\phi_{2i} = \phi_{3i} = \phi_{4i} = 0$ in equation (4) with *k*=3							
US	1.361	1.193	**1.877***	**1.813**	**1.711**	1.236	0.940	0.915
(p^*=15)	(0.071)	(0.196)	**(0.001)**	**(0.002)**	**(0.005)**	(0.154)	(0.585)	(0.629)
Northeast	**1.485**	0.982	0.972	1.180	**1.519**	**1.458**	1.257	**1.600***
(p^*=14)	**(0.033)**	(0.507)	(0.525)	(0.217)	**(0.026)**	**(0.039)**	(0.142)	**(0.014)**
Midwest	1.229*	1.038	1.131	1.107	1.094	0.919	1.113	1.166
(p^*=17)	(0.151)	(0.410)	(0.264)	(0.298)	(0.317)	(0.633)	(0.289)	(0.218)
South	**1.785***	1.427	0.994	1.029	0.751	0.725	0.724	0.899
(p^*=13)	**(0.004)**	(0.054)	(0.486)	(0.428)	(0.860)	(0.888)	(0.889)	(0.646)
West	1.082	1.211	1.185	1.136	1.211*	0.846	0.630	1.001
(p^*=14)	(0.344)	(0.184)	(0.211)	(0.269)	(0.184)	(0.741)	(0.965)	(0.475)
	LM$_4$: LM Test of H'''_{01} : $\phi_{2i} = \phi_{3i} = \phi_{4i} = \phi_{5i} = 0$ in equation (4) with *k*=4							
US	1.186	**1.442**	**1.784**	**1.591**	**1.619**	1.178	0.886	0.94
(p^*=15)	(0.182)	**(0.026)**	**(0.001)**	**(0.007)**	**(0.005)**	(0.191)	(0.709)	(0.602)
Northeast	1.367	1.069	1.092	1.151	**1.486**	**1.763**	**1.331**	**1.840**
(p^*=14)	(0.053)	(0.355)	(0.316)	(0.229)	**(0.020)**	**(0.001)**	**(0.070)**	**(0.001)**
Midwest	1.043	0.865	1.109	0.892	0.974	0.809	1.094	1.196
(p^*=17)	(0.398)	(0.761)	(0.279)	(0.710)	(0.538)	(0.851)	(0.305)	(0.161)
South	**1.582**	1.296	1.163	1.093	0.647	0.761	0.791	1.258
(p^*=13)	**(0.010)**	(0.096)	(0.219)	(0.318)	(0.971)	(0.884)	(0.847)	(0.123)
West	1.057	1.210	1.245	1.051	**1.486**	1.065	0.811	0.962
(p^*=14)	(0.376)	(0.161)	(0.128)	(0.387)	**(0.020)**	(0.362)	(0.828)	(0.555)

Note: The numbers in parenthesis equal the lowest *p*-values associated with the corresponding null hypothesis. The minimum p-value associated with the LM$_3$ test (H''_{01} : $\phi_{2i} = \phi_{3i} = \phi_{4i} = 0$ in equation (4) with the corresponding *d*) is marked with *. Bold values indicate significance at the 5-percent level. p^* equals the lag order in the linear AR model selected by the AIC.

Table 2 Test of the appropriate STAR model

	US	Northeast	Midwest	South	West
H_{04} : $\varphi_{4i} = 0$,	1.137	0.993	1.142	0.946	0.861
i* = 1,...,*p	(0.322)	(0.460)	(0.313)	(0.505)	(0.602)
H_{03} : $\varphi_{3i} = 0$,	1.397	0.479	0.608	0.956	1.242
given $\varphi_{4i} = 0$	(0.147)	(0.943)	(0.885)	(0.495)	(0.243)
H_{02} : $\varphi_{2i} = 0$,	**1.579**	**1.475**	**1.658**	**1.085**	**1.453**
given $\varphi_{3i} = \varphi_{4i} = 0$	**(0.077)**	**(0.118)**	**(0.049)**	**(0.371)**	**(0.127)**
H_{0E} : $\varphi_{3i} = 0$,	1.272	**1.171**	1.089	1.235	0.976
$\varphi_{5i} = 0$	(0.161)	**(0.256)**	(0.344)	(0.203)	(0.504)
H_{0L} : $\varphi_{2i} = 0$,	**1.329**	0.997	**1.118**	**1.369**	**1.199**
$\varphi_{4i} = 0$	**(0.123)**	(0.473)	**(0.307)**	**(0.112)**	**(0.230)**
Optimal delay *d*	3	8	1	1	5
Optimal lag *p*	15	14	17	13	14
Selection model	LSTAR	LSTAR	LSTAR	LSTAR	LSTAR

Note: The values in parentheses equal the *p*-values for the nested tests H_{04}, H_{03}, and H_{02}; and the H_{0E} and H_{0L} tests. H_{0E} and H_{0L} equal the model selection tests recommended in Escribano and Jordá (1999) and obtained from equation (4) with *k*=4 for the corresponding restrictions. Bold values indicate the lowest *p*-value for the nested and the H_{0E} and H_{0L} tests. The model selection reflects the nested H_{04}, H_{03}, and H_{02} tests.

Table 3 Estimation of the STAR Model

US: Adjusted R-Square = 0.897 , SER = 0.000112 , LLV = 1154.496

$$r_t = 0.018(0.003) + 0.608(0.053)r_{t\text{-}1} + 0.284(0.067)r_{t\text{-}2} - 0.214(0.116)r_{t\text{-}3} - 0.327(0.100)r_{t\text{-}5} + 0.204(0.105)r_{t\text{-}6}$$
$$+ 0.247(0.086)r_{t\text{-}10} - 0.191(0.088)r_{t\text{-}11} - 0.550(0.056)r_{t\text{-}12} + 0.282(0.057)r_{t\text{-}13} + 0.068(0.045)r_{t\text{-}15}$$
$$+ [-0.015(0.004) + 0.224(0.111)r_{t\text{-}3} + 0.431(0.125)r_{t\text{-}5} - 0.203(0.121)r_{t\text{-}6} - 0.365(0.104)r_{t\text{-}10}$$
$$+ 0.332(0.109)r_{t\text{-}11} + 0.137(0.071)r_{t\text{-}14}]\times[1+\exp\{-2.751(1.179)[r_{t\text{-}3} - 0.037(0.003)]\}]^{-1} + \varepsilon_t$$

Northeast: Adjusted R-Square = 0.799 , SER = 0.000797 , LLV = 795.017

$$r_t = 0.004(0.002) + 0.545(0.067)r_{t\text{-}1} + 0.200(0.057)r_{t\text{-}2} + 0.125(0.049)r_{t\text{-}4} + 0.144(0.049)r_{t\text{-}8} - 0.488(0.053)r_{t\text{-}12}$$
$$+ 0.182(0.061)r_{t\text{-}13} + 0.217(0.059)r_{t\text{-}14} + [0.095(0.033) + 0.195(0.098)r_{t\text{-}1} - 0.705(0.177)r_{t\text{-}6}$$
$$+ 0.119(0.083)r_{t\text{-}12}] \times [1+\exp\{-7.523(1.482)[r_{t\text{-}6} - 0.153(0.008)]\}]^{-1} + \varepsilon_t$$

Midwest: Adjusted R-Square = 0.831 , SER = 0.000192 , LLV = 1049.457

$$r_t = 0.007(0.008) + 1.241(0.308)r_{t\text{-}1} - 0.823(0.321)r_{t\text{-}2} + 0.534(0.280)r_{t\text{-}4} + 0.639(0.300)r_{t\text{-}5}$$
$$- 0.709(0.314)r_{t\text{-}6} + 0.477(0.214)r_{t\text{-}7} - 0.377(0.222)r_{t\text{-}10} + 0.389(0.118)r_{t\text{-}11}$$
$$+ 0.368(0.055)r_{t\text{-}13} - 0.844(0.277)r_{t\text{-}14} + 0.414(0.200)r_{t\text{-}17} + [-0.006(0.009) - 0.631(0.302)r_{t\text{-}1}$$
$$+ 1.137(0.328)r_{t\text{-}2} - 0.536(0.291)r_{t\text{-}4} - 0.643(0.317)r_{t\text{-}5} + 0.801(0.327)r_{t\text{-}6} - 0.453(0.230)r_{t\text{-}7}$$
$$+ 0.473(0.233)r_{t\text{-}10} - 0.409(0.133)r_{t\text{-}11} - 0.586(0.054)r_{t\text{-}12} + 0.937(0.286)r_{t\text{-}14}$$
$$- 0.427(0.215)r_{t\text{-}17}] \times [1+\exp\{-5.949(3.115)[r_{t\text{-}1} - 0.010(0.005)]\}]^{-1} + \varepsilon_t$$

South: Adjusted R-Square = 0.885 , SER = 0.000158 , LLV = 1097.234

$$r_t = 0.030(0.006) - 0.250(0.210)r_{t\text{-}1} + 0.504(0.151)r_{t\text{-}2} + 0.121(0.049)r_{t\text{-}4} + 0.135(0.091)r_{t\text{-}5}$$
$$+ 0.097(0.043)r_{t\text{-}6} - 0.478(0.072)r_{t\text{-}11} - 0.448(0.077)r_{t\text{-}13} + [-0.028(0.006)$$
$$+ 1.095(0.208)r_{t\text{-}1} - 0.489(0.167)r_{t\text{-}2} - 0.219(0.101)r_{t\text{-}5} + 0.429(0.092)r_{t\text{-}11} - 0.505(0.070)r_{t\text{-}12}$$
$$+ 0.971(0.092)r_{t\text{-}13}]\text{x}[1+\exp\{-4.481(3.280)[r_{t\text{-}1} - 0.003(0.004)]\}]^{-1} + \varepsilon_t$$

West: Adjusted R-Square = 0.878 , SER = 0.000474 , LLV = 891.141

$$r_t = 0.005(0.003) + 0.744(0.069)r_{t\text{-}1} + 0.185(0.049)r_{t\text{-}3} - 0.265(0.066)r_{t\text{-}12} + 0.239(0.045)r_{t\text{-}14}$$
$$+ [-0.006(0.005) - 0.310(0.087)r_{t\text{-}1} + 0.331(0.075)r_{t\text{-}2} + 0.133(0.060)r_{t\text{-}4} - 0.338(0.096)r_{t\text{-}12}$$
$$+ 0.276(0.073)r_{t\text{-}13}]\text{x}[1+\exp\{-5.474(4.172)[r_{t\text{-}5} - 0.039(0.008)]\}]^{-1} + \varepsilon_t$$

Note: The values in the parenthesis correspond to the standard errors. SER stands for the standard error of the regression, while LLV stands for the log likelihood value. We include only significant lags following Teräsvirta (1994) and Sarantis (2001).

Table 4 Estimation of the AR Model

US: Adjusted R-Square = 0.886 , SER = 0.000131 , LLV = 1125.966

$r_t = 0.002(0.001) + 0.677(0.05)r_{t\text{-}1} + 0.209(0.051)r_{t\text{-}2} + 0.097(0.051)r_{t\text{-}3} - 0.516(0.046)r_{t\text{-}12} + 0.395(0.053)r_{t\text{-}13} + 0.104(0.043)r_{t\text{-}15} + \varepsilon_t$

Northeast: Adjusted R-Square = 0.787 , SER = 0.000876 , LLV = 777.345

$r_t = 0.004(0.002) + 0.607(0.051)r_{t\text{-}1} + 0.191(0.059)r_{t\text{-}2} + 0.106(0.05)r_{t\text{-}3} + 0.074(0.042)r_{t\text{-}6} - 0.471(0.049)r_{t\text{-}12} + 0.237(0.059)r_{t\text{-}13} + 0.19(0.052)r_{t\text{-}14} + \varepsilon_t$

Midwest: Adjusted R-Square = 0.823 , SER = 0.000217 , LLV = 1035.921

$r_t = 0.003(0.002) + 0.629(0.047)r_{t\text{-}1} + 0.206(0.049)r_{t\text{-}2} + 0.136(0.047)r_{t\text{-}4} + 0.088(0.044)r_{t\text{-}10} - 0.555(0.05)r_{t\text{-}12} + 0.358(0.052)r_{t\text{-}13} + 0.085(0.039)r_{t\text{-}16} + \varepsilon_t$

South: Adjusted R-Square = 0.876 , SER = 0.000179 , LLV = 1070.874

$r_t = 0.002(0.001) + 0.832(0.032)r_{t\text{-}1} + 0.118(0.047)r_{t\text{-}4} - 0.087(0.055)r_{t\text{-}5} + 0.119(0.045)r_{t\text{-}6} - 0.48(0.045)r_{t\text{-}12} + 0.455(0.044)r_{t\text{-}13} + \varepsilon_t$

West: Adjusted R-Square = 0.873 , SER = 0.000514 , LLV = 876.167

$r_t = 0.003(0.002) + 0.555(0.05)r_{t\text{-}1} + 0.268(0.056)r_{t\text{-}2} + 0.177(0.047)r_{t\text{-}3} - 0.503(0.047)r_{t\text{-}12} + 0.241(0.056)r_{t\text{-}13} + 0.221(0.05)r_{t\text{-}14} + \varepsilon_t$

Note: The values in the parenthesis correspond to the standard errors. SER stands for the standard error of regression, while LLV stands for the log likelihood value. We include only significant lags following Teräsvirta (1994) and Sarantis (2001).

Table 5 Out-of-sample point forecast evaluation, linear AR and STAR models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
h[a]	**AR**[b]	**STAR/AR**[c]	**M-DM**[d]	**MW-DM**[e]	**AR**[b]	**STAR/AR**[c]	**M-DM**[d]	**MW-DM**[e]
	US, 2001:1-2010:5 out of sample period				***Northeast, 2001:1-2010:5 out of sample period***			
1	0.02	1.26	-3.71 [1.00] {0.58}	-3.12 [1.00] {0.50}	0.05	1.21	-2.58 [0.99] {0.27}	-2.33 [0.99] {0.21}
2	0.03	1.11	-1.27 [0.90] {0.73}	-1.20 [0.88] {0.54}	0.06	1.03	-0.39 [0.65] {0.24}	-1.38 [0.92] {0.43}
3	0.04	1.09	-1.01 [0.84] {0.75}	-1.06 [0.85] {0.60}	0.06	1.05	-0.44 [0.67] {0.46}	-1.57 [0.94] {0.63}
6	0.07	1.02	-0.18 [0.57] {0.90}	-0.35 [0.64] {0.82}	0.06	0.89	0.87 [0.19] {0.48}	-0.10 [0.54] {0.50}
9	0.09	0.92	0.66 [0.26] {0.95}	0.38 [0.35] {0.90}	0.08	0.83	1.53 [**0.06**] {0.53}	0.76 [0.22] {0.52}
12	0.14	0.63	2.28 [**0.01**] {0.41}	2.18 [**0.02**] {0.26}	0.09	0.68	2.13 [**0.02**] {0.35}	1.52 [**0.07**] {0.45}
18	0.37	0.26	1.75 [**0.04**] {0.22}	1.74 [**0.04**] {0.16}	0.09	0.55	1.53 [**0.07**] {0.23}	1.39 [**0.08**] {0.29}
24	0.83	0.12	1.32 [**0.09**] {0.36}	1.35 [**0.09**] {0.25}	0.10	0.67	1.33 [**0.09**] {0.18}	1.34 [**0.09**] {0.17}
30	0.85	0.12	1.25 [0.11] {0.40}	1.26 [0.11] {0.35}	0.11	0.86	1.48 [**0.07**] {0.15}	1.40 [**0.08**] {0.16}
36	0.48	0.23	1.31 [**0.10**] {**0.06**}	1.31 [**0.10**] {**0.05**}	0.13	0.92	1.76 [**0.04**] {**0.07**}	1.74 [**0.04**] {**0.08**}
42	0.22	0.52	1.43 [**0.08**] {**0.06**}	1.44 [**0.08**] {**0.05**}	0.14	0.92	2.11 [**0.02**] {**0.08**}	2.24 [**0.01**] {**0.08**}
48	0.13	0.9	1.35 [**0.09**] {0.11}	1.14 [0.13] {0.11}	0.15	0.92	2.88 [**0.00**] {**0.09**}	2.92 [**0.00**] {**0.09**}
	Midwest, 2001:1-2010:5 out of sample period				***South, 2001:1-2010:5 out of sample period***			
1	0.03	1.55	-4.96 [1.00] {0.88}	-4.86 [1.00] {0.92}	0.02	1.50	-4.20 [1.00] {0.56}	-3.91 [1.00] {0.34}
2	0.04	1.46	-2.55 [0.99] {0.91}	-2.68 [1.00] {0.88}	0.03	1.39	-2.46 [0.99] {0.83}	-2.50 [0.99] {0.73}
3	0.04	1.30	-1.84 [0.97] {0.86}	-2.13 [0.98] {0.82}	0.03	1.30	-1.92 [0.97] {0.81}	-2.23 [0.99] {0.70}
6	0.05	0.98	0.12 [0.45] {0.82}	-0.21 [0.58] {0.75}	0.04	0.95	0.37 [0.36] {0.87}	-0.42 [0.66] {0.80}
9	0.06	0.60	2.01 [**0.02**] {0.78}	1.94 [**0.03**] {0.51}	0.05	0.42	3.07 [**0.00**] {0.90}	2.51 [**0.01**] {0.87}
12	0.07	0.26	1.74 [**0.04**] {0.70}	1.73 [**0.04**] {0.45}	0.05	0.12	2.59 [**0.01**] {0.89}	2.34 [**0.01**] {0.82}
18	0.07	0.15	1.22 [0.11] {0.63}	1.26 [**0.10**] {0.37}	0.05	0.01	1.37 [**0.09**] {0.68}	1.43 [**0.08**] {0.81}
24	0.07	0.36	1.12 [0.13] {0.54}	1.15 [0.13] {0.33}	0.07	0.00	1.18 [0.12] {0.46}	1.20 [0.12] {0.44}
30	0.07	0.91	0.99 [0.16] {0.77}	1.17 [0.12] {0.31}	0.08	0.01	1.10 [0.14] {0.57}	1.11 [0.14] {0.53}
36	0.09	0.98	1.07 [0.14] {0.30}	1.76 [**0.04**] {**0.08**}	0.09	0.01	1.17 [0.12] {**0.06**}	1.20 [0.12] {**0.05**}
42	0.10	0.95	2.10 [**0.02**] {**0.09**}	2.71 [**0.00**] {**0.07**}	0.10	0.06	1.38 [**0.09**] {**0.03**}	1.42 [**0.08**] {**0.03**}
48	0.10	0.90	4.13 [**0.00**] {**0.08**}	5.10 [**0.00**] {**0.06**}	0.10	0.34	1.64 [**0.05**] {**0.06**}	1.66 [**0.05**] {**0.05**}

Table 5 Out-of-sample point forecast evaluation, linear AR and STAR models (continued)

(1) h^a	(2) AR^b	(3) $STAR/AR^c$	(4) $M\text{-}DM^d$	(5) $MW\text{-}DM^e$
	West, 2001:1-2010:5 out of sample period			
1	0.04	1.17	-2.99 [1.00] {0.82}	-2.18 [0.98] {0.85}
2	0.05	1.10	-1.54 [0.94] {0.91}	-1.46 [0.93] {0.92}
3	0.06	1.06	-0.81 [0.79] {0.90}	-0.95 [0.83] {0.91}
6	0.08	1.05	-0.50 [0.69] {0.84}	-0.61 [0.73] {0.86}
9	0.10	1.00	0.03 [0.49] {0.43}	-0.18 [0.57] {0.35}
12	0.12	0.95	0.57 [0.29] {0.26}	0.46 [0.32] {0.26}
18	0.14	0.89	1.38 [**0.09**] {0.59}	0.99 [0.16] {0.38}
24	0.16	0.73	1.60 [**0.06**] {0.54}	1.45 [**0.08**] {0.40}
30	0.18	0.65	1.38 [**0.09**] {0.83}	1.33 [**0.09**] {0.61}
36	0.21	0.65	1.33 [**0.09**] {0.34}	1.33 [**0.09**] {0.23}
42	0.23	0.82	1.60 [**0.06**] {0.11}	1.56 [**0.06**] {**0.08**}
48	0.24	0.91	1.33 [**0.09**] {0.13}	1.27 [0.10] {0.11}

Note: The p-values use the Student's *t* distribution with (P_h – 1) degrees of freedom and appear in square brackets. Bootstrapped *p*-values appear in braces and obtained with 2000 bootstrap simulations. Bold *p*-values indicate significance at the 10-percent level. Finally, 0.00 indicates less than 0.005 and 1.00 indicates greater than 0.995.

a Forecast horizon (in months).

b Linear AR model RMSFE.

c Ratio of the STAR model RMSFE to the linear AR model RMSFE.

d Modified Diebold and Mariano (1995) test statistic for the null hypothesis that the linear AR model MSFE equals the STAR model MSFE against the alternative hypothesis that the linear AR model MSFE exceeds the STAR model MSFE.

e Modified weighted Diebold and Mariano (1995) test statistic for the null hypothesis that the linear AR model weighted MSFE equals the STAR model weighted MSFE against the alternative hypothesis that the linear AR model weighted MSFE exceeds the STAR model weighted MSFE.

Table 6: Out-of-sample interval forecast evaluation, linear AR and STAR models

(1) Model	(2) h^{a}	(3) 0.10/*h*	(4) χ^2_{UC} [b]	(5) χ^2_{IND} [c]	(6) χ^2_{CC} [d]
A. US, 2001:1-2010:5 out of sample period					
Linear AR	1	0.10	**63.94** [0.00]	**3.78** [0.00]	**64.65** [0.00]
Linear AR	2	0.050	**44.64** [0.00], **34.57** [0.00]	0.34 [1.00], 0.34 [1.00]	**40.25** [0.00], **40.25** [0.00]
Linear AR	3	0.033	**25.97** [0.00], **19.70** [0.00], **29.43** [0.00]	6.89 [0.05], 6.89 [0.05], 6.89 [0.05]	**24.50** [0.00], **24.50** [0.00], **24.50** [0.00]
Linear AR	4	0.025	**14.29** [0.00], **11.57** [0.00], **13.37** [0.00], **16.33** [0.00]	0.55 [0.59], 0.55 [0.59], 0.46 [0.60], 0.46 [0.60]	**8.71** [0.01], **8.71** [0.01], **7.87** [0.01], **7.87** [0.01]
STAR	1	0.10	**21.25** [0.00]	1.00 [1.00]	23.13 [0.00]
STAR	2	0.050	**10.29** [0.00], **12.07** [0.00]	0.46 [0.55], 0.46 [0.55]	6.97 [0.03], 6.97 [0.03]
STAR	3	0.033	**7.81** [0.01], 4.57 [0.05], **7.81** [0.01]	5.66 [0.04], 5.66 [0.04], 5.66 [0.04]	**11.65** [0.00], **11.65** [0.00], **11.65** [0.00]
STAR	4	0.025	2.29 [0.18], 0.57 [0.46], 3.00 [0.09], **8.33** [0.01]	0.17 [0.71], 0.17 [0.71], 0.39 [0.69], 0.39 [0.69]	1.09 [0.58], 1.09 [0.58], 1.00 [0.64], 1.00 [0.64]
B. Northeast, 2001:1-2010:5 out of sample period					
Linear AR	1	0.10	**9.64** [0.00]	0.00 [1.00]	**10.32** [0.01]
Linear AR	2	0.050	**4.57** [0.04], **8.64** [0.00]	2.19 [0.18], 2.19 [0.18]	2.21 [0.36], 2.21 [0.36]
Linear AR	3	0.033	0.68 [0.51], 1.32 [0.32], 3.27 [0.10]	5.36 [0.04], 5.36 [0.04], 5.36 [0.04]	5.73 [0.06], 5.73 [0.06], 5.73 [0.06]
Linear AR	4	0.025	1.29 [0.26], 5.14 [0.04], 0.04 [0.85], **6.26** [0.01]	2.97 [0.13], 2.97 [0.13], 2.48 [0.24], 2.48 [0.24]	3.26 [0.20], 3.26 [0.20], 2.62 [0.32], 2.62 [0.32]
STAR	1	0.10	**3.19** [0.09]	2.07 [0.18]	**5.58** [0.06]
STAR	2	0.050	3.50 [0.08], **7.14** [0.01]	2.19 [0.18], 2.19 [0.18]	2.21 [0.36], 2.21 [0.36]
STAR	3	0.033	0.68 [0.51], 0.68 [0.51], 3.27 [0.10]	5.46 [0.04], 5.46 [0.04], 5.46 [0.04]	5.56 [0.08], 5.56 [0.08], 5.56 [0.08]
STAR	4	0.025	0.57 [0.46], 5.14 [0.04], 0.33 [0.70], 3.00 [0.09]	2.77 [0.13], 2.77 [0.13], 3.55 [0.11], 3.55 [0.11]	3.07 [0.24], 3.07 [0.24], 4.08 [0.14], 4.08 [0.14]
C. Midwest, 2001:1-2010:5 out of sample period					
Linear AR	1	0.10	**70.10** [0.00]	0.08 [1.00]	**69.17** [0.00]
Linear AR	2	0.050	**31.50** [0.00], 28.57 [0.00]	0.25 [1.00], 0.25 [1.00]	**40.23** [0.00], **40.23** [0.00]
Linear AR	3	0.033	**19.70** [0.00], **16.89** [0.00], **25.97** [0.00]	2.68 [0.24], 2.68 [0.24], 2.68 [0.24]	**25.82** [0.00], **25.82** [0.00], **25.82** [0.00]
Linear AR	4	0.025	**28.00** [0.00], **20.57** [0.00], **13.37** [0.00], **16.33** [0.00]	5.71 [0.15], 5.71 [0.15], 5.46 [0.15], 5.46 [0.15]	**21.16** [0.00], **21.16** [0.00], **20.17** [0.00], **20.17** [0.00]
STAR	1	0.10	**35.12** [0.00]	0.74 [1.00]	**35.53** [0.00]
STAR	2	0.050	**25.79** [0.00], **16.07** [0.00]	0.09 [1.00], 0.09 [1.00]	**30.61** [0.00], **30.61** [0.00]
STAR	3	0.033	**7.81** [0.01], **14.30** [0.00], **7.81** [0.01]	5.76 [0.05], 5.76 [0.05], 5.76 [0.05]	**19.20** [0.00], **19.20** [0.00], **19.20** [0.00]
STAR	4	0.025	**14.29** [0.00], **14.29** [0.00], 4.48 [0.05], **13.37** [0.00]	1.88 [0.22], 1.88 [0.22], 1.72 [0.24], 1.72 [0.24]	**11.84** [0.00], **11.84** [0.00], **10.91** [0.00], **10.91** [0.00]

Table 6: Out-of-sample interval forecast evaluation, linear AR and STAR models (continued)

(1) Model	(2) h^a	(3) 0.10/h	(4) χ^2_{UC} ^{b}	(5) χ^2_{IND} ^{c}	(6) χ^2_{CC} ^{d}
D. South, 2001:1-2010:5 out of sample period					
Linear AR	1	0.10	**66.98** [0.00]	0.14 [1.00]	**69.20** [0.00]
Linear AR	2	0.050	**28.57** [0.00], **25.79** [0.00]	1.17 [0.58], 1.17 [0.58]	**31.08** [0.00], **31.08** [0.00]
Linear AR	3	0.033	**19.70** [0.00], **22.73** [0.00], **29.43** [0.00]	0.19 [1.00], 0.19 [1.00], 0.19 [1.00]	**28.48** [0.00], **28.48** [0.00], **28.48** [0.00]
Linear AR	4	0.025	**20.57** [0.00], **11.57** [0.00], **23.15** [0.00], **16.33** [0.00]	0.82 [0.60], 0.82 [0.60], 0.86 [0.59], 0.86 [0.59]	**13.78** [0.00], **13.78** [0.00], **12.91** [0.00], **12.91** [0.00]
STAR	1	0.10	**32.9**3 [0.00]	0.01 [1.00]	**34.33** [0.00]
STAR	2	0.050	**12.07** [0.00], **4.57** [0.04]	0.48 [0.71], 0.48 [0.71]	15.64 [0.00], 15.64 [0.00]
STAR	3	0.033	**14.30** [0.00], **9.76** [0.00], **9.76** [0.00]	1.66 [0.39], 1.66 [0.39], 1.66 [0.39]	10.25 [0.01], 10.25 [0.01], 10.25 [0.01]
STAR	4	0.025	**7.00** [0.01], 0.57 [0.46], **10.70** [0.00], 4.48 [0.05]	0.06 [1.00], 0.06 [1.00], 0.02 [1.00], 0.02 [1.00]	1.87 [0.43], 1.87 [0.43], 1.40 [0.58], 1.40 [0.58]
E. West, 2001:1-2010:5 out of sample period					
Linear AR	1	0.10	**55.23** [0.00]	**17.58** [0.00]	**65.75** [0.00]
Linear AR	2	0.05	**14.00** [0.00], **18.29** [0.00]	1.67 [0.23], 1.67 [0.23]	**8.03** [0.02], **8.03** [0.02]
Linear AR	3	0.033	**11.92** [0.00], **9.76** [0.00], **9.76** [0.00]	3.85 [0.07], 3.85 [0.07], 3.85 [0.07]	**7.42** [0.02], **7.42** [0.02], **7.42** [0.02]
Linear AR	4	0.025	**7.00** [0.01], **9.14** [0.00], **10.70** [0.00], 4.48 [0.05]	0.27 [0.71], 0.27 [0.71], 0.13 [1.00], 0.13 [1.00]	0.60 [0.76], 0.60 [0.76], 0.29 [0.95], 0.29 [0.95]
STAR	1	0.10	**28.75** [0.00]	**10.17** [0.00]	**37.48** [0.00]
STAR	2	0.05	**12.07** [0.00], **8.64** [0.00]	1.53 [0.27], 1.53 [0.27]	3.67 [0.17], 3.67 [0.17]
STAR	3	0.033	4.57 [0.05], 3.27 [0.10], 4.57 [0.05]	**9.03** [0.01], **9.03** [0.01], **9.03** [0.01]	**9.03** [0.01], **9.03** [0.01], **9.03** [0.01]
STAR	4	0.025	1.29 [0.26], 3.57 [0.09], 3.00 [0.09], 1.81 [0.25]	**7.40** [0.01], **7.40** [0.01], 6.25 [0.03], 6.25 [0.03]	**8.72** [0.01], **8.72** [0.01], **8.12** [0.02], **8.12** [0.02]

Note: Statistics appear for each of the h subgroups. The exact *p*-value appears in brackets. Bold values mean significant at the 0.10/*h* level, according to the exact p-value. Finally, 0.00 indicates less than < 0.005.

[a] Forecast horizon (in months).

[b] Pearson χ^2 test statistic for the null hypothesis that the prediction intervals exhibit the correct unconditional coverage.

[c] Pearson χ^2 test statistic for the null hypothesis that the "hits" relating to the prediction intervals are independent.

[d] Pearson χ^2 test statistic for the null hypothesis that the prediction intervals exhibit the correct conditional coverage.

Table 7 Out-of-sample density forecast evaluation, linear AR and STAR models

(1) Model	(2) h^a	(3) 0.10/h	(4) KS[b]	(5) DH[c]	(6) LB, k=1[d]	(7) LB, k=2[d]	(8) LB, k=3[d]	(9) LB, k=4[d]
A. US, 2001:1-2010:5 out of sample period								
Linear AR	1	0.10	**0.14**	**8.01**	**8.16**	**39.84**	**6.00**	**17.42**
Linear AR	2	0.05	**0.21, 0.26**	2.38, 4.30	**4.77, 13.85**	**32.39, 25.71**	**9.31, 9.99**	**22.56, 19.52**
Linear AR	3	0.033	**0.23, 0.25, 0.29**	3.44, 3.40, 0.65	**6.10, 9.60, 5.92**	**23.43, 21.48, 12.55**	**8.17, 11.82, 4.33**	**16.57, 14.24, 7.22**
Linear AR	4	0.025	0.27, 0.24, 0.28, **0.34**	1.28, 4.78, 3.44, 0.50	**8.39, 8.09, 2.27, 11.15**	**12.33, 16.93, 13.28, 16.68**	**7.05, 7.59,** 2.69, **7.41**	**7.57, 12.00, 8.36, 10.68**
STAR	1	0.10	0.09	4.55	**8.91**	**32.92**	**5.85**	**11.81**
STAR	2	0.05	0.15, 0.18	2.87, 5.38	**7.63, 13.69**	**27.31, 24.14**	**11.48, 10.00**	**15.59, 18.54**
STAR	3	0.033	0.20, 0.17, 0.25	6.32, 4.03, 1.73	**7.87, 11.29, 7.09**	**21.30, 17.51, 11.19**	**8.39, 11.79,** 4.61	15.39, 12.03, 5.22
STAR	4	0.025	0.20, 0.21, 0.20, 0.28	3.65, 4.16, 5.09, 7.11	**9.16, 8.72,** 3.04, **10.85**	**11.16, 15.63, 13.01, 14.71**	**7.59, 6.63,** 3.30, **4.70**	**7.72, 11.67, 8.64, 6.54**
B. Northeast, 2001:1-2010:5 out of sample period								
Linear AR	1	0.10	0.10	**10.96**	0.55	**44.25**	0.91	**22.58**
Linear AR	2	0.05	0.14, 0.17	1.03, 5.05	1.93, 0.06	**32.51, 31.50**	0.30, 0.02	**23.76, 23.36**
Linear AR	3	0.033	0.14, 0.12, 0.20	7.18, **10.37,** 3.39	0.01, 0.04, 0.01	**10.90,10.52, 16.48**	0.00, 0.82, 1.27	**4.66, 5.46, 10.32**
Linear AR	4	0.025	0.17, 0.21, 0.15, 0.25	3.36, 7.18, 7.33, **8.00**	2.07, 2.48, 2.53, 0.04	**9.06, 9.73, 5.09, 22.73**	0.89, 1.20, 1.58, 0.26	4.15, **5.52,** 1.28, **18.06**
STAR	1	0.10	0.06	**11.83**	0.70	**38.24**	1.07	**17.09**
STAR	2	0.05	0.09, 0.13	4.51, 6.94	1.44, 0.03	**29.42**, **28.52**	0.03, 0.18	**19.01**, **20.11**
STAR	3	0.033	0.15, 0.10, 0.15	**10.63**, 7.16, 1.53	0.02, 0.14, 0.07	**8.84**, **8.62**, **14.07**	0.09, 1.46, 1.82	3.19, 4.18, **7.60**
STAR	4	0.025	0.13, 0.18, 0.14, 0.22	7.12, 5.58, 3.00, 1.91	1.83, 2.05, 2.28, 0.01	**7.68**, **8.65**, 3.36, **22.29**	0.37, 0.25, 0.82, 0.03	3.10, 4.96, 0.50, **17.21**
C. Midwest, 2001:1-2010:5 out of sample period								
Linear AR	1	0.10	0.18,	3.84,	0.09,	**59.79**,	0.82,	**33.93**
Linear AR	2	0.05	0.24, 0.24,	1.79, 2.41,	0.03, 3.37,	**22.49**, **20.02**,	0.05, 3.17,	**11.25**, **8.36**
Linear AR	3	0.033	0.26, 0.24, 0.25,	3.55, 1.55, 4.91,	0.20, 7.36, 0.01,	6.21, **25.53**, **18.90**,	0.22, **7.86**, 0.09,	0.47, **18.29**, **10.63**
Linear AR	4	0.025	0.24, 0.27, 0.28, 0.30,	**9.84**, 0.72, 3.41, 1.75,	1.17, 0.90, 0.12, 0.53,	**14.86**, **10.18**, **15.22**, **21.02**,	2.73, 1.25, 0.42, 0.13,	**10.97**, 4.55, **8.69**, **15.69**
STAR	1	0.10	0.13,	5.83	0.20	**53.02**	0.97,	**28.53**
STAR	2	0.05	0.21, 0.21,	4.15, 1.68	0.04, 3.63,	**21.53**, **18.82**	0.05, 3.06,	**11.34**, **7.88**
STAR	3	0.033	0.23, 0.21, 0.23	3.58, 3.07, 3.38	0.16, 7.48, 0.03	**6.01**, **23.83**, **17.35**	0.20, 7.18, 0.23	0.49, 15.74, **8.51**
STAR	4	0.025	0.22, 0.26, 0.26, 0.27,	3.38, 4.64, 4.70, 0.65	1.35, 1.03, 0.12, 0.49	**14.14**, **9.86**, **13.86**, **20.44**	3.25, 1.24, 0.29, 0.07,	**9.75**, 4.37, **7.19**, **14.27**

Table 7 Out-of-sample density forecast evaluation, linear AR and STAR models (continued)

(1) Model	(2) h^a	(3) 0.10/h	(4) KS[b]	(5) DH[c]	(6) LB, k=1[d]	(7) LB, k=2[d]	(8) LB, k=3[d]	(9) LB, k=4[d]
D. South, 2001:1-2010:5 out of sample period								
Linear AR	1	0.10	0.13	**48.33**	1.20	**28.30**	0.16	**8.21**
Linear AR	2	0.05	0.17, 0.19	**22.28**,**13.03**	0.03, 4.66	**22.50**, **30.64**	0.45, 6.26	**8.71**, **21.08**
Linear AR	3	0.033	0.21, 0.26, 0.24	**10.33**, **9.05**, 5.67	0.36, 0.91, 0.13	**16.79**, **16.30**, **11.05**,	0.00, 0.72, 0.00	**12.92**, 5.90, 1.68
Linear AR	4	0.025	0.32, 0.24, 0.29, 0.26,	2.92, 3.62, 3.53, 1.81,	0.91, 2.29, 0.30, 1.34,	**11.36**, **17.27**, **10.90**, **10.26**,	1.22, 2.57, 0.27, 1.85	2.53, **11.32**, 3.90, 5.73
STAR	1	0.10	0.11,	**12.96**	1.22	**24.15**	0.00,	**8.34**
STAR	2	0.05	0.15, 0.17,	3.57, 5.43	0.03, 5.00	**21.73**, **29.75**	0.46, 6.73	**8.17**, **19.45**
STAR	3	0.033	0.18, 0.23, 0.21	4.30, 6.01, 5.21	0.28, 1.06, 0.20	**16.15**, **15.59**, **10.25**,	0.08, 0.61, 0.01	**12.21**, 4.93, 1.41
STAR	4	0.025	0.28, 0.21, 0.25, 0.23	5.93, 3.22, 2.14, **9.19**	0.77, 2.28, 0.39, 1.72	**9.59**, **16.53**, **9.70**, **9.95**	0.69, 3.17, 0.32, 2.63	1.69, **10.30**, 2.90, 6.10
E. West, 2001:1-2010:5 out of sample period								
Linear AR	1	0.10	0.16	5.17	5.02	**52.77**	5.01	**37.17**
Linear AR	2	0.05	0.22, 0.20	**6.06**, 1.04	5.07, **7.79**	**24.71**, **30.47**	4.10, **9.36**	**11.39**, **20.83**
Linear AR	3	0.033	0.22, 0.26, 0.28	1.26, 0.63, 1.41	2.55,12.43, 3.37,	**20.14**, **23.66**, **20.16**	4.09, 11.03, 2.15	**10.82**, **16.87**, **9.77**
Linear AR	4	0.025	0.26, 0.27, 0.30, 0.29,	0.13, 0.17, 2.90, 2.39,	**7.21**,**12.68**, **7.66**, **6.82**,	**17.98**, **18.51**, **16.16**, **16.22**	**10.46**, **9.03**, 4.45, 4.99	**15.31**, **11.66**, **9.86**,
STAR	1	0.10	0.13	3.89	5.31	**49.99**	6.40	**34.09**
STAR	2	0.05	0.19, 0.18	3.72, 1.07	5.37, **8.05**	**23.33**, **28.38**	3.91, **9.28**	**10.47**, **18.03**
STAR	3	0.033	0.20, 0.24, 0.27	1.63, 0.62, 2.19	3.10, 2.77, 3.57	**18.42**, **22.36**, **18.63**	4.57, **10.99**, 2.22	**10.02**,**15.24**, **8.62**,
STAR	4	0.025	0.24, 0.24, 0.28, 0.26	0.90, 0.29, 4.13, 5.33	**7.90**,**12.70**, **7.70**, **6.86**	**17.68**, **17.28**, **14.89**, **16.05**	**11.05**, **8.45**, 4.09, 4.75	**14.64**,**10.60**, **7.93**, **11.91**

Note: Statistics appear for each of the h subgroups. Bold statistics indicate significance at the 0.10/h level. Finally, 0.00 indicates less than 0.005.

[a] Forecast horizon (in months).

[b] Kolmogorov–Smirnov test statistic for the null hypothesis that $z_t \sim U(0,1)$.

[c] Doornik and Hansen (1994) test statistic for the null hypothesis that $z_t \sim N(0,1)$.

[d] Ljung–Box test statistic for the null hypothesis of no first-order autocorrelation in $(z_t - \bar{z})^k$, $k = 1, \ldots, 4$.

Table 8 In-sample comparison of conditional densities corresponding to fitted STAR and linear AR models

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		Block bootstrap Z_T critical values				Block bootstrap $R-Z_T$ critical values		
Segment	Z_T [a]	**10%**	**5%**	**1%**	$R-Z_T$ [b]	**10%**	**5%**	**1%**
US	0.0201	0.0364	0.0415	0.0502	0.0000	0.0001	0.0001	0.0002
Northeast	0.0319	0.0491	0.0547	0.0660	0.0085	0.0131	0.0160	0.0214
Midwest	0.0169	0.0272	0.0295	0.0344	0.0000	0.0000	0.0001	0.0001
South	0.0148	0.0198	0.0230	0.0301	0.0001	0.0001	0.0002	0.0003
West	0.0158	0.0325	0.0396	0.0631	0.0034	0.0085	0.0111	0.0246

Notes: Bolded bootstrapped critical values indicate statistical significance for the test statistic at the corresponding significance level. Bootstrapped critical values are obtained using 2000 block bootstrap simulations.

[a] The Corradi and Swanson (2003) test statistic for the null hypothesis that the conditional densities corresponding to the STAR and linear AR models give equal accuracy relative to the true conditional density against the alternative hypothesis that the conditional density corresponding to the STAR model proves more accurate than the conditional density corresponding to the linear AR model.

[b] The Corradi and Swanson (2003) test statistic for the null hypothesis that the conditional densities corresponding to the STAR and linear AR models give equal accuracy relative to the true conditional density against the alternative hypothesis that the conditional density corresponding to the STAR model proves more accurate than the conditional density corresponding to the linear AR model for values of q_t in the upper and lower quartiles of the in-sample observations.

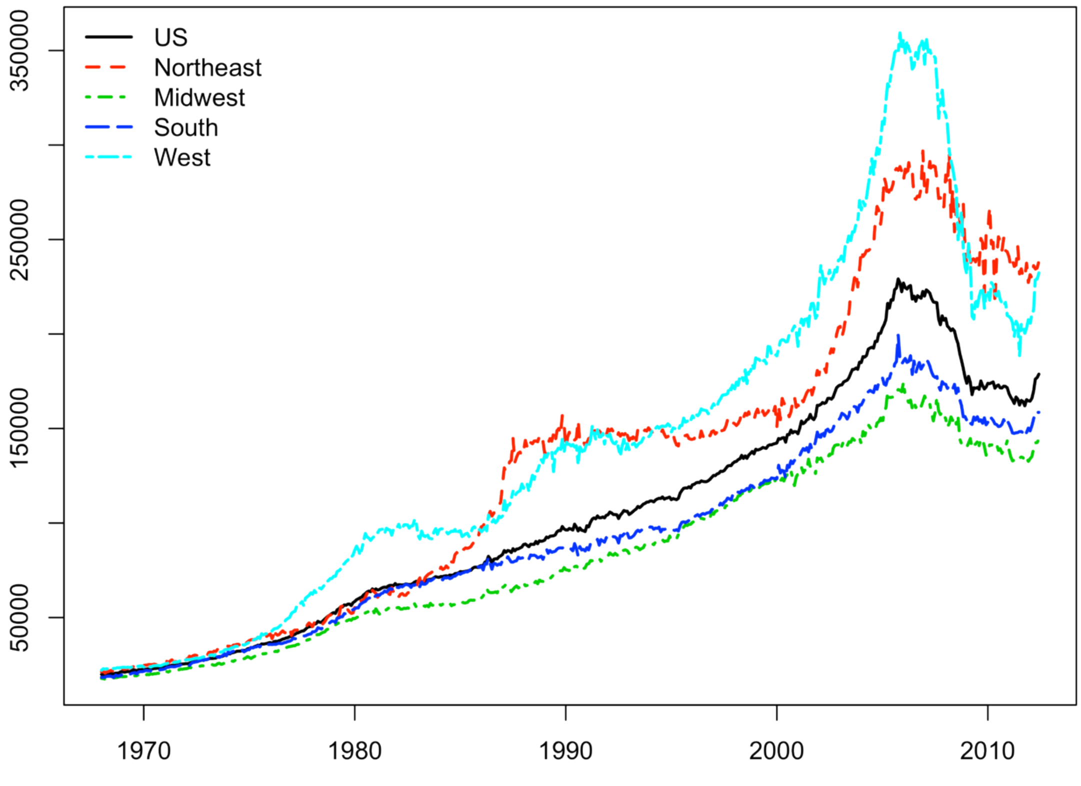

Figure 1. Median Home Price in US and the Four Regions, 1968:1–2012:6. The figure plots median home prices in dollars. All series are seasonally adjusted by the authors using X-12 filter. Source: National Association of Realtors.

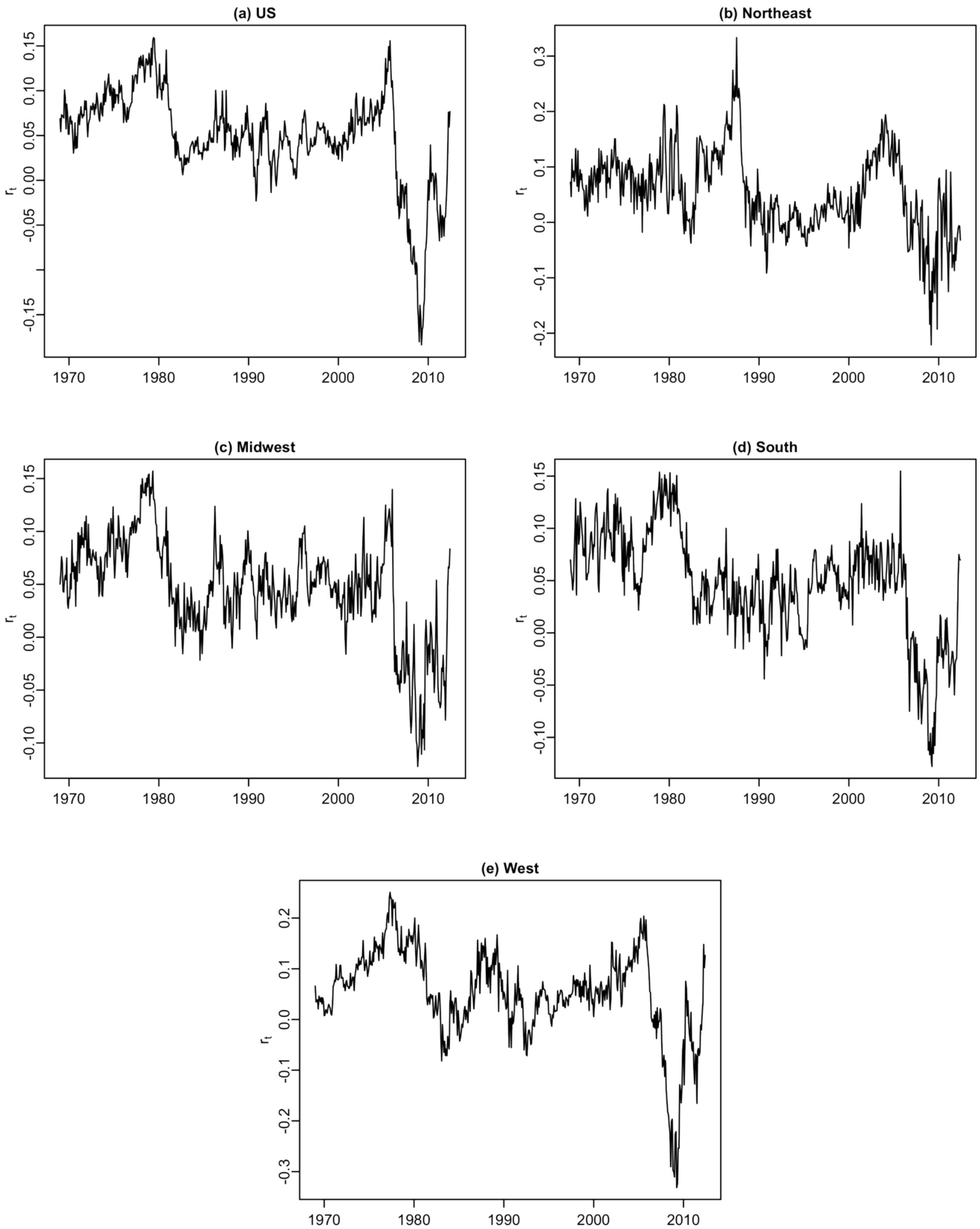

Figure 2. Seasonal first differences of logarithm of median home prices, 1969:1-2012:6. The figure plots $r_t = \Delta_{12} \log P_t = \log P_t - \log P_{t-12}$, where P_t is the median home price. The data corresponds to annual growth rate of median home prices, which were actually analyzed in the paper.

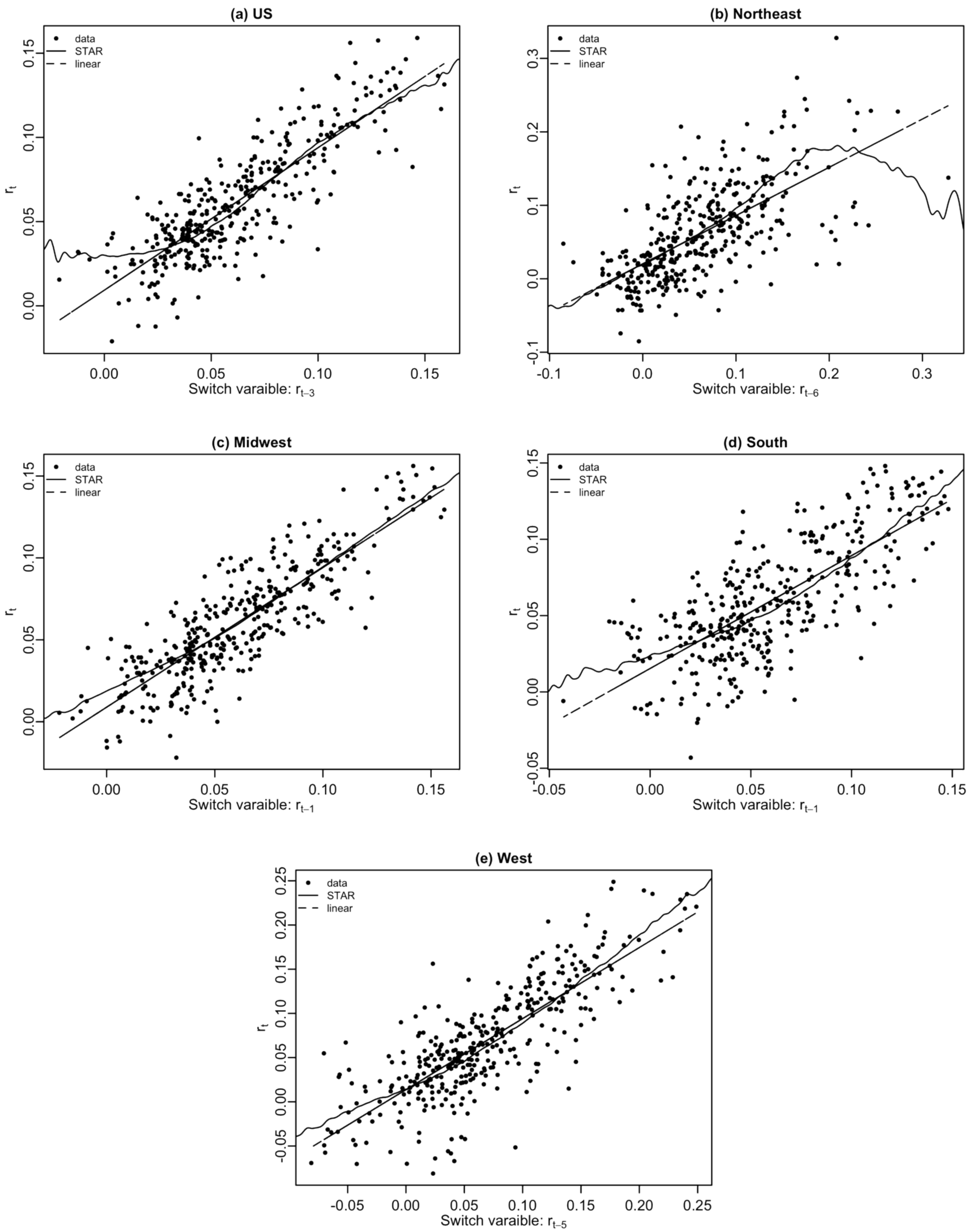

Figure 3. Scatterplot of annual growth rate of home price r_t and switch variable r_{t-d} of the estimated STAR model. Dashed straight line is the conditional expectation function of the fitted linear AR(p). Solid line is the conditional expectation function of the fitted STAR model, which is obtained by 60000 bootstrap simulations of the fitted model and estimated using Nadaraya-Watson kernel regression. The kernel regression bandwidth is chosen using the least-squares cross validation and a second order Gaussian kernel is used.

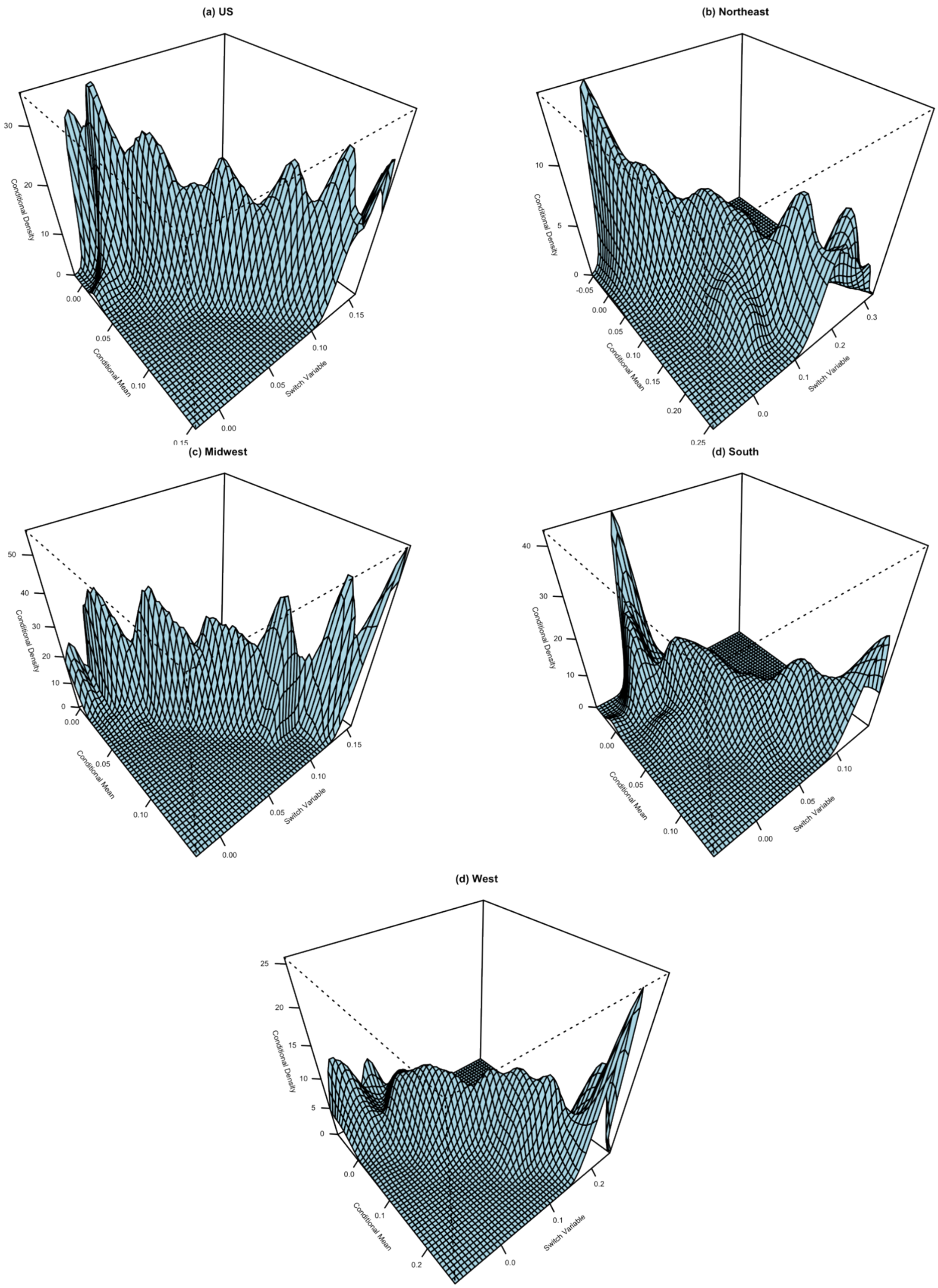

Figure 4. Kernel density estimate of the conditional expectation function of the fitted STAR and the switch variable r_{t-d}. The conditional expectation function of the fitted STAR model and the kernel density are obtained by 60000 bootstrap simulations of the fitted model and estimated using Nadaraya-Watson kernel estimator. The kernel regression bandwidth is chosen using the least-squares cross validation and a second order Gaussian kernel is used.

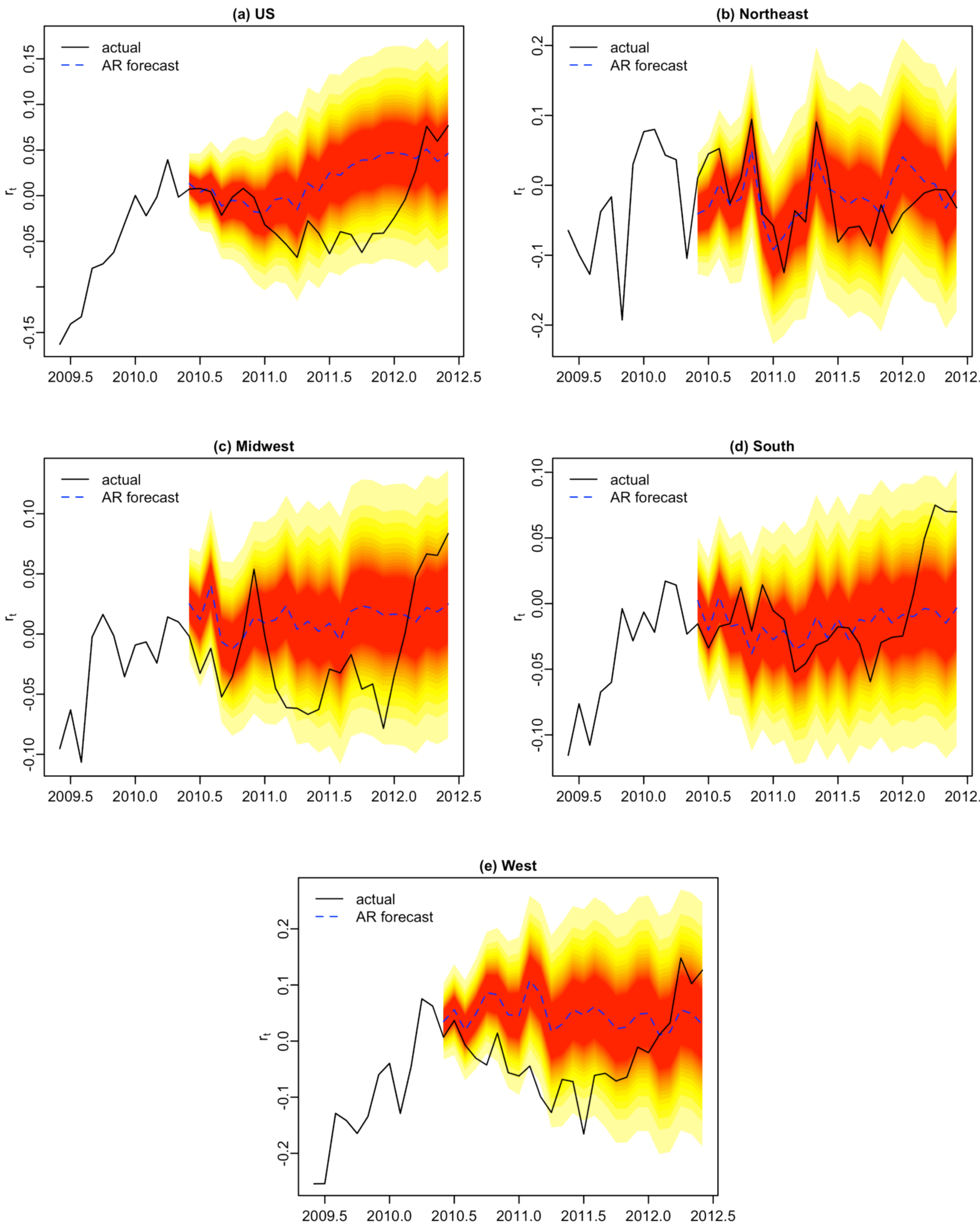

Figure 5. Point Forecast of the annual growth rate of home price r_t from the estimated linear AR(p) models for the period 2010:6-2012:6 and 50 to 95 percent interval forecasts. Dashed lines show the dynamic 25-step forecasts and solid lines show the actual data over the 2009:5-2012:6.

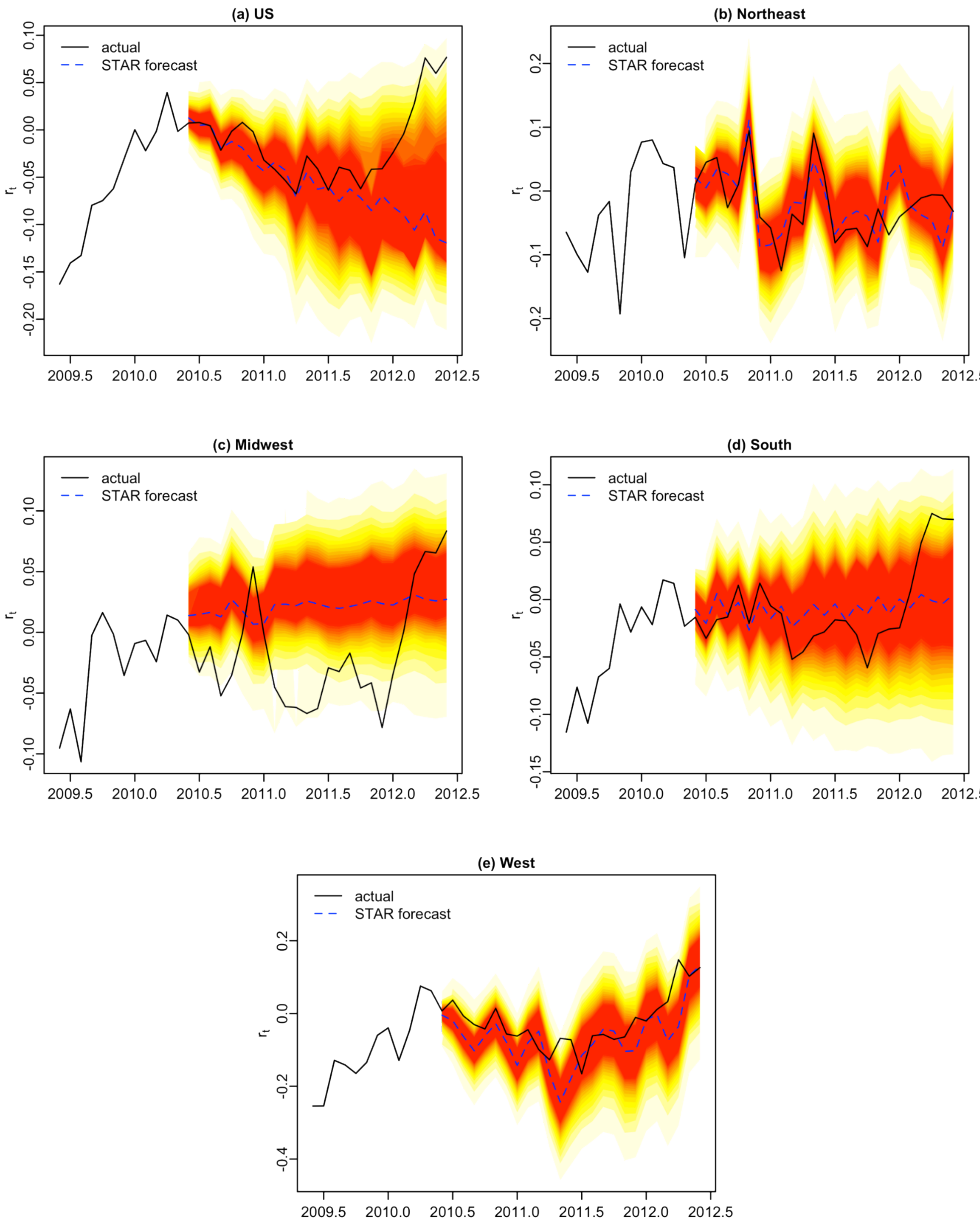

Figure 6. Point Forecast of the annual growth rate of home price r_t from the estimated nonlinear AR models for the period 2010:6-2012:6 and 50 to 95 percent interval forecasts. Dashed lines show the dynamic 25-step forecasts and solid lines show the actual data over the 2009:5-2012:6. Each point forecast is obtained by 2000 bootstrap and an additional 2000 bootstrap simulations are used to obtain interval forecast for each time point. The interval forecasts are calculated using highest density region estimator of Hyndman (1996).