FORECAST OF NEW HOME SALES AND PRICES: A CASE STUDY FOR THE UNITED STATES

by

JING LIU

Submitted in partial fulfilment of the requirements
for the degree of Master of Arts

at

Dalhousie University
Halifax, Nova Scotia
August 2010

DALHOUSIE UNIVERSITY

DEPARTMENT OF ECONOMICS

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "FORECASRT OF NEW HOME SALES AND PRICES: A CASE STUDY FOR THE UNITED STATES" by JING LIU in partial fulfillment of the requirements for the degree of Master of Arts.

Dated: August 18, 2010

Supervisor: _________________________________

Readers: _________________________________

____

DALHOUSIE UNIVERSITY

DATE: August 18, 2010

AUTHOR: JING LIU

TITLE: FORECAST OF NEW HOME SALES AND PRICES: A CASE STUDY FOR THE UNITED STATES

DEPARTMENT OR SCHOOL: Department of Economics

DEGREE: MA CONVOCATION: October YEAR: 2010

Signature of Author

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

A nation's housing sector has been the cornerstone of economic activity over the past several years. Right now, the economy of the United States is in a recession. To recover the economy, activity in the U.S. housing market deserves more attention, especially the new home market. Economists in the United States believe that if new home sales could keep increasing in the future, recovery of the whole housing market, even the whole economy in the U.S., would be hopeful. To bring the hope closer to the reality, forecasting changes in the new home market is important. An accurate forecast can provide useful information for the future, so that proper planning can take peace.

The purpose of this thesis is to look for an appropriate method which can accurately forecast changes in new home sales and prices in the U.S. housing market, so that policy makers base decisions on reliable information.

LIST OF ABBREVIATIONS USED

ACF	Autocorrelation Function
ADF	Augmented Dickey Fuller
ADL	Autoregressive Distributed Lag
AIC	Akaike's Information Criterion
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information Criteria
CONS	Construction Material Sales
CPI	Consumer Price Index
FPE	Final Prediction Error
HE	Household Estimated
HPI	Housing Price Index
HQIC	Hannan and Quinn Information Criterion
LP	Logarithmic Value of New Home Prices
LS	Logarithmic Value of New Home Sales
MAD	Mean Absolute Deviation
ME	Mideast
MSA	Metropolitan Statistical Areas
MSE	Mean Square Error
NHP	New Home Price
NW	Northwest
PACF	Partial Autocorrelation Function
PE	Personal Expenditure
RMSE	Root Mean Square Error
S	South
SBIC	Schwarz's Bayesian Information Criterion
SES	Single Exponential Smoothing
SP	Stock Price
VAR	Vector Autoregressive
VECM	Vector Error Correction Model
W	West

ACKNOWLEDGEMENTS

I would like to convey my gratitude to the following individuals without whom this thesis might not have been finished.

My heartfelt thanks go to my supervisor, Dr. Yonggan Zhao, who has been an amazing teacher and mentors me throughout the whole process. Meanwhile, I am thankful to Dr. Daniel Rosenblum and Dr. Leonard MacLean for their time and valuable feedback.

CHAPTER 1 INTRODUCTION

A nation's housing sector has been a cornerstone of economic activity over the past several years. The housing sector and growth of homeownership have been a foundation for the nation, and for people working toward the American dream. Taking the United States' housing market as an example, this thesis focuses on the forecast of changes in new home sales and prices.

Why does forecasting new home sales and prices play an important role in a nation's economy? Four main sectors of a nations' economy that benefits from accurate and timely forecast of new housing sales and prices answer this question.

The first economic sector that benefits from forecasting the housing market is business growth. Forecasting is an important part of a business's long term strategy and helps a business examine its current state and give information about what may happen in the future so that enough preparations can be made.

Second, the accurate and timely forecast of housing sales and prices has proven to be a good predictor of inflation. Forecasting housing prices can give policy makers an idea about the direction of CPI inflation in the future, and hence, can provide a better control for making appropriate policies.

Third, the housing market is highly related to the labor market. In an economy, if new home sales fall, employment of construction workers, real estate brokers and sales agents will be affected. Duffy, Gerald and Kearney (2005) found that rising housing prices, could significantly reduce the growth potential of the economy and shift the balance of labor market growth from employment to wages. Johnes and Hyclak (1999) found some evidence that housing prices have a significant effect on the size of the labor force and changes in unemployment.

Finally, the housing market is related to the financial market. When the housing market shows a downturn, the volume of mortgage applications fall. Fratantoni (2007) showed that a shrinking U.S. housing market correlated with declining mortgage applications in 2006. Goetzmann, Liang and Yen (2009) found that past price appreciation increased the number of mortgage applications, the leverage of borrowers, the prices of purchased homes, and that past price changes evidently affected the approvals of prime and subprime applications.

In summary, the housing market plays an important role in a nation's economy. An accurate prediction of the housing sales and prices can provide critical information about the future, so that enough preparations can be made for the nation's economy as a whole.

So far, most studies about forecasting housing sales and prices concentrate on housing price index (HPI), which measures the average price changes in repeat sales or refinancings of the same properties (existing homes). Few researchers or economists pay attention to forecasting new home sales and prices. Although existing home sales occupy the biggest part of the U.S. housing market, influences of new home sales and prices on the U.S. housing market and on the U.S. economy should not be ignored.

In Chapter 2 of this thesis, the U.S. housing market, including both the new home market and the existing home market during the recession period from 2007 to 2009, is described. From the description, we find that a recession has a bigger effect on the new home market than that on the existing home market. To ensure that the housing market develops in a healthy way during a recession, economists should closely concentrate on changes in the new home sales and prices.

With the importance of forecasting housing sales and prices to the development of an economy, the question with regard to accurate forecasting of housing sales and prices is raised. Chapter 3 provides a review of methods applied by different researchers.

In Chapter 4 of this thesis, time-series and dynamic structure models are adopted to forecast changes in new home sales and prices in the U.S. housing market. These models include the ARIMA model, the single exponential smoothing (SES) model, the Holt-Winters nonseasonal smoothing model, and the ADL model. To apply these forecasting models, data must be stationary. In other words, growth rates or changes of the data are stationary. For the ARIMA, the SES and the Holt-Winters nonseasonal smoothing models, only historical data are included. For the ADL model, forecasting changes in new home sales and prices is not only based on historical data, but also based on macro factors, such as changes in GDP, in personal disposable income, in stock prices, in prime loan rate, and in CPI. The results of out-of-sample forecasting and MSE of the ARIMA, the SES, the Holt-Winters nonseasonal smoothing and the ADL models are compared. According to the minimum MSE, the ADL models perform best on out-of-sample forecasting, because the ADL models not only take the effects of time into account, but also the effects of external factors.

So far, most forecasts about the housing market are country wide rather than a specific state or city. This can largely be explained by the existence of ripple effects. Although the ripple effect is not identified in the U.S. housing market, it deserves more attention. If direction, strength and origin of the ripple effect can be clarified, forecast based on regions or on metropolitan statistical areas (MSA) will be much more accurate. Therefore, Chapter 5 focuses on proving the existence of the ripple effect in the U.S. home market by a VAR model, which is used to identify the relationships among the four regions in the short run, and a VECM model, which is adopted to discuss the long-run relationships among the four regions. Results indicate that ripple effects exist in the U.S. housing market. The Northwest is the initial region of the ripple effect, and changes in home prices in the Northwest have positive effects on other three regions. Changes in home prices in the West do not have any effects on other three regions. Therefore, government and investors should pay more attention to changes in home prices in the Northwest in the U.S. housing market.

CHAPTER 2 HOUSING MARKET IN THE U.S.

2.1 The U.S. Housing Market in the Recession: 2007-2008

During the recession, the U.S. housing market was unstable because of the collapse of the subprime lending industry. The Office of Federal Housing Enterprise Oversight, which tracks mortgages loans bought or backed by Fannie Mae and Freddie Mac, stated that home prices in the first quarter of 2008 fell by 3.1% compared with the same period in 2007 and 1.7% as compared with that in the fourth quarter of 2007. This is the steepest decline in history. Another report from S&P showed a 14.1% drop in existing home prices in the first quarter of 2008. Moreover, a report from the National Association of Realtors said that home sales fell by 2.6% in June 2008, reaching a ten-year low. July brought further bad news. The number of new home construction projects started in July 2008 fell by 11%, reaching its lowest level in 17 years, while the number of new foreclosure notices shot up by 55% from 2007 and banks repossessed 184% more homes than in July 2007.

Homes are durable goods, and a recession can lead to lower demand for home sales, construction, appliances, and even furniture. Because homes are durable goods, new home construction and sales are often highly correlated with economic cycles. Economic cycles have a number of significant repercussions for the economy. The most notable of these is the fact that household disposable income rises during booms and falls during recessions. Therefore, the average American's purchasing power rises and falls in tune with these economic cycles. However, durable goods, or larger purchases that are generally meant to last a while, are very hard hit by recessions. For example, if one's income is halved, their food consumption will probably not change that much; they will be more likely to put off buying a new washing machine, car, or house. Since these goods are durable, those that they already have will probably last until their finances improve. If people intend to buy new homes, they must be confident that they'll have enough income to pay for it, so economic downturns can depress the new home market considerably.

The housing market was the main catalyst of the most painful downturn in the past 70 years in the U.S. and renewed weakness could hobble the economic recovery, because the housing market is extremely sensitive to economic cycles.

2.2 New Home Market in the U.S.

New home market is much more susceptible than existing home market in a recession. For the whole year of 2009, new home sales not only slumped by 22.9 percent to a record low of 374,000 units, but they were also losing out from the existing home sales market, with a flood of foreclosures depressing prices.

However, new home sales rose by 14.8 percent in April 2010. More buyers in the U.S. rushed to purchase new residential property, partly to take advantage of an expiring tax credit. The government was offering an $8,000 credit for first-time buyers and $6,500 for current homeowners who buy and move into another property. To qualify for the credit, buyers must have a signed contract complete by the end of April 2010 and must complete the transaction by the end of June 2010. Nearly 1.8 million households have used the credit at a cost of $12.6-billion, according to the Internal Revenue Service.

New home sales surged with help from government incentives, giving the economic recovery a jolt in April 2010. Along with the increase of new home sales, factory orders rose sharply in March 2010. Some economists pointed out that the decline in the U.S. mortgage rates in April 2010 could help support housing recovery's momentum, even without the temporary tax credit policy. Since surging sales are occurring at rock-bottom prices, others doubt the sustainability of the housing market recovery. Therefore, distressed owners and discouraged homebuilders continue to lower expectations of their properties' value for the future. If future changes in the new home market could be predicted accurately so that policy makers could make the right decisions to develop the housing market better, recovery of the economy would be more hopeful.

CHAPTER 3 LITERATURE REVIEW AND DATA DESCRIPTION

3.1 Time- Series and Dynamic Structure Methods

A key sector of the U.S. economy has been its housing market, which has served as a strong equity base on its bank balance sheet. The equity of the housing market increased until 2006. The recession occurred when housing prices started to crumble in 2007. Now, housing starts are still at the bottom, and housing prices still declining. Unemployment is steadily rising at 10% and showing no signs of a downward trend and this will further push banking failures and asset price suppression. Based on these statistics, consumers, and some economists are not confident about the current and future economy in the United States.

To make consumers and economists confident about the future economy, an accurate forecast is an important step. Forecasts can tell us which preparations and the extent of preparations that we need to make for changes to the future economy.

Forecasting the housing market is not an easy task to complete, but any forecasting approach that consistently provides better odds at making the correct investment and policy decisions than those from flipping a coin should merit careful examination.

Researchers and economists focus on two areas of forecast. One is to choose an optimal forecasting model to predict some specific target, and the other is to select the factors which may play important roles in forecasting. Zellner and Palm (1974) argued that for forecasting purposes, time-series models generally performed as well as or better than dynamic structural econometric specifications. They propose that dynamic structure models prove most effective in performing policy analysis, while, time-series models prove most effective at forecasting, because fewer errors enter the time-series models.

Usually, three kinds of time-series models are used to make forecasts: ARIMA models, exponential smoothing models, and ADL models. For example, Raymond (1997) applied the ARIMA model to forecast real-estate prices in Hong Kong. He argued that, by looking at autocorrelation coefficients for time lags of more than one period, one could determine additional information on how values of a given time series were related. An ARIMA model produces forecasts that are likely to be more accurate than those produced by other approaches. Moreover, ARIMA models have proved to be excellent short-term forecasting models for a wide variety of time series because short-term factors are expected to change slowly. In addition, he suggests that ARIMA model and econometric models can be combined to give an improved forecast. However, econometric methods are not always as accurate as time-series models. Econometric methods help in understanding causal relationships between variables and can provide evidence of the validity of economic theory.

Crawford and Fratantoni (2003) have focused on assessing the forecasting performance of Regime-Switching, ARIMA and GARCH models of housing prices. They found that when price changes on any particular home were difficult to predict, aggregate home prices were accurately predicted. Moreover, they compared the in- and out-of-sample forecasting performance of Regime-Switching, ARIMA and GARCH models. They found that home price growth was correlated with the probability of a regime change, but that Regime-Switching models performed better in-sample forecasting, while, simple ARIMA models generally performed better in out-of-sample forecasting.

Abeysinghe, Balasooriya and Tsui (2000) thought that if a sample size was not big enough, the ARIMA model would not perform well on forecasting. As opposed to the ARIMA model, a regression model uses more than one variable, which adds explanatory power to the model. Based on a Monte Carlo exercise, their results imply that even with autoregressive forecasts on the exogenous variables, the regression models forecast much better than the ARIMA models, especially as the forecast horizon increases. Therefore, they believed that when constrained by a small sample it would be better to resort to a regression model to produce better forecasts, rather than an ARIMA model.

The exponential smoothing method is another forecasting method frequently used by economists. There are three different forms of exponential smoothing: single or simple exponential smoothing, double or Holt exponential smoothing, and triple or Holt-Winters exponential smoothing. Single exponential smoothing finds a smooth approximation to a noisy signal. Double exponential smoothing allows for extracting a linear trend, and triple exponential smoothing takes into account periodic variations. The three different forms are based on different assumptions. To decide which model is most appropriate, it first must be determined whether the assumptions of the model fit the data set. Therefore, to apply exponential smoothing methods, there is always a problem of selection.

There are a large number of ways to solve the selection problem and choose an optimal model.

Janert (2006), for example, stated that simple exponential smoothing could work well for time series data without an overall trend. The single exponential smoothing assumes that the system will stay steady at its last value. The double exponential smoothing assumes that the system will continue to grow linearly at its most recent rate. Both of the two assumptions will miss drastic departures from the previous behavior. But, he believed that when we had a good reason to believe that either of the two models was applied, exponential smoothing could be helpful in predicting future states.

Sinha (2008) wrote a note on some exponential smoothing forecasting methods, comparing SES, double exponential smoothing and triple exponential smoothing. He observed that double or triple exponential smoothing methods might give more accurate forecasts, but that there might not be a significant difference among these methods. On the contrary, Gardner and McKenzie (2009) believed that damped trend method of exponential smoothing creates more accurate predictions.

Billah, King, Snyder and Koehler (2006) also studied how to select the appropriate exponential smoothing model. They used a method which was based on an information criterion when maximum likelihood methods were used in conjunction with exponential smoothing to estimate the smoothing parameters. Their results indicate that there is little to distinguish the various information criteria, and the information criterion approaches provide the best basis for automated method selection. However, the Akaike information criterion (AIC) has a slight edge over its counterparts. Snyder and Ord (2009) also found that the Akaike information criterion (AIC) worked slightly better than prediction validation on withheld data for forecasting, based on an innovation state space approach.

An additional factor important to determining the optimal exponential smoothing method is whether the seasonality exists in time-series data or not. For example, triple exponential smoothing works well when seasonality exists in the series. Many methods can be used to test the seasonality, such as regression analysis, nonparametric, chi-square methods, correlation analysis and graphical techniques. Among those methods, the graphical technique is the easiest one, but the seasonality unit roots test provides the most persuasive results. The unit roots test which was used by Franses (1990, 1991) will be used in our analysis.

Usually, most of ARIMA models and exponential smoothing models are based on a target's own historical data, so the external noise may have strong effects on the accuracy of the forecast results. Therefore, taking external factors into account may reduce noise. Pre-selection of the macroeconomic indicators under investigation is important for the analytical setup. Sutton (2002) believed that income, stock market wealth and interest rates were three major drivers of housing prices. He thought that income and stock prices increase housing prices by increasing the demand for houses due to the "wealth effect".

Abraham and Hendershott (1996) have focused on the determinants of real housing price appreciation, based on the variation in real housing price movements in 30 cities over the period from 1977 to 1992. They found that growth in real income, real construction costs, real after-tax interest rates, lagged real appreciation, and the difference between the actual and equilibrium real housing price levels could explain three-fifths of the variation in their sample. Sutton (2002), on the other hand, explained housing price changes by shocks to GNP growth, real interest rates, and equity prices. He found that in the United States, housing prices increase by about 1% over three years following a 10% rise in equity prices. In addition, the impact of stock price fluctuations on the U.S. housing prices appears to be smaller than in some of other countries. He proposes two possible explanations: one is that households in the United States might not have regarded their particularly large stock market gains as permanent; two is that investments may be a substitute for housing assets in the United States stock market.

Holly, Pesaran, and Yamagata (2006) modeled the dynamic adjustment of real housing prices by examining the extent to which real housing prices at the state level are driven by factors such as real income and common shocks, in other words determining the speed of adjustment of home prices to macroeconomic and local disturbances. Similarly Rapach and Strauss (2007) forecast real housing price growth for the individual states of the Federal Reserve's Eighth District. They used an ADL model to analyze the forecasting ability of a large number of potential predictors of real housing price growth. The potential predictors include the state housing price-to-income ratio, state unemployment rate, consumer confidence, and national inflation rate.

On the same topic, Tsatsaronis and Zhu (2004) studied the determinants of housing price dynamics based on 17 industrialized economies between 1970 and 2003. They found that for all included countries, the GDP growth affected the housing price growth by 7.6%.

Moreover, of all modeled predictors, inflation plays the most important role in housing price changes. Across countries, average inflation accounts for more than half of the total variation in housing prices at the five-year horizon. They also found that GDP growth summarized the information contained in other more direct measures of household income, such as unemployment and wages.

For the average American, housing is both the largest financial asset and the greatest expense in his or her lifetime. According to U.S. government statisticians, average Americans spend about 40% of disposable income on their houses by way of rent or mortgage payments. The 40% ratio is also used as the weighting for housing in the U.S. Consumer Price Index (CPI), which is the official measurement of the rate of price inflation in the U.S. economy. Many economists support the relationship between housing prices and CPI. For example, Ahearne, Ammer, Doyle, Kole and Martin (2005) studied housing prices and monetary policy based on 18 major industrial countries. They found that real housing prices were pro-cyclical and co-moving with each of real GDP, consumption, investment, CPI inflation, budget and current account balances, and output gaps. Moreover, since CPI can be used to index rents, the relationship between housing prices and rents can reflect the relationship between housing prices and CPI.

Taken together, for different data and scenarios, forecasting models are different. However, a plain fact is that comparison is the basic way to decide which forecasting model can yield accurate results.

3.2 Ripple Effects in Housing Market

Housing price shocks in one area are likely to have an impact on other areas (MacDonald and Taylor1993). This is so called housing price diffusion or ripple effect. The ripple effect in housing markets affects the accuracy of forecasts, which are based on a specific state or region. This is because it is hard to tell which is the initial state or region, how many initial states or regions there are in the whole U.S. housing market, and how different ripple effects interact. However, the existence of ripple effects in the U.S. housing market is still uncertain today.

Giussani B. and Hadjimatheou G.（1991） first tested the hypothesis with the housing price data of 10 areas in UK and pointed out there was housing price ripple effect in the Southeast UK and it preceded toward the Northwest. Alexander and Barrow（1994） asserted that the immobility of housing and divergence in housing prices were inevitable. But the migration of households owing to the economic changes within regions caused the possibility of convergence in regional housing prices. Ashworth and Parker（1997）, and Pollakowski and Ray（1997） also mentioned the ripple effect of housing prices. These studies described how the housing prices rose first in the South East and gradually spread out over the rest of UK.

Tu (2000) tested the dynamics of the Australian housing market at both the national and the sub-national level, and found two housing price diffusion paths: from Brisbane to Sydney then to Melbourne, or from Brisbane to national then to Melbourne. The research indicates that Brisbane is the first disseminator of this diffusion pattern. Luo et al (2007）developed an approach to quantitatively examine the diffusion patterns of housing prices in cities of Australia based on econometrics principles of the cointegration test and the error correction model, and confirmed Tu's results.

However, the ripple effect hypothesis does not receive much support in the U.S. economy. Gupta and Miller (2009) examined the time-series relationship among the housing prices in Los Angeles, Las Vegas, and Phoenix. They found that Los Angeles housing prices directly affected Las Vegas housing prices and indirectly affected Phoenix housing prices through their effect on Las Vegas housing prices, which indicates that the ripple effect exists in the U.S. housing market.

Until now, the studies on ripple effects of housing prices are insufficient, and the research about the existence of ripple effects of housing prices in the United States is limited. In our analysis, an intensive study about ripple effects will not be made, but the existence of ripple effects in the U.S. housing market will be discussed.

3.3 Data Description

To make sure that the sample size is large enough, monthly data on new home sales and prices in the United States are obtained from Economic Indicators website (http://www.census.gov/cgi-bin/briefroom/BriefRm#home_sales). The monthly data are from Jan 1963 to Jan 2010, a sample of 565 months.

Estimates of new single-family homes sold and for sale are obtained from the Survey of Construction (SOC), and the data are collected by Census field representatives. Two different surveys comprise SOC: (1) Survey of Use Permits (SUP) which estimates the amount of new home sales in areas which require building permits and (2) Nonpermit Survey (NP), which estimates the amount of new home sales in areas which do not require building permits. Less than 1% of all new single-family home sales take place in nonpermit areas.

The data on new home sales only include new single-family residential structures. Sales of multi-family units are excluded. Moreover, only houses sold prior to being built or built for sale are included. Excluded from these estimates are houses built for rent, houses built by the owner, and houses built by a general contractor on the owner's land. Estimates prior to 1999 include an upward adjustment of 3.3 percent made to account for houses for sale in permit-issuing areas that will never have a permit authorization.

The data on new home prices are also obtained from SOC, including average prices and median prices. All prices include the value of lands. The new home average price is the price agreed upon between purchaser and seller at the time when the first sales contract is signed or a deposit made, and it does not reflect any subsequent price changes resulting from change orders or from any other factors affecting the price of the house. But some data on average new home prices are missing. Therefore, the median sales prices are used. The median sales price is the sales price of the house which falls on the middle point of the total number of houses sold.

To take external factors into account and apply the ADL model, some other data will be needed, such as GDP, CPI, stock prices, prime loan rate, personal disposable income, households estimated and construction material sales. All these external factors are recorded monthly. Because monthly data obtained from Macroeconomic Advisor website (http://www.macroadvisers.com) on GDP are from Apr 1992 to Jan 2010, the data on CPI, stock prices, and prime loan rate and personal disposable income are all from Apr 1992 to Jan 2010, a sample of 214 months. Macroeconomic Advisers' index of Monthly GDP is a monthly indicator of real aggregate output that is conceptually consistent with real Gross Domestic Product in National Income and Product Accounts' (NIPA). The consistency derives from two sources. First of all, monthly GDP is calculated using much of the same underlying monthly source data that is used in the calculation of annual GDP. Secondly, the method of aggregation to arrive at monthly GDP is similar to that for official annual GDP.

The monthly data on CPI and stock prices are obtained from the Irrational Exuberance website (http://www.irrationalexuberance.com/). CPI can be seen as a price inflation indicator. The influences on changes in new home prices or sales from inflation can be shown, through the correlation between CPI and new home prices or sales.

The housing market and stock market are related. Taking stock prices into account can reduce some noise in prediction. The data on stock prices are open prices of the S&P 500 Index and are obtained from Yahoo Finance website (http://finance.yahoo.com/). The stocks included in the S&P 500 Index are those of large publicly held companies that trade on either of the two largest stock market companies in the U.S.: NYSE Euronext and NASDAQ OMX.

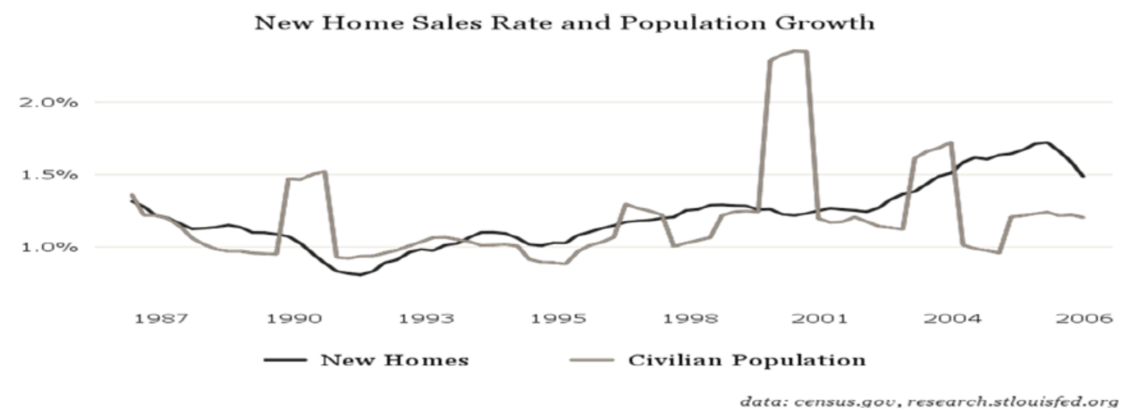
Monthly data on prime loan rates are from Federal Reserve Statistical Release (http://www.federalreserve.gov/). The prime loan rate is posted by a majority of the top 25 (by assets in domestic offices) insured U.S.-chartered commercial banks, and it is a short-term loan rate.

Data in billions of dollars on personal disposable income are from the U.S. Department of Commerce: Bureau of Economic Analysis. Personal disposable income is personal income less personal current taxes and the income available to persons for spending or saving. Personal disposable income is an important indicator of consumers' demand for new homes.

Figure 3.1 shows that new home sales tracked population growth for most of the last quarter of century. In other words, new home sales are a decent proxy for household formation in the last quarter of century. To a large extent, estimated number of households can tell us the future demand for new homes. The data on estimated households are collected by Housing Vacancies and Homeownership by computer-assisted telephone and personal visit interviews, and all values are in thousands.

Construction material sales, to some extent, indicate how many new homes and buildings are under construction, so construction material sales reflect the supply of new homes indirectly. Since supply of new homes is determined by demand for new homes,

Figure 3.1



construction material sales are related to new home sales. The data on construction material sales are estimated based on the data from the Monthly Wholesale Trade Survey, and have been benchmarked using results of the 2008 Annual Wholesale Trade Survey and the preliminary results of the 2007 Economic Census. Material sales estimates are in millions of dollars.

In our analysis, logarithmic values of all variables are used. Because all data and models are time series, stationarity of the data must be tested. If the data are non-stationary, first differences should be tested by ADF test with 12 lags because of monthly data. Results indicate that all data are stationary at first differences, so first differences are used to apply the ARIMA, the SES, the Holt-Winters nonseasonal smoothing and the ADL models.

To check for the existence of ripple effects in the U.S. housing market, regional and quarterly data of new home prices will be used, since monthly data cannot be obtained. The quarterly data are obtained from New Residential Sales of U.S. Census Bureau and from Q1:1963 to Q1:2010， including 189 samples. All statistics in the New Residential Sales release are tabulated only for the United States and four Census Regions, including Northwest, Mideast, South, and West. Ripple effects may happen when there are more than two groups, so the quarterly data of the four census regions are included. Because the existence of ripple effects will be checked by the VAR and the VECM models which require stationary data, first differences of the quarterly data are used.

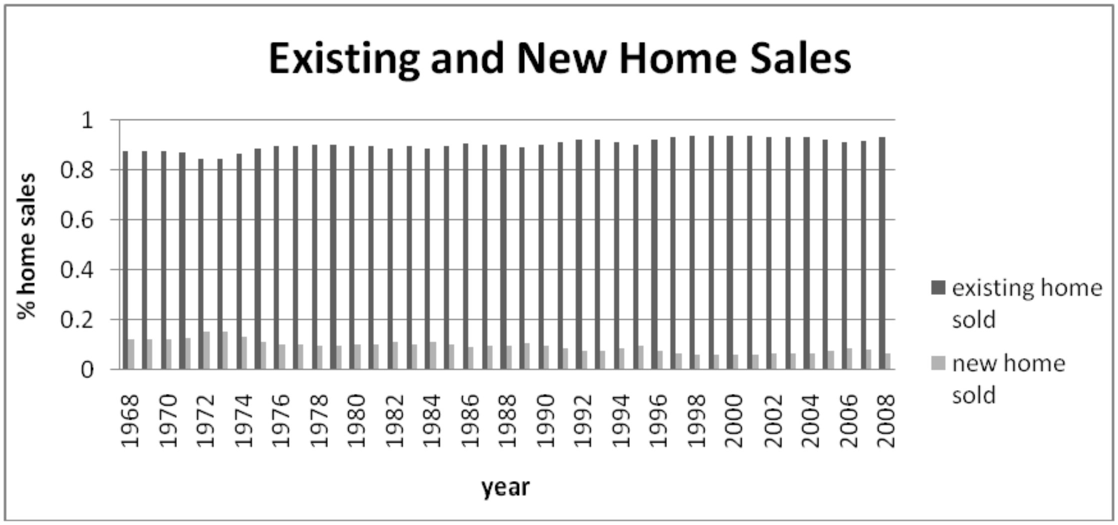
CHAPTER 4 FORECASTING NEW HOME SALES AND PRICES

Housing market adjustments play an important role in the economic cycle, not only because housing investment provides a volatile component of demand, but also because housing price changes generate important wealth effects on consumption and investment. The existing home market and new home market are two main parts of the housing market. But economists or researchers usually pay attention to existing home market in the U.S. housing market. One possible reason is that demand for existing homes plays a more important role than new homes in the U.S. housing market (see Figure 4.1). Among advantages of existing homes, lower sales price is the main reason why most people prefer existing homes. According to Figure 4.2, existing home sales prices are lower than new home sales prices in almost every single year. However, new home market is also an important sector in the housing market, and as mentioned in Chapter 1, housing market is a cornerstone of economic activity. Therefore, forecasting new home sales and prices is meaningful.

In this chapter, a time-series model, which is more appropriate to forecast new home sales and prices, will be discussed. Usually, the time-series models perform better than the structure models on out-of-sample forecasting. Different time-series approaches will be discussed, including ARIMA model, exponential smoothing model and ADL model.

Figure 4.1

Existing Home Sales and New Home Sales in the U.S.:1968-2008

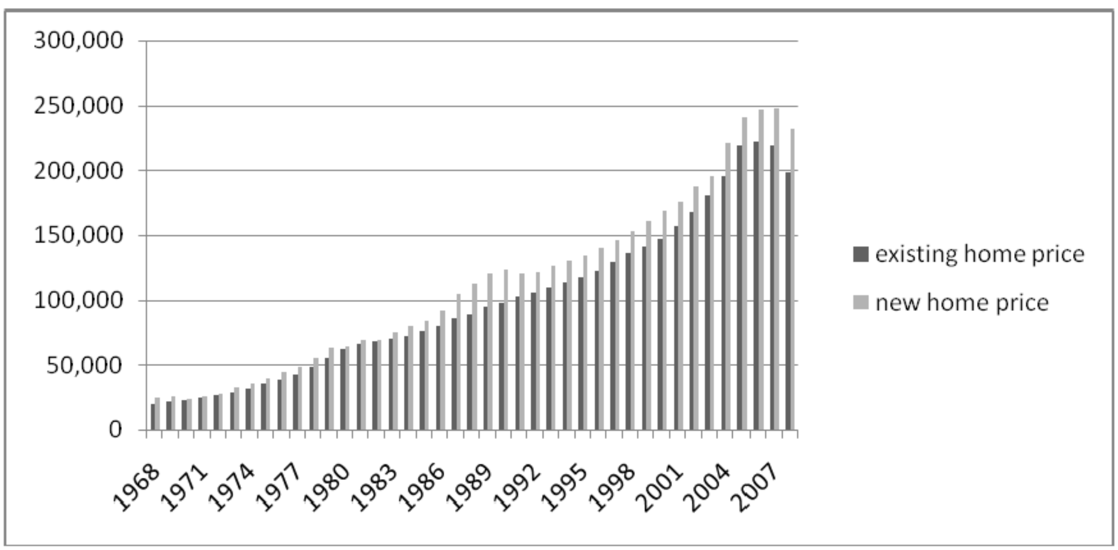


Note: All data is about annually single-family home sales, and assume that the whole housing market is composed of existing single-family homes and new single-family homes.

Source: Housing Sales Historic Press Release, U.S. Census Bureau

Figure 4.2

Existing Home Prices and New Home Prices in the U.S.:1968-2008



Note: Here, price is sales median price.

Source: National Association of Realtors and Economic Indicators, U.S. Census Bureau

4.1 Autoregressive Integrated Moving Average Model

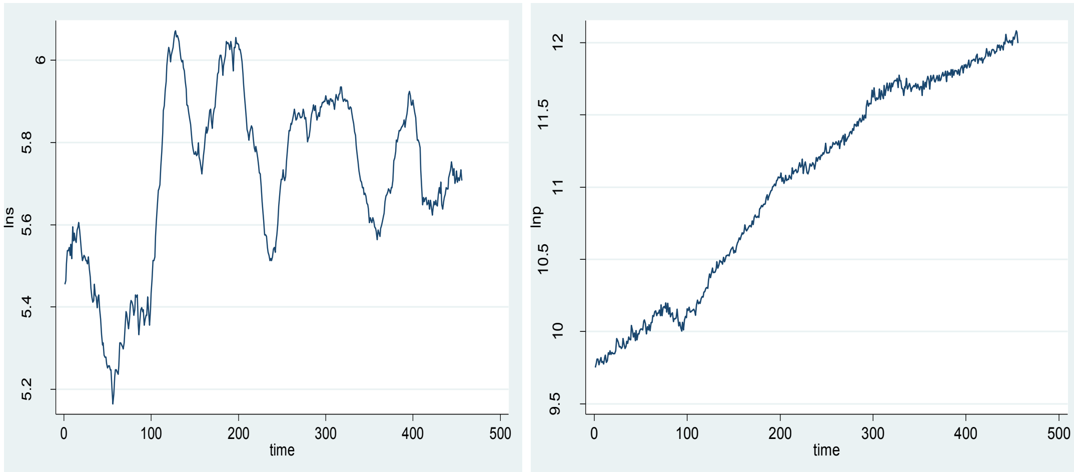
Generally, forecast models on real housing sales and price inflation or changes provide an idea about the future direction of the whole macroeconomy to policy makers. In other words, the housing sector acts as a leading indicator for economic activity.

ARIMA models are the most general type of models for time-series data which can be made stationary by differencing and logarithmic transformations. In our analysis, monthly data on new housing sales and prices are studied. The ARIMA model is fitted to the long-term historical time-series data. ARIMA econometric modeling takes into account historical data and decomposes it into three components: an autoregressive (AR) process, characterizing memory of past events; an integrated (I) process, accounting for stationarity of data; a moving average (MA), indicating lags of the forecast errors. Generally, ARIMA has a basic form like ARIMA (p, d, q), where p is AR (p), q is MA (q) and d is the number of differences.

To apply the ARIMA model for out-of-sample forecasting, stationarity of data from Jan 1963 to Dec 2000 (including 456 samples) is tested before specifying an ARIMA model. According to Figure 4.3, logarithmic values of new home sales (LS) look non-stationary and logarithmic values of new home prices (LP) look stationary. To make sure that the graphs show correct results, Augmented Dickey-Fuller (ADF) and Phillips-Perron (PP) tests are used to check stationarity of the time series data for lags 1 to lags 12. The results indicate that LS and LP are both non-stationary, because almost all p-values are greater than 0.05 (see Table 4.1).

Figure 4.3

Logarithmic Values of New Home Sales and Prices in the U.S.: Jan-1963 to Dec-2000



Source: Economic Indicators, U.S. Census Bureau

Since LS and the LP are non-stationary, their first differences are tested. Figure 4.4 and Figure 4.5 show that the first differences of LS and the first differences of LP look stationary. Meanwhile, the ADF test and the PP test are also used to check stationarity of the first differences of LS and the first differences of LP. The results are presented in Table 4.1. We can see that all p-values for first differences of the data are smaller than 0.05, which indicates that the first differences of LS and the first differences of LP are stationary.

In addition, to select the values of p and q, autocorrelation function (ACF) and partial autocorrelation function (PACF) are used for the first difference of the monthly time-series data. Figure 4.6 shows the results for the first differences of LS which indicate that the q in MA (q) should be equal to 10 or 11 and the p in AR (p) should be equal to 5. Figure 4.7 indicates that the q should be equal to 1 or 2 and the p should be 1 or 2 for first differences of LP.

Figure 4.4
First Difference of Logarithmic Values of New Home Sales in the U.S.: Feb-1963 to Dec-2000

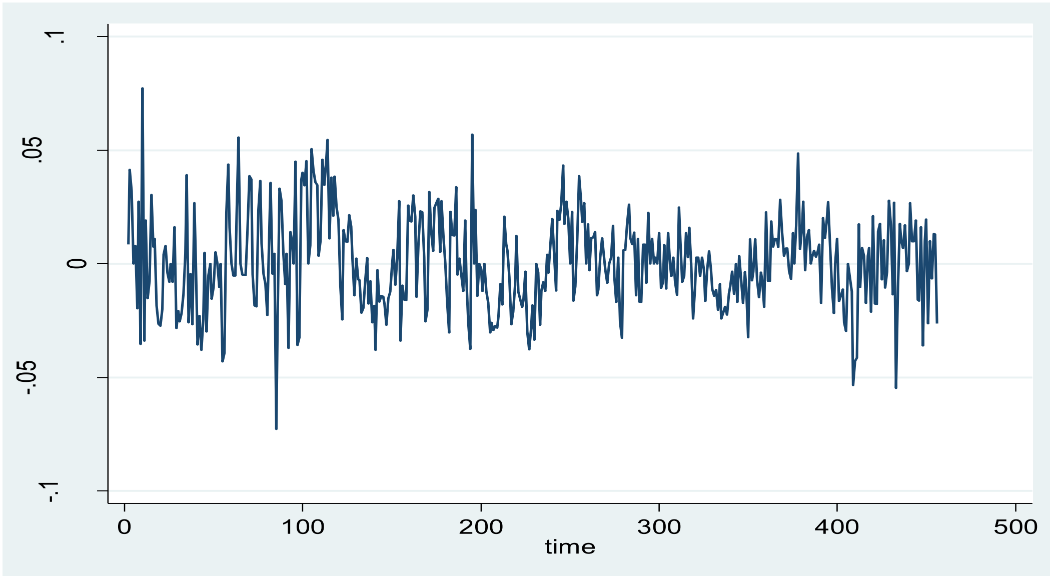


Figure 4.5
First Difference of Logarithmic Values of New Home prices in the U.S.: Feb-1963 to Dec-2000

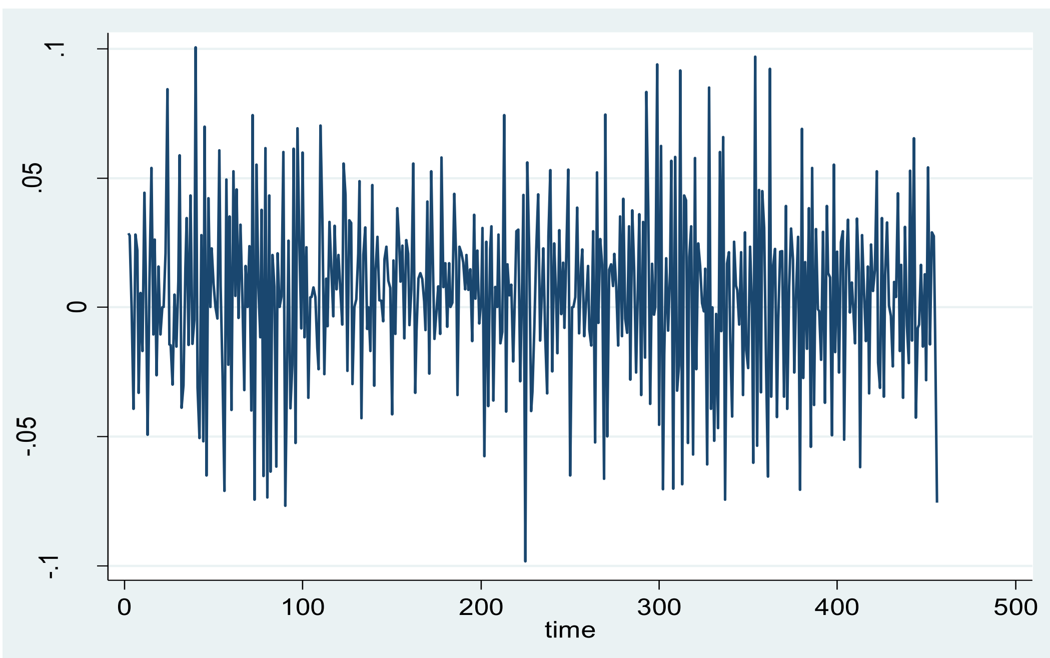


Table 4.1
Stationarity of Data about New Sales and Prices by ADF test

lags	LS trend	LS p-value	d.LS trend	d.LS p-value	LP trend	LP p-value	d.LP Trend	d.LP p-value
1	-1.15×10^{-6}	0.8451	-3.68×10^{-6}	0.0000	6.12×10^{-5}	0.9155	-1.2×10^{-5}	0.0000
2	-1.1×10^{-7}	0.7844	-2.85×10^{-6}	0.0000	3.58×10^{-5}	0.9668	-1.21×10^{-5}	0.0000
3	2.2×10^{-6}	0.7491	-3.26×10^{-6}	0.0000	3.16×10^{-5}	0.9721	-1.46×10^{-5}	0.0000
4	1.86×10^{-6}	0.7454	-2.05×10^{-6}	0.0003	1.87×10^{-5}	0.9816	-1.54×10^{-5}	0.0000
5	4.45×10^{-6}	0.4795	-1.73×10^{-6}	0.0014	1.24×10^{-5}	0.9852	-1.36×10^{-5}	0.0000
6	6.72×10^{-6}	0.1397	-1.13×10^{-6}	0.0012	2.33×10^{-5}	0.9790	-1.6×10^{-5}	0.0000
7	7.29×10^{-6}	0.1579	-3.11×10^{-6}	0.0005	9.57×10^{-6}	0.9864	-1.52×10^{-5}	0.0000
8	4.83×10^{-6}	0.2221	-7.1×10^{-6}	0.0015	2.18×10^{-5}	0.9790	-1.58×10^{-5}	0.0000
9	7.44×10^{-6}	0.1817	-2.45×10^{-6}	0.0009	2.54×10^{-5}	0.9754	-1.55×10^{-5}	0.0000
10	5.27×10^{-6}	0.2457	-1.16×10^{-6}	0.0085	2.23×10^{-5}	0.9787	-1.47×10^{-5}	0.0000
11	7.64×10^{-6}	0.1161	-1.88×10^{-6}	0.1688	2.48×10^{-5}	0.9776	-1.59×10^{-5}	0.0000
12	1.06×10^{-5}	0.0036	-1.39×10^{-6}	0.086	2.46×10^{-5}	0.9768	-1.6×10^{-5}	0.0000

Note: ***significant at 1% level;**significant at 5% level;*significant at 10% level.

Source: Economic Indicators, U.S. Census Bureau

Figure 4.6
ACF and PACF for First Difference of LS: Jan 1963-Dec 2000

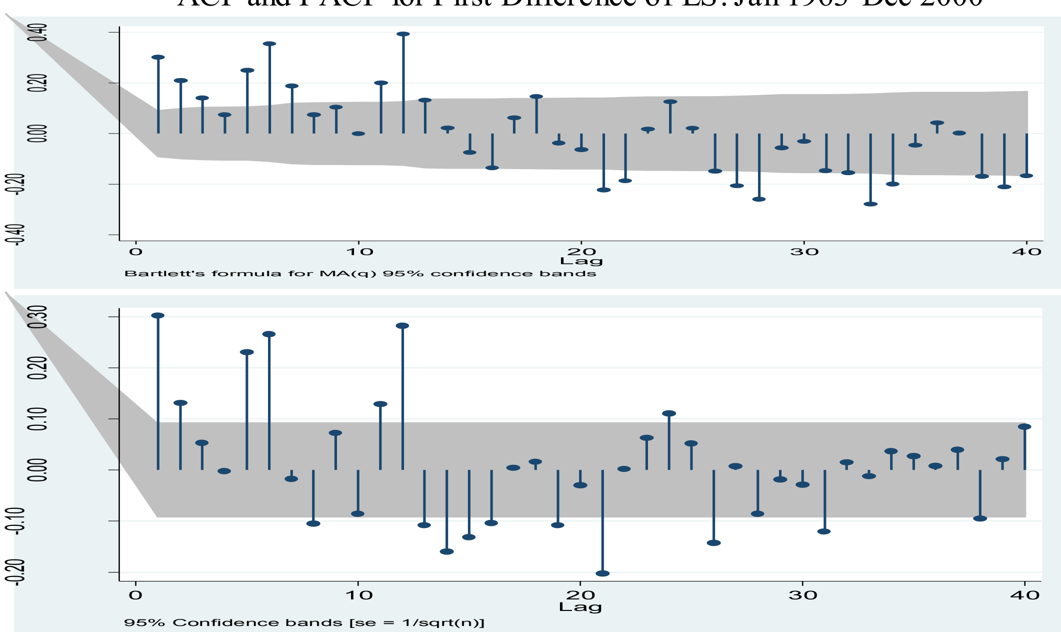
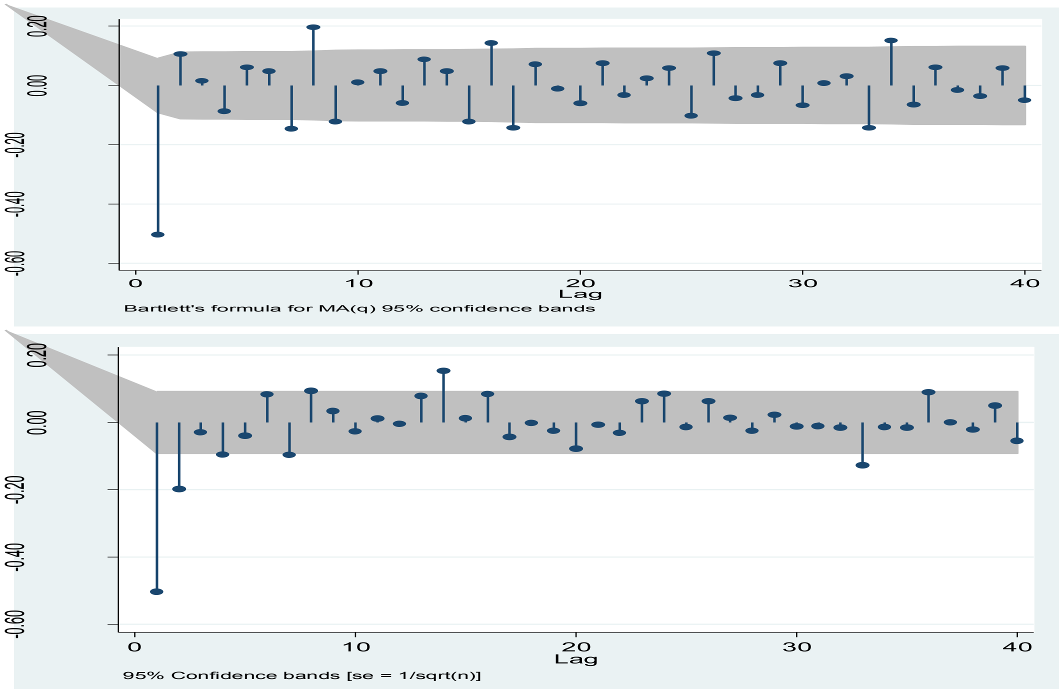


Figure 4.7
ACF and PACF for First Difference of LP: Jan 1963-Dec 2000



Although ACF and PACF could provide information about the number of p and q, preliminary estimates of p and q from ACF and PACF may overestimate the true values of these parameters. Therefore, to decide which ARIMA model is the most appropriate one for the monthly data of new home sales, different ARIMA models are tried in fitting the data. For first differences of LS, ARIMA (5, 0, 11) and ARIMA (5, 0, 10) are tried (see Table 4.2). All coefficients are less than one. In the AR component, coefficients of L1 and L5 are significant, and in the MA component, coefficients of L1, L10 and L11 are significant. In other words, the two ARIMA models are stationary and invertible. Because problem of overestimation, other five ARIMA models are tried, ARIMA (4, 0, 10), ARIMA (4, 0, 9), ARIMA (5, 0, 9), ARIMA (3, 0, 7) and ARIMA (3, 0, 6). Some coefficients of ARIMA (4, 0, 9) in AR component are greater than one, which means the process is not stationary. ARIMA (4, 0, 10) and ARIMA (4, 0, 11) have the same problem, so they are eliminated from the model selection process. However, ARIMA (3, 0, 7) and

ARIMA (3, 0, 6) are both feasible, because they are stationary and invertible. ARIMA (5, 0, 12) is also tried. According to the minimum AIC and BIC criteria, an optimal ARIMA model can be chosen from those feasible models. The results indicate that ARIMA (5, 0, 12) is the optimal one. But the optimal ARIMA model has to satisfy the condition that the residuals must be white noise. Therefore, Ljung-Box test is run. The Ljung-Box test is based on the autocorrelation plot. However, instead of testing randomness at each distinct lag, it tests the "overall" randomness based on a number of lags. If the Q-statistic value is smaller than critical value of chi-square, the data implies a white noise and vice versa.

In our model, the Q-statistic value is smaller than its critical value, which means that there is white noise in the residuals of the ARIMA (5, 0, 12). Hence, ARIMA (5, 0, 12) is absolutely feasible for forecast of changes in new home sales.

Table 4.2
ARIMA Models for First Differences of LS

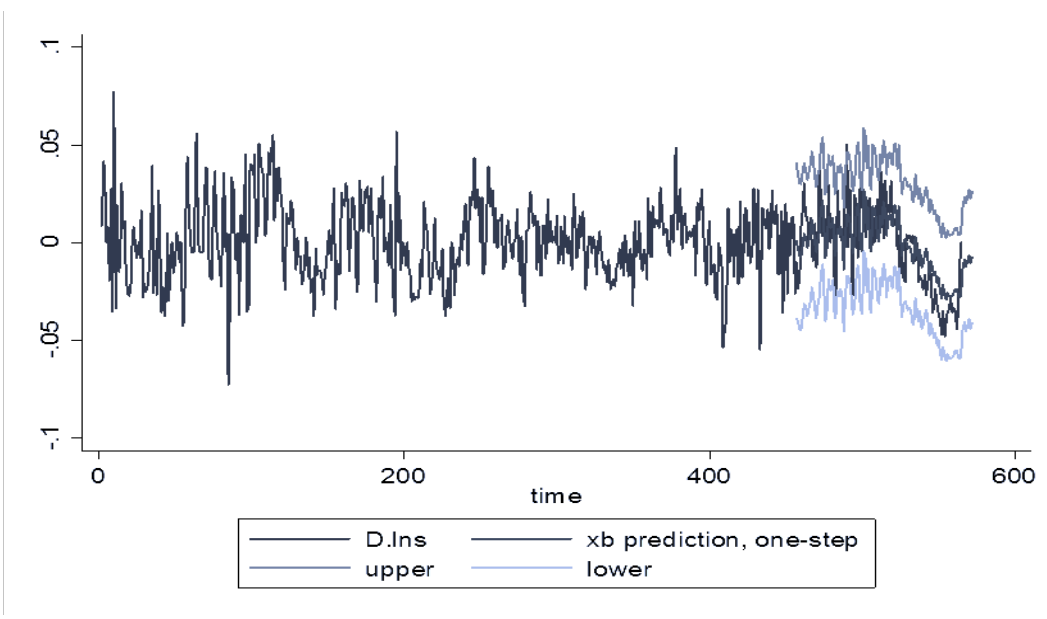
	ARIMA(5,0,11)	ARIMA(5,0,10)	ARIMA(5,0,12)	ARIMA(3,0,7)	ARIMA(3,0,6)
AR					
L1	***0.923 (0.085)	***0.805 (0.082)	***0.546 (0.12)	***-0.223 (0.053)	*-0.109 (0.059)
L2	*** -0.673 (0.13)	***-0.728 (0.054)	-	-	-
L3	*0.261 (0.143)	***0.601 (0.078)	-	***0.857 (0.046)	***0.825 (0.052)
L4	**-0.338 (0.12)	***-0.799 (0.052)	**-0.294 (0.098)		
L5	***0.548 (0.083)	*** 0.863 (0.074)	***0.582 (0.081)		
MA					
L1	*** -0.79 (0.084)	***-0.608 (0.093)	**-0.384 (0.123)	***0.469 (0.063)	*** 0.309 (0.072)
L2	***0.798 (0.122)	*** 0.859 (0.07)	**0.365 (0.124)	**0.183 (0.071)	** 0.204 (0.075)
L3	*-0.269	***-0.515	-	*** -0.852	*** -0.815

	(0.14)	(0.118)		(0.073)	(0.077)
L4	**0.436 (0.129)	***0.867 (0.093)	***0.36 (0.102)	***-0.277 (0.065)	*-0.105 (0.062)
L5	***-0.447 (0.107)	***-0.663 (0.125)	***-0.385 (0.091)	-	-
L6	-	-	-	***0.3 (0.051)	***0.261 (0.046)
L7	-	-	-	***0.228 (0.054)	
L8	-	-	-		
L9	**0.259 (0.083)	-	***0.273 (0.064)		
L10	***-0.34 (0.062)	**-0.137 (0.062)	* -0.143 (0.084)		
L11	***0.349 (0.052)		* 0.136 (0.074)		
L12			**0.249 (0.078)		
AIC	-2418.182	-2401.947	-2420.146	-2353.647	-2343.668
BIC	-2344.056	-2331.939	-2432.172	-2304.23	-2298.369

Note: ***significant at 1% level; **significant at 5% level; *significant at 10% level."-" not significant.

Figure 4.8

Estimation of Confidence Bounds of ARIMA (5, 0, 12) for d.LS: Jan 2001-Jan 2010



Note: D.lns=d.LS, which represents first differences of logarithmic values of new home sales.

To make sure of the validity of the optimal ARIMA model, confidence bounds are used. Figure 4.8 shows that the actual line (d.lns) does not exceed its confidence bounds, which means that the forecasting based on the ARIMA (5, 0, 12) is valid. Meanwhile, the one-step prediction line is very close to the line of actual data.

For first difference of LP (d.LP), ARIMA (1, 0, 1), ARIMA (1, 0, 2), ARIMA (2, 0, 1) and ARIMA (2, 0, 2) are tested. Among those models, only ARIMA (1, 0, 1) is feasible, other models are either non-stationary or not invertible. In addition, other ARIMA models are tried, such as ARIMA (1, 0, 3), ARIMA (1, 0, 4), ARIMA (1, 0, 7), ARIMA (2, 0, 3), ARIMA (2, 0, 5), ARIMA (4, 0, 4) and so on, but none is feasible. After that, large values of p and q are tried, and results show that ARIMA (12, 0, 4) and ARIMA (9, 0, 5) are feasible (see Table 4.3). The minimum AIC implies ARIMA (12, 0, 4), however, the minimum BIC implies ARIMA (1, 0, 1). Here, the optimal model cannot be decided by AIC and BIC. Residuals for ARIMA (1, 0, 1) and ARIMA (12, 0, 4) are checked by Ljung-Box test. According to the results of Table 4.4, residuals for ARIMA (1, 0, 1) are not white noise, but residuals for ARIMA (12, 0, 4) are white noise. In other words, ARIMA (12, 0, 4) is the optimal model for forecasting changes in new home prices in the U.S. housing market.

Table 4.3
ARIMA Models for First Differences of LP

	ARIMA(1,0,1)	ARIMA(1,0,6)	ARIMA(9,0,5)	ARIMA(12,0,4)
AR				
L1	**-0.217 (0.077)	***-0.963 (0.05)	**-0.345 (0.126)	***-0.422 (0.061)
L2			-	*** 0.547

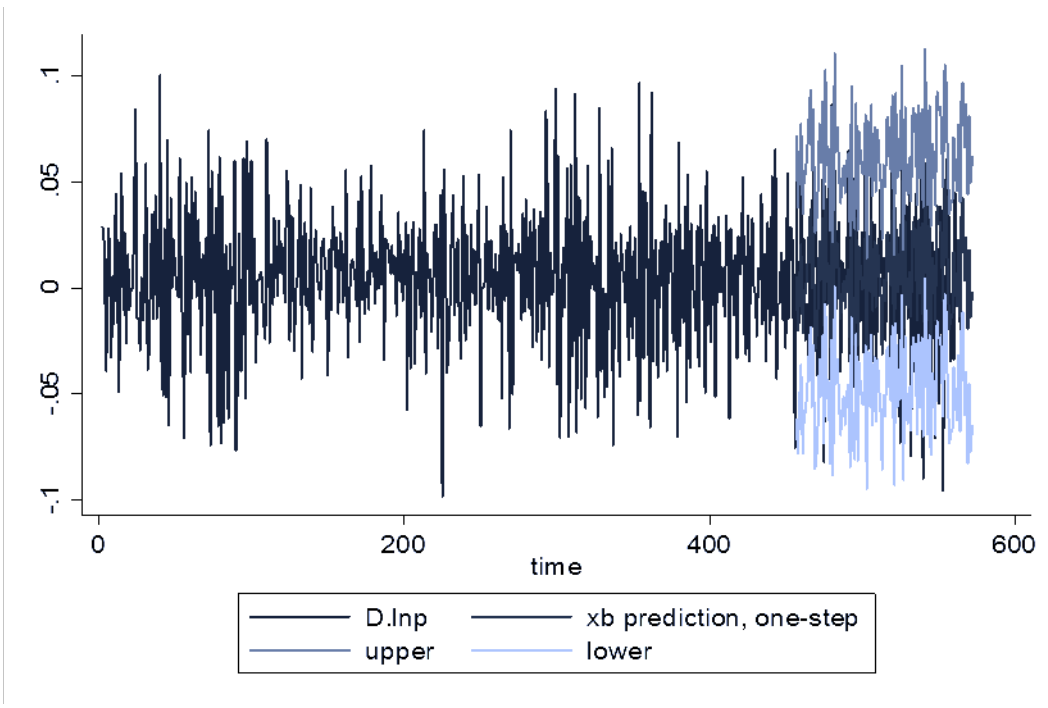
				(0.065)
L3			-	***0.508 (0.07)
L4			**-0.369 (0.134)	***-0.687 (0.071)
L5			***0.661 (0.158)	***-0.398 (0.071)
L6			***0.479 (0.087)	-
L7			**0.152 (0.053)	-
L8			**0.163 (0.049)	-
L9			* 0.096 (0.056)	-
L10				-
L11				-
L12				** 0.123 (0.049)
MA				
L1	***-0.398 (0.074)	***0.362 (0.07)	-0.0276	***-0.177 (0.037)
L2		***-0.461 (0.057)	-	***-0.693 (0.038)
L3		-	-	***-0.179 (0.037)
L4		-	***0.285 (0.08)	***0.922 (0.039)
L5		-	***-0.847 (0.109)	
AIC	-1919.393	-1915.873	-1927.018	-1933.151
BIC	-1902.921	-1878.81	-1861.128	-1859.025

Note: ***significant at 1% level; **significant at 5% level; *significant at 10% level.

To identify the validity of ARIMA (12, 0, 4), confidence bounds is used again. Figure 4.9 shows that among 109 data points, eight points exceed the lower bound and no point exceeds the upper bound. The frequency of the actual first differences of LP lying within its confidence bounds is so high. To provide stronger evidence, the data is set back to Apr 1996, and then, the confidence bounds are checked. According to Figure 4.10, only eight points exceed its lower confidence bound among the 165 data points. The frequency of the actual first differences of LP lying within its confidence bounds is much higher. In other words, the more historical data are included, the more accurate forecast is. Therefore, the ARIMA (12, 0, 4) with enough historical data can be used to forecast changes in new home prices in the U.S. housing market.

Figure 4.9

Estimation of Confidence Bounds of ARIMA (12, 0, 4) for d. LP: Jan 2001-Jan 2010



To sum up, ARIMA (5, 0, 12) can forecast changes in new home sales and ARIMA (12, 0, 4) can forecast changes in new home prices, but, until now we cannot say they are the best models which can forecast changes in new home prices and sales. To judge which model is the most appropriate one to make a prediction, accuracy is the best standard. To compare accuracy of different models, other methods will be discussed in next section.

Figure 4.10
Estimation of Confidence Bounds of ARIMA (12, 0, 4) for d.LP: May 1996-Jan 2010

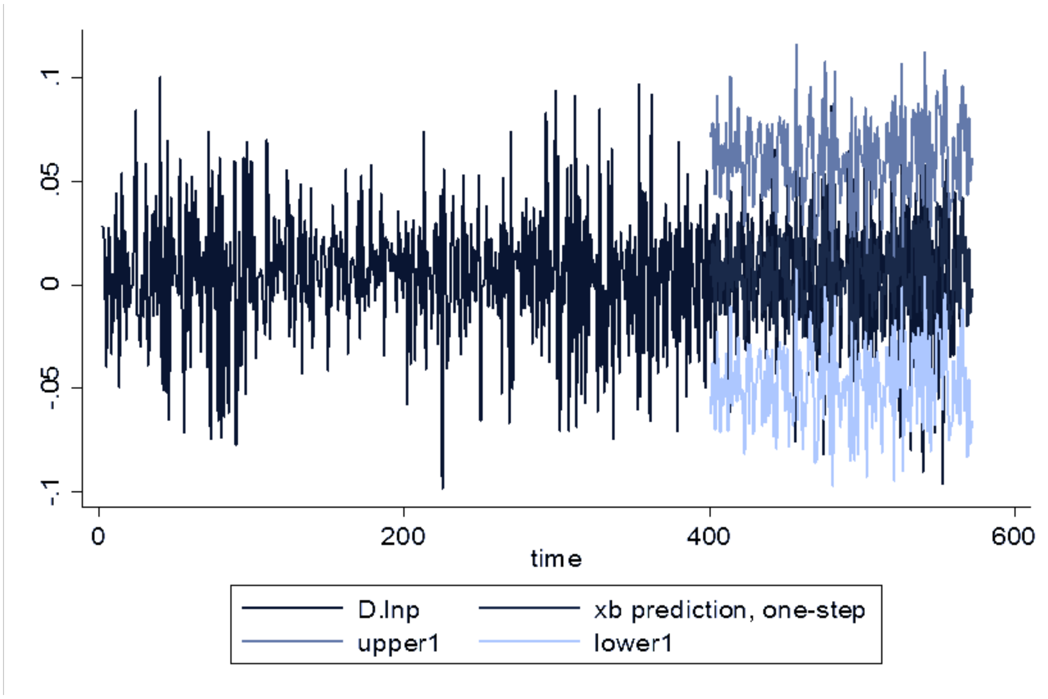


Table 4.4
Portmanteau Test for White Noise of Residuals

	res of ARIMA (5,0,12)	Res of ARIMA (1,0,1)	res of ARIMA (12,0,4)
Q-statistics	47.3298	61.0234	21.2435
chi(40) (5% level)	55.758	55.758	55.758

4.2 Simple Exponential Smoothing

Exponential smoothing is the most common model for time-series data. It requires the specification of a smoothing factor which is usually chosen from the time-series data to minimize the average squared residual of previous one-step-ahead forecasts. Meanwhile, exponential smoothing is a procedure for continually revising a forecast in light of relatively recent experience. In exponential smoothing, recent observations are given relatively more weight in forecasting than the older observations. Generally, there are three categories: SES, double exponential smoothing and triple exponential smoothing.

To decide which kind of exponential smoothing model is suitable, seasonality of the time-series data has to be checked initially. Here, the model adopted by Franses (1990, 1991) and Song, Martin and Bob (2009) is:

$$\varphi(L)y_{8,t} = \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{3,t-1} + \pi_5 y_{4,t-2} + \pi_6 y_{4,t-1} + \pi_7 y_{5,t-2} + \pi_8 y_{5,t-1} + \pi_9 y_{6,t-2} + \pi_{10} y_{6,t-1} + \pi_{11} y_{7,t-2} + \pi_{12} y_{7,t-1} + \mu_t + \varepsilon_t$$

Where φ (L) is a polynomial function of L for which the usual assumptions apply, including the assumption that the errors ε_t are independent and identical distributed in standard normal form, and the assumption that the deterministics are not absent. y_t is the new home price or the new home sale, π_is (i=1……12) are coefficients for seasonal roots, μ_t is the deterministic component which might include a constant, seasonal dummy variables or a trend, and

$$y_{1,t}=(1+L)(1+L^2)(1+L^4+L^8)y_t$$

$$y_{2,t}=-(1-L)(1+L^2)(1+L^4+L^8)y_t$$

$$y_{3,t}=-(1-L^2)(1+L^4+L^8)y_t$$

$$y_{4,t}=-(1-L^4)(1-\sqrt{3}L+L^2)(1+L^2+L^4)y_t$$

$$y_{5,t}=-(1-L^4)(1+\sqrt{3}L+L^2)(1+L^2+L^4)y_t$$

$$y_{6,t}=-(1-L^4)(1-L^2+L^4)(1-L+L^2)y_t$$

$$y_{7,t}=-(1-L^4)(1-L^2+L^4)(1+L+L^2)y_t$$

$$y_{8,t}=(1-L^{12})y_t$$

If all π_i (i=1……12) are equal to 0, there is seasonality in the monthly data. If one of π_i is not equal to 0, there is no seasonality. Results are shown in Table 4.5. According to the Table 4.5, almost all values of π_i are significantly different from 0, which means that there is no seasonality in the monthly data from Jan 1963 to Dec 2000. This result is consistent with reality.

Because the monthly data about new home sales and prices does not have seasonality, triple exponential smoothing may be not appropriate. Meanwhile, according to the Table 4.1, all coefficients of trend are not significantly different from 0, which means that there is no trend in the time-series data. Hence, the double exponential smoothing should not be applicable, either.

The SES is a method for smoothing a time series or for forecasting where the mean is either stationary or changes only slowly with time, and it is used for short-range forecasting based on the condition that there is no trend. This method assumes that the

Table 4.5

Seasonality Unit Roots in Monthly Changes in New Home Prices and Sales

	d.LP	d.LS
π_1	***-0.166	**-0.013
π_2	***-0.24	***-0.118
π_3	***-0.152	***-0.166
π_4	***0.112	**-0.064
π_5	***-0.085	**-0.172
π_6	***-0.055	***-0.157
π_7	***-0.272	***-0.105
π_8	***0.319	***0.059
π_9	***-0.084	***-0.098
π_{10}	***0.065	-0.016
π_{11}	***-0.184	***-0.065
π_{12}	***0.202	0.007

Note: ***significant at 1% level;**significant at 5% level.

data fluctuates around a reasonably stable mean (no trend or consistent pattern of growth).Generally, the specific formula for SES is:

$$Y_t = \alpha X_{t-1} + (1-\alpha)Y_{t-1} \qquad (1)$$

where Y_t stands for smoothed observation at time t, X_t stands for the original observation at time t, and α is the smoothing constant which is the weighted average of the previous observations. When the value of smoothing constant is close to zero, it will lead to a stable model, while, if the value of smoothing constant is close to one, it will lead to a highly reactive model.

Except for the smoothing constant, the initial value plays an important role in computing all the subsequent values. Usually, two methods are used to set the initial value: one is that set $Y_1=X_0$; the other method is averaging the first four or five observations. Here, the second method is used.

The value of α has to be determined. In this model, the optimal value for α is the value which results in the smallest mean squared errors (MSE).

STATA is used to calculate the optimal α, which minimizes the MSE for first difference of LS and LP for the period of Jan 1963 to Dec 2000. The computational result shows that the optimal α is 0.1692 for d.LS and it is 0.0001 for d.LP. Bowerman and O'Connell (1979) suggest that values of α around 0.10 to 0.30 work quite well in single exponential smoothing. If the "best" smoothing constant is greater than 0.3, then it is possible that the values in the time series are dependent upon each other. This dependency may be captured by time series methods which analyze the autocorrelations of data. In our analysis for the value of α is so close to 0, which indicates that dampening is slow. In other words, the smoothing effect is greater and less responsive to recent changes.

To make sure that the SES model is good at forecasting the growth rate of new home sales and prices in the United States, an out-of-sample forecasting for Jan 2001 to Jan 2010 is made. The initial value is 0.00067 which is the average value of the actual growth rate of new home sales from Aug 2000 to Dec 2000. The result is shown in Figure 4.11.

According to Figure 4.11, the SES model for the first difference of LS is good at out-of-sample forecasting, since the predicted line captures the majority of the actual observations and the predicted line indicates the same direction of changes with the actual line. The predicted or fitted values do not show a large variation as the actual values embedded with noise, because a lot of external factors may affect the growth rate of new home sales, such as construction costs, household income and so on. To provide stronger evidence, the data is set back to Dec 1994 and then a new out-of-sample forecasting is made. The first five observations are from Aug 1994 to Dec 1994. For first difference of LS, the initial value is 0.01338. With the initial value and the optimal α, the SES model is fitted. Figure 4.12 shows the result. The direction of growth rate is forecasted accurately, but the fluctuations are not big enough. This is still can be explained by influences of external factors.

The same method is used for forecasting changes in new home prices (see Figure 4.13). From Figure 4.13, the fitted values are almost equal, because the optimal α is pretty close to 0. According to the equation (1), if the optimal α is very close to 0, $Y_t \approx Y_{t-1}$. The predicted line is too flat, so the SES model neither captures direction of changes nor strength of changes. The SES model is not good at forecasting changes in new home prices in the U.S. housing market, at least with the monthly data in our analysis. The SES model can only be used to forecast changes in growth rate of new home sales, but it is not very effective in forecast of the changes in new home sales in the future.

According to the literature review in Section 3.1, the SES model with damped drift may be better than the standard SES model at out-of-sample forecasting. However, the damped trend method is usually appropriate when there is a trend in the time series. In the monthly data about growth rate of new home sales and prices, there is no trend. Therefore, the damped trend method cannot be adopted in our analysis.

To choose between the ARIMA model and the SES model, results of out-of-sample forecasting are compared. Based on Figure 4.8 and Figure 4.11, the standard SES model with α=0.1692 is a little more precise than the ARIMA (5, 0, 12) model in forecast of changes in new home sales. Comparing Figure 4.9 with Figure 4.13, ARIMA (12, 0, 4) is much more precise than the SES model with α=0.0001 in forecast of changes in new home prices. However, the comparisons are only based on visual effects, so in the end of this Chapter, MSE will be used to compare different models.

4.3 Holt-Winters Nonseasonal Smoothing

Everette and Joaquin (2008) find that the Holt and damped trend methods can improve accuracy of exponential smoothing model by trimming the time series to eliminate irrelevant early data, fitting the exponential smoothing method to minimize mean absolute deviation (MAD) rather than MSE, and optimizing the parameters. Moreover, in their study, the results suggest that the damped trend is more accurate than Holt's method. Because damped trend is not appropriate for our dataset, Holt's method will be applied.

Holt-Winters nonseasonal smoothing is for forecasting a series without seasonality. The series can be modeled as a linear trend in which the intercept and the coefficient vary over time. For a time series y_1, y_2, … , y_n, the forecast function, which gives an estimate of the series t steps ahead is written as:

$$Y_{t+1}=a_t+b_t$$

where Y_t is observed data, and a_t is the level component and drifts over time. b_t is the component of slope (or trend) on time that also drifts. Furthermore,

$$a_t=\alpha Y_t+(1-\alpha)Y_t' \quad 0<\alpha<1$$

where, Y_t' is the fitted value of Y_t, which provides the level at time t. Since the level at time t-1 is already known, it is possible to update the estimate of the slope as:

$$b_t=\beta(a_t-a_{t-1})+(1-\beta)b_{t-1} \qquad 0<\beta<1 \qquad (2)$$

In this model, two important things have to be done for estimation. The starting values of a_t and b_t for the level and slope must be provided at the beginning of the series in order to initiate updating procedure. Many different ways of doing this are proposed by some economists, but one simple method (Granger & Newbold, 1986, Chapter 5) is to set the starting value equal to the average observation in the first year. Secondly, the smoothing parameters α and β must be estimated. There are two general ways of selecting the optimal α and β. The first is to estimate them by minimizing some function of the forecast errors of historical data, while the second is to simply "guestimate" them. For the first approach, the usual procedure is to minimize the sum of squared one- step-ahead forecast errors over a suitable fitting period for which historical data are available. Given specific values of α and β, the updating procedure yields a sequence of one-step-ahead error, et, so that Σe_t^2 can be calculated.

However, STATA can provide optimal values of α and β, so the optimal values are used directly. In this section, the data spans from Jan 1963 to Dec 2000. For d.LS, the minimum sum-of-squared is obtained when α=0.1686 and β=0.0000. For d.LP, there is no optimal values of α and β, because the STATA program goes to an endless loop. Because β=0 cannot satisfy the condition in the equation (2) which requires $0<\beta<1$, the Holt-Winters nonseasonal smoothing model becomes a standard SES model with a constant term.

Figure 4.14 shows the results of in-sample and out-of-sample forecasting of the SES model with a constant term for d.LS. We can see that the standard SES model with a constant term performs very well on in-sample forecasting, but not on the out-of-sample forecasting for d.LS. To provide stronger evidence, the data is set back to Dec 1994, and a new out-of-sample forecasting is shown in Figure 4.15. The standard SES model with a constant term still displays a bad result of out-of-sample forecasting for d.LS.

Comparing the results of out-of-sample forecasting presented in Figure 4.11 and Figure 4.14, the difference between the predicted values and the actual values in the SES model without a constant term is smaller than the difference in the standard SES model with a constant term. Moreover, the predicted line of the SES model without constant term captures the direction of changes. The SES model with a constant term only shows a straight forecast line. Comparing Figure 4.12 and Figure 4.15, the same conclusion can be drawn. In other words, the SES model without a constant term is better than the SES model with a constant term in the forecast of changes in new home sales.

Although, the Holt-Winters nonseasonal smoothing method is appropriate for in-sample forecasting, the optimal value of α is very close to 0 (0.0001) and β is pretty close to 1 (0.9994), so the out-of-sample forecasting line is flat. Therefore, the Holt-Winters nonseasonal smoothing method is not used to forecast changes in new home prices in our analysis.

Based on the monthly data, a conclusion is drawn among those time-series models that in out-of-sample forecasting for changes in new home sales, the standard SES model without a constant term is better. But the ARIMA (12, 0, 4) performs better than the SES model without a constant term in forecasting changes in new home prices in the U.S. housing market. The Holt-Winters nonseasonal smoothing model is not appropriate to forecast changes in new home sales and prices, because it is only good at in-sample forecasting. This result supports the conclusion that dynamic structure models only perform well on in-sample forecasting which drawn by Zellner and Palm (1974).

Accuracy of forecasting changes in new home sales and prices of the ARIMA model and the standard SES model without a constant term is not good enough, so linear prediction model with external factors will be tested in next section.

4.4 Linear Prediction Model

In the previous sections, we discuss different time-series models for the forecast of new home sales and prices based on their historical data. However, the results of out-of-sample forecasting of all those time-series models are not as accurate as expected. Usually, the noise of out-of-sample forecasting is mainly from external factors, so, to some extent, taking external factors into account may reduce noise. According to the Section 3.1, some researchers have shown that GDP, stock price, CPI, income, interest rates may have influences on housing prices. Therefore, a few macro factors will be included into the linear prediction models.

The goal of this section is to identify a linear prediction model with external macro factors which is expected to perform better on out-of-sample forecasting than time-series models. All data in this section is monthly data from Apr 1992 to Jan 2010 and all data is transformed to logs. There are 214 observations. The data from Apr 1992 to Dec 2007 is used for model calibration. Data stationarity is tested at first, and we find all data is non-stationary, but the first differences are stationary. The depending variable is the change in new home prices, and the additional regressors include the change in GDP, in stock price, in personal disposable income, in CPI and in prime loan rates.

We need to first specify a linear model. A lot of studies have established the relationship between housing price changes and GDP changes, stock price (SP) changes, income (IN) changes and interest rates (R) changes, so different simple linear models based on current values are tried. Good results of some models are shown in Table 4.6. R1 is a simple linear regression model which takes changes in GDP, in SP, in CPI, in IN and in R into account. The coefficients of changes in CPI and changes in GDP are very insignificant, but the coefficient of changes in SP is just a little insignificant. Therefore, variables of changes in CPI and changes in GDP are eliminated in R2, which gives a new linear regression model. In R2, the coefficient of changes in SP is still insignificant, but the adjusted R-square increases and root MSE declines. R3 is run without variables of changes in GDP, in CPI and in SP. According to the maximum adjusted R-square and the minimum root MSE, R2 is the optimal linear regression model. The results in R2 indicate that changes in personal disposable income have positive effects on changes in new home prices, and changes in prime loan rates have negative effects on new home prices changes. However, growth rates of GDP, CPI and SP do not have any influences on the growth rate of new home prices.

The positive correlation between income and home prices suggests that households enter financial markets with a greater exposure to risk. Higher prime loan rates bring heavier economic burden to household, so that the demand for new homes may decrease and new home prices may decline. Usually, changes in GDP, in CPI and in SP should have effects on changes in housing prices. But, why do these effects disappear here? Three reasons may answer this question. First, the sample size is too small to capture the weak influences from changes in GDP, in CPI and in SP. Second, effects of changes in GDP, in CPI, and in stock prices on new home prices are not immediate. Current changes in GDP, in CPI and in SP may affect new home prices in a few months. Finally, almost all empirical research on these relationships are relevant to the HPI measure which averages price changes in repeat sales or refinancing on the same properties (existing homes). Besides sales prices, a big difference between existing homes and new homes is the correlation with housing construction. The new home market has a direct and strong relationship with housing construction, but existing home market has an indirect and weak relationship with housing construction. Housing construction is directly related to start permits, construction material costs, labor costs and so on. To a great extent, those costs decide new home final sales prices. However, existing home sales prices are greatly decided by current economic environment and characteristics of existing homes. Therefore, GDP, CPI, SP may have significant effects on HPI, but not on new home sales prices.

Over all, the simple linear regression models with current changes are good because of higher adjusted R-square. To test the effects of changes in those macro factors in the previous month on current changes in new home prices, the ADL models are tested. In the ADL models, the regressors include lagged values of the dependent variable and current and lagged values of one or more explanatory variables. Generally, in the ADL (p, q) model, p is the number of lags of dependent variable and q is the number of lags of additional regressors.

Table 4.6
Different Simple Linear Regressions for First Differences of LP

	d.LP		
	R1	R2	R3
d.lnGDP	-		
d.lnSP	-	-	
d.lnCPI	-		
d.lnIN	***1.33 (0.071)	***1.33 (0.07)	***1.323 (0.07)
d.lnR	** -0.044 (0.019)	**-0.043 (0.019)	**-0.043 (0.019)
constant	***-0.006 (0.0008)	***-0.006 (0.0006)	***-0.006 (0.0006)
adjusted R-square	0.653	0.6566	0.6562
root MSE	0.00675	0.00671	0.00672

Note: ***significant at 1% level;**significant at 5% level;*significant at 10% level. Ri is regression i (i=1, 2, 3). "-" indicate insignificant in this table. SP is stock price and R is prime loan rate.

$R1: d.LP=\alpha_1 d.lnGDP+\alpha_2 d.lnSP+\alpha_3 d.lnCPI+\alpha_4 d.lnIN+\alpha_5 d.lnR+Constant$

$R2: d.LP=\beta_1 d.lnGDP+\beta_2 d.lnCPI+\beta_3 d.lnIN+\beta_4 d.lnR+Constant$

$R3: d.LP=\gamma_1 d.lnIN+\gamma_2 d.lnR+Constant$

First of all, VARSOC (a command in STATA) is used to determine the lag order of the ADL model. VARSOC reports the final prediction error (FPE), Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (SBIC), and the Hannan and Quinn information criterion (HQIC) lag-order selection statistics for a series of vector autoregressions. All those criteria indicate that the optimal lag length p should be 4, and the optimal lag length q should be 1, 2 or 3. Therefore different combinations of the ADL models are tested. The results of the ADL models are displayed in Table 4.7. According to Table 4.7, the last two lags of d.LP should be eliminated, so ADL (2, 1), ADL (2, 3) and ADL (2, 2) are tested. Based on the minimum AIC, ADL (2, 1) is the optimal one among those ADL models.

Table 4.7
Different ADL models for First Differences of LP

	ADL (4,1)	ADL (4,3)	ADL (4,2)	ADL (2,3)	ADL (2, 2)	ADL (2,1)
d.LP						
L	**0.72 (0.223)	***-1.474 (0.078)	***-1.469 (0.077)	***0.533 (0.127)	***0.522 (0.125)	***0.495 (0.122)
L2	**0.435 (0.176)	***-1.788 (0.138)	**0.459 (0.18)	**0.251 (0.075)	**0.253 (0.073)	**0.244 (0.072)
L3	-	-	-			
L4	-	-	-			
d.lnGDP						
L1	-	-	*0.185 (0.107)	*0.184 (0.109)	*0.18 (0.105)	-
L2		-	-	-	-	
L3		-		-		

d.lnSP						
L1	-	-	-	-	-	-
L2		-	-	-	-	
L3		-		-		
d.lnCPI						
L1	-	-	-	-	-	-
L2		-	-	-	-	
L3		-		-		
d.lnR						
L1	*0.038 (0.02)	-	-	-	-	*0.034 (0.019)
L2		-	-	-	-	
L3		-		-		
d.lnIN						
L1	***-2.172 (0.264)	-	-	***-1.975 (0.167)	***-1.953 (0.164)	***-1.91 (0.161)
L2		-	***-2.234 (0.271)	-	-	
L3		***-2.211 (0.278)		-		
adjusted R-square	0.6807	0.6783	0.6778	0.6744	0.6732	0.6758
root MSE	0.00654	0.00657	0.00657	0.00659	0.00658	0.00656
AIC	-1319.07	-1310.32	-1313.7	-1318.11	-1329.29	-1334.55

Note:***significant at 1% level;**significant at 5% level;*significant at 10%leveL;”-“insignificant. p=new home price, SP=stock prices, R=prime loan rate, IN=personal disposable income.

For the five external economic factors, changes in personal disposable income in the previous month have the strongest and negative explanatory power for the changes in current new home prices. Moreover, if new home prices increases in last two months, it will keep increasing in this month. But, in the optimal ADL model, its adjusted R-square is just little higher than the simple linear regression model.

In the various ADL and simple linear regression models, there is an important commonality that changes in CPI and in SP do not have any influences on growth rate of new home prices, currently or previously. Moreover, in the optimal ADL and simple linear models, the results indicate that changes in GDP cannot also affect the new home prices.

Taken together, changes in current and previous disposable income, in previous new home prices and in prime loan rates may have effects on current changes in new home prices. Based on this conclusion, new linear regression models are tried. The new optimal model is ADL (3, 1) (see equation (3))

$$d.LP_t = -\ 0.009 + 0.731 d.LP_{t-1} + 0.452 d.LP_{t-2} + 0.148 d.LP_{t-3} - 2.168 d.\ln IN_{t-1} + 0.037 d.\ln R_{t-1} + \xi_t \quad (3)$$

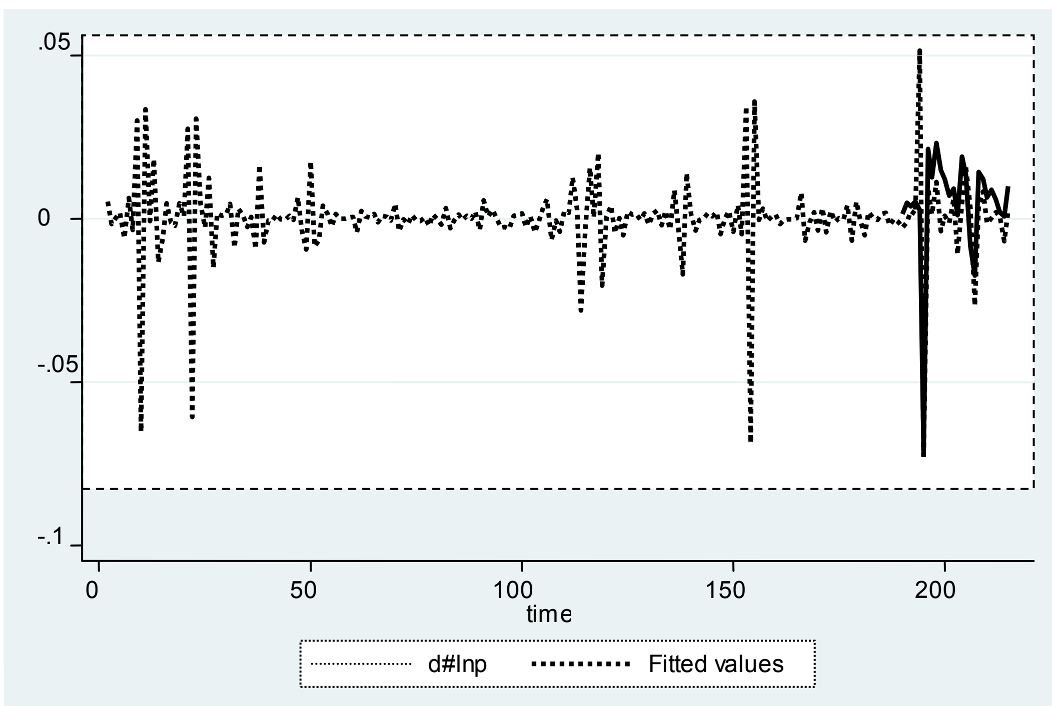
In equation (3), adjusted R-square is 0.6801, the AIC is -1330.754, and the root MSE is 0.00653. Comparing all different models, although the AIC in equation (3) is not the minimum one, ADL (3, 1) yields the maximum adjusted R-square and the minimum root MSE. Hence, among those models, equation (3) is the most appropriate one to forecast changes in new home prices. In addition, Figure 4.16 which displays an out-of-sample forecasting from Jan 2008 to Jan 2010 shows that the differences between the changes in actual new home prices and the predicted new home prices are small, which means that the ADL (3, 1) model not only captures the direction of changes in new home prices, it also shows the strength of changes in new home prices.

To make sure that equation (3) can be used to forecast changes in new home prices in the United States, the data is set back to Dec 2005, and a new out-of-sample forecasting is made (see Figure 4.17).

According to Figure 4.17, this linear model can predict the changes in new home prices very well, no matter on direction or on strength.

Figure 4.16

Linear graph of equation (3) for forecasting changes in new home price: Jan 2008 to Jan 2010



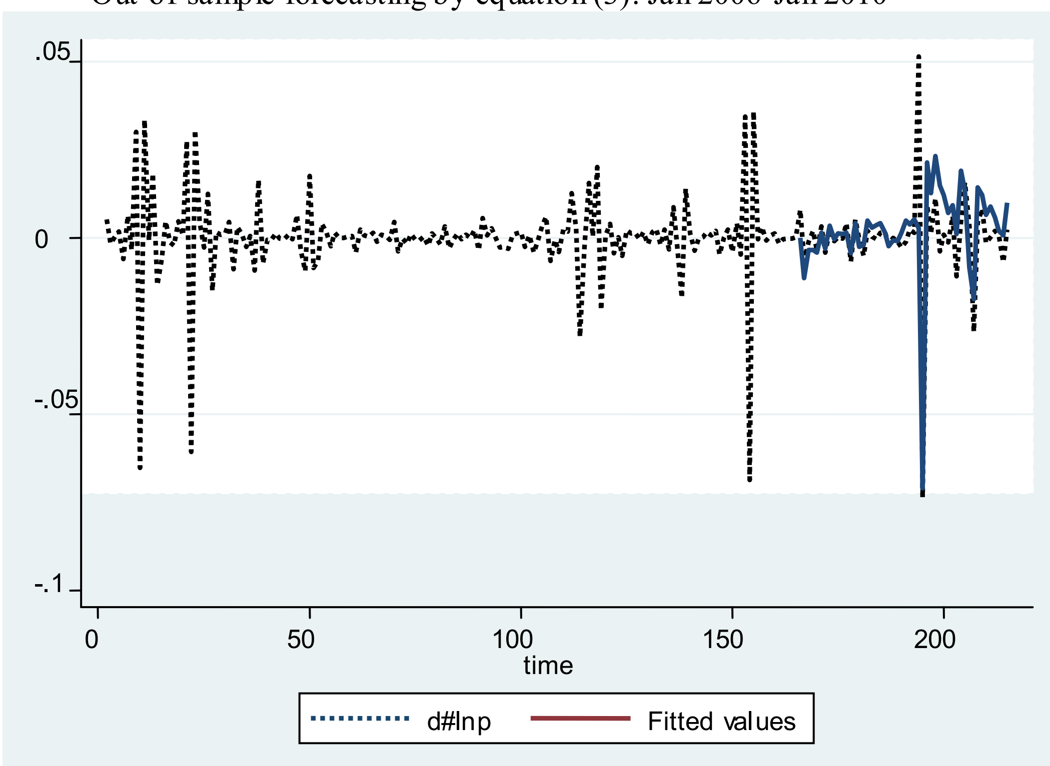
Note: d#lnp= first difference of LP.

To forecast changes in new home sales in the United States, the same method is used. But two more regressors are included, household estimates (HE) and construction material sales (CONS). Some relatively better results are shown in Table 4.8.

According to Table 4.8, the model R3 has the maximum adjusted R-square, the minimum AIC and the minimum root MSE, so it is the optimal model for changes in new home sales, based on the three criteria. However, adjusted R-square of R3 is not high, which means that the goodness of fit of R3 does not perform well. Therefore, some

Figure 4.17

Out-of-sample forecasting by equation (3): Jan 2006-Jan 2010



Note: d#lnp= first difference of LP.

Table 4.8
Different ADL Models for First Differences of LS

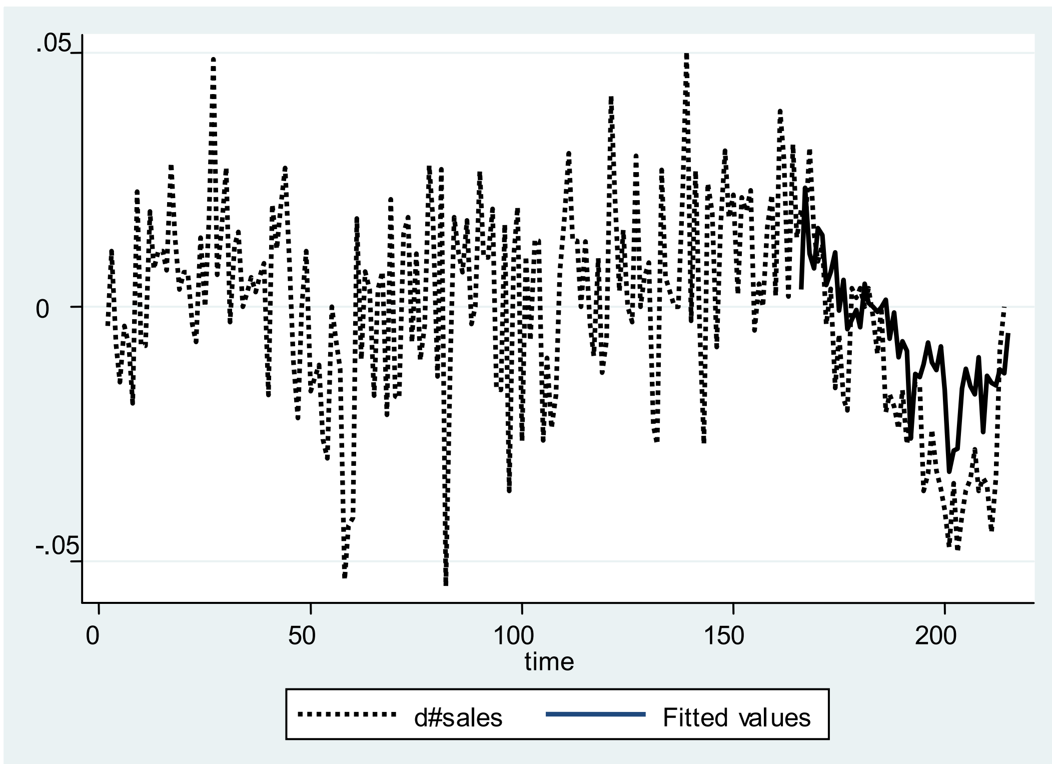
	R1	R2	R3
d.LS			
L	**0.188 (0.075)	**0.196 (0.072)	**0.194 (0.071)
L3	**0.226 (0.075)	**0.22 (0.071)	**0.221 (0.071)
L4	**-0.192 (0.073)	**-0.192 (0.072)	**-0.193 (0.072)
L6	**0.219 (0.073)	**0.228 (0.071)	**0.229 (0.07)
d.lnR			
L1	**0.14 (0.066)	**0.129 (0.062)	**0.1 (0.047)
L2	-	-	
d.lnHE			
L1	**0.786 (0.393)	**0.802 (0.386)	**0.788 (0.384)
L2	-	-	
Adjusted R-square	0.1941	0.2107	0.2172
AIC	-974.569	-982.098	-985.51
root MSE	0.01607	0.01591	0.01584

Note: ***significant at 1% level;**significant at 5% level;*significant at 10% level;"-"insignificant.
sales=new home sales, R=prime loan rates, HE=household estimated, SP=stock prices.

other factors are taken into account, such as changes in construction price index, personal disposable income, new home prices, unemployment rate and so on. But all of these factors cannot increase the adjusted R-square or decrease AIC. Therefore, more explainatory power is attributed to the charactersistics of new homes, existing home sales and existing home prices. But because monthly data on those factors cannot be obtained, practical evidence cannot be provided in our analysis. However, this guess is reasonable, because empircal reseach on hedonic pricing models suggest that characteristics of homes affect their prices and sales. In addition, existing homes are substitutes of new homes, so existing home sales can affect the demand for new homes. In a word, changes in new home sales may be mainly affected by characteristics of new homes and existing home sales and prices, rather than by these macro factors.

Figure 4.18

Out-of-sample forecasting for changes in new home sales by R3: Dec 2005-Jan 2010



Note: d#sales=first difference of LS.

An out-of-sample forecasting is made for comparision. The data is also set back to Dec 2005 (see Figure 4.18).Comparing Figure 4.18 and Figure 4.11, visually, the predicted values in the SES model without a constant are much closer to the actual values than those in the linear regression model R3. Therefore, to forecast changes in new home sales in the United States, the standard SES model without a constant may be better.

To get an accurate result, MSE is used to select the most appropriate model to forecast changes in new home sales and prices in the U.S. housing market. Table 4.9 displays MSE for different forecasting models which are used in our analysis, and linear regressions yield the minimum MSE, which means that linear regressions based on external factors and internal factors are the best models to forecast changes in new home sales and prices.

Table 4.9

MSE for Different Forecasting Models

	d.LS			
	ARIMA(5,0,12)	Linear Regression for d.LS	SES without a constant	SES with a constant
MSE	0.00026	0.000251*	0.000374	0.000374
	d.LP			
	ARIMA(12,0,4)	Linear Regression for d.LP	SES without a constant	SES with a constant
	0.000765	0.0000426*	0.001186	0.001332

Note: * indicates minimum MSE. Meanwhile, all those models are optimal ones in each section.

CHAPTER 5 RIPPLE EFFECTS IN THE U.S. HOUSING MARKET

From the previous section, we know that linear time-series models are good at forecasting growth rate of new home sales and prices for the whole United States. However, if changes in new home sales and prices of a specific state or region are predicted, the forecasting process will be complex because of ripple effects. The ripple effect indicates that the influences from other states or regions affect the direction and strength of changes in new home sales and prices. Therefore, if the ripple effect exists in the U.S. housing market, forecasting changes in new home sales and prices on national level is more advisable than forecasting those on a specific state or region. In this chapter, the existence of ripple effects is discussed.

5.1 Vector Auto Regression Model

A VAR model will be discussed to analyze the Granger causality among the four regions. To apply this method, it is also important to test stationarity of the quarterly data. ADF test is applied again, with the AIC for the selection of the appropriate numbers of lags. Results of ADF test suggest that all data be stationary at first differences, and results of AIC are shown in Table 5.1. Except Mideast, one lag is better for other regions with the VAR model.

Table 5.1

AIC for First Difference of New Home Prices by Regions

	AIC			
lags	NW	ME	S	W
1	*-473.5385	-571.573	*-735.306	*-697.8433
2	-470.756	*-573.3385	-734.789	-692.46
3	-466.458	-569.194	-735.26	-691.187
4	-461.855	-571.457	-729.741	-687.386

Note: "*" indicates the minimum AIC for each region. NW=Northwest, ME=Mideast, S=South, W=West.

In a VAR model, seasonality is also a problem. If the time series is seasonally integrated, it will require the application of a higher order of seasonal differencing, rather than the use of seasonal dummy variables in a VAR model for tackling the seasonality effect. Testing for seasonality in the quarterly data is given in Depalo (2008) as follows:

$$\varphi(L)y_{4,t} = \pi_1\, y_{1,t-1} + \pi_2\, y_{2,t-1} + \pi_3\, y_{3,t-1} + \pi_4\, y_{3,t-2} + \varepsilon_t \quad (4)$$

Where $\varphi(L)$ is some polynomial function of L for which the usual assumption applies, y_t is the quarterly new home price, ε_t is the deterministic component which might include a constant, seasonal dummy variables or a trend, π_i are coefficients for seasonal roots, and

$$y_{1,t} = (1+L + L^2 + L^3)\, y_t$$

$$y_{2,t} =- (1 -L + L^2 - L^3)\, y_t$$

$$y_{3,t} = (1- L^2)\, y_t$$

$$y_{4,t} = (1- L^4)\, y_t$$

Applying ordinary least squares to the above equation gives the estimates of π_i. If there are seasonal unit roots, the corresponding parameters (π_i) in the auxiliary regression are all zero, and then seasonal differences can be applied, otherwise, seasonal dummies should be applied.

Table 5.2

Seasonality Checked by the Equation (4)

NW		ME		S		W	
π_1	***-0.5	π_1	***-0.416	π_1	***-0.22	π_1	***-0.181
π_2	***-0.578	π_2	***-0.481	π_2	***-0.6	π_2	***-0.645
π_3	***-0.612	π_3	***-0.506	π_3	***-0.646	π_3	***-0.629
π_4	**0.244	π_4	**0.207	π_4	*0.147	π_4	***0.347

Note: ***significant at 1% level;**significant at 5% level;*significant at 10% level.

Results of seasonality are displayed in Table 5.2. All coefficients for seasonal roots are significantly different from 0, which means that there is no seasonality in the data. Therefore, seasonal dummy variables are used.

A bivariate VAR equation is constructed as follows:

$$x_t = a_0+\sum_{i=1}^{l} a_i x_{t-i}+\sum_{i=1}^{l} b_i y_{t-i}+\varepsilon_t \quad (5)$$

where x and y are logarithmic and differenced values of new home prices in the first difference for two different regions, a_0 is constant, l is the number of lags and ε is white noise. The null hypothesis H_0 is that past y does not Granger-cause current x.

The VAR equation for the Granger causality test with one lag is formulated as following:

$$\Delta x_t=a_0+a_1\Delta x_{t-1}+b_1\Delta y_{t-1}+3 \text{ seasonal dummies}+\varepsilon_t \quad (6)$$

where x and y are logarithmic values of new home prices, a_0 is a constant and ε is white noise. The null hypothesis H_0 is that past Δy does not Granger cause current Δx. Table 5.3 displays the results of the standard Granger causality test.

Table 5.3

Granger Causality Tests of Growth Rate of New Home Price Co-movements by VAR

	NW	ME	S	W
NW	***-0.422 (0.066)	**0.117 (0.047)	*0.059 (0.032)	**0.095 (0.035)
ME	-	***-0.409 (0.067)	**0.094 (0.044)	**0.097 (0.049)
S	-	-	***-0.279 (0.071)	*0.141 (0.078)
W	-	-	-	**-0.158 (0.071)

Note: *significant at 1% level;**significant at 5% level;***significant at 10% level;”-”insignificant. Meanwhile, the VAR model includes three seasonal dummies.

It becomes very clear that the ripple effect commences at the Northwest and then spreads out over other regions. There is an obvious feature of the ripple effect in Table 5.3 is that changes in new home prices in the Northwest affect other three regions, but changes in new home prices in other three regions do not have any influences on changes in new home prices in the Northwest. In addition, changes in new home prices in the West are affected by changes in new home prices in other three regions, but changes in new home prices in the West do not affect changes in new home prices in other three regions. Another clear feature of the ripple effect is that the effects of changes in new home prices from other regions are positive, while, the effects of changes in new home prices from themselves are negative.

5.2 Vector Error Correlation Model

To identify whether new home prices are interlinked in the long run, a bivariate cointegration test is concluded. Usually, if two integrated of order-one variables are cointegrated, an error correction model can be used to specify the dynamic between them. Therefore, a bivariate vector error correlation model (VECM) is applied. VECM is illustrated as follows:

$$\Delta x_t = a_0 + \sum_{i=1}^{l} a_i \Delta x_{t-i} + \sum_{i=1}^{l} b_i \Delta y_{t-i} + c_1 \mu_{t-1} + \varepsilon_t \quad (7)$$

where x and y are assumed to be integrated order-one processes, μ is the error correction term from the cointegration regression and others are the same as defined for equation (5).

Because there are more than two price series in regions for cointegration analysis, the multivariate cointegration test may be more meaningful than usual bivariate cointegration test. In this section, the Johansen cointegration trace test is used to obtain the cointegration rank. Because all statistic values are greater than their 5% critical values, the null hypothesis can be rejected. In other words, the number of cointegrating vectors is greater than 0, so there is cointegration.

The Granger causality test is designed by Song (2009) based on a VECM model as follows:

$$\Delta x_t = a_0 + \sum_{i=1}^{l} a_i \Delta x_{t-i} + \sum_{i=1}^{l} b_i \Delta y_{t-i} + c_1 \mu_{t-1} + 3 \text{ seasonal dummies} + \varepsilon_t$$

where μ is a k-vector of the error correction terms from the cointegration regression and the other terms are as defined in equation (7). The same procedure is used as Song (2009) did. The procedure for testing the F-statistics for the Granger causality is set out as follows: first of all, a regression is run on the first-order difference of the variable (Δx) against a constant, 3 seasonal dummies and 1 lag of Δx itself. Secondly, error terms are added to the above regression in step 1 and an F-statistic is calculated to test the joint significance of the error terms. If the error terms are jointly significant, they are to be retained in the regression model. If the error terms are not jointly significant they will be ignored. Finally, 1 lag of Δy is added to the regression in step 2. An F-statistic of Δy indicates whether or not the past Δy help explain the current Δx. Results are shown in Table 5.4.

Results presented in Table 5.4 suggest that causality be most likely through a long-run relationship among the four regions in the U.S. housing market, rather than short-run dynamics among the four regions. Because the significance of the error terms represents the long-run relationship in the regression, the more significant the error term, the more

Table 5.4

F-statistics for Granger causality test of price co-movements based on VECM – regions

	Error	NW	ME	S	W
NW	**-0.365 (0.173)		**0.117 (0.048)	*0.059 (0.032)	**0.093 (0.035)
ME	-0.165 (0.183)	-		**0.094 (0.045)	*0.095 (0.05)
S	-0.269 (0.321)	-	-		*0.138 (0.081)
W	**1.182 (0.571)	-	-	-	

Note: ***significant at 1% level;**significant at 5% level;*significant at 10% level;"-" insignificant.

dependent the explained variable is on other variables in the regression. The significance of the error term clearly indicates that changes in new home price in the West are dependent upon other three regions. Although the error term of the Northwest is significant at 5% level, other coefficients are not significantly different from 0, which means that the long-run relationship exists but not dependent upon the other three regions in the regression, maybe upon other factors.

In a word, the ripple effect of new home co-movements across regions exists in the U.S. housing market. All results suggest that the short-run dynamic relationships among four regions be clear, and the initial region of the ripple effect is the Northwest. The long-run dynamic relationships only exist in the Northwest and the West: changes in new home prices in the West are affected by other three regions, and changes in new home prices in the Northwest are not affected by other three regions in the long run.

CHAPTER 6 CONCLUSION

After being a laggard in the global recovery, the country's crucial service and housing sectors are signalling a major rebound. In the United States, the government extended and expanded an $8000 tax credit in November 2009 to stimulate new home sales. Although new-home sales in March 2010 exceeded analyst expectations, sales are still trending near record lows. Meanwhile, the number of new homes on the market fell 7 % to 211,000 units in April 2010, the lowest since October 1968, but new home prices are still under pressure due to oversupply. The average price of a new home was $258,600, according to the Census Bureau. That was virtually flat compared to a year earlier, and 12% below average prices in 2008. "Firming home prices and an improving jobs market will make recovery felt on Main Street as well as Wall Street," said Robert Dye.

However, the tax credit expired on the end of April 2010. Without the stimulation from the tax credit, what may happen in the new home market in the U.S.? Will the recovery be on Main Street? Forecasting can provide information on answering those questions. In our analysis, different types of forecasting models are applied to predict changes in new home sales and prices, incorporation historical data and external factors. Since new home sales and prices are different data, models which are appropriate for them are different.

The ARIMA (5, 0, 12) can predict changes in new home sales. Because this model only takes historical growth rates of new home sales into account, noise must exist. The result of out-of-sample forecasting is very close to the actual values, and all actual values do not exceed the confidence bounds at 95% level. Meanwhile, the standard SES model with α=0.1692 and without a constant also performs well on out-of-sample forecasting. The Holt-Winters nonseasonal model with α=0.1686 and β=0 only performs well on in-sample forecasting.

All those models do not perform very well in accuracy of forecast. Usually, adding external variables can help solve the problem of accuracy. In our analysis, some macro variables are incorporated to forecast changes in new home sales. But, unfortunately, these macro variables cannot greatly improve the accuracy of forecast, even if changes in new home prices and unemployment rates are taken as regressors. Although the problem about accuracy of forecasting is still not solved, one conjecture is raised that housing characteristics (such as location, number of rooms, number of bedrooms and so on) and existing home sales or prices may have greater influences on new home sales. This aspect can be studied in future. Based on the minimum MSE criterion, one conclusion is drawn that among these models, a simple linear regression with external factors performs best on forecasting changes in new home sales in the U.S. housing market.

The ARIMA (12, 0, 4) performs well on out-of sample forecasting for changes in new home prices, but the Holt-Winters nonseasonal model with α=0.0001 and β=0.9994 only performs well on in-sample forecasting for changes in new home prices. However, the standard SES model with α=0.0001 cannot be used to forecast changes in new home prices, because it gives a very flat predicted line. Taking external factors into account can make the forecasting more accurate. The linear regression model which includes historical changes in new home prices, in personal disposable income, and in prime loan rates, makes the prediction much more precise than the forecasting models which only include historical changes in new home prices. Although the value of adjusted R-square in the linear regression is not very close to 1, it is high enough. Moreover, no matter in direction or strength of changes, the result of out-of-sample forecasting is pretty close to the actual data.

In a word, combining external factors into a time-series model can make forecast of changes in new home sales and prices in the U.S. housing market much more accurate. Time-series models perform better on out-of-sample forecasting than dynamic structure models do.

In addition, to show good grounds for forecasting based on a whole country, the existence of ripple effects in the U.S. housing market is confirmed. Both in the short term and in the long term, changes in new home prices in the Northwest is the origin of the ripple effect, so changes in new home prices in the Northwest have positive influences on other three regions. On the contrary, changes in new home prices in the West do not affect other three regions, but changes in new home prices in other three regions affect changes in new home prices in the West positively, both in the short run and in the long run.

So far, the ripple effect in the U.S. housing market is a field which is not intensives studied, so forecasting changes in new home sales and prices based on a country is more accurate and persuasive than that on a state or region level. In future, economists or researchers should find a sophisticated way to study origin, direction and strength of ripple effects, so that policy makers in different regions or states can take actions with respect to local circumstances.

REFERENCES

Achintya Ray, & Indrani Ray (2009). Stability of Housing Prices in Major US Cities: A Time Series Analysis of S&P/Case-Shiller Housing Price Indices. Journal of Money, Investment and Banking, ISSN 1450-288X Issue 7 (2009).

Alan G. Ahearne, John Ammer, Brian M. Doyle, Linda S. Kole, & Robert F. Martin (2005),. House Prices and Monetary Policy: A Cross-Country Study. Board of Governors of the Federal Reserve System, September, No.841.

Alexander, & M.Barrow (1994). Seasonality and co-integration of regional house prices in the UK. Urban Studies, vol.31, Oct.1994, pp.1667-1689.

Andreas Hornstein (2009). Problems for a Fundamental Theory of House Prices. Economic Quarterly Volume 95, No. 1, Winter 2009: 1–24.

Andrew Light (2008). FORECAST THE EXISTING-HOME SALES IN THE U.S. http://www.sedsi.org/Proceedings/2009/proc/p081010039.pdf

Baki Billah, Maxwell L. King, Ralph D. Snyder, & Anne B. Koehler (2006). Exponential smoothing model selection for forecasting. International Journal of Forecasting 22: 239–247.

B.Ashuri, PhD, & J. Lu (2010). Exponential Smoothing Time Series Models for Forecasting ENR Construction Cost Index. http://ascpro.ascweb.org/chair/paper/CPRT182002010.pdf

B.Giussani, & G.Hadjimatheou (1991). Modeling regional house prices in the United Kingdom. The Journal of Regional Science Associational International, vol.70,Feb. 1991,pp.201-219.

Bowerman B. L., & Richard T. O'Connell (1979). Forecasting & Time Series. Duxbury Press, Belmont, CA.

David E. Rapach, & Jack K. Strauss (2007). Forecasting Real Housing Price Growth in the Eighth District States. Federal Reserve Bank of St. Louis Regional Economic Development, 2007, 3(2), pp. 33-42.

David E. Rapach, & Jack K. Strauss (2009). Differences in Housing Price Forecastability across US States. International Journal of Forecasting 25 (2009) 351–372.

David Duffy, John Fitz Gerald, & Ide Kearney (2005). Rising House Prices in an Open Labour Market. The Economic and Social Review, Vol. 36, No. 3, Winter 2005, pp. 251-272.

Dean Baker (2008). The Key to Stabilizing House Prices: Bring Them Down. Center for Economic and Policy Research, December 2008.

Domenico Depalo (2008). A seasonal unit root test with STATA. http://www.stata.com/meeting/italy08/depalo_2008.pdf

Eric Belsky & Joel Prakken (2004). Housing Wealth Effects: Housing's Impact on Wealth Accumulation, Wealth Distribution and Consumer Spending. Harvard University, December, W04-13.

Everette S. Gardner, Jr., & Eddie McKenzie (2009). Why the damped trend works. Journal of the Operational Research Society, October 22, 2009

Everette S. Gardner Jr., & Joaquin Diaz-Saiz (2008). Exponential smoothing in the telecommunications data. International Journal of Forecasting, Volume 24, Issue 1, January-March 2008, Pages 170-174.

Franses, P.H. (1990). Testing for seasonal unit roots in monthly data. Econometric Institute Report No. 9032/A. Rotterdam: Erasmus University.

Franses, P.H. (1991). Seasonality, nonstationarity and the forecasting of monthly time series. International Journal of Forecasting, 7, 199–208.

Fratantoni & Micheal (2007). The Residential Mortgage Market and Its Economic Context in 2007. MBA Research Monograph Series, January 30, 2007.

Geraint Johnes, & Thomas Hyclak (1999). House Prices and Regional Labor Markets. The Annuals of Regional Science (1999) 33:33-49.

Gordon W. Crawford, & Michael C. Fratantoni (2003). Assessing the Forecasting Performance of Regime-Switching, ARIMA and GARCH Models of House Prices. Real Estate Economics, V31 2: pp. 223-243.

Granger, C.W.J., & Paul Newbold (1986). Forecasting Economic Time Series (Second Edition). San Diego, CA: Academic Press.

Gregory D Sutton (2002). Explaining changes in house prices. BIS Quarterly Review, September 2002: 46-55.

J. Ashworth, & S.C.Parker (1997). Modeling regional house prices in the UK. Scottish Journal of Political Economy, vol.44, Mar.1997, pp.225-246.

Jack Healy. New-Home Sales Jump as Prices Fall Sharply. Published July 27, 2009, from The New York Times: http://www.nytimes.com/2009/07/28/business/economy/28econ.html

James H. Stock, & Mark W. Watson (2000). FORECASTING OUTPUT AND INFLATION: THE ROLE OF ASSET PRICES. http://www.frbsf.org/economics/conferences/0103/conf7.pdf

James W. Taylor (2004). Smooth Transition Exponential Smoothing. Journal of Forecasting, 2004, Vol. 23, pp. 385-394.

James A. Kahn (2008). What Drives Housing Prices. Federal Reserve Bank of New York Staff Reports no. 345 September 2008.

Jesse M. Abraham, & Patric H. Hendershott (1996). Bubbles in Metropolitan Housing Market. Journal of Housing Research, Volume 7, Issue 2: 191-207.

Jianhua Huang, & Charles Lindsey. VAR & VECM Diagnostics (with Stata). Stat689: Statistical Methods for Finance. http://www.stat.tamu.edu/~panhj/689-08sp/project/charlesfinal.pdf

Joachim Zietz , Emily Norman Zietz, & G. Stacy Sirmans (2007). Determinants of House Prices: A Quantile Regression Approach. Journal of Real Estate Finance Econ (2008) 37:317–333 DOI 10.1007/s11146-007-9053-7.

Joshua Gallin (2006). The Long-Run Relationship between House Prices and Income: Evidence from Local Housing Markets. REAL ESTATE ECONOMICS 2006 V34 3: pp. 417–438.

Kostas Tsatsaronis, & Haibin Zhu (2004). What Drives Housing Price Dynamics: Cross-country Evidence. BIS Quarterly Review, March 2004.

Karen E. Dynan, & Dean M. Maki (2001). Does Stock Market Wealth Matter for Consumption. FEDS Discussion Paper No. 2001-23.

Liu Lanhai, & Zhang Hong Jing (2008). Housing Price Ripple Effect Within Cities : A Perspective of Regional Economy Three-sector Equilibrium Model. International Seminar on Business and Information Management 2008.

Luo Zhen qiang, Liu Chunlu,& D. Picken (2007). Housing price diffusion pattern of Australia's state capital cities. International journal of strategic property management, Nov.2007, pp227-242.

Mak Kaboudan, & Avijit Sarkar (2008). Forecasting prices of single family homes using GIS-defined neighborhoods. Journal of Geographical system, 10: 23-45.

Max Neukirchen, & Helen Lange (2005). Characteristics and Macroeconomic Drivers of House Price Changes in Australia. U21 Global Working Paper, No.016/2005.

Philipp K. Janert (2006). Exponential Smoothing. February 2006. http://www.toyproblems.org

Pollakowski, H.O., & Ray T.S. (1997) Housing price diffusion patterns at different aggregation levels: an examination of housing market efficiency. Journal of Housing Research, 8(1), 107-124.

Prajakta S. Kalekar (2004). Time Series Forecasting using Holt-Winters Exponential Smoothing. http://www.docstoc.com/docs/21211855/Time-series-Forecasting-using-Holt-Winters-Exponential-Smoothing.

Pritibhushan Sinha (2008). A Note on some Exponential Smoothing Forecasting Methods. IIMK/WPS/34/OM&QM/2008/04.

R. Macdonald, & M. P. Taylor (1993). Regional house prices in Britain: Long-run relationships and short-run dynamics. Scottish journal of political economy, vol.40, Jan. 1993, pp. 43-55.

Rangan Gupta, Alain Kabundi, & Stephen M. Millern (2006). Using Large Data Sets toForecast Housing Prices: A Case Study of Twenty US States. University of Nevada, Las Vegas , Department of Economics, working papers, No.916, May.

Rangan Gupta, & Stephen M. Miller (2009). 'Ripple Effects'and Forecasting Home Prices in Los Angeles,Las Vegas, and Phoenix. University of Connecticut working paper, January, 05R.

Raphael Bostic, Stuart Gabriel, & Gary Painter (2009). Housing wealth, financial wealth, and consumption: New evidence from micro data. Regional Science and Urban Economics 39:79–89.

Ralph D. Snyder, & J. Keith Ord (2009). Exponential Smoothing and the Akaike Information Criterion. Department of Econometrics and Business Statistics. Working paper 4/09.

Raymond Y.C. Tse (1997). An application of the ARIMA model to real-estate prices in Hong Kong. Journal of Property Finance Vol. 8 No. 2, 1997, pp. 152-163.

Reuters. US New Home Sales at 2-Year High in April. Published 26 May 2010, from CNBC:
http://www.cnbc.com/id/37354168/US_New_Home_Sales_at_2_Year_High_in_April

Rob J Hyndman (2002). Forecasting based on state space models for exponential smoothing.
http://www.diku.dk/hjemmesider/ansatte/pisinger/production/ForecastingBasedOnStateSpaceModelsForExponentialSmoothing.pdf

Sean Holly, M. Hashem Pesaran, & Takashi Yamagata (2006). A Spatio-Temporal Model of House Prices in the US. IZA Discussion Paper No. 2338, September 2006.

Song Shi, Martin Young, & Bob Hargreaves (2009). The ripple effect of local house price movements in New Zealand. Journal of Property Research Vol. 26, No. 1, March 2009, 1–24.

Tilak Abeysinghe, Uditha Balasooriya, & Albert Tsui (2000). Small-Sample Forecasting Regression or ARIMA Models. National University of Singapore, March.

Tirtirglou, D. (1992). Efficiency in Housing Markets: temporal and Spatial Dimensions. Journal of Housing Economics, vol. 2, 276-292.

William Miles (2008). Volatility Clustering in U.S. Home Prices. JRER, Vol.30:73-90.

William N. Goetzmann, Liang Peng, & Jacqueline Yen (2009). THE SUBPRIME CRISIS AND HOUSE PRICE APPRECIATION. NBER WORKING PAPER SERIES:15334. September 2009.

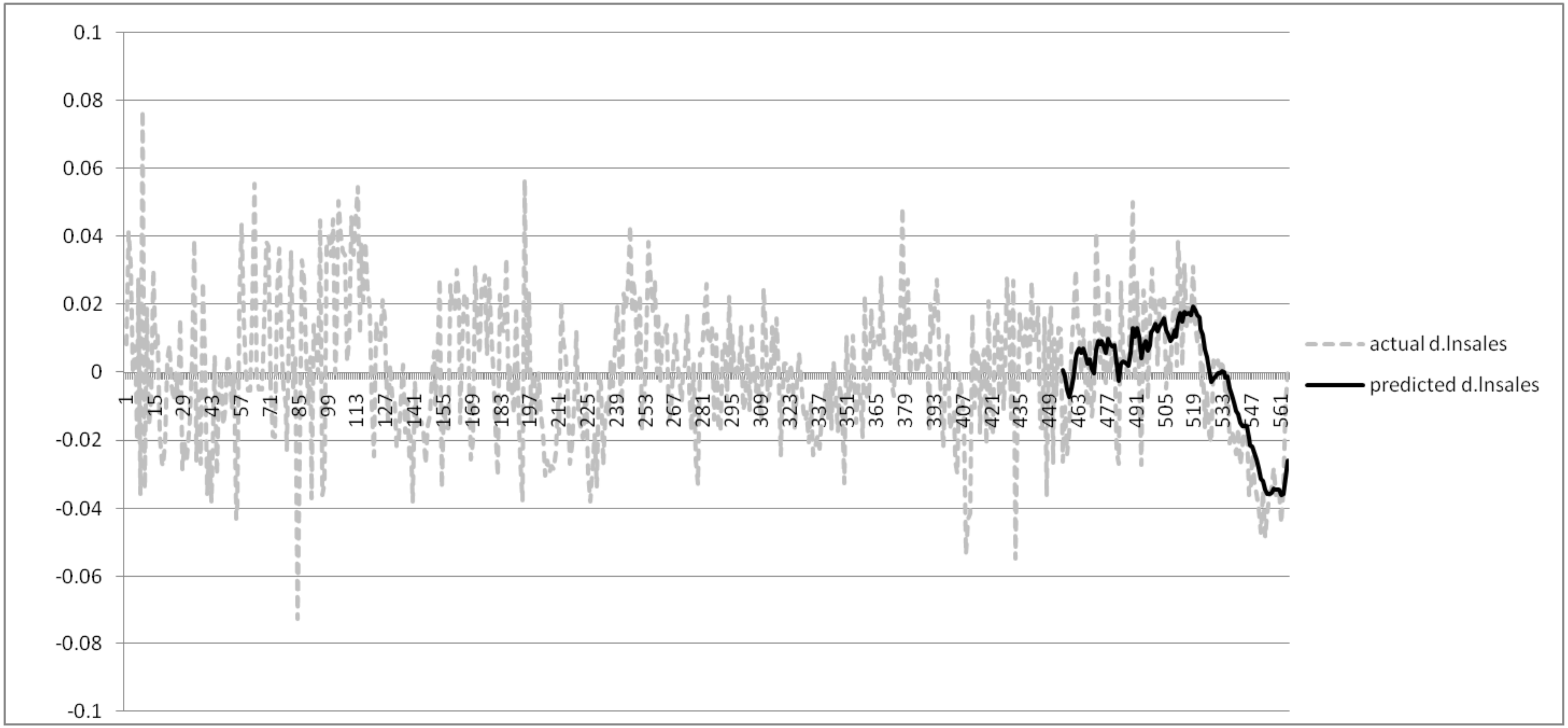
Y.Tu (2000). Segmentation of Australia housing market: 1989-98. Journal of property research, vo.17,Apr.2000, pp.311-327.

Zan Yang (2005). Co-integration of housing prices and property stock prices: evidence from the Swedish market. Journal of Property Research, March 2005, 22(1) 1–17.

Zellner, A., & Palm, F. (1974). Time Series Analysis and Simultaneous Equation Econometric Models. Journal of Econometrics 2(1), 17-54.

Figure 4.11

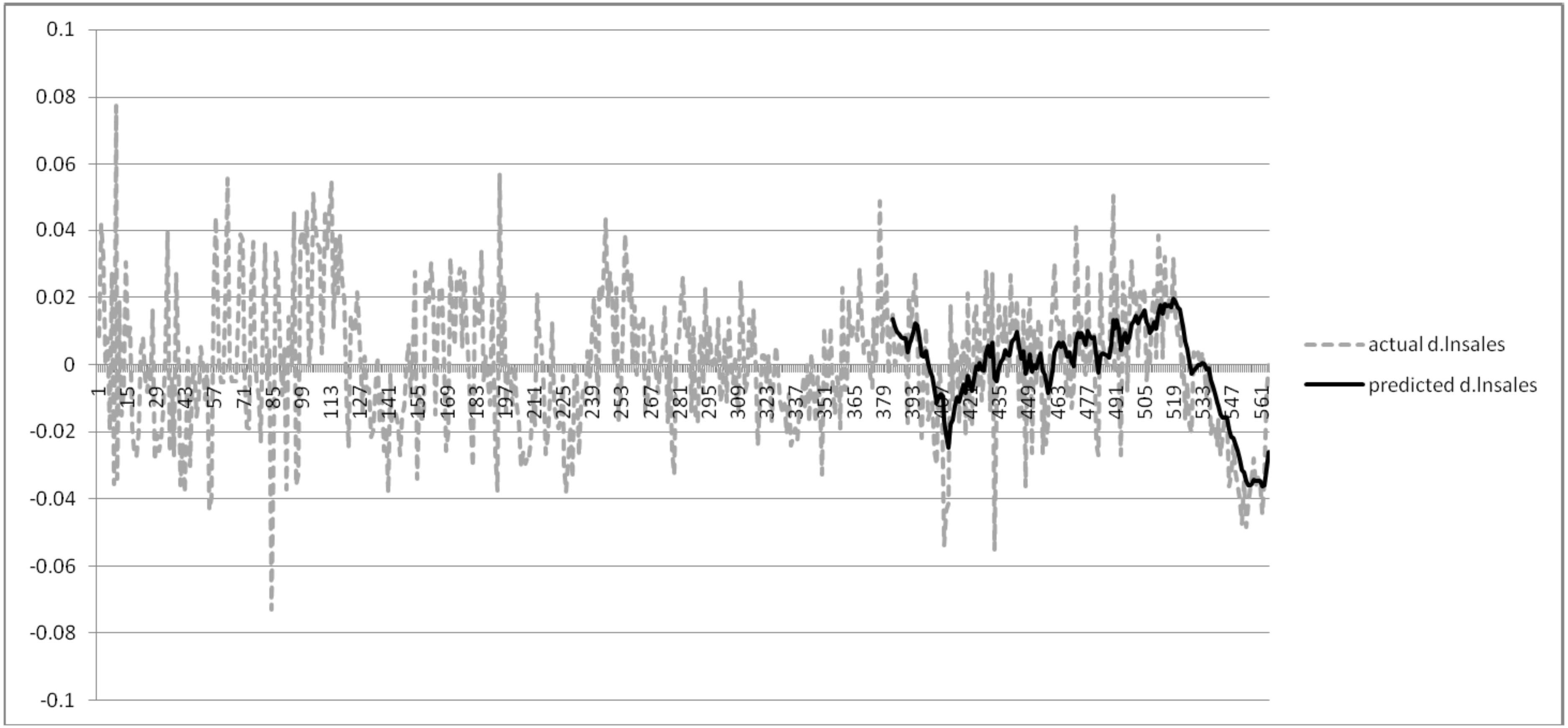
Forecasting d.LS by Single Exponential Smoothing with α=0.1692: Jan 2001-Jan 2010



Note: d.lnsales=d.LS.

Figure 4.12

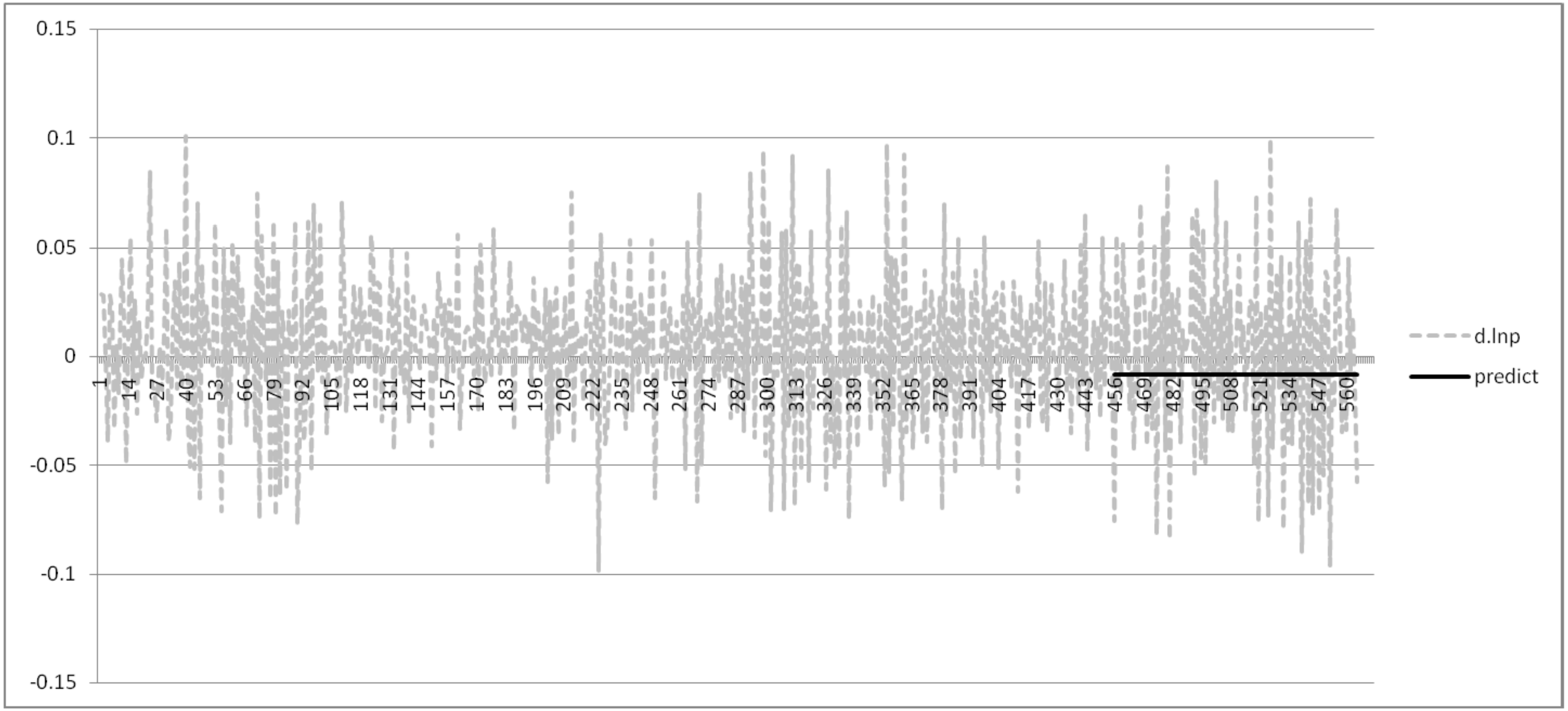
Forecasting d.LS by Single Exponential Smoothing with α=0.1692: Jan 1995-Jan 2010



Note: d.lnsales=d.LS.

Figure 4.13

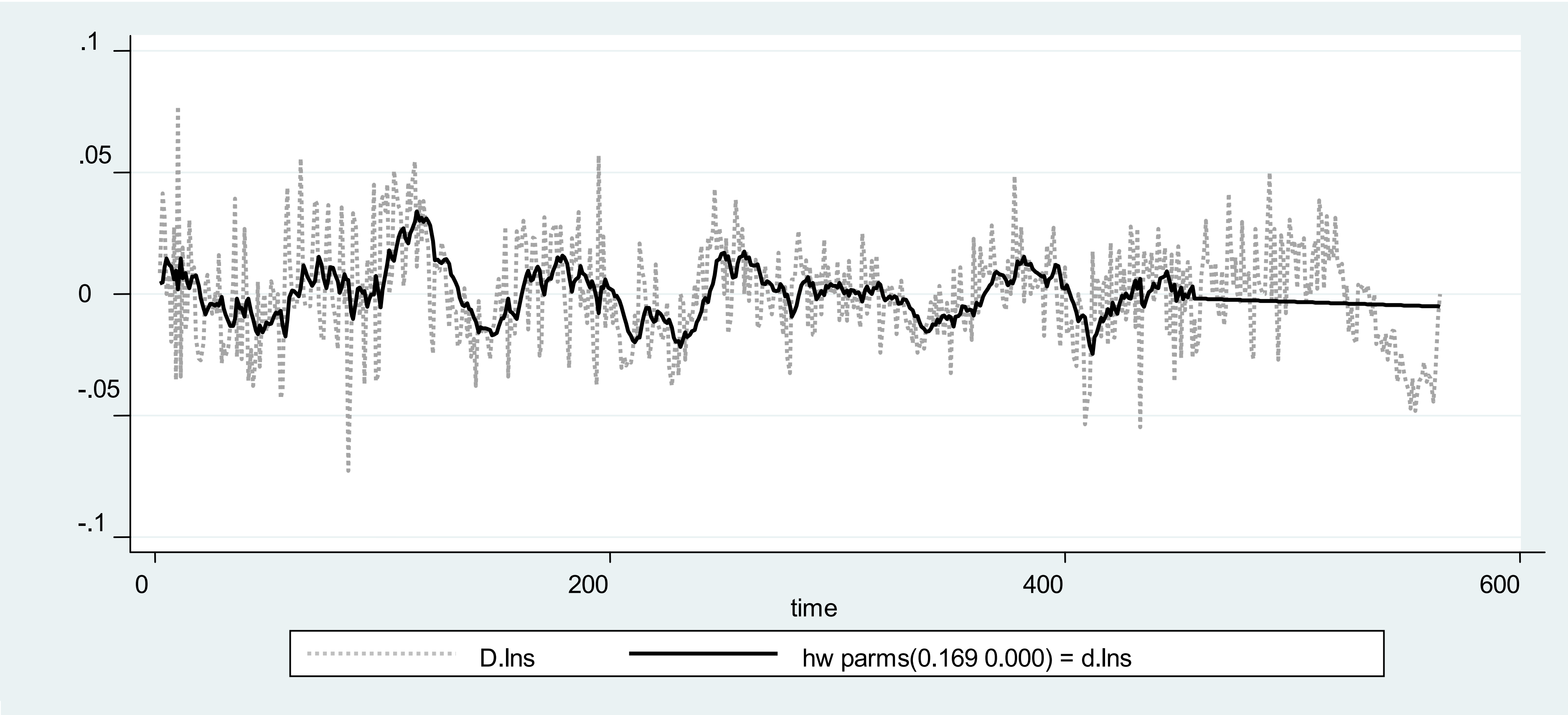
Forecasting d.LP by Single Exponential Smoothing with α=0.0001 : Jan 2001-Jan 2010



Note: d.lnp=d.LP

Figure 4.14

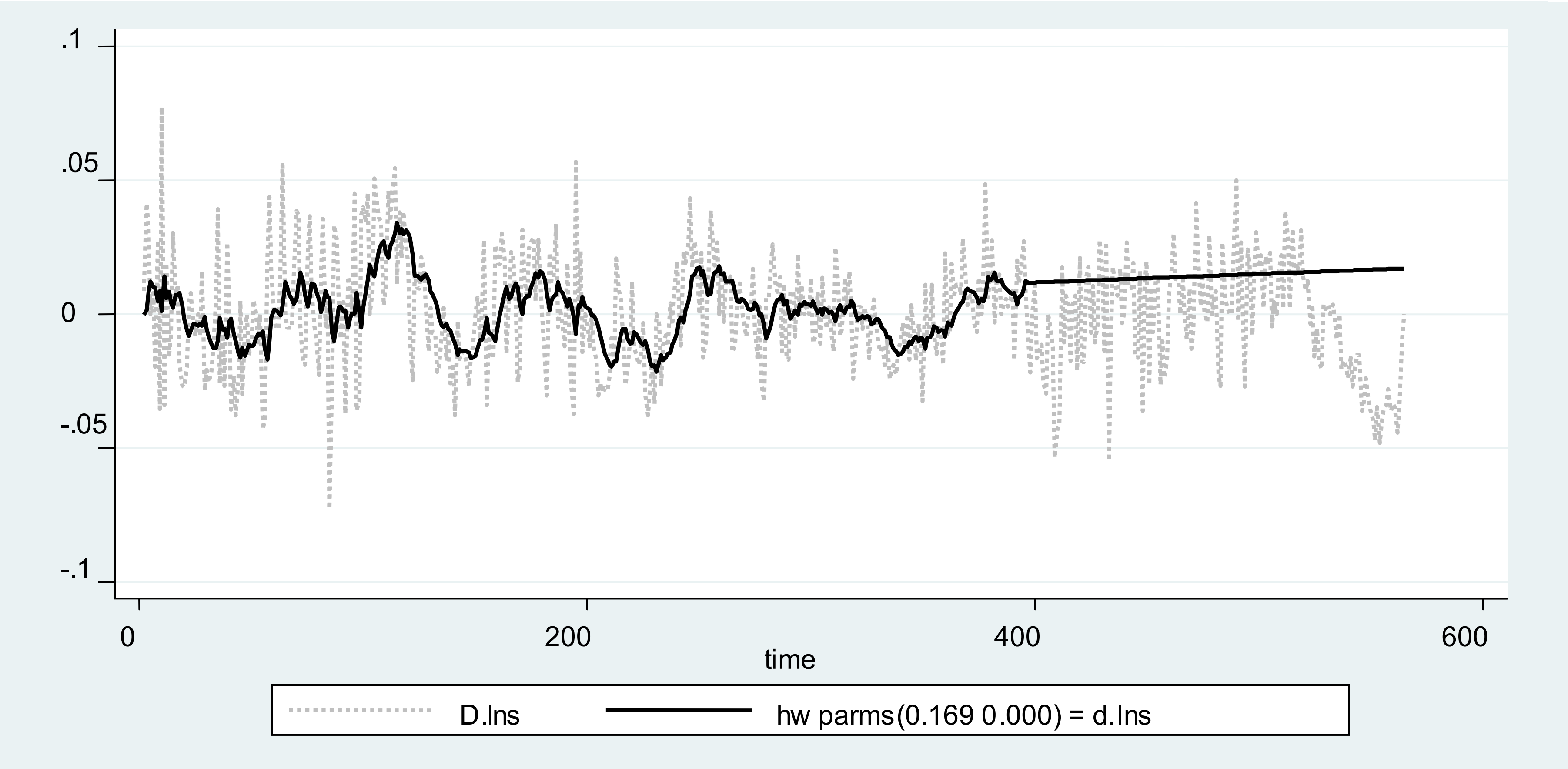
Forecasting d.LS by Holt-Winters Nonseasonal Smoothing Model with α=0.1686 and β=0.0000: Jan 1963-Jan 2010



Note: Out-of-sample forecasting is from Jan 2001 to Jan 2010. This figure includes in-sample and out-of-sample forecasting.

Figure 4.15

Forecasting d.LS by Holt-Winters Nonseasonal Smoothing Model with α=0.1686 and β=0.0000: Jan 1963-Jan 2010



Note: Out-of-sample forecasting is from Jan 1995 to Jan 2010. This figure includes in-sample and out-of-sample forecasting.