An in-sample and out-of-sample empirical investigation of the nonlinearity in house prices of South Africa ☆

Mehmet Balcilar [a,1], Rangan Gupta [b,*], Zahra B. Shah [b]

[a] Department of Economics, Eastern Mediterranean University, Famagusts, North Cyprus, via Mersin 10, Turkey
[b] Department of Economics, University of Pretoria, Pretoria, 0002, South Africa

ARTICLE INFO

ABSTRACT

This paper tests whether housing prices in the five segments of the South African housing market, namely large–middle, medium–middle, small–middle, luxury and affordable, exhibit non-linearity based on smooth transition autoregressive (STAR) models estimated using quarterly data from 1970:Q2 to 2009:Q3. Findings point to an overwhelming evidence of non-linearity in these five segments based on in-sample evaluation of the linear and non-linear models. We next provide further support for non-linearity by comparing one- to four-quarters-ahead out-of-sample forecasts of the non-linear time series model with those of the classical and Bayesian versions of the linear autoregressive (AR) models for each of these segments, for the out-of-sample horizon 2001:Q1 to 2009:Q3, using the in-sample period 1970:Q2 to 2000:Q4. Our results indicate that barring the one-, two and four-step(s)-ahead forecasts of the small segment, the non-linear model always outperforms the linear models. In addition, given the existence of strong causal relationship amongst the house prices of the five segments, the multivariate versions of the linear (classical and Bayesian) and STAR (MSTAR) models were also estimated. The MSTAR always outperformed the best performing univariate and multivariate linear models. Thus, our results highlight the importance of accounting for non-linearity, as well as the possible interrelationship amongst the variables under consideration, especially for forecasting.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

The objectives of this paper are twofold: First, we want to address whether house prices in the five segments of the South African housing market, namely, luxury, large–middle, medium–middle, small–middle (henceforth large, medium and small) and affordable,[2] exhibit non-linearity based on smooth transition autoregressive (STAR) models estimated using quarterly data from 1970:Q2 to 2009:Q3; and second, whether the five housing segments exhibit non-linearity for which we compare one- to four-quarters-ahead out-of-sample forecasting performances of the non-linear time series models with those of the classical and Bayesian versions of the linear autoregressive (AR) over an out-of-sample horizon of 2001:Q1 to 2009:Q3, using an in-sample of 1970:Q2 to 2000:Q4. Note that the choice of the in-sample period, especially the starting date, depended on data availability for all the five housing segment. The end-point of the out-of-sample horizon is data-driven, with the starting point of the same preceding the rapid run-up and then slowdown of house prices experienced over the last decade in the South African economy, as indicated in Fig. 1.

Related to the two stated objectives, two questions arise immediately, and define the motivations for our paper: First, why should one expect house prices to exhibit non-linearity? And second, why should one be interested in forecasting house prices? To answer to the first question, it must be realized that the behaviour of the housing market is not the same across phases of expansion and contraction of the swings that characterize the real estate sector (Kim and Bhattacharya, 2009). Seslen (2004) argues that households exhibit forward looking behaviour and have higher probability of trading up, during the upswing with equity constraints being less binding. However, the same is not true during the downswing of the housing market cycle, since when house prices are on the decline, households are less likely to trade implying downward rigidity of house prices. The aversion to loss during the downswing is more than likely to reduce the mobility of households as far as trading is concerned. Further, as pointed out by Muellbauer and Murphy (1997), the presence of lumpy transaction costs in the housing market can also cause non-linearity. Given these issues, it is important to test whether house prices are in fact non-linear.

As far as forecasting house prices are concerned, a large number of papers show a strong link between the housing market and economic activity (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
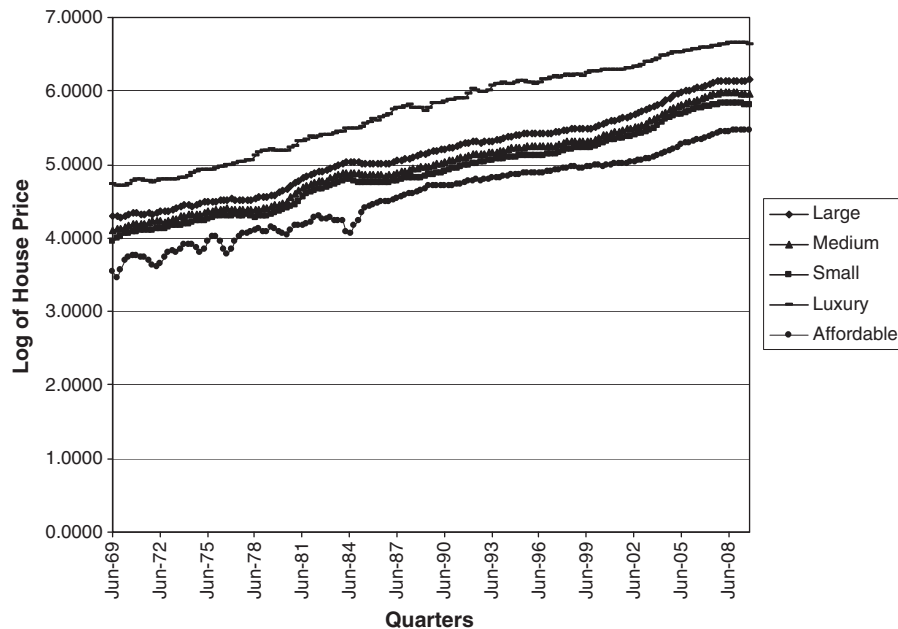

Fig. 1. Home prices in the five different segments of the South African housing market.

Owyang, 2009; Pavlidis et al., 2009; Iacoviello and Neri, 2010). This is understandable because housing contributes towards a large percentage of private sector wealth (Cook and Speight, 2007), and house price impacts on economic stability is vital in explaining household consumption and saving patterns (Englund and Ioannides, 1997). Campbell and Coco (2006) contend that housing can be considered as a consumption good and hence, house prices and consumption are strongly correlated (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 real activity, inflation, or both, and hence, can indicate to where the economy will head. Moreover, the recent emergence of boom–bust cycles in house prices are a cause of much concern and interest amongst policy makers (Borio et al., 1994; Bernanke and Gertler, 1995, 1999), since the bust of house price bubbles always leads to significant contractions in the real economy, vouched for by the current economic downturn. Given this, models that forecast house price can give policy makers an idea about the direction of the economy, and hence, can provide a better control on designing appropriate policies. As such, it is of paramount importance to deduces the underlying nature of the data-generating process for house price, i.e., whether it is linear or non-linear.

To the best of our knowledge, this is the first attempt to simultaneously test for non-linearity in the housing market and compare forecasts generated from STAR and classical and Bayesian AR models. The only other study that we are aware of that tests for non-linearity in the housing market is by Kim and Bhattacharya (2009) in which the authors examined the STAR models based on non-linear properties of house prices for the aggregate US economy and its four census regions (North East, Midwest, South and West) using monthly data over the period of 1969:01–2004:12. The authors concluded that house prices for the entire US and all four census regions, barring the Midwest, were non-linear. However, their paper does not look at the ability of the non-linear framework in forecasting house prices relative to their linear counter-part.[3] One must realize that the forecasting exercise is crucial in reaching

a firm and overwhelming conclusion about modelling house prices using non-linear models as indicated by Rapach and Wohar (2006a), while forecasting real exchange rates using (S)TAR models, the gains from using a non-linear framework in forecasting relative to the linear model can be quite small, even when there is strong evidence on non-linearity in the in-sample. Beck et al. (2000, 2004) note that forecasting provides the root of inference and prediction in time-series analysis. Further, Clements and Hendry (1998) argue that in time-series models, estimation and inference essentially means minimizing of the one-step (or multi-step) forecast errors, thereby, establishing a model's superiority would boil down to showing that it produces smaller forecast errors than its competitors. Thus, it is typically believed that out-of-sample comparisons, over and above in-sample results, provides a better measure of the appropriate data generating process, as statistical models are tested using out-of-sample observations that are not used in the estimation stage (Rapach and Wohar, 2006b). If the forecasts generated by the STAR model are superior to those generated by classical and Bayesian AR models, it can be considered as strong evidence in favor of the STAR model. The remainder of the paper is organized as follows: Section 2 discusses the specification of the STAR models. Section 3 outlines the details regarding the data used, while Section 4 presents the formal test of non-linearity in the five segments of the South African housing market. Section 5 compares the forecasting performance of the appropriate version of the non-linear STAR model in relation to the classical and Bayesian AR models. In addition, this section tests for causal relationships amongst the five segments of the housing market and repeats the forecasting analysis with the multivariate versions of the linear and non-linear models. Finally, Section 6 concludes.

2. Specification and estimation of STAR models

We use the STAR framework developed by Luukkonen et al. (1988), to model house price growth rates[4] as a non-linear and state-

---

[3] In addition to this, Kim and Bhattacharya (2009) highlighted that the dynamic properties implied by the nonlinear estimation was in line with the typical patterns that have been observed in the aggregate US and regional housing markets, and also, based on non-linear Granger causality tests, the authors found housing price to cause employment and mortgage rates to cause housing price.

[4] Note non-linear estimation, just like linear estimation, requires one to ensure that the variables used are stationary to avoid spurious estimates. Hence, house prices of all the five segments were converted to their yearly growth rates, the stationarity of which, in turn, were confirmed by the Augmented-Dickey–Fuller (ADF) test, the Dickey–Fuller test with GLS Detrending (DF-GLS), the Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test and the Phillips–Perron (PP) test. The results are available upon request from the authors.

dependent variable. The STAR framework allows for smooth transition across regimes to describe the dynamics of long-horizon house price growth rates, and, hence, is preferred over the threshold autoregressive (TAR) (Tsay, 1989) and the Markov switching (Hamilton, 1989) models, since the latter two frameworks specify discrete jumps across regimes implying sudden jumps. In addition, our decision to employ the STAR model is governed by our belief that house price growth rates are better described by the STAR model rather than the TAR or Markov switching models. The low speeds of transition obtained in the estimation of the non-linear model, vindicate our decision as well.

The STAR model of order p, for variable r_t, is specified as follows[5]:

$$r_t = \left[\phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i}\right] + \left[\rho_0 + \sum_{i=1}^{p} \rho_i r_{t-i}\right] \cdot F(r_{t-d}) + u_t \qquad (1)$$
$$= [\phi_0 + \phi(L)r_t] + [\rho_0 + \rho(L)r_t] \cdot F(r_{t-d}) + u_t$$

where r_t is the house price growth rate, and; $F(r_{t-d})$ is the transition function controlling the regime shift mechanism and is a smooth and continuous function of past realized house price growth rates. Thus, house price growth rates evolve with a smooth transition between regimes that depends on the sign and magnitude of past realization of the house price growth rates. The non-linearities are obtained by conditioning the autoregressive coefficients, $\rho(L)$, to change smoothly with past house price growth rates in such a way that the past realized home price growth rate r_{t-d} is the transition variable with d being the delay parameter, which, in turn, indicates the number of periods r_{t-d} leads the switch in dynamics.

Teräsvirta and Anderson (1992) define the transition function $F(.)$ by using two alternative forms, namely the logistic smooth transition autoregressive (LSTAR) model and the exponential smooth transition autoregressive (ESTAR) model. In the LSTAR model, $F(.)$ is defined by a logistic function, so that:

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

while, $F(.)$ in the ESTAR framework is captured as follows by an exponential function:

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

In the above equations γ is the speed of transition between regimes and c measures the halfway point or threshold between the two regimes. Eq. (1) combined with Eq. (2) produces the LSTAR(p) model and Eq. (1) combined with Eq. (3) yields the ESTAR(p) model. In STAR models, expansion and contraction are a representation of two different economic phases, but transition between the two regimes is smooth, controlled by r_{t-d} (Sarantis, 2001). The LSTAR and ESTAR models describe different dynamic behaviour. The LSTAR model allows the expansion and contraction regimes to have different dynamics whereas the ESTAR model suggests that the two regimes have similar dynamics (Sarantis, 2001). We also take note that when $\gamma \to \infty$, then the model degenerates into the conventional TAR(p), while, when $\gamma \to 0$, then the model degenerates to the linear AR(p) model (Teräsvirta and Anderson, 1992).

The procedure for constructing an appropriate STAR model for a specific variable, comprises three stages:

a) Firstly a linear AR model needs to be specified, with the value of p being chosen based on the unanimity of at least two of the popular lag-length tests, namely, LR test statistic, Akaike information criterion (AIC), Schwarz information criterion (SIC), the final prediction error (FPE) criterion and the Hannan–Quinn (HQ) information criterion;

b) Then we test for linearity against a nonlinear STAR model, for different values of the delay parameter d using the linear model in (a) as the null, based on a Lagrange multiplier smooth transition (LM-STR) test for linearity. This boils down to estimating the following auxiliary regression, as proposed by Teräsvirta and Anderson (1992):

$$r_t = \phi_0 + \sum_{i=1}^{p} \varphi_{1,i} \cdot r_{t-i} + \sum_{i=1}^{p} \varphi_{2,i} \cdot r_{t-i} r_{t-d} \qquad (4)$$
$$+ \sum_{i=1}^{p} \varphi_{3,i} \cdot r_{t-i} r_{t-d}^2 + \sum_{i=1}^{p} \varphi_{4,i} \cdot r_{t-i} r_{t-d}^3 + u_t$$

with the null hypothesis of linearity being $H_{01}: \phi_{2i} = \phi_{3i} = \phi_{4i} = 0$ for all i. To get the appropriate values of the delay parameter d, the estimation of Eq. (4) is carried out for a range of values, $1 \leq d \leq D$. In the scenario, where linearity is rejected for more than one value of d, then d is chosen such that $d = arg\ min\ p(d)$ for $1 \leq d \leq D$.[6]

c) The choice between the LSTAR and ESTAR model, when linearity is rejected, is then conducted by applying the sequence of nested tests,
   1. $H_{04}: \phi_{4i} = 0, \qquad i = 1, \dots, p$
   2. $H_{03}: \phi_{3i} = 0 | \phi_{4i} = 0, \qquad i = 1, \dots, p$
   3. $H_{02}: \phi_{2i} = 0 | \phi_{3i} = \phi_{4i} = 0, \qquad i = 1, \dots, p.$

A standard procedure, as discussed in Teräsvirta and Anderson (1992), is then followed in the selection of the appropriate STAR model. There are three possible sequential outcomes, given d:

   i. The rejection of $H_{04}: \phi_{4i} = 0$, implies the selection of the LSTAR model.
   ii. If H_{04} is not rejected, then we move to the second part of the test, which tests if $H_{03}: \phi_{3i} = 0 | \phi_{4i} = 0$,. Rejection of H_{03} implies the selection of the ESTAR model.
   iii. If H_{03} is not rejected, then we move to the last component of the test which tests $H_{02}: \phi_{2i} = 0 | \phi_{3i} = \phi_{4i} = 0$,. Rejection of H_{02} implies selection of the LSTAR model.

Various authors (Granger and Teräsvirta, 1993; Teräsvirta, 1994; Eitrheim and Teräsvirta, 1996; Sarantis, 2001) argue that if this sequence of tests is strictly applied then it may lead to wrong conclusions since the higher order terms of the Taylor expansion used in the derivation of these tests are ignored. Thus it is recommended that one computes p-values for all the F-tests of Eqs. (1)–(3) above. Then one can make the choice of the appropriate STAR model based on the lowest p-value or highest F-statistic.

3. Data

We use quarterly house price data obtained from the ABSA[7] Housing Price Survey, for the period 1970:Q2 to 2009:Q3. The survey distinguishes between three price categories as: affordable (R430,000 and area below 40 m²–79 m²), middle (R430,000 to R3.1 million) and luxury (R3.1 million to R11.5 million). The data is further subdivided for the middle segment of the housing market, based on sizes (square meters), into small (80 m²–140 m²), medium (141 m²–220 m²) and large (221 m²–400 m²). Given that Genesove and Mayer (2001) and Engelhardt (2001) point out that sellers are averse to realizing losses in nominal and not real terms, we follow Kim and Bhattacharya (2009) and use nominal house prices, since it is nominal changes in house prices that cause asymmetric effects on mobility and the housing market in general. We use annual growth rate of house prices, which are measured with respect to the same quarter in the previous year.

---

[5] This part of the paper relies heavily on the discussion available in Kim and Bhattacharya (2009), and, hence we retain their symbolic representation of the equations.

[6] $p(d)$ is the p-value of the test.
[7] ABSA is one of the leading private banks in South Africa.

4. Empirical results[8]

In this section, we first start of with the LM-STR test for linearity of house price growth rates, and then we conduct 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 the period of 1970:Q2 till 2009:Q3. When conducting the (LM-STR) test for linearity, as discussed above, the optimal lag, p, was selected based on the unanimity of at least two of the popularly-used lag-length tests. We allowed the delay lag, d, to vary over the range $1 \leq d \leq 8$. The optimal delay lag d is estimated on the basis of lowest p-value or highest F-statistic associated with the null hypothesis: $H_{01}: \phi_{2i} = \phi_{3i} = \phi_{4i} = 0$ for all i.

From Table 1, we can conclude that the null hypothesis of linearity can be rejected for all the five segments, with the null being rejected at the 1% level for the small and affordable sections and at the 5% level for the large, medium and luxury segments of the South African housing market. Since all five categories of house price growth rates are nonlinear, we now need to specify the appropriate STAR model to capture accurately the non-linear dynamics (Table 2). As proposed by Teräsvirta and Anderson (1992) and outlined in Section 2, 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.

As observed from Table 3, we can deduce that the LSTAR model is best-suited in describing the non-linear dynamics of all the five segments of the South African housing market. The choice of the LSTAR model suggests that the dynamics of the house price growth rates is characterized by asymmetric dynamics during the phases of contraction and expansion. Next, we provide further evidence of nonlinearity by providing in-sample comparison based on the estimation of the linear AR model, given in Eq. (5), and the nonlinear LSTAR model described in Eq. (6):

$$r_t = \left[\phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} \right] + u_t \tag{5}$$

and

$$r_t = \left[\phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} \right] + \left[\rho_0 + \sum_{i=1}^{p} \rho_i r_{t-i} \right] \cdot \left[1 + \exp\left\{ -\frac{\gamma}{\sigma(r_t)} (r_{t-d} - c) \right\} \right]^{-1}. \tag{6}$$

Teräsvirta (1994) contends that for LSTAR models it is not easy to carry out the joint estimation of $(\gamma, c, \phi_0, \rho_0, \phi_i, \rho_i)$ since one faces difficulties with the estimation of c and γ. When γ is large then c is steep and a large number of observations in the neighbourhood of c would be required to estimate γ, i.e., relatively large changes in γ would have only a minor effect on shape of $F(.)$. Thus the sequence of estimates for γ may converge slowly. Note if γ is statistically insignificant then Eq. (6) becomes the linear AR model. In accordance with 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 house price growth rate r_t.

The results of the estimation of the LSTAR model and the AR model have been tabulated in Tables 3 and 4, where we use the nonlinear least-squares (NLS) to estimate the LSTAR model.[9] Note, it

Table 1
LM-STR test for linearity.

	Large $p^* = 8$	Medium $p^* = 7$	Small $p^* = 8$	Luxury $p^* = 8$	Affordable $p^* = 6$
Optimal delay d	3 (0.0213)	3 (0.0303)	3 (0.0000)	4 (0.0347)	4 (0.0000)

Notes: The numbers in parenthesis are the lowest p-values associated with the H_{01}: $\phi_{2i} = \phi_{3i} = \phi_{4i} = 0$ in Eq. (4) with the corresponding d.

Table 2
Test of the appropriate STAR model.

	Optimal delay d	H_{04}: $\varphi_{4i} = 0$, $i = 1, \ldots, p$	H_{03}: $\varphi_{3i} = 0$, given $\varphi_{4i} = 0$	H_{02}: $\varphi_{3i} = 0$, given $\varphi_{3i} = \varphi_{4i} = 0$	Selection of model	Optimal lag p
Large	3	0.0090*	0.0862	0.7736	LSTAR	8
Medium	3	0.0039*	0.3477	0.6431	LSTAR	7
Small	3	0.0014*	0.0310	0.0103	LSTAR	8
Luxury	4	0.0037*	0.6221	0.4193	LSTAR	7
Affordable	4	0.0038*	0.0000	0.0000	LSTAR	6

Notes: The values in the table are the p-values for the nested tests H_{04}, H_{03}, and H_{02}. A '*' indicates the lowest p-value for the three tests.

is the logistic function which conditions the autoregressive parameters to change smoothly with lagged realized changes in the growth rates of home prices in the LSTAR model that generates the endogenous nonlinearity. When we compare the estimation results over the period 1970:Q2 to 2009:Q3 for the AR and the LSTAR models, the following features pointing to the superiority of the non-linear estimation emerge: (a) The measures of the standard error and the log likelihood value of the nonlinear regression show a significant improvement over the corresponding values obtained from the linear regression; (b) The adjusted R^2 value in the nonlinear regression is always higher than the corresponding value under the linear regression, implying that a large portion of variance in the house price growth rates in the long-run is associated with nonlinear dynamics; (c) As reported, most of the estimates of the coefficients of the nonlinear portion of Eq. (6), i.e., ρ_is, are statistically significant, and; (d) The value of γ, which governs the speed of transition between regimes, is always positive as expected and is statistically significant at the 10% level or better. The statistical significance of γ confirms the presence of nonlinearity outlined by the LSTAR model. These results together provide strong evidence that the LSTAR model appropriately captures the inherent non-linearity in the long horizon house price growth rates in the five segments of the South African housing market. Thus a linear model would clearly be misspecified 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.[10]

It is important to note that γ is relatively small for all the categories of house price growth rates except for the small segment. Relatively small estimates of γ suggests a slower transition from one regime to another, which is, in turn, in contrast with the TAR or

---

[8] All computations in this paper have been carried out with the RSTAR package (Version 0.1–1) in R developed by the first author of the paper.

[9] Following the suggestions of van Dijk et al. (2002), a battery of misspecifications tests, namely, no residual autocorrelation parameter constancy, no remaining non-linearity, no autoregressive conditional heteroskedasticity (ARCH), besides the test of normality, were carried out for the LSTAR model. The estimated LSTAR models for all the five segments were found to be free from any type of misspecification. These results are available upon request from the authors.

[10] The Ramsey model specification test provides further evidence of nonlinearity in the housing price growth rates of the five segments. The null hypothesis that the correct specification is a linear AR model, against a nonlinear LSTAR model, is rejected at the 1% level of significance for all the five 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 large, medium, small, luxury and affordable sections were respectively: 21.3594, 11.7915, 34.8803, 11.6411 and 124.4432.

Table 3
Estimation of the LSTAR model.

Large: adjusted $R^2 = 0.942435$, SER $= 2.1282$, LLV $= -326.1296$
$r_t = 0.9527(0.0172) + 1.9295r_{t-1}(0.0000) - 1.7276r_{t-2}(0.0000) + 1.3858r_{t-3}(0.0000)$
$\quad - 1.4175r_{t-4}(0.0000) + 1.2626r_{t-5}(0.0000) - 0.6366r_{t-6}(0.0000) + 0.1037r_{t-8}(0.0021)$
$\quad + (5.9853(0.0013) - 0.5839r_{t-1}(0.0001) + 1.3606r_{t-2}(0.0000) - 1.5065r_{t-3}(0.0000)$
$\quad + 1.1312r_{t-4}(0.0046) - 1.0456r_{t-5}(0.0009) + 0.7290r_{t-6}(0.0019)$
$\quad - 0.3042r_{t-8}(0.0111)) \times [1 + \exp\{-8.8080(r_{t-3}(0.0032) - 19.8674(0.9981))\}]^{-1}$

Medium: adjusted $R^2 = 0.965983$, SER $= 1.7804$, LLV $= -301.3587$
$r_t = 0.39504(0.1399) + 2.0503r_{t-1}(0.0000) - 1.71736r_{t-2}(0.0000) + 1.2947r_{t-3}(0.0000)$
$\quad - 1.5771r_{t-4}(0.0000) + 1.5711r_{t-5}(0.0000) - 0.8225r_{t-6}(0.0000)$
$\quad + 0.1725r_{t-7}(0.0462) + (9.7308(0.0090) - 0.9474r_{t-1}(0.0000) + 1.2618r_{t-2}(0.0000)$
$\quad - 1.2712r_{t-3}(0.0002) + 1.4053r_{t-4}(0.0047) - 0.9035r_{t-5}(0.0077))$
$\quad \times [1 + \exp\{-5.5690(r_{t-3}(0.0156) - 30.7287(0.0000))\}]^{-1}$

Small: adjusted $R^2 = 0.956223$, SER $= 2.0193$, LLV $= -318.2516$
$r_t = 1.3205(0.0010) + 1.4527r_{t-1}(0.0000) - 0.7125r_{t-2}(0.0000) + 0.8697r_{t-3}(0.0000)$
$\quad - 1.4370r_{t-4}(0.0000) + 0.9296r_{t-5}(0.0000) - 0.3852r_{t-6}(0.0504) + 0.5676r_{t-7}(0.0000)$
$\quad - 0.3978r_{t-8}(0.0000) + (-2.1363(0.2096) + 0.3755r_{t-1}(0.0096) - 0.4890r_{t-2}(0.0093)$
$\quad + 0.8497r_{t-5}(0.0005) - 0.4908r_{t-6}(0.0712)$
$\quad - 0.3101r_{t-8}(0.0017)) \times [1 + \exp\{-126.8074(r_{t-3}(0.0188) - 21.5245(0.0000))\}]^{-1}$

Luxury: adjusted $R^2 = 0.721693$, SER $= 5.2717$, LLV $= s -465.2763$
$r_t = 3.4527(0.0006) + 1.1511r_{t-1}(0.0000) - 0.5182r_{t-2}(0.0000) + 0.3160r_{t-3}(0.0008)$
$\quad - 0.7613r_{t-4}(0.0000) + 0.3923r_{t-5}(0.0000) + (3.9178(0.0347)$
$\quad - 0.2680r_{t-7}(0.0007)) \times [1 + \exp\{-29.4707(r_{t-4}(0.0000) - 11.7210(0.9995))\}]^{-1}$

Affordable: adjusted $R^2 = 0.9008$, SER $= 6.5703$, LLV $= -501.8283$
$r_t = -13.4349(0.0002) + 1.2648r_{t-1}(0.0000) - 1.1212r_{t-2}(0.0000) - 2.0792r_{t-4}(0.0000)$
$\quad + 1.1639r_{t-5}(0.0000) - 0.7423r_{t-6}(0.0000) - 0.3852r_{t-6}(0.0000)$
$\quad + (16.6382(0.0000) + 0.2220r_{t-1}(0.0706) + 0.7424r_{t-3}(0.0000) + 1.4184r_{t-4}(0.0000)$
$\quad - 0.6588r_{t-5}(0.0066) + 0.5022r_{t-6}(0.0000))$
$\quad \times [1 + exp\{-7.7768(r_{t-4}(0.0537) - 11.5396(0.0000))\}]^{-1}$

Notes: The values in the parenthesis correspond to the *p* value; SER, standard error of regression; LLV, log likelihood value. We include only significant lags following Teräsvirta (1994) and Sarantis (2001).

Markov switching models where one witnesses sudden switch between regimes, given that the estimate of γ tends to infinity. The fact that the parameter γ takes a value of 126.8074 for the small

Table 4
Estimation of the AR model.

Large: adjusted $R^2 = 0.937389$, SER $= 2.365968$, LLV $= -337.3782$
$r_t = 0.849136 \ (0.0275) + 1.780462r_{t-1}(0.0000)$
$\quad - 1.359759r_{t-2}(0.0000) + 0.987180r_{t-3}(0.0000)$
$\quad - 0.994519r_{t-4}(0.0000) + 0.846320r_{t-5}(0.0000) + 0.172149r_{t-8}(0.0234)$

Medium: adjusted $R^2 = 0.965060$, SER $= 1.908279$, LLV $= -307.7264$
$r_t = 0.627505(0.0306) + 2.041448r_{t-1}(0.0000)$
$\quad - 1.749766r_{t-2}(0.0000) + 1.319690r_{t-3}(0.0000)$
$\quad - 1.570302r_{t-4}(0.0000) + 1.575341r_{t-5}(0.0000)$
$\quad - 0.882169r_{t-6}(0.0000) + 0.208929r_{t-7}(0.0104)$

Small: adjusted $R^2 = 0.947878$, SER $= 2.339952$, LLV $= -335.7197$
$r_t = 1.460725(0.0001) + 1.580604r_{t-1}(0.0000) - 0.907154r_{t-2}(0.0000)$
$\quad + 0.864273r_{t-3}(0.0000) - 1.420194r_{t-4}(0.0000)$
$\quad + 1.072038r_{t-5}(0.0000) - 0.413761r_{t-6}(0.0096)$
$\quad + 0.499351r_{t-7}(0.0007) - 0.398282r_{t-8}(0.0000)$

Luxury: adjusted $R^2 = 0.698159$, SER $= 5.663563$, LLV $= -468.3081$
$r_t = 4.980659(0.0000) + 1.121342r_{t-1}(0.0000)$
$\quad - 0.511161r_{t-2}(0.0001) + 0.378676r_{t-3}(0.0052)$
$\quad - 0.726343r_{t-4}(0.0000) + 0.498261r_{t-5}(0.0002)$

Affordable: adjusted $R^2 = 0.819200$, SER $= 9.309189$, LLV $= -551.2078$
$r_t = 5.418556(0.0000) + 1.538784r_{t-1}(0.0000) - 1.325532r_{t-2}(0.0000)$
$\quad + 0.945101r_{t-3}(0.0000) - 1.000029r_{t-4}(0.0000)$
$\quad + 0.782214r_{t-5}(0.0000) - 0.351645r_{t-6}(0.0000)$

Notes: The values in the parenthesis correspond to the *p* value; SER, standard error of regression; LLV, log likelihood value. We include only significant lags following Teräsvirta (1994) and Sarantis (2001).

segment is an indication of the households in this category to trade up very quickly especially during the upswing phase of the housing market. The parameter *c* representing the half-way point between regimes is positive for all the five segments of the housing market, and, hence, indicates that similar value of house price growth rate shock triggers a shift in regimes. However note, the value of *c* is not significant for the large and the luxury housing categories.

5. Forecast accuracy

Having established that house price growth rates in the South African housing market should be modelled via a LSTAR framework, in this section we compare the forecast performances of the non-linear model with those of the classical and Bayesian versions of the linear AR models. The decision to use a Bayesian variant of the classical linear AR framework emanates from the general finding in the forecasting literature that Bayesian models which impose prior restrictions on the mean and variance of the classical AR models tend to perform better.[11] For this forecasting exercise, we choose an out-of-sample horizon of 2001:Q1 to 2009:Q3 and thus, use an in-sample of 1970:Q2 to 2000:Q4. This provides 35 observations for evaluating the out-of-sample forecast performance. As mentioned before, the end-point of the out-of-sample horizon is data-driven, the starting point precedes the rapid run-up and then collapse of house prices experienced over the last decade in the South African economy, as indicated in Fig. 1.

Before we present the forecast performances of the alternative models, it is important to outline the restrictions on the priors in the Bayesian AR (BAR). The Bayesian method imposes restrictions on the coefficients across different lag lengths, assuming that the coefficients of longer lags may approach zero more closely than the coefficients on shorter lags. If however, stronger effects come from longer lags, the data can override this initial restriction. Researchers impose the constraints by specifying normal prior distribution with zero mean and small standard deviation for most coefficients, where the standard deviation decreases as the lag length increases, implying that the zero-mean prior holds with more certainty. The first own-lag coefficient in each equation is the exception with a unitary mean. Finally, a diffuse prior is imposed on the constant. We employ this "Minnesota prior" in our analysis, where we implement the Bayesian variants of the classical AR models.

Formally, the means of the Minnesota prior take the following form:

$$\phi_i \sim N\left(1, \sigma_{\phi_i}^2\right) \text{ and } \phi_j \sim N\left(0, \sigma_{\phi_j}^2\right) \tag{7}$$

where ϕ_i equals the coefficients associated with the lagged dependent variable in each equation of the AR model (i.e., the first own-lag coefficient), while ϕ_j equals any other coefficient. In sum, the prior specification reduces to a random-walk with drift model for each variable, if we set all variances to zero. The prior variances, $\sigma_{\phi_i}^2$ and $\sigma_{\phi_j}^2$, specify uncertainty about the prior means, $\overline{\phi}_i = 1$, and $\overline{\phi}_j = 0$. However, given that in our case, all the five yearly house price growth rates are stationary, we adopt the specification in Banbura et al. (2010) and Bloor and Matheson (2008), and set a white-noise prior (i.e., $\overline{\phi}_i = 0$).

Doan et al. (1984) propose a formula to generate standard deviations that depend on a small numbers of hyper-parameters: *w*, *d*, and a weighting matrix $f(i, j)$. The specification of the standard deviation of the distribution of the prior imposed on variable *j* in equation *i* at lag *m*, for all *i*, *j* and *m*, equals $S(i, j, m)$, defined as follows:

$$S(i, j, m) = [w \times g(m) \times f(i, j)] \frac{\hat{\sigma}_i}{\hat{\sigma}_j}, \tag{8}$$

---

[11] Refer to Das et al. (2008, forthcoming-a) and Gupta and Das (2008) for further details.

where $f(i,j) = 1$, if $i = j$ and k_{ij} otherwise, with $(0 \leq k_{ij} \leq 1)$, and $g(m) = m^{-d}$, with $d > 0$. The estimated standard error of the univariate autoregression for variable i equals $\hat{\sigma}_i$. The ratio $\hat{\sigma}_i / \hat{\sigma}_j$ scales the variables to account for differences in the units of measurement and, hence, causes the specification of the prior without consideration of the magnitudes of the variables. The term w indicates the overall tightness, with the prior getting tighter as the value falls. The parameter $g(m)$ measures the tightness on lag m with respect to lag 1, and equals a harmonic shape with decay factor d, which tightens the prior at longer lags. The parameter $f(i,j)$ equals the tightness of variable j in equation i relative to variable i, and by increasing the interaction (i.e., the value of k_{ij}), we loosen the prior. We estimate the alternative BARs using Theil's (1971) mixed estimation technique. Essentially, the method involves supplementing the data with prior information on the distribution of the coefficients. The number of observations and degrees of freedom increase artificially by one for each restriction imposed on the parameter estimates. For the BAR models, we start with a value of $w = 0.1$ and $d = 1.0$, and then increase the value to $w = 0.2$ to account for more influences from variables other than the first own lags of the dependant variables of the model. In addition, as in Dua and Ray (1995), Gupta and Sichei (2006), Gupta (2006), Gupta and Miller (2009) and Gupta et al., forthcoming, we also estimate the BAR with $w = 0.3$ and $d = 0.5$. We also introduce $d = 2$ to increase the tightness on lag m. Finally, k_{ij} is set equal to 0.001 (Gupta and Sichei, 2006).

In Table 5, we compare the forecast performances of the LSTAR, classical AR and the BAR models, with the latter being estimated under different parameterization of w and d to account for various degrees of

Table 5
One- to four-quarters-ahead RMSEs of house price growth rates from univariate models (2001:Q1–2009:Q3).

Segments	Models	QA			
		1	2	3	4
Large	AR	2.3064	4.6731	6.6471	8.4807
	BAR1	2.3279	4.7211	6.7797	8.6807
	BAR2	2.6149	4.9080	6.8485	8.5741
	BAR3	2.9304	5.1872	7.0297	8.5801
	BAR4	2.6828	4.8503	6.6489	8.1957
	BAR5	2.9498	5.1455	6.8889	8.3117
	LSTAR	2.2385	4.6511	6.2433	7.4299
Medium	AR	1.7399	4.1979	6.6155	8.7820
	BAR1	1.7956	4.2276	6.6694	8.9093
	BAR2	2.0599	4.5086	6.8608	8.9309
	BAR3	2.4757	4.9960	7.2755	9.2147
	BAR4	2.2659	4.7665	7.0346	8.9682
	BAR5	2.7259	5.3257	7.5269	9.3277
	LSTAR	1.3021	3.3834	5.5225	7.6528
Small	AR	1.9980	4.4696	6.8375	9.0661
	BAR1	2.0772	4.5347	6.9668	9.3587
	BAR2	2.3815	4.9128	7.2607	9.2880
	BAR3	2.8626	5.5046	7.7538	9.5257
	BAR4	2.6718	5.3166	7.5621	9.3572
	BAR5	3.0905	5.7886	7.9398	9.5765
	LSTAR	2.0152	4.5746	6.7427	9.2772
Luxury	AR	3.4589	5.6949	6.6357	7.0589
	BAR1	3.4620	5.6302	6.5750	6.9553
	BAR2	3.7279	5.7850	6.6256	6.8933
	BAR3	4.1238	5.9977	6.6285	6.7480
	BAR4	3.7173	5.5956	6.2984	6.5621
	BAR5	4.0765	5.8547	6.4217	6.5728
	LSTAR	3.2237	5.2773	5.8535	6.5774
Affordable	AR	2.9728	6.4199	8.6318	9.2449
	BAR1	3.2095	6.6379	8.7709	9.3920
	BAR2	3.3060	6.2919	8.1404	8.6139
	BAR3	3.5814	6.0777	7.4544	7.8011
	BAR4	2.9433	5.2845	6.6985	7.2397
	BAR5	3.1979	5.2267	6.3584	6.8426
	LSTAR	2.2766	5.0528	7.0189	8.0698

Notes: QA: Quarter(s) Ahead; BAR1: $w = 0.3$, $d = 0.5$; BAR2: $w = 0.2$, $d = 1.0$; BAR3: $w = 0.1$, $d = 1.0$; BAR4: $w = 0.2$, $d = 2.0$; BAR5: $w = 0.1$, $d = 2.0$.

tightness (from most loose to most tight) of the prior-structure. We estimate the univariate models over the period 1970:Q2 to 2000:Q4, and then forecast from 2001:Q1 through 2009:Q3. Since we use eight, seven, eight, eight and six lags for the large, medium, small, luxury and affordable segments respectively, the initial eight, seven, eight, eight and six quarters from 1970:Q2 for the respective categories feed the lags. We re-estimate the models each quarter over the out-of-sample forecast horizon in order to update the estimate of the coefficients, before producing the four-quarters-ahead forecasts. We implement this iterative estimation and the four-quarters-ahead forecast procedure for 35 quarters, with the first forecast beginning in 2001:Q1. This produces a total of 35 one-quarter-ahead forecasts, 35 two-quarter-ahead forecasts, …, up to 35 four-quarters-ahead forecasts.[12] We calculate the root mean squared errors (RMSE)[13] for the 35 one-, two-, three-, and four-quarters-ahead forecasts for the real house price index of the models. We then examine the average of the RMSE statistic for one-, two-, three-, and four-quarters ahead forecasts over the 2001:Q1 to 2009:Q3 period.

As can be seen from Table 5, barring the one-, two and four-quarters-ahead forecast obtained from the classical AR model for the small–middle segment, the LSTAR model always outperforms the AR models whether estimated under classical or Bayesian models. Interestingly, unlike in the literature, the classical AR model outperforms all the BAR models. The result is in line with the fact that many of the coefficients on the longer lags in the linear AR model is significant and different from zero, as seen from Table 4, and, hence, tends to invalidate the Bayesian prior assumption on the mean of the coefficients. The evidence on the best performing BAR models is mixed. Following the literature (see for example, Das et al., 2008, 2010, forthcoming, and references cited there in for further details) on forecasting with Bayesian (V)AR models, we use the average RMSEs over one- to four-quarters-ahead to find that BAR1 ($w = 0.3$ and $d = 0.5$) performs the best for the medium and small segments, the BAR4 ($w = 0.2$ and $d = 2.0$) outperforms other BAR models for the large and luxury segments, while, for the affordable section of the housing market, the best suited BAR model is the BAR 5 model, with ($w = 0.1$ and $d = 2.0$).

Though the main objective of this paper is to analyze non-linearity individually in the house prices of the five segments, the possibility of causal influence amongst these house prices cannot be ignored. In fact, as shown in Table 6, causality tests revealed strong evidence of the same. In light of this, we also estimate and forecast the multivariate versions of the linear and non-linear models. For the Bayesian VAR, we retained the parameterization of w and d used in the Bayesian AR models, but change $k_{ij} = 0.5$, as popularly used in the literature. The basics of a multivariate LSTAR (MLSTAR) model can be summarized as follows:

The specification of MSTAR models follows the procedure used for the univariate STAR models. Define $y_t = (y_{1t}, y_{2t}, \dotsc, y_{kt})'$ as $(k \times 1)$ vector time series. In our case, $y = (Large, Medium, Small, Luxury, Affordable)'$, where all variables are in logarithms. We specify the k-dimensional MSTAR model as follows:

$$\Delta_4 y_t = \left(\Theta_{1,0} + \sum_{j=1}^{p} \Theta_{1,j} \Delta y_{t-j} \right) + \left(\Theta_{2,0} + \sum_{j=1}^{p} \Theta_{2,j} \Delta_4 y_{t-j} \right) G(s_t; \gamma, c) + \varepsilon_t, \qquad (9)$$

where Δ_4 denotes the quarterly difference operator such that $\Delta_4 x_t = x_t - x_{t-4}$, $\Theta_{i,0}$, $i = 1,2$, are $(k \times 1)$ vectors, $\Theta_{i,j}$, $i = 1,2$, $j = 1,2,\dotsc,p$, are $(k \times k)$

---

[12] For this, we used the Kalman filter algorithm in RATS (Version 7.0) for the AR and BAR models. While, the recursive forecasts from the LSTAR model is based on 2000 bootstrap replications, since analytical point forecasts are not available for non-linear AR models when the disturbance term is Gaussian even when $h \geq 2$, as $E[f(x)] \neq f[E(x)]$, where h is the number of steps-ahead for the forecasts. Details of the bootstrapping procedure are available upon request from the authors.

[13] Note that if A_{t+h} denotes the actual value of a specific variable in period $t+h$ and $_t F_{t+h}$ equals the forecast made in period t for $t+h$, the RMSE statistic equals the following: $\sqrt{\sum_{1}^{N} (_t F_{t+h} - A_{t+h})^2 / N}$ where N equals the number of forecasts.

Table 6
VAR Granger causality/block exogeneity Wald tests.

Excluded	Chi-sq	df	Prob.
Dependent variable: LARGE			
MEDIUM	34.32604	8	0.0000
SMALL	13.35981	8	0.1001
LUXURY	13.05156	8	0.1101
AFFORDABLE	13.93679	8	0.0834
All	95.57204	32	0.0000
Dependent variable: MEDIUM			
LARGE	37.48630	8	0.0000
SMALL	23.51406	8	0.0028
LUXURY	4.812892	8	0.7774
AFFORDABLE	2.578906	8	0.9579
All	86.61967	32	0.0000
Dependent variable: SMALL			
LARGE	6.000619	8	0.6472
MEDIUM	24.87303	8	0.0016
LUXURY	5.419474	8	0.7119
AFFORDABLE	13.45155	8	0.0972
All	110.6922	32	0.0000
Dependent variable: LUXURY			
LARGE	14.38929	8	0.0722
MEDIUM	20.46393	8	0.0087
SMALL	6.858180	8	0.5520
AFFORDABLE	20.86293	8	0.0075
All	56.00612	32	0.0054
Dependent variable: AFFORDABLE			
LARGE	10.73026	8	0.2175
MEDIUM	7.917962	8	0.4415
SMALL	11.51856	8	0.1740
LUXURY	10.16040	8	0.2539
All	42.92864	32	0.0939

matrices, and $\varepsilon_t = (\varepsilon_{1t}, \varepsilon_{2t}, \cdots, \varepsilon_{kt})$ is a k-dimensional vector of white noise processes with zero mean and nonsingular covariance matrix Σ, $G(\cdot)$ is the transition function that controls smooth moves between the two regimes, and s_t is the transition variable. In both univariate and multivariate cases, we allow the transition variable s_t to equal any lagged component of y_t. The MSTAR model in Eq. (9) defines two regimes, one associated with $G(s_t; \gamma, c) = 0$ and another associated with $G(s_t; \gamma, c) = 1$. The transition from one regime to the other is smooth and determined by the shape of the $G(\cdot)$ function. In this paper, we consider a logistic transition function, Eq. (2):

$$G(s_t; \gamma, c) = \frac{1}{1 + exp\{-\gamma(s_t - c)/\hat{\sigma}_s\}}, \quad \gamma > 0,$$

where $\hat{\sigma}_s$ is the estimate of the standard deviation of transition variable s_t. The parameter c is the threshold determining the midpoint between two regimes at $G(c; \gamma, c) = 0.5$. The speed of transition between the regimes is determined by the parameter γ, with higher values corresponding to faster transition. To specify the MSTAR models, we follow the procedure presented in Terasvirta (1998) (see also van Dijk, et al., 2002; Lundbergh and Terasvirta, 2002). The first step specifies the lag order of $p = 8$. The next step tests linearity against the MSTAR alternative. Since the MSTAR model contains parameters not identified under the alternative, we follow the approach of Luukkonen et al. (1988) and replace the transition function $G(\cdot)$ with a suitable Taylor approximation to overcome the nuisance parameter problem. The testing procedure selects a logistic MSTAR model with a single threshold. The third step in our MSTAR model identification selects the transition variable s_t. Although a wider selection is available in the multivariate case, we use the transition variables determined in the univariate case.

The results form the forecasting exercise obtained from the multivariate models are presented in Tables 7 and 8. Table 7 compares across

Table 7
One- to four-quarters-ahead RMSEs of house price growth rates from multivariate models (2001:Q1–2009:Q3).

Segments	Models	QA			
		1	2	3	4
Large	VAR	1.8342	4.2076	6.5169	8.8474
	BVAR1	2.1983	4.4559	6.4477	8.3087
	BVAR2	2.3720	4.3253	6.0770	7.7498
	BVAR3	2.6385	4.5647	6.2745	7.8759
	BVAR4	2.4448	4.2959	5.9492	7.5220
	BVAR5	2.7107	4.6432	6.2984	7.7886
	MSTAR	1.2315	2.8660	4.2998	5.6596
Medium	VAR	1.5230	4.0192	6.7186	9.5419
	BVAR1	1.7737	4.1264	6.5441	8.8587
	BVAR2	1.9245	4.2364	6.4681	8.4925
	BVAR3	2.2916	4.7131	6.9345	8.8551
	BVAR4	2.1467	4.5782	6.7544	8.5875
	BVAR5	2.5987	5.1517	7.3019	9.0353
	MSTAR	0.7950	2.1364	3.6089	5.1556
Small	VAR	2.2468	5.0138	7.2608	9.9139
	BVAR1	2.2535	4.5907	6.8039	9.0062
	BVAR2	2.3335	4.5987	6.6337	8.4876
	BVAR3	2.7302	5.1179	7.1483	8.8699
	BVAR4	2.6013	5.0531	7.0903	8.7820
	BVAR5	3.0500	5.6553	7.7317	9.3765
	MSTAR	1.1606	2.7100	4.2094	5.8529
Luxury	VAR	3.6252	5.8442	6.9924	7.3924
	BVAR1	3.9107	6.2952	7.3144	7.6294
	BVAR2	3.8819	6.0285	6.9600	7.2427
	BVAR3	4.0858	6.0023	6.7304	6.8889
	BVAR4	3.8076	5.6874	6.3714	6.5275
	BVAR5	4.1344	5.9491	6.5300	6.6257
	MSTAR	1.8810	3.3623	4.5137	5.6317
Affordable	VAR	4.9578	11.5256	15.1037	16.4163
	BVAR1	5.5046	11.1813	14.4068	15.2496
	BVAR2	5.2487	9.9510	12.4820	12.6338
	BVAR3	5.5468	9.3403	11.0710	10.9772
	BVAR4	4.8294	8.7656	10.9426	11.4140
	BVAR5	5.0296	8.3231	10.0755	10.6504
	MSTAR	1.3754	3.5085	5.2056	6.1756

Notes: QA: Quarter(s) Ahead; BVAR1: $w = 0.3$, $d = 0.5$; BVAR2: $w = 0.2$, $d = 1.0$; BVAR3: $w = 0.1$, $d = 1.0$; BVAR4: $w = 0.2$, $d = 2.0$; BVAR5: $w = 0.1$, $d = 2.0$.

the multivariate models, while Table 8 compares the best performing univariate model relative to all the multivariate models. Table 7 indicates that the MSTAR always outperforms the linear classical and Bayesian VAR models. Within, the linear VAR models, there always exists a BVAR model that beats its classical counterpart on average. Specifically, the optimal Bayesian models are as follows: BVAR4 ($w = 0.2$, $d = 2.0$) for large and luxury segments, BVAR2 ($w = 0.2$, $d = 1.0$) for middle and small segments and BVAR5 ($w = 0.1$, $d = 2.0$) for the affordable segment. Further, as can be seen from Table 8, though there is no sufficiently strong evidence of the multivariate linear models outperforming the best-performing linear and non-linear univariate models, the MLSTAR model is clearly the overwhelming favorite. Thus from a forecasting perspective, our results highlight the importance of accounting for non-linearity, as well as the possible interrelationship amongst the variables under consideration.[14]

6. Conclusion

A large number of recent papers have shown that there exists a strong link between the housing market and economic activity. In addition, these papers have also highlighted the fact that house price movements lead real activity or inflation, or both. In this backdrop, models that forecast house price movements can provide policy makers with an idea about the direction of the economy, and hence, a better

---

[14] We would like to thank an anonymous referee for bringing this to our notice.

Table 8
One- to four-quarters-ahead ratio of RMSEs of multivariate models with best-performing univariate models of house price growth rates (2001:Q1–2009:Q3).

Segments	Models	QA 1	2	3	4
Large	VAR	0.8194	0.9046	1.0438	1.1908
	BVAR1	0.9821	0.9580	1.0327	1.1183
	BVAR2	1.0597	0.9299	0.9734	1.0431
	BVAR3	1.1787	0.9814	1.0050	1.0600
	BVAR4	1.0922	0.9236	0.9529	1.0124
	BVAR5	1.2110	0.9983	1.0088	1.0483
	MLSTAR	0.5502	0.6162	0.6887	0.7617
Medium	VAR	1.1697	1.1879	1.2166	1.2469
	BVAR1	1.3622	1.2196	1.1850	1.1576
	BVAR2	1.4781	1.2521	1.1712	1.1097
	BVAR3	1.7600	1.3930	1.2557	1.1571
	BVAR4	1.6487	1.3531	1.2231	1.1221
	BVAR5	1.9958	1.5227	1.3222	1.1807
	MLSTAR	0.6106	0.6314	0.6535	0.6737
Small	VAR	1.1245	1.1217	1.0768	1.0935
	BVAR1	1.1278	1.0271	1.0091	0.9934
	BVAR2	1.1679	1.0289	0.9838	0.9362
	BVAR3	1.3665	1.1450	1.0602	0.9784
	BVAR4	1.3019	1.1305	1.0516	0.9687
	BVAR5	1.5265	1.2653	1.1467	1.0342
	MLSTAR	0.5809	0.6063	0.6243	0.6456
Luxury	VAR	1.1246	1.1074	1.1946	1.1239
	BVAR1	1.2131	1.1929	1.2496	1.1599
	BVAR2	1.2042	1.1423	1.1890	1.1011
	BVAR3	1.2674	1.1374	1.1498	1.0473
	BVAR4	1.1811	1.0777	1.0885	0.9924
	BVAR5	1.2825	1.1273	1.1156	1.0073
	MLSTAR	0.5835	0.6371	0.7711	0.8562
Affordable	VAR	2.1778	2.2811	2.1519	2.0343
	BVAR1	2.4180	2.2129	2.0526	1.8897
	BVAR2	2.3056	1.9694	1.7783	1.5656
	BVAR3	2.4365	1.8486	1.5773	1.3603
	BVAR4	2.1213	1.7348	1.5590	1.4144
	BVAR5	2.2093	1.6472	1.4355	1.3198
	MLSTAR	0.6042	0.6944	0.7417	0.7653

Notes: The best-performing univariate models is the LSTAR model for all housing segments and quarters, barring 1, 2 and 4 quarter(s) ahead forecasts for the small middle-segment; QA: Quarter(s) Ahead; BVAR1: $w = 0.3$, $d = 0.5$; BVAR2: $w = 0.2$, $d = 1.0$; BVAR3: $w = 0.1$, $d = 1.0$; BVAR4: $w = 0.2$, $d = 2.0$; BVAR5: $w = 0.1$, $d = 2.0$.

control for designing appropriate policies. Thus, it is of paramount importance that one can deduce the underlying nature of the data-generating process for house price, i.e., whether it is linear or non-linear, as the presumption of house prices being linear could lead to incorrect forecasts for not only house prices, but the economy in general.

Against this backdrop, this paper first tests whether house prices in the five segments of the South African housing market, namely, large, medium, small, luxury and affordable, exhibit non-linearity based on the smooth transition autoregressive (STAR) models estimated using quarterly data from 1970:Q2 to 2009:Q3. Second, we compare the one- to four-quarters-ahead out-of-sample forecasting performances of the non-linear time series models with those of the classical and Bayesian versions of the linear autoregressive (AR) models for each of these segments, over an out-of-sample horizon of 2001:Q1 to 2009:Q3, using an in-sample of 1970:Q2 to 2000:Q4. We find overwhelming evidence of non-linearity in these five segments based on in-sample statistical tests. Specifically, we obtain that the LSTAR framework is statistically preferable over the ESTAR framework in describing house price growth rate movements in all the five segments of the South African housing market, implying that the expansion and contraction regimes have different dynamics. More importantly, our results imply that a linear model would clearly be misspecified 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. We then provide further support for non-linearity by comparing one- to four-quarters-ahead out-of-sample

forecasts of the LSTAR model with those of the classical and Bayesian versions of the linear autoregressive (AR) models over an out-of-sample horizon of 2001:Q1 to 2009:Q3, using an in-sample of 1970:Q2 to 2000:Q4. Our results indicate that barring the one-, two and four-step(s)-ahead forecasts of the small–middle-segment the LSTAR model always outperforms the linear models. In addition, given the existence of strong causal relationship amongst the house prices of the five segments, multivariate versions of both linear (classical and Bayesian) and non-STAR models were estimated. Though there is not enough evidence of the multivariate linear models to outperform the best-performing linear and non-linear univariate models, the MLSTAR model is by far the overwhelming favorite. Our paper thus highlights the importance of not presuming the movements of house prices as linear.

Rapach and Strauss (2007, 2009) and Das et al. (2008, 2010, forthcoming) report evidence that numerous economic variables, such as income, interest rates, construction costs, labor market variables, stock prices, industrial production, and consumer confidence index potentially predict movements in house prices and the housing sector, and hence, show that large-scale models like the dynamic factor model (DFM), factor-augmented VAR (FAVAR) or large-scale Bayesian VAR forecast much better than small-scale pure time series models of house prices (i.e., models that contain only the house price in levels or growth rates). Given this, further research could focus on comparing the forecast performance of these large-scale models, involving hundreds of variables, with an appropriate pure non-linear time series model, especially if there is evidence to suggest that the underlying data generating process for house price is non-linear. In addition, one can go beyond point forecasts and look at interval and density forecast comparisons across linear and non-linear models as in Rapach and Wohar (2006a).

References

Banbura, M., Gianonne, D., Reichlin, L., 2010. Bayesian VARs with large panels. Journal of Applied Econometrics 25 (1), 71–92.

Bao, Y.C., Guay, C.L., Li, S.M., 2009. A small open economy DSGE Model with a housing sector. Preliminaly Draft Paper Prepared for the Conference of Economists. Department of Economics, the University of Melbourne.

Beck, N., King, G., Zeng, L., 2000. Improving quantitative studies of international conflict: a conjecture. American Political Science Review 94, 21–36.

Beck, N., King, G., Zeng, L., 2004. Theory and evidence in international conflict: a response to de Marchi, Gelpi, and Grynaviski. American Political Science Review 98, 379–389.

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.

Bloor, C., Matheson, T., 2008. Analysing shock transmission in a data-rich environment: a large BVAR for New Zealand. Reserve Bank of New Zealand, Discussion Paper Series DP2008/09.

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.

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

Case, K., Shiller, R., Quigley, J., 2005. Comparing wealth effects: the stock market versus the housing market. Advances in Macroeconomics 6, 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.

Clements, M., Hendry, D., 1998. Forecasting Economic Time Series. Cambridge University Press, New York.

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

Das, S., Gupta, R., Kabundi, A., 2008. Could we have predicted the recent downturn in the South African housing market? Journal of Housing Economics 18, 325–335.

Das, S., Gupta, R., Kabundi, A., forthcoming. Is a DFM well-suited for forecasting regional house price inflation? Journal of Forecasting.

Das, S., Gupta, R., Kabundi, A., 2010. Forecasting real house price growth in the nine census divisions of the US. Journal of Housing Research 19 (1), 89–109.

Doan, T., Litterman, R., Sims, C., 1984. Forecasting and conditional projections using realistic prior distributions. Econometric Reviews 3, 1–100.

Dua, P., Ray, S.C., 1995. A BVAR model for the Connecticut economy. Journal of Forecasting 14 (3), 167–180.

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

Engelhardt, G., 2001. Nominal loss aversion, housing equity constraints, and household mobility: evidence from the United States. Center for Policy Research Working Paper 42, 1–61 (September).

Englund, P., Ioannides, Y.M., 1997. House price dynamics: an international empirical perspective. Journal of Housing Economics 6 (2), 119–136.

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 1233–1260 (November).

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., 2006. Forecasting the South African economy with VARs and VECMs. South African Journal of Economics 74 (4), 611–628.

Gupta, R., Miller, S.M., 2009. "Ripple Effects" and Forecasting Home Prices in Los Angeles, Las Vegas, and Phoenix. University of Nevada, Las Vegas. Working Paper No. 0902 [Online] Available from http://ideas.repec.org/p/nlv/wpaper/0902.html.

Gupta, R., Sichei, M.M., 2006. A BVAR model for the South African economy. South African Journal of Economics 74 (3), 391–409.

Gupta, R., Das, S., 2010. Predicting Downturns in the US Housing Market: A Bayesian Approach. Journal of Real Estate Finance and Economics 41 (3), 294–319.

Gupta, R., Miller, S.M., Stephen, M., forthcoming. The Time-Series Properties on Housing Prices: A Case Study of the Southern California Market. Journal of Real Estate Finance and Economics.

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

Iacoviello, M., 2005. House prices, borrowing constraints, and monetary policy in the business cycle. The 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 (2), 125–164.

Leamer, E.E., 2007. Housing is the business cycle. Working Paper 13428. National Bureau of Economic Research.

Lundbergh, S., Terasvirta, T., 2002. Forecasting with smooth transition autoregressive models. In: Clements, M.P., Hendry, D.F. (Eds.), A Companion to Economic Forecasting. Blackwell, Oxford, pp. 485–509.

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

Muellbauer, J., Murphy, A., 1997. Booms and busts in the UK housing market. The Economic Journal 107 (1), 701–727.

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.

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., Strauss, J.K., 2007. Forecasting real housing price growth in the Eighth District states. Federal Reserve Bank of St. Louis. Regional Economic Development 3, 33–42.

Rapach, D.E., Strauss, J.K., 2009. Differences in housing price forecast ability across U.S. states. International Journal of Forecasting 25 (2), 351–372.

Rapach, D.E., Wohar, M.E., 2006a. The out-of-sample forecasting performance of nonlinear models of real exchange rate behavior. International Journal of Forecasting 22 (2), 341–361.

Rapach, D.E., Wohar, M.E., 2006b. In-sample vs. out-of-sample tests of stock return predictability in the context of data mining. Journal of Empirical Finance 13 (2), 231–247.

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).

Kim, Sei-Wan, Bhattacharya, Radha, 2009. Regional Housing Prices in the USA: An Empirical Investigation of Nonlinearity, Journal of Real Estate Finance and Economics, 38, 443–460.

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

Terasvirta, T., 1998. Modelling economic relationships with smooth transition regressions. In: Ullah, A., Giles, D.E.A. (Eds.), Handbook of Applied Economic Statistics. Marcel Dekker, New York, pp. 507–552.

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

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

Theil, H., 1971. Principles of Econometrics. Wiley, New York.

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

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.