ARIMA modelling of Lithuanian house price index

Author(s):
Arvydas Jadevicius (The Royal Agricultural University)

Abstract:
Purpose
– This paper aims to investigate Lithuanian house price changes. Its twin motivations are the importance of information on future house price movements to sector stakeholders and the limited number of related Lithuanian property market studies.

Design/methodology/approach
– The study employs ARIMA modelling approach. It assesses whether past is a good predictor of the future. It then examines issues relating to an application of this univariate time-series modelling technique in a forecasting context.

Findings
– As the results of the study suggest, ARIMA is a useful technique to assess broad market price changes. Government and central bank can use ARIMA modelling approach to forecast national house price inflation. Developers can employ this methodology to drive successful house-building programme. Investor can incorporate forecasts from ARIMA models into investment strategy for timing purposes.

Research limitations/implications
– Certainly, there are number of limitations attached to this particular modelling approach. Firm predictions about house price movements are also a challenge, as well as more research needs to be done in establishing a dynamic interrelationship between macro variables and the Lithuanian housing market.

Originality/value
– Although the research focused on Lithuania, the findings extend to global housing market. ARIMA house price modelling provides insights for a spectrum of stakeholders. The use of this modelling approach can be employed to improve monetary policy oversight, facilitate planning for infrastructure or social housing as a countercyclical policy and mitigate risk for investors. What is more, a greater appreciation of Lithuania housing market can act as a bellwether for real estate markets in other trade-exposed small country economies.

Keywords:
Housing, Prices, Index, Modelling, Lithuania, ARIMA
Article

In the past few decades, academic interest in property market modelling and forecasting expanded (Case and Shiller, 1990; Zhou, 1997; Barras, 2009; Brooks and Tsołacos, 2010). At the same time, the real estate profession responded to the growing need for better property market research (McDonald, 2002). As the field evolved, econometric models specifications improved (Barras, 2009; Lizieri, 2009). Multi-equation with stationary data models replaced initial simple single-equation specifications (Tonelli et al., 2004; Tsołacos, 2006; Brooks and Tsołacos, 2010).

According to Stevenson (2007), despite the range of modelling and forecasting techniques available to the researchers, many still employ regression, i.e. ordinary least squares (OLS)-based models. In case of Lithuanian residential property market research, regression analysis also dominates (see inter alia, Ambrasas and Stankevičius, 2007; Egert and Mihaljek, 2007; Ivanauskas et al., 2008). However, as Stevenson (2007) argued regression techniques can employ explanatory variables, without necessarily generating good modelling results. His solution was simple univariate models. Especially for short-term forecasts, Stevenson (2007) favoured the ARIMA technique, to mine information embedded in the series. Other useful assessments of model forecasting performance include Crawford and Fratantoni’s (2003) evaluation of Regime-Switching, ARIMA and GARCH house prices models for the USA. Crawford and Fratantoni found that although Regime-Switching model performs better in-sample, less complex ARIMA model generate better out-of-sample results. Hepsen and Vatansever’s (2011) validated ARIMA modelling superiority to forecast future trends of the Dubai housing market.

Despite its successes, ARIMA critics urge caution. McGough and Tsołacos (1995), Stevenson and McGrath (2003), Stevenson (2007), Stevenson and Young (2007) and Miles (2008) point out unreliability of some ARIMA long-term forecasts or turning point predictions. Canarella et al. (2012) discussed the modelling constraints when structural breaks are present in the series. In short,
notwithstanding its limitations, ARIMA time-series modelling is a useful tool for property market forecasting.

Having noted the predominance of regression-based international housing market analysis, the current study uses ARIMA modelling approach to investigate the Lithuanian housing market. Subsequently, the study assesses ARIMA’s application for forecasting. A greater appreciation of Lithuania housing market can act as a bellwether for real estate markets in other trade-exposed small country economies.

The remainder of this paper is as follows. The next section reviews the previous studies on housing market modelling. The data are presented in the third section. The fourth section provides with the empirical estimates. The final section concludes the study.

Previous studies

Section: Choose

**Housing market studies**

Housing market is a significant component of the nation’s economy. In the USA, housing stock is the largest element of household wealth (Guirguis et al., 2005). In the UK, Savills (2013) estimates its total value is around £5 trillion. In Lithuania, the housing market (EUR 5.9 billion) is worth twice the stock market (USD 3.96 billion) (Hypostat, 2013; World Bank, 2014). Even before the Global Financial Crisis, Shiller (1998) pointed out the critical role an efficient housing market plays for sustaining the economy. After the crisis, central banks and policymakers realised the importance of housing market analysis. Academic (Case and Shiller, 1990, 2003; Quigley, 1999) and professional research (Egert and Mihaljek, 2007; Goodhart and Hofmann, 2008) intensified.

Historically, Mankiw and Weil (1989) conducted one of the first statistically robust empirical analysis of housing markets. Researchers investigated how major demographic changes affected the US housing market. Mankiw and Weil (1989) employed a multiple-regression approach to quantify effect of age on housing demand and found that the “Baby Boom” generation sparked house price inflation in the 1970s.

Subsequent analysis of US housing markets by Case and Shiller (1990) employed multiple regression to forecast house prices in four US cities including Atlanta, Chicago, Dallas and Oakland. Significant independent price drivers were construction costs, income and population growth.

Malpezzi (1996) examined US house prices using a hedonic model. The regression results identified population change and income as the most significant regression determinants.

Cho (1996) reviewed theoretical and empirical issues surrounding house price dynamics at that time. For Cho, the main theoretical issue was whether real estate markets were efficient. He concluded – they were not. On the methodological side, “house price indexing” and “excess returns” were the key areas of contention. Cho advocated for model frameworks which incorporated
rational expectations, and tighter empirical measurements which included housing transaction costs.

Zhou (1997) examined the accuracy of a single vector autoregression (VAR) forecasting model with error correction term in predicting sales and prices of single-family houses in the USA. Quigley (1999) employed an autoregressive distributed lag modelling approach to examine the housing prices in US metropolitan areas. Case and Shiller (2003) performed linear and log-linear reduced-form regressions with three dependent variables, including the level of home prices, the quarter-to-quarter change in home prices and the price-to-income ratio to examine the relationship between house prices and explanatory variables. Gupta and Miller (2009a) assessed the relationship between house price series in Los Angeles, Las Vegas and Phoenix. The authors generated forecasts for each housing market using VAR and vector error correction (VEC) modelling approaches. In their subsequent paper, Gupta and Miller (2009b) examines the interrelationship between eight metropolitan statistical areas in Southern California, including Bakersfield, Los Angeles, Oxnard, Riverside, San Diego, San Luis Obispo, Santa Ana and Santa Barbara. The authors estimated VAR and VEC models as well as Bayesian (BVAR and BVEC), Spatial (SBVAR and SBVEC), as well as Causality (CBVAR and CBVEC) model specifications. More recently, Bork and Møller (2012) employed multiple regression approach to assess US house price forecastability.


Internationally, Englund and Ioannides (1997) examined house price changes in 15 The Organisation for Economic Co-operation and Development (OECD) countries. To explain future house price changes, the authors employed OLS-based equation with three explanatory variables. Goodhart and Hofmann (2008) assessed the interrelationship between money, credit markets, economic activity and house prices for 17 countries using fixed-effects panel VAR model. More recently, Hirata et al. (2013) examined house price fluctuations and their determinants for 18 OECD economies using the factor-augmented VAR (FAVAR) modelling approach.

The literature above shows dominance of regression in various permutations within housing market research. Mankiw and Weil (1989) and Case and Shiller

In all these studies, an OLS (regression) was used to quantify the impact on housing market from a set of explanatory variables. The gap in modelling research is univariate and ARIMA modelling.

**Lithuanian housing market studies**

Lithuania emerged from its Soviet shadow in the 1990s (Jadevicius and Parsa, 2014). Like other Eastern European countries, independence and privatisation unleashed a flood of investment, which stimulated its property markets. After the initial euphoria, investors turned to more sober modelling, hoping to better predict market fluctuations. The regression-based models only brought partial clarity. Ambrasas and Stankevicius (2007) analysed dwelling market in Vilnius using simple regression model. Their estimates suggested that credit worthiness, economic position of the state, living standards and migration influenced capital city housing market.

In their empirical study, Egert and Mihaljek (2007) used panel dynamic OLS modelling approach to assess the determinants of house prices in Central and Eastern Europe. They also used error-correction framework to assess the existence of long-term relationships between house prices and explanatory variables. Egert and Mihaljek (2007) found that conventional fundamentals drove Central and Eastern Europe (CEE) house prices. Unsurprising but significant variables were gross domestic product (GDP) per capita, real interest rates, housing credit and demographic indicators. In CEE, housing finance, transition-specific factors and institutional developments also exerted strong influences on housing market evolution.

When Ivanauskas et al. (2008) investigated the Lithuanian housing market with Cointegration and Granger causality tests, they contradicted Egert and Mihaljek’s (2007) earlier results. Fundamental economic factors such as interest rates, GDP and average income did not drive Lithuanian housing price inflation. According to Ivanauskas et al. (2008), house prices reacted to nation-specific factors, including historical background, cultural heritage and mentality.

Burinskiene et al. (2011) produced “a model of the public opinion on the factors” influencing real estate prices in Vilnius. The econometric specification the authors computed was based on OLS estimates. The study suggested hedonic factors including proximity to home, work and recreation amenities, as well safety and aesthetics were all significant factors in explaining house price changes in Lithuanian capital.

To bring greater clarity to the contested Lithuanian housing market, the current research adopts univariate modelling approach. As discussed, whilst ARIMA models have some limitations, the technique can strengthen property market forecasting. The current study, therefore, deploys the ARIMA approach.

Data and methodological framework
Data and its acquisition

The Ober-House Lithuanian House Price Index (OHBI) is adopted as the dependent variable for the current research. The OHBI is the longest house price index in Lithuania (Ober-Haus, 2014). The series covers the period January 1994 through July 2014. The OHBI is a weighted transaction-based index.Weights are allocated to each city and each property category accordingly. On the city level, weighting depends on the number of transactions in particular location. In terms of individual assets, weights are based on the determinants of the property, e.g. size, age and number of bedrooms. The data are then cross-referenced with the official statistics from the national registry.

Unfortunately, the index suffers from a number of limitations. A major criticism of the index is that its universe is unknown. There is no information available on the size and the number of transaction the index contains at each period of time. The other serious weakness of the data set is its weighting methodology. The appendix does not provide with any quantitative estimates noting as to how weights are allocated.

Notwithstanding these limitations, OHBI is considered an adequate benchmark of the national housing market. Figure 1 graphs monthly index performance over the research period. An examination of the housing index trends reveals several interesting observations. The OHBI values saw an interchanging periods of rise and decline over the 20 year period. In its early years, the index reflects three major market corrections, which all related to events in Russia. The first was trigger by the aftermath of the Russian constitutional crisis in 1993 (Brown, 1993; Treisman, 2001). The conflict between Boris Yel’tsin and his opponents resulted in the bloody military combat, which had a knock-on effect on all post-soviet economies. The second coincided with the economic crisis in Russia in 1998, when the rouble and Russian stock markets collapsed. Russia defaulted on both domestic and foreign debts (Brown, 1999, Lokshin and Ravallion, 2000). The crisis subsequently spread to the Baltic States (Dudzinska, 2013) but its aftershocks were global (Kotz, 1999).

The third notable series correction again emanated from Russia. In the 2000s, for political and geostrategic reasons, Moscow employed strong-arm tactics to pressure the Baltic States. Baltic goods were boycotted and sanctions slapped on Baltic shipments. Energy suppliers were cut. Markets realised Lithuania’s energy vulnerability. Moscow controlled energy security in the three Baltic nations who imported 90 per cent of their oil and almost all their gas from Russia (Grigas, 2012).

The most obvious from Figure 1 is the impact of the Global Financial Crisis. The OHBI for Lithuania reflects the broad situation in most other East European countries. During the worst part of the downturn, in 2009, Lithuania’s GDP declined by around 15 per cent (Macys, 2012). Internal consumption and incomes fell as unemployment rose. Foreign Direct Investment and bank lending declined. Government pruned public expenditures. The knock-on effect on property market was obvious. The OBHI almost halved from its peak of 802 points in December 2007 to 480 points in April 2010. To date, the recovery has
been shaky. The Index now deviates around 500 points mark, which is close to 40 per cent reduction from its peak.

**ARIMA model specification**

ARIMA is the econometric specification which combines an autoregressive operator (AR) of order $p$, the $d$-th difference and the moving average (MA) operator of order $q$. Here, the AR component of the specification implies that future values of the times-series can be approximated and predicted from the current and past values of the time-series itself. The MA component appreciates both current and past effects of random shocks, i.e. error terms, in the series. In combination, it becomes ARIMA model of order $(p,d,q)$ (Box et al., 1994; Stevenson and McGarth, 2003; Karakozova, 2004; Brooks and Tsonacos, 2010).

The basic equation of AR framework is as follows (Makridakis et al., 1998; Brooks and Tsonacos, 2010): Equation 1

Where the current value of variable $Y_t$ depends on past values of the variable itself $Y_{t-p}$, constant term $\mu$, $j$-th autoregressive parameter $\Phi_j$ and the error term $u_t$ at a time period $t$.

The equation for the MA process is presented in the equation below: Equation 2

Here, $\mu$ is a constant term, $b_j$ is $j$-th MA parameter and $u_t$ the error term at a time $t$.

ARIMA model is subsequently generated by combining both AR and MA processes into one equation (Box et al., 1994; Makridakis et al., 1998; Brooks and Tsonacos, 2010). The full model specification is as follows: Equation 3

Model evaluation

To correctly specify the ARIMA model, the dependent variable is first assessed for stationarity. Makridakis et al. (1998) suggest that visual analysis of time series is often enough to appreciate whether series contains a unit root. The current study, however, performs a formal testing. The study considers four unit-root tests, including augmented Dickey-Fuller (ADF) test, Elliott-Rothenberg-Stock (ERS) test, Phillips-Peron (PP) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test (Kwiatkowski et al., 1992).

Table I reports unit-root test results for OBHI house price index series. The tests results demonstrate that levels series are non-stationary. Hence, the series is differenced for model parameterisation.

Based on these results, 20 ARIMA specifications ranging from ARIMA (1,0,0) to ARIMA (4,0,4) are computed. The use of four autoregressive and four moving-average components comes from similar studies on the subject (inter alia, McGough and Tsonacos, 1995; Stevenson, 2007), as well as it is considered as being optimal for monthly series. A greater ARIMA order would incorporate extraneous information from a previous quarter.
All ARIMA models are estimated over January 1994-July 2013 period, leaving last 12 months for out-of-sample performance measurement. This time-series separation into initialisation and holdout periods allows testing models fit and their true forecasting performance. The 12-month out-of-sample period is considered as being adequate for model forecasting accuracy assessment. For government and private property market participants, one year period gives enough time to produce adequate strategies in relation to expected future housing market behaviour.

Table II reports in-sample model fit ($R^2$-squared) and likelihood ratio test statistics akaike information criterion (AIC). As the empirical estimates suggest, ARIMA (4,0,4) models fits monthly series best following ARIMA (4,0,3) and ARIMA (3,0,3) models. ARIMA (4,0,4) model has the largest $R^2$-square value amongst competing specifications. However, AIC estimates suggest ARIMA (3,0,3) as the best parameterised model. ARIMA (2,0,2) and ARIMA (4,0,4) specifications have the second and third lowest AIC values. In combination, these estimates make ARIMA (3,0,3) the best in-sample performing model, following ARIMA (4,0,4) specification. The least accurate is ARIMA (1,0,0) model. Its in-sample fit statistical values are the lowest of all competing specifications.

Table III contains main estimates for out-of-sample model performance and their ability to forecast house price index movements. The specified models forecast series in a 12-month ex-post horizon. The accuracy of each specification is assessed by computing Theil’s second inequality coefficient “U”. The use of Theil’s U allows for a relative comparison of the selected forecasting methods with a naïve modelling approach. As Chaplin (1998, 1999) and Makridakis et al. (1998) commented, if $U$ is equal to zero, then predictions are perfect. If it is equal to one, then the forecasts are the same as those that would be obtained using naïve forecasting approach. If, however, $U$ is greater than 1, then there is no point in using formal forecasting method. In such a case, a naïve approach would produce better results.

The modelling results suggest that ARIMA (3,0,3) model produced the best out-of-sample estimates. This particular specification had the smallest $U$ values amongst competing specifications. This model consistently outperformed competing specifications in all 12 out-of-sample periods. The second best performing specification was ARIMA (4,0,1) model with its accuracy estimates being close to those of ARIMA (3,0,3) specification. The lest accurate of all models was ARIMA (3,0,4) model. It had the lowest model fit statistical values (Figure 2).

An interesting observation which emerges from these estimates is that ARIMA (3,0,3) model was the best-performing specification in- and out-of-sample. This contradicts previous studies in some respect. Chaplin’s (1999), Newell’s et al.’s (2002) and Stevenson and McGarth’s (2003) estimates suggested that in-sample model fit does not necessarily imply good out-of-sample forecasting performance and vice-versa. In the latter three studies, the best fitting models were the least accurate ex-post. In the current study, however, ARIMA (3,0,3) outperformed all other specifications ex-post and ex-ante, suggesting finer modelling results.
The ARIMA (3,0,3) model was then used to generate 12 months ahead forecasts. The model was re-estimated on the whole time-period from January 1994 to July 2014 and then forecasts were produced up to July 2015. The estimates suggest Lithuanian house price growth over the next year. It is expected that OHBI index will, ceteris paribus, increase by around 8 per cent over the coming year (Figure 3).

Conclusions
Over recent years, extensive research has been undertaken to refine and tighten housing market forecasting. From these endeavours, various econometric specifications emerged. However, despite the range of modelling and forecasting techniques available, regression-based modelling still dominates the empirical side of the subject. Despite its obvious parsimonious advantages, ARIMA models have remained less popular. However, a recent spate of property market volatility has sparked fresh interest in the ARIMA approach. Volatile markets undermine confidence in the stability of model demand fundamentals. Eastern European, the Baltic States and Lithuania, in particular, offer a useful milieu to investigate the robustness of alternate models in volatile housing markets. To-date, Lithuanian property market econometric research reflects wider modelling bias. Regression-based specifications dominate. The current study fills the gap. It applied ARIMA technique to assess Lithuanian house prices.

Twenty ARIMA models were generated ranging from ARIMA (1,0,0) to ARIMA (4,0,4). The assessment of model accuracy suggested that ARIMA (3,0,3) model produced the best modelling results in- and out-of-sample. This differs from previous studies which commented that model fit does not necessarily imply good forecasting performance. The current study shies away from drawing any hasty conclusions around the best model specification solely on the basis of model fit statistics. However, in modelling Lithuanian house price index, the ARIMA (3,0,3) outperformed all other specifications ex-post and ex-ante. This particular specification was accurate in tracking the dependent variable for 12 months ex-post. Trivially, the model forecasts (ceteris paribus), 8 per cent house price inflation in the coming year, but the more substantive conclusion is that the ARIMA modelling technique plays a useful role in forecasting or analysing national housing markets.

Although the research focused on Lithuania, the findings extend to global housing markets. This is in line Hepsen and Vatansever’s (2011) suggestions that ARIMA house price modelling provides insights for a spectrum of stakeholders. Given the relative ease of specify and populating ARIMA, property market stakeholders should incorporate this modelling approach to analyse national housing markets. ARIMA models could tighten policy oversight in monetary, infrastructure or planning and social housing spheres. Government and central bank can use ARIMA modelling to forecast national house price inflation and dampen speculative excess. Commercially, ARIMA could prove invaluable for investors, financiers or developers. ARIMA forecasts can help tighten property investment analytics and feasibility forecasts.

Equation 1
Equation 2

Equation 3

Figure 1. OBHI index (January 1994-July 2014)

Figure 2. ARIMA (3,0,3) ex-post model fit

Figure 3. OBHI index forecasts for August 2014 – July 2015

Table I. Unit-root test results

Table II. In-sample model fit statistical estimates

Table III. Out-of-sample model fit statistical estimates

References


9.


10.


11.

Case, K.E. and Shiller, R.J. (1990), “Forecasting prices and excess returns in the housing market”, *Real Estate Economics*, Vol. 18 No. 3, pp. 253-273. [Crossref], [ISI], [Google Scholar] [Infotrieve]

12.


13.


14.


21.
22.


23.


24.


25.


26.


27.

28.


29.


30.


31.


32.


33.

34.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P. and Shin, Y. (1992), “Testing the null hypothesis of stationarity against the alternative of a unit root*: how sure are we that economic time series have a unit root?”, *Journal of Econometrics*, Vol. 54 Nos 1/3, pp. 159-178. [Crossref], [ISI], [Google Scholar] [Infotrieve]

35.


36.


37.


38.


39.

40.


41.


42.


43.

Mankiw, N.G. and Weil, D.N. (1989), “The baby boom, the baby bust, and the housing market”, Regional Science and Urban Economics , Vol. 19 No. 2, pp. 235-258. [Crossref], [ISI], [Google Scholar] [Infotrieve]

44.


45.

46.


47.


48.


49.

Savills (2013), “UK housing stock value climbs £5,000,000,000,000 as London becomes more dominant”, available at: www.savills.co.uk/_news/article/72418/144360-0/2/2013/uk-housing-stock-value-climbs-%C2%A35-000-000-000-000-000-as-london-becomes-more-dominant (accessed 8 August 2014). [Google Scholar]

50.


51.

52.


53.


54.


55.


56.

Tsai, I.C., Chen, M.C. and Ma, T. (2010), “Modelling house price volatility states in the UK by switching ARCH models”, *Applied Economics*, Vol. 42 No. 9, pp. 1145-1153. [Crossref], [ISI], [Google Scholar] [Infotrieve]

57.

58.


59.