

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 four Census subregions' 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 and non-linear smooth-transition autoregressive models perform equally 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. In sum, we conclude that when forecasting regional housing prices in the US, generally the additional costs associated with nonlinear forecasts outweigh the benefits for forecasts only a few months into the future.

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

JEL classification: C32, R31

* We acknowledge the helpful comments of two anonymous referees.

** *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 subregions – 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 to 2012:6 to examine the real world success of the forecasts generated from the linear AR and non-linear STAR models.

We find that the use of non-linear models to forecast housing prices at the US and four Census subregion levels typically does not generate improvements in forecast performance, especially at short horizons, to justify the additional costs of non-linear forecasts. That is, the use of misspecified linear models may still make sense, even if the data conform to a non-linear model specification.

The housing market plays a key role in the business cycle. Leamer (2007) strongly links residential housing and the business cycle, since residential investment and durable consumption prove important in explaining recessions. He argues that the stock-flow nature of the housing market and the reluctance of home owners to lower their prices in a weak market provide the setting for cyclical movement in sales volume, leading to cyclical movements in housing construction and employment. That is, construction and employment in the housing sector expand, along with increases in nominal house prices, when the economy booms, whereas nominal house prices fall sluggishly in recessions, leading to decreases in sales volume and, thus, in construction and employment activity in housing. In

sum, developers overbuild the supply of new housing during a boom, where the over building will partially determine the length of the next recession.

Good monetary policy requires action before the overbuilding goes too far and necessitates central bank intervention early in the boom period, when political pressure probably weighs against monetary policy restraint. That is, understanding and forecasting movements in the housing market plays a critical role for monetary policy authorities and their willingness to “lean against the wind.”

Further, existing evidence implies that asset prices help to forecast both inflation and output (Forni *et al.*, 2003; Stock and Watson, 2003). Moreover, for many households, their homes provide the major component of household wealth. Thus, house price adjustments can signal impending adjustments in consumption, output, and inflation. That is, movements in the housing market importantly affect the business cycle (Vargas-Silva, 2008a; Iacoviello and Neri, 2010), not only because housing investment proves a volatile component of demand (Bernanke and Gertler, 1995), but also because house price changes generate important wealth effects on consumption (International Monetary Fund, 2000) and investment (Topel and Rosen, 1988). Leamer (2007) notes that the housing market predicted 8 of the 10 post World War II recessions. If writing today, he probably would argue that the housing market predicted 9 of the 11 post World War II recessions. In other words, the housing sector acts as a leading indicator for the real sector of the economy. The recent world-wide credit crunch began with the burst of the house-price bubble, which, in turn, led the real sector of the world’s economy toward an economic slump.

Conventional wisdom argues that US housing prices adjust asymmetrically – prices adjust more quickly when rising than when falling. Recent studies (Genesove and Mayer

2001, Engelhardt 2003, Seslen 2004, Kim and Bhattacharya 2009, Balcilar *et al.* 2011) document evidence of such nonlinearity in housing prices.

Kim and Bhattacharya (2009), for example, show that housing prices in the US and three of the four Census subregions exhibit non-linearity. The Midwest, the exception, exhibits linear movements. They conclude that the behavior of the housing market does not differ across phases of expansion and contraction of the residential real estate sector, but does differ between these two phases. In the first part of our analysis, we attempt to replicate the findings of Kim and Bhattacharya (2009) in developing the non-linear model with which to perform our forecasting exercises. Our analysis, however, chooses different non-linear specifications. That is, we find that the behavior of the housing market does differ across phases of expansion and contraction of the residential real estate sector.

We then consider whether forecasting with the non-linear model leads to important improvements in forecast performance over forecasting with an incorrectly specified linear model. Moreover, we also examine this issue of the forecast performance of non-linear versus linear models for interval and density forecasts, in addition to point forecasts.

Miles (2008) considers linear and nonlinear forecasts of house prices in five US states – California, Florida, Massachusetts, Ohio, and Texas – using the generalized autoregressive (GAR) model. He concludes that the “GAR does a better job at out-of-sample forecasting ... in many cases, especially in those markets traditionally associated with high home-price volatility.” (p. 249). Cabrero, Wang, and Yang (2011) compare the out-of-sample forecasting performance of international securitized real estate returns using linear and nonlinear models. They compare the performance of a number of nonlinear models -- exponential generalized autoregressive conditional heteroskedasticity, functional coefficient, feed forward artificial

neural network, and nonparametric models -- to the benchmark linear autoregressive model. They conclude that nonlinear models produce better out-of-sample forecasts.¹

Several possible explanations for intrinsic nonlinearity in house prices exist. First, as noted above, 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 aversion with falling house prices. 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.

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

¹ They also consider combined forecasts that produce even better forecasts as well as adjust the tests for data snooping (White 2000).

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 Mariano (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).

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 Diebold *et al.* (1998). 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 statistic from Corradi and Swanson (2003) that relies 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.

Unlike the huge existing literature on forecasting house prices using linear models,² we rely on a nonlinear approach, given the theoretical reasons outlined above, and the evidence that we provide below based on statistical tests, for the inherent nonlinear data generating process of housing prices. Of courses as discussed above, a few authors do use nonlinear models. Unlike these papers, however, we not only rely on point forecasts, but also

² Gupta (2013) and Plakandaras et al., (forthcoming) provide detailed literature reviews

interval and density forecasts, both of which provide a description of forecast uncertainty. Also, we analyze the forecasting ability of our models using an ex-ante dynamic forecasting experiment to examine the real-world success of the forecasts generated, something rarely done in the literature. Also, unlike the literature that generally uses many predictors, we rely on univariate models of housing prices. We believe, given the evidence that house price is a leading indicator, we should be analyzing univariate models to forecast house prices independent of the information content of the economic fundamentals.

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 and evaluates the empirical findings. Section 5 examines forecast accuracy. Section 6 compares the in-sample conditional densities and ex-ante forecasts. Section 7 concludes.

2. Methodology

We adopt the STAR framework, developed by Luukkonen *et al.* (1988) as extended by Escribano and Jordá (1999), to model house price growth rates as non-linear and state-dependent.³ 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

³ 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 subregions 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.

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:⁴

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

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 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:

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

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

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

⁴ 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.

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 with different dynamics in the middle between the expansionary and contractionary regimes (Sarantis 2001). When $\gamma \rightarrow \infty$, the model degenerates into the conventional $TAR(p)$, while when $\gamma \rightarrow 0$, the model degenerates to the linear $AR(p)$ model (Teräsvirta and Anderson 1992).⁵

3. Point, Interval, and Density Forecasts: Method and Analysis

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 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 and consider whether these 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.

⁵ A longer version of the current paper provides details on the testing procedure for choosing between the ESTAR and LSTAR models. See http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2138980.

Generating point, interval, and density forecasts for linear AR models with Gaussian errors proves straightforward.⁶ 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 significance.

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.

⁶ 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.

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. 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 also modify the above procedure to accommodate autocorrelation in the optimal forecasts at horizon h . 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.

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 *iid*. $U(0, 1)$. 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 (2008) 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. Finally, 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.

4. Data and Empirical Findings

Data:

Following Kim and Bhattacharya (2009), we use the National Association of Realtors (NAR) median prices for the nation and four census subregions on a monthly basis. We seasonally adjust the data in levels using the Census X-12 method. Since home price data are 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. Figure 1 plots the seasonally adjusted level of the median home sale prices. 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

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. We then

⁷ The four Census subregions 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.

estimate the appropriate 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 in the LM_3 test: $H''_{01} \phi_{2i} = \phi_{3i} = \phi_{4i} = 0$ for all i . As noted by van Dijk et al. (2002), since LM tests of linearity may prove sensitive to outliers, then outlier robust estimation is preferred. Therefore, we estimate all test regressions using outlier robust M-estimation.

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 tests with the following null hypothesis: $H'_{01} \hat{f}_{2i} = 0$, $H'_{01} \hat{f}_{2i} = \hat{f}_{3i} = 0$, and $H'''_{01} \phi_{2i} = \phi_{3i} = \phi_{4i} = \phi_{5i} = 0$ for all i , respectively. In this case, we add the West to the Census subregions 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

⁸ In this case, however, the delay lag changes for the Midwest to 3. We can still reject linearity for the Midwest at the 5-percent level.

ESTAR alternatives. Then we implement the H_{OE} and H_{OL} tests proposed by Escribano and Jordá (1999). Table 2 reports the findings. The Teräsvirta and Anderson (1992) method selects the LSTAR model for the entire 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 (4), and the nonlinear LSTAR model described in equation (5):

$$r_t = [\phi_0 + \sum_{i=1}^p \phi_i r_{t-i}] + u_t, \text{ and} \quad (4)$$

$$r_t = [\phi_0 + \sum_{i=1}^p \phi_i r_{t-i}] + [\rho_0 + \sum_{i=1}^p \rho_i r_{t-i}] [1 + \exp\{-\frac{\gamma}{\sigma(r_t)}(r_{t-d} - c)\}]^{-1} + u_t, \quad (5)$$

Following Teräsvirta (1994), we standardize the exponent of the function $F(\cdot)$ of the LSTAR model by multiplying it by the term $\frac{1}{\sigma(r_t)}$, where $\sigma(r_t)$ is the standard deviation of the corresponding yearly housing price growth rate r_t .¹⁰

⁹ We test regressions estimated using outlier robust M-estimation. We also select ESTAR models for all Census subregions, using the sample period and estimation method (OLS) in Kim and Bhattacharya (2009). Thus, differences in findings reflect different estimation techniques and sample periods. The details of these results are available upon request from the authors.

¹⁰ The results from estimating LSTAR and AR models are available on request. Van Dijk *et al.* (2002) suggest 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

We use a general-to-specific method to drop insignificant (worse than 10-percent 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 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 (5) (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 subregion. 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 subregions housing markets. Thus, a linear model would introduce misspecification, since it does not allow the dynamics of home price growth rates to evolve

LSTAR model. Our estimated LSTAR models for the US and its four Census subregions do not exhibit any misspecification.

smoothly between regimes depending on the sign and magnitude of past realization of home price growth rates.¹¹

Kim and Bhattacharya (2009) find that the ESTAR model provides the best in-sample fit over their sample period from 1969:1 to 2004:7. Our in-sample period effectively runs from 1969:1 to 2000:12, since we take the annual log difference in the house price. Our findings of the LSTAR model indicate that the behavior of the housing market differs between expansionary and contractionary regimes, whereas the ESTAR models indicate similar dynamics.

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 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 subregions 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

¹¹ The Ramsey model specification test provides further evidence of nonlinearity in the housing price growth rates of the US and the four Census subregions. 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.

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 .

This subsection reports the out-of-sample point forecast evaluation results for the LSTAR and linear AR models for the US and the four Census subregions.¹³ The relative root mean squared forecast error (RMSFE) exceeds one for short horizons for the US and each Census subregion, 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. More specifically, the relative RMSFEs generally increase to the end of the forecast horizon at 48 months, except for the US, which reaches a peak earlier.

We adopt the modified (M-DM) and modified weighted (MW-DM) Diebold and Mariano (1995) to test for significant differences in forecast performance. Our null hypothesis states that the MSFE or the weighted MSFE of the linear AR equals the respective values of the LSTAR model. The alternative hypothesis states that the MSFE or weighted MSFE of the linear AR model exceeds the respective values of the LSTAR model. 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. Overall, robust evidence exists that the LSTAR model offers forecasting gains at long horizons relative to simple linear AR models for the US and the four Census subregions – 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 subregions.

Comparison to Existing Results on Nonlinear Point Forecasting

In the introduction, we listed two papers that consider nonlinear point forecasting within the housing market – Miles (2008) and Cabrero, Wang, and Yang (2011). Cabrero, Wang, and Yang (2011) forecast international securitized real estate returns. As such, this paper falls

¹³ Tables for the results of point, interval, and density forecasts appear in a longer version of this paper. See http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2138980.

most closely into the forecasting of financial assets traded in organized markets. The Miles (2008) paper forecasts house prices in five US states – California, Florida, Massachusetts, Ohio, and Texas – and, thus, most directly relates to our paper. In their longest sample that runs only through 2005, their (GAR) models outperform linear models in states with high house-price volatility such as California, whereas linear models prove the best in states with low house-price volatility such as Ohio.

Our findings at the Census region and national levels averages house prices across the states contained in the various regions. As a result, the aggregation will attenuate the volatility of house price movements in our sample. This may help to explain the differences in our findings. To wit, we find that linear models generally perform the best for shorter forecast horizons and nonlinear models sometimes perform the best for longer forecast horizons. Note that our paper considers a sample period that includes the financial crisis and Great Recession as well as interval and density forecast, to which we now turn.

Interval Forecasts

We evaluate interval forecasts for the LSTAR and linear AR models for lags 1, 2, 3, and 4 months. 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 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 regions tell different stories. The best performance occurs for the Northeast. Here, we cannot reject independence at any horizon except for the 3-month 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 performances occur 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

We consider density forecast evaluation findings for the linear AR and LSTAR models for the US and the four Census subregions across lags of 1, 2, 3, and 4 months. The results differ across the US and its Census subregions, suggesting deficiencies exist in the density forecasts for both the linear AR and LSTAR models. 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.

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 that for nonlinear models, a large number of

simulations from the fitted model evaluate the conditional mean function and that a useful informal evaluation fits a nonlinear model and 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 2 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 2 gives a good idea on why the LSTAR models do not generate superior forecasts. For the Midwest Census subregion, 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 from 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 subregion for growth rates higher than 10-percent.

Although Figure 2 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 3 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 3 reveal that a highly dense region exists at the low growth rates for the Northeast, Midwest, and South Census subregions, 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 subregion, 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 subregions.

Combining the information from Figure 2 and 3, we clearly see why nonlinear AR models do not strongly dominate linear ones. Except for the South and West Census subregions, we observe the nonlinearity more in those periods where extremes house price changes occur. Also for the South and West Census subregions, 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, 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 3 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 subregions 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 3 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 4 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 5 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 subregions. 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. Good policy requires that one first deduce the underlying nature of the data-generating process for house prices (i.e., whether linear or nonlinear), 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 subregions 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 and nonlinear models perform about the same at short horizons, but the non-linear model dominates at longer horizon.

Third, both the linear AR and LSTAR models produce similar findings with regard to interval forecasts. 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 subregion level. That is, the linear AR and LSTAR models produce similar forecasting performance.

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.¹⁴

In sum, we conclude that when forecasting regional housing prices, generally the additional costs associated with nonlinear forecasts outweigh the benefits when forecasting only a few months into the future. That is, researchers do not sacrifice much forecast

¹⁴ The ex ante forecast provides a case study of the difference between the linear AR and LSTAR models. The results may not generalize to other sample periods.

performance by adopting a linear model when, in fact, the data suggest a non-linear model. Our analysis examined the US and four Census subregions. Future research can examine the issues at a more disaggregated level -- states and metropolitan areas.

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.
- Berkowitz, J., 2001. Testing density forecasts, with applications to risk management. *Journal of Business and Economic Statistics* 19, 465-474.
- Bernanke, B., and Gertler, M., 1995. Inside the black box: the credit channel of monetary transmission. *Journal of Economic Perspectives* 9, 27-48.
- 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.
- Cabrera, J. F., Wang, T., and Yang, J., 2011. Linear and nonlinear predictability of international securitized real estate returns: A reality check. *Journal of Real Estate Research* 33, 565-594.
- Christoffersen, P., 1998. Evaluating interval forecasts. *International Economic Review* 39, 841-862.
- 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.
- Corradi, V., & Swanson, N. R., 2003. A test for comparing multiple misspecified conditional distributions. Rutgers University, [WP No. 2003-14 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 Economic 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., 2008. An omnibus test for univariate and multivariate normality. *Oxford Bulletin of Economics and Statistics* 70, Supplement, 927-939.
- Enders, W. & Siklos, P., 2001, Cointegration and threshold adjustment. *Journal of Business & Economic Statistics* 19, 166-176.
- Engelhardt, G. V., 2003. Nominal loss aversion, housing equity constraints, and household mobility: Evidence from the United States. *Journal of Urban Economics* 53, 171–195.
- 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., and Reichlin, L., 2005. The generalized dynamic factor model, one sided estimation and forecasting. *Journal of the American Statistical Association* 100, 830–840.
- Genesove, D., & Mayer, C. J., 2001. Nominal loss aversion and seller behavior: Evidence from the housing market. *Quarterly Journal of Economics* 116, 1233–1260.
- Gupta, R., 2013. Forecasting house prices for the four census regions and the aggregate US economy in a data-rich environment. *Applied Economics* 45, 4677-4697.
- 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., and 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, [Kansas City, MO](#), 149-233.
- 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.
- Miles, W., 2008. Boom-bust cycles and the forecasting performance of linear and non-linear models of house prices. *Journal of Real Estate Finance and Economics* 36, 249-264.
- 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.
- Plakandaras, V., Gupta, R., Gogas, P., and Papadimitriou, T., forthcoming. Forecasting the U.S. Real House Price Index. *Economic Modelling*.
- 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. N., 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.
- Stock, J. H., & Watson, M. W., 1999. A comparison of linear and nonlinear univariate models for forecasting macroeconomic time series. In Engle R, White H *Cointegration, Causality and Forecasting: A Festschrift for Clive W.J. Granger* Oxford: Oxford University Press, 1-44.
- Stock, J. H., and 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.
- Topel, R. H., and Rosen, S., 1988. Housing investment in the United States. *Journal of Political Economy* 96, 718–740.
- 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 signs 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.
- White, H., 2000. A reality check for data snooping. *Econometrica* 68, 1097-1126.

Table 1 LM-STR test for linearity

	US (p*=15)	Northwest (p*=14)	Midwest (p*=17)	South (p*=13)	West (p*=14)
LM₁: LM Test of $H_{01} : \phi_{2i} = 0$ in equation (A1) with $k=1$					
<i>d</i> (p-value)	1 (0.041)	8 (0.023)	3 (0.049)	2 (0.107)	3 (0.127)
LM₂: LM Test of $H'_{01} : \phi_{2i} = \phi_{3i} = 0$ in equation (A1) with $k=2$					
<i>d</i> (p-value)	4 (0.000)	5 (0.011)	8 (0.096)	1 (0.068)	1 (0.112)
LM₃: LM Test of $H''_{01} : \phi_{2i} = \phi_{3i} = \phi_{4i} = 0$ in equation (A1) with $k=3$					
<i>d</i> (p-value)	3 (0.001)	8 (0.014)	1 (0.151)	1 (0.004)	5 (0.184)
LM₄: LM Test of $H'''_{01} : \phi_{2i} = \phi_{3i} = \phi_{4i} = \phi_{5i} = 0$ in equation (A1) with $k=4$					
<i>d</i> (p-value)	3 (0.001)	8 (0.001)	8 (0.161)	1 (0.010)	5 (0.020)

Note: The delay parameter *d* is followed by the p-value in parentheses. 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
$H_{04} : \phi_{4i} = 0,$ $i = 1, \dots, p$	1.137 (0.322)	0.993 (0.460)	1.142 (0.313)	0.946 (0.505)	0.861 (0.602)
$H_{03} : \phi_{3i} = 0,$ given $\phi_{4i} = 0$	1.397 (0.147)	0.479 (0.943)	0.608 (0.885)	0.956 (0.495)	1.242 (0.243)
$H_{02} : \phi_{2i} = 0,$ given $\phi_{3i} = \phi_{4i} = 0$	1.579 (0.077)	1.475 (0.118)	1.658 (0.049)	1.085 (0.371)	1.453 (0.127)
$H_{0E} : \phi_{3i} = 0,$ $\phi_{5i} = 0$	1.272 (0.161)	1.171 (0.256)	1.089 (0.344)	1.235 (0.203)	0.976 (0.504)
$H_{0L} : \phi_{2i} = 0,$ $\phi_{4i} = 0$	1.329 (0.123)	0.997 (0.473)	1.118 (0.307)	1.369 (0.112)	1.199 (0.230)
<i>Optimal delay d</i>	3	8	1	1	5
<i>Optimal lag p</i>	15	14	17	13	14
<i>Selection model</i>	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 (A1) 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 In-sample comparison of conditional densities corresponding to fitted STAR and linear AR models

(1)	(2)	(3) (4) (5) Block bootstrap Z_T critical values			(6)	(7) (8) (9) 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.

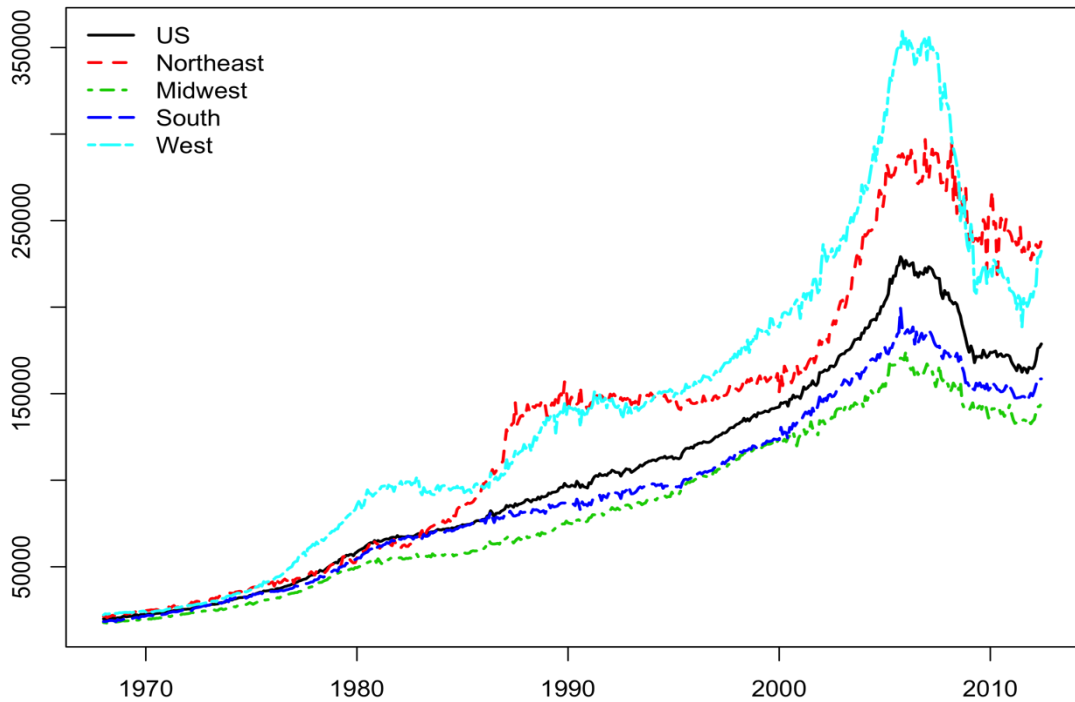


Figure 1. Median Home Price in US and the Four Regions, 1968:1–2012:6. The figure plots median home prices in dollars. All series are seasonally adjusted by the authors using X-12 filter. Source: National Association of Realtors.

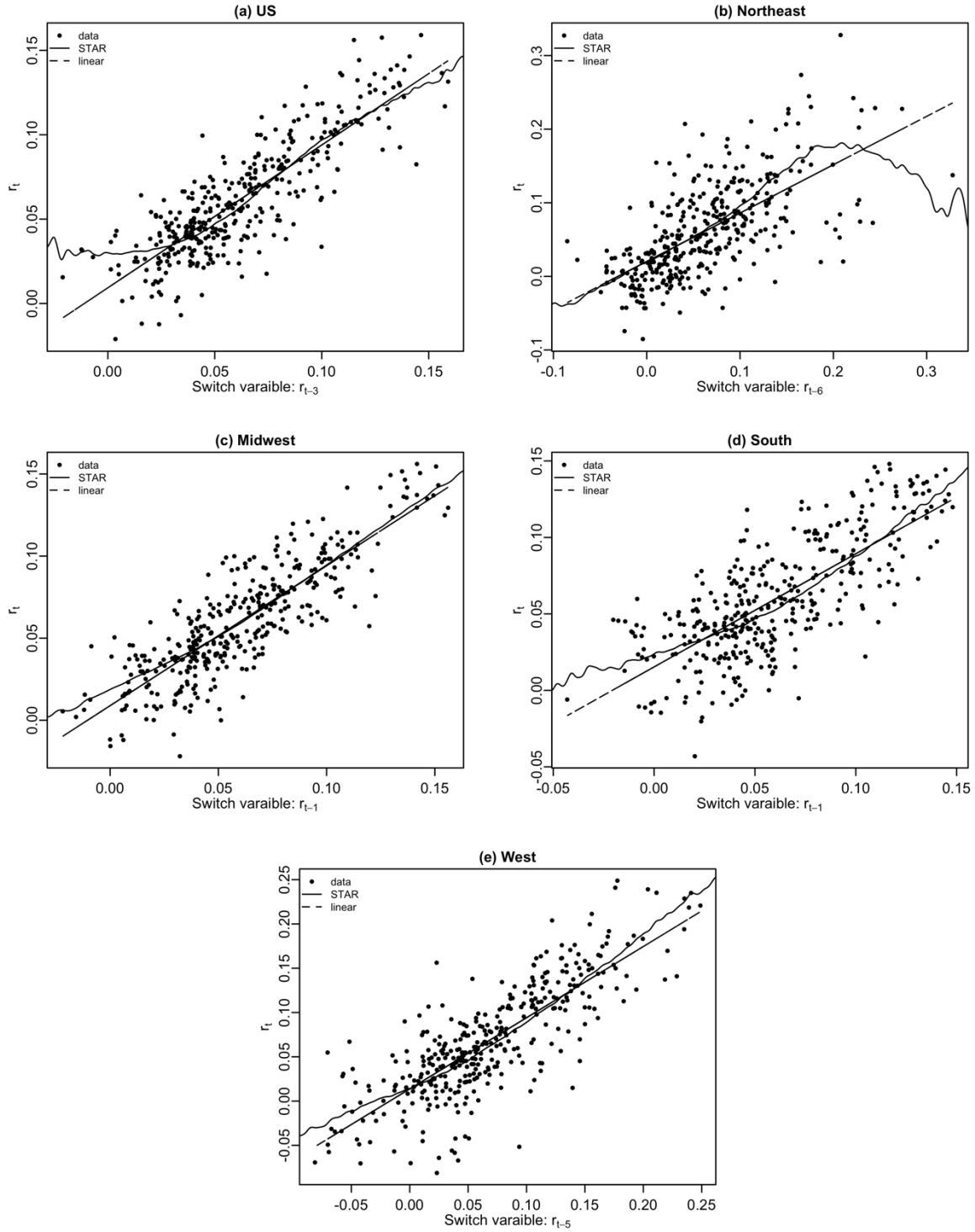


Figure 2. Scatterplot of annual growth rate of home price r_t and switch variable r_{t-d} of the estimated STAR model. Dashed straight line is the conditional expectation function of the fitted linear $AR(p)$. Solid line is the conditional expectation function of the fitted STAR model, which is obtained by 60,000 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.

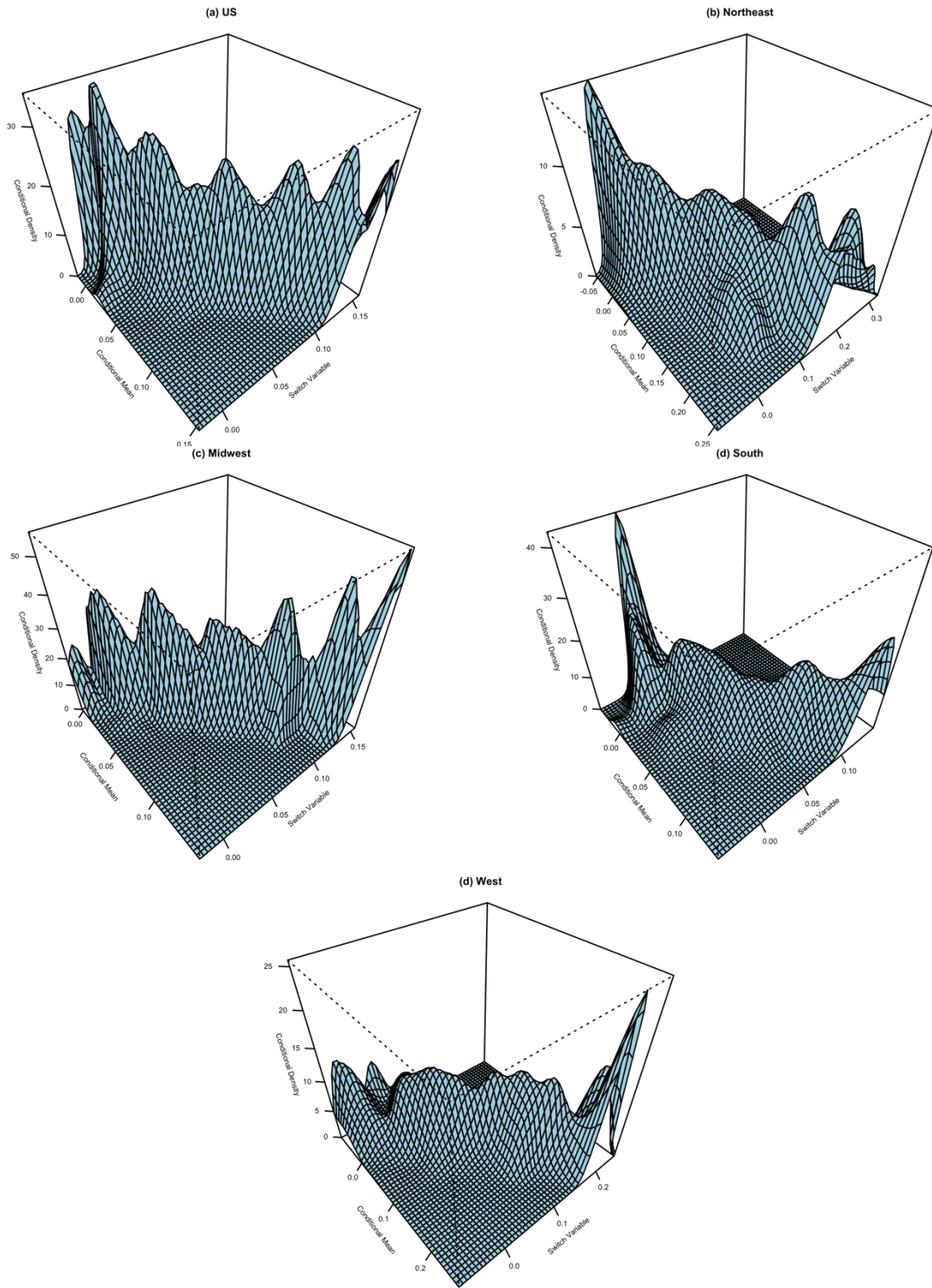


Figure 3. Kernel density estimate of the conditional expectation function of the fitted STAR and the switch variable r_{t-d} . The conditional expectation function of the fitted STAR model and the kernel density are obtained by 60,000 bootstrap simulations of the fitted model and estimated using Nadaraya-Watson kernel estimator. The kernel regression bandwidth is chosen using the least-squares cross validation and a second order Gaussian kernel is used.

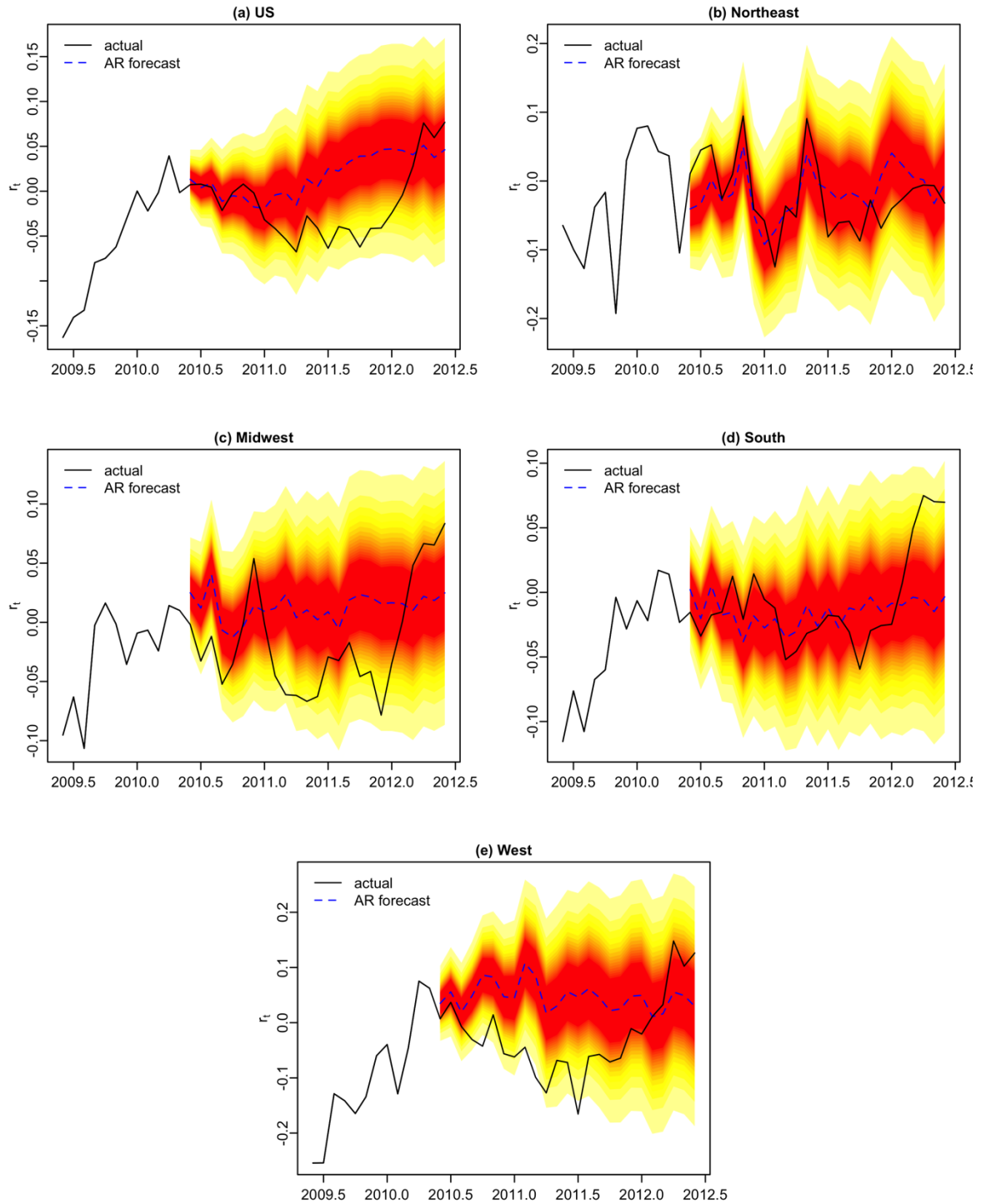


Figure 4. Point Forecast of the annual growth rate of home price r_t from the estimated linear $AR(p)$ models for the period 2010:6 to 2012:6 and 50 to 95 percent interval forecasts. Dashed lines show the dynamic 25-step forecasts and solid lines show the actual data over the 2009:5 to 2012:6.

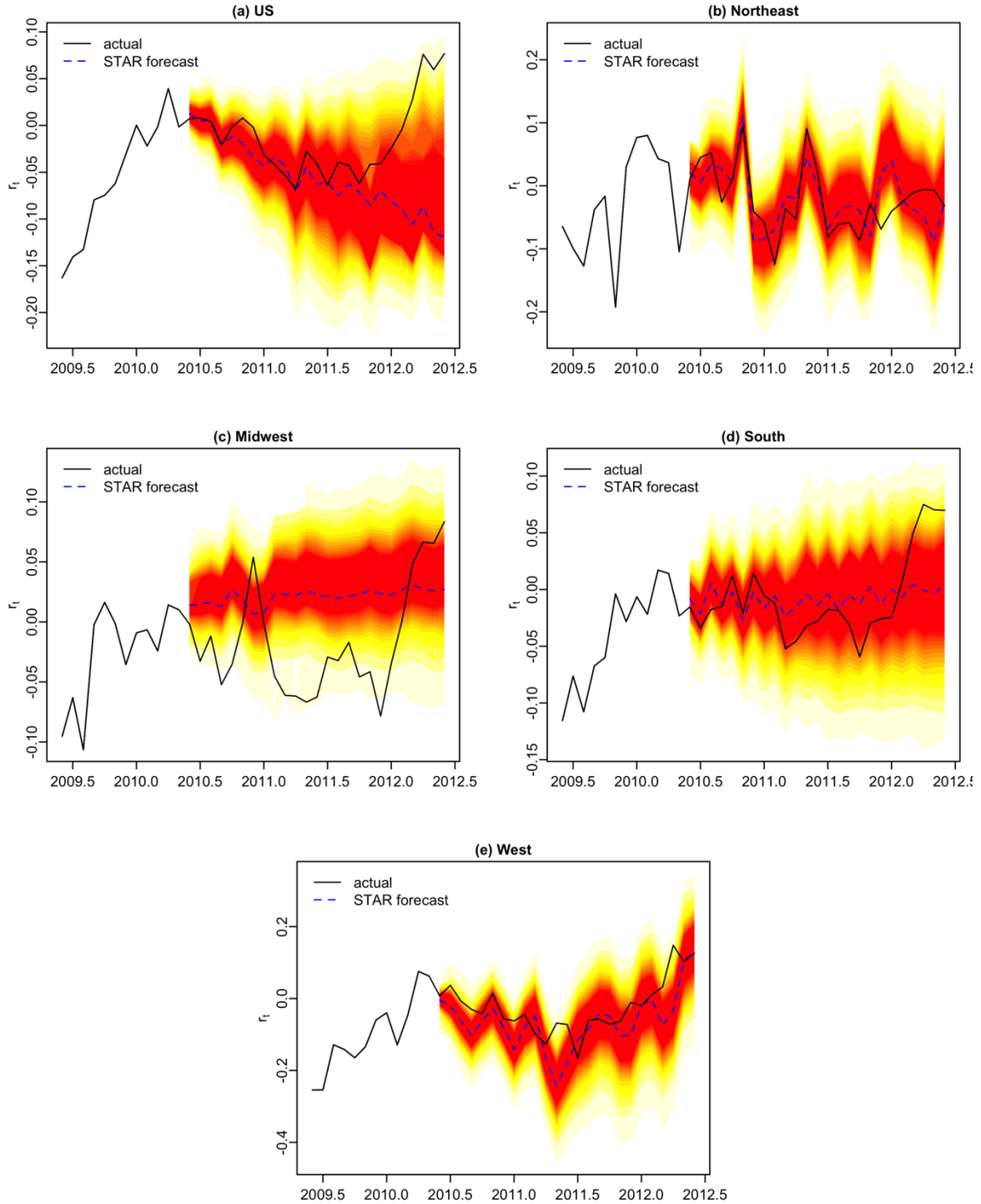


Figure 5. Point Forecast of the annual growth rate of home price r_t from the estimated nonlinear AR models for the period 2010:6 to 2012:6 and 50 to 95 percent interval forecasts. Dashed lines show the dynamic 25-step forecasts and solid lines show the actual data over the 2009:5 to 2012:6. Each point forecast is obtained by 2,000 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).