Forecasting US Home Prices with Artificial Neural Networks and Fuzzy Methods Combination and Single Forecasts

Pejman Bahramianfar

Submitted to the
Institute of Graduate Studies and Research
in partial fulfillment of the requirements for the Degree of

Master of Science
in
Economics

Eastern Mediterranean University
July 2013
Gazimağusa, North Cyprus

Approval of the Institute of Graduate Studies and Research

Prof. Dr. Elvan Yılmaz
Director

I certify that this thesis satisfies the requirements as a thesis for the degree of Master of Science in Economics.

Prof. Dr. Mehmet Balcılar
Chair, Department of Economics

We certify that we have read this thesis and that in our opinion it is fully adequate in scope and quality as a thesis for the degree of Master of Science in Economics.

Prof. Dr. Mehmet Balcılar
Supervisor

Examining Committee

1. Prof. Dr. Mehmet Balcılar
2. Prof. Dr. Salih Turan Katircioglu
3. Assist. Prof. Dr. Kamil Seroglu

ABSTRACT

Recent studies have shown that there is a link between the housing market and economic activity. Also, they suggest that house-price fluctuations lead to real activity, inflation, or both. Therefore the existence of good model to forecast is very crucial for policy makers.

The main objective of this thesis is to forecast the housing price indices for US and four Census regions of the US, namely, Northeast, South, Midwest and West by using relevant time series techniques. The purpose is to forecast out-of-sample period, from 2001:1 to 2010:5 according to the monthly data covering the in-sample period from 1968:1 to 2000:12 by using four advanced valuation method artificial neural networks and fuzzy methods multi layer perception (MLP), nonlinear autoregressive neural network (NAR), adaptive Neuro-fuzzy inference systems (ANFIS) and genetic algorithm (GA) as well as the forecast combination method. Also, the 24-step-ahead price indices will be predicted covering 2010:6-2012:6 period.

The result of this study showed that both MLP and NAR separately had better answer in all parts of the data (US and four census regions) and they could have better forecast accuracy. Similarly, the results of ANFIS have a better forecast power especially in the initial steps than MLP and NAR.

The results of this research also posits that both the neural network (MLP and NAR) and ANFIS have a suitable ability to model and forecast especially when there is a non- linear relationship between the data .On the other hand the results of the GA (as a linear model) in all parts of the data were not desirable. The results also showed that the nonlinear models like neural networks are better at longer horizons while the GA (as a linear model) is better at short horizons.

Keywords: Forecasting, Neural Networks, US and Census Housing Price Indexes, Adaptive Neuro Fuzzy Inference Systems (ANFIS), Genetic Algorithm

ÖZ

Günümüzdeki son çalışmalar, emlak piyasasıyla ekonomik aktiviteler arasında bir ilişki olduğunu göstermektedir. Ayrıca bu çalışmalar, konut fiyatlarındaki dalgalanmaların piyasalarda gerçek aktiviteler, enflasyon veya her ikisine de öncülük ettiğini göstermektedir. Bu nedenle, öngörülerde bulunabilmek için iyi modelin belirlenmesi. politika yapıcıları için çok önemlidir.

Bu tezin ana amacı, zaman serileri teknikleri uygulayarak Amerika Birleşik Devletleri ile ona bağlı kuzey doğu, güney, orta batı ve batı bölgelerindeki konut fiyatlarını öngerebilen en yi modelleri belirlemektir. Bu çalışmadaki diğer amaç, dört gelişmiş sinir ağı ve bulanık öngörü yönetimlerini çok katmanlı algılama (MLP) doğrusal olmayan otoregresif sinir ağı (NAR) uyarlanan bulanık sinir ağı çıkarım sistemi (ANFIS) ve genetik algoritma (GA) kullanarak 1968:1 ile 2000:12 tarihli iç örneklem tahminlerini ve 2001:1 ile 2010:5 tarihli dış örneklem öngörülerini yapabilmektir. Ayrıca, 2010:6 ile 2012:6 periyotlarını içeren 24-adım fiyat endeksi de öngürül müştür.

Bu çalışma; hem MPL hem de NAR öntemlerinin Amerika ve ona bağlı bölgelerde daha iyi sonuçlar verdiğini göstermiş ve aynı zamanda öngörü doğruluk oranlarının daha iyi olduğu görülmüştür. Benzer olarak, ANFİS yönteminin sonuçlarına göre, bu yöntem özellikle ilk adımlarda MLP ve NAR'a nazaran daha iyi tahmin gücüne sahiptir. Yine bu çalışmanın sonuçları; özellikle veriler arasında doğrusal olmayan bir ilişki olduğunda MPL, NAR ve ANFİS yöntemlerinin modelleme ve öngörü için

uygun olduğunu göstermiştir. Aynı zamanda, GA'nın (doğrusal model olarak) sonuçları, çalışmadaki bütün veriler için yeteri ölçüde iyi değildir.

Yine bu çalışma, NAR ve MPL gibi doğrusal olmayan modellerin uzun zaman aralığında ve GA gibi doğrusal modellerin kısa zaman aralıklarında daha iyi olduğunu göstermiştir.

Anahtar Kelimeler: Öngörü, Sinir Ağı, Amerika Birleşik Devletleri Konut Fiyat Endeksi, Uyarlanan Bulanık Sinir Ağı Çıkarım Sistemi (Anfis), Genetik Algoritma

To My Life, Andisheh

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to my supervisor, Prof. Dr. Mehmet Balcilar, who patiently guided me from the initial phases of this study. His continuing supervision, motivated me and put me back on track when I occasionally went astray, making this task an interesting learning experience. Without his encouragement and support, it would not have been possible to complete this thesis in such a short time.

Special thanks go to Assoc. Prof. Sevin Ugural for her support and insightful comments and her suggestion were valued with utmost attention since they contributed profoundly toward my understanding of complex issues throughout my study.

Special thanks go to my life, Andisheh, whose unconditional support and encouragement were amazing her faith in me made me feel confident and proud of my work.

I also thank my best friends HesamAldin Shahrivar, Shahryar Mirzaalikhani, Fru Asaba, Samir Orujov, Isah Wada for their continuous brotherly care, support and their friendly encouragement.

I also thank all the academic and administrative staff of the Economics Department of the Eastern Mediterranean University for their friendship and for providing the best environment and facilities to study.

Finally, my heartfelt gratitude and absolute respect go to my parents, Khosro Bahramianfar and Azar Bahramian, for their support and understanding in all my educational endeavors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

The main objective of this thesis is to forecast the housing price indices for US and four Census regions of the US, namely, Northeast, South, Midwest and West The purpose is to forecast out-of-sample period, from 2001:1 to 2010:5 according to the monthly data covering the in-sample period from 1968:1 to 2000:12 by using four advanced valuation method (artificial neural networks and fuzzy methods) and the 24-step-ahead price indices will be analyzed covering 2010:6-2012:6 period.

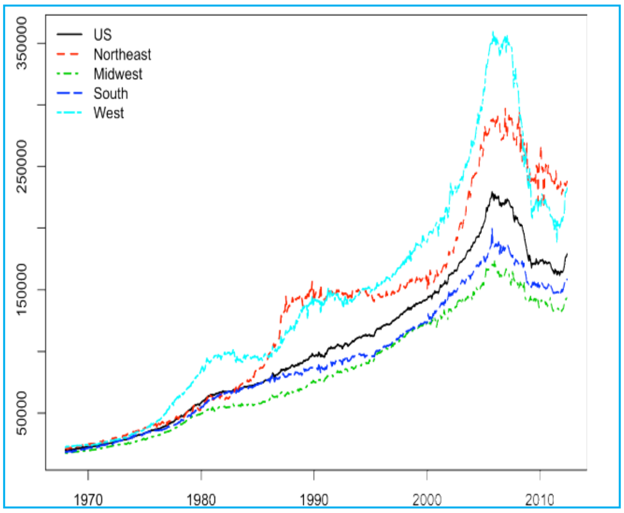


Figure 1. Median Home Price in US and the Four Regions, 1968:1–2012:6.The Figure Plots Median Home Prices in Dollars. All Series Are Seasonally Adjusted By Using X-12 Filter. Source: NAR

In spite of the limited capability of linear models to predict the real world dynamics most of the researchers used linear models in their analysis. Up until now limited number of scholars has preferred non-linear models in their studies such as Rapach and Wohar (2006) that used non-linear techniques in order to forecast exchange rate dynamics.

Housing prices are less intense, and partially dependent on the market events that decrease the equilibrium price than those events which increase it. The historical analysis of the latter would clarify the non-linearity of housing price indices, especially, the Great Recession period would be of great significance for this analysis throughout which housing prices fell slowly rather than quickly slowing down the restoration of the US economy. Moreover, Kim and Bhattachya (2009) did a great deal of work to elucidate the nonlinear dynamics of housing prices in three of the aforementioned regions, the Midwest is an exception. Their research showed that housing market responses differ throughout the expansion and contraction period of real estate sector. Seslen (2004) mentioned that households are more inclined to trading up during expansion rather than contraction of the market. She goes further stressing that loss aversion make households less mobile and active during the downswings of the mentioned market. Furthermore, Murphy and Muellbauer (1997) underline the non-linear movements caused by the lumpy transaction costs. Considering all the points mentioned so far checking for non-linearity seems to be logical and promising.

Housing market plays a paramount role in the business cycle dynamics, and we would stress that the volatility of the latter owes to the movements in the mentioned sector. The Leamer's statement is very interesting in the sense he said that housing itself is the business cycle and stressing further that "any attempt to control the business cycle needs to focus especially on residential investment". In (p150), he concluded that construction booms led to overstock of new homes, which in turn led to building hiccup. He argued that monetary policy should be regulated to prevent construction booms which will lead to eventual slumps. Smets emphasized that interest rates and monetary policy are the main variables setting up the dense connection between housing and business cycles.

Residential housing is the compartment of investment demand which, in turn is an important particle of GDP. However, including housing to GDP via investment demand per se, and not considering its effects on other segments of GDP would not be correct. Case, et al. (2005) provided a good insight to this issue. The problem is that in usual life-cycle theories, no classification of different types of wealth is provided, conversely a unique marginal propensity to consume is assumed. However, the case mentioned above will examine five rationalizations for different MPC out of various sorts of wealth, namely ”differing perceptions about the impacts of permanent and transitory components, differing bequest motives, differing wealth accumulation motives, differing abilities to measure wealth accumulation, and differing physiological “framing” effects. A Balcilar in his paper (2012) provided another rationalization not mentioned in the former case. He argued that housing as well as durable goods also enables consumption services for households, and this should be taken into account because consumers may accommodate their nondurables and services consumption differently to the changes in market prices of housing and durables already been under exploitation of consumers.

According to Fromlet (2010, p. 64), at the heart of the neoclassical economic doctrine, with full and concise information at the disposal of humans, they are able to make rational decisions.

From the foregoing, the axiom of maximization of utility is made clear (Ibid, p. 64)

In any market scenario where rational behavior is satisfied, prices generally the true value of the product sold and sensitivity of the market to avail information is accelerated (Kindleberger and Aliber, 2005, p. 38). Thus, the way in which we interact in the market, orders for products and work directly correlates with the availability of robust and integrated information system.

The advantage here ranges from the easiest to ascertain quality information about the availability of goods and services, determine product prices and engage in cost saving transactions. These advantages encapsulate what is regarded as a naked economy (Hessius, 2000, p. 2)

A continuous steady price rise supports the idea of non-negative trend in the market in general. Shiller (2002, p. 19), in spite of the threat of rising prices, good return help signal optimism and stimulate rising price. In the wake of the global financial crisis 2008-2009, the price of housing soared astronomically despite the high leverage in some European countries like Sweden (Central Bank, 2011. p. 7). One basic feature of the mortgage sector is seemingly short interest rate duration and a reduced loan level for amortized assets.

There is the tendency the payment might be tough to affect either due to the loss of jobs or a rise in unemployment. Thus, the result of this would imply a decrease in the housing demand and downward pressure on prices (Englund, 2011, p. 28).

Empirical evidence has shown that economic crises have been triggered in part by the falling prices of houses. The health of market and the rising debt profile of an economy as a result the mix of fiscal and monetary policies expansion has been identified as some of the possible factors giving rise to such crises. By and large, housing demand is directly correlated with income and employment variables and this tend to have also influenced and affect price in general. The quantity of newly constructed homes and the building cost outlays also affects housing prices (Englund, 2011, p. 28)

Housing prices also change in response to the repo rate. This is simply the rate of interest determined by the central bank. This response of the housing price to the repo raises the threat of possible risk and instability in the financial system when prices drop to a very low level. These give rise to challenges in offsetting mortgage loans by the individual household. With falling prices, evidence suggests that this may trigger a crisis scenario. Home owners then experience a fall in the level below the mortgage for them triggering indebtedness (Jönsson, Nordberg, & Fredholm, 2011, p. 136). The objective of this study centers on the need to reliably forecast the movement in prices in the housing market in the US and four census regions.

Chapter 2

LITERATURE REVIEW

In the scientific literature there are various empirical studies with differing results regarding the impact of housing prices on consumption , however most of them suggesting a significant positive correlation between the two. Elliot (1980) and Levin (1998) found no important significant impact; however, Peek (1983), Bhatia (1987), Case (1992) and Case, et al. (2005) could manage it. Furthermore, Engelhardt observed asymmetry between the two that is, only negative news has an influence on consumption.

The strong relationship between housing markets and economic activity has been provided in studies of different scholars so far, Green (1997), Iacoviello (2005), Case et al. (2005), Rapach and Strauss (2006), Leamer (2007), Paries and Notarpietro (2008), Vargas-Silva (2008), Bao et al. (2009) Christensen et al. (2009), Ghent (2009), Ghet and Owyang (2009), Pavlidis et al. (2009), Iacoviello and Neri (2010) being the most considerable of them. As far as housing is an important component of private wealth accumulation (Cook and Speight 2007), said that price movements in the housing market have a paramount impact on consumption and saving behavior of consumers (Englund and Ioannides 1997). Housing prices and consumption expenditure have a strong positive relationship because consumers consider housing as consumption good (Pavlidis et al. 2009). Volatilities in some macroeconomic variables such as GDP can be predicted by the comprehensive analysis of housing price indices according to Forni et al. (2003), Stock and Watson (2003), and Gupta Das (2010) because of the correlation between housing sector dynamics and volume of economic activity.

As far as housing price movements are crucial for the future direction of the economy, we believe that forecasting the latter would be very helpful in making better economic policies. However, while doing so, we should first analyze the nature of the data that is whether it follows linear or non-linear trend. The reason is that predictions of non-linear models will be biased if the data possesses linear adjustment or vice versa.

As mentioned before housing prices mostly follow a non-linear trend (Genesove and Mayer 2001, Engelhardt 2001, Seslen 2004, Kim and Bhattacharya 2009, Balcilar et al. 2011). A large number of further studies can provide a useful evidence of nonlinearity of macro-variables. Few out of many examples might be Neftchi (1984), Falk (1986), and Bradley and Jansen (1997) that proved the non-linear and asymmetric trend in macro variables. Skalin (1999) and Terasvirta (2002) showed the relevance of GDP and unemployment to the STAR model. On the one hand, according to the aforementioned there is a dense relationship between the housing sector and economic activity, and GDP, the asymmetric nature of macro-variables can be exposed. On the other hand, the non-linearity in housing prices might be caused by the asymmetric nature of GDP and interest rates (Neftchi 1984, Enders and Skilos 2001). Stated that there are two approaches to the asymmetric trend of housing price indices first of which stresses that because of the behavior of households change asymmetrically that is they are more inclined to buy when prices rise than prices decrease due to risk aversion (Abelson et al. 2005). Equity constraints and transaction costs are reasons for non-linearity in housing prices argued by the authors such as Seslen (2004), and Muellbauer and Murphy (1997).

To conclude, housing prices can be seen as an important tool affecting business cycles via its impact on investment and consumption spending. Also, local specifications allow for differences in regional business cycles.

In order to elucidate the non-linear, in this thesis, I will present out-of sample forecast of housing price indices of four aforementioned regions and US itself by comparison. Afterwards, evaluate the performance of both models to ascertain linearity or non-linearity.

Root mean square error (RMSE) criterion is usually used in the out-of-sample forecasting, however, found unreliable. In the analysis, the neural network methods (MLP _NAR) and ANFIS as Neuro _fuzzy systems and GA examined and later tested for superiority by Diebold and Marino (1995) test..

Furthermore, comparison of the superiority of out-of-sample forecasting performance of those methods is emphasized using and Debold et al (1998). Eventually, 24 step-ahead forecasts of housing price indices for four regions are developed by using advanced valuation methods following the ex-ante forecast design which covers the period from 2010:6 to 2012:6.

Various approaches and methodology have been utilized in studying price dynamics for housing market researches. Crawford and Fratantoni (2003) on the one hand and Miles (2008) both focused on the price dynamism in the housing sector using non-linear price movements obtained from the asymmetry of price relations for housing business cycle. The famous Markov regime switching technique was adopted by Crawford and Fratatoni (2003) to capture the tendencies for the repetition of exchanges on levels for housing price indicators for a cross section states in the United State i.e. California, Florida, Massachusetts, Texas and Ohio. The aim was to draw a comparison other time series models such as ARIMA and GARCH. The conclusion showed that the Regime switching model by Markov outperformed the ARIMA model for in-sample forecasting while the reverse is held for out-sample forecasting. Following to Crawford and Fratantoni paper, Miles (2008) also utilized state wise level data set. The inadequacy of the Markov models motivated miles (2008) to opt for a different approach for the family of the nonlinear modeling such as TAR and GAR model. This approach did not support evidences for the presence of TAR syndrome in the housing data set.

Using the ARIMA and the GARCH technique, the GAR approach was reported to be adequate for forecasting out of sample.

The final deductions from Mile (2008), showed that in those geographical areas where housing prices recorded fluctuations, the GAR comes in handy and are more adequate than compared with the Switching model attributed to Markov .

The review of the literature showed that the above mentioned studies did not test for the possibility of a structural break in the price series for the housing market. Guirguis et.al. (2005), opined that for a series of data covering a long period of time, the justification of observing coefficient with fixed estimates is difficult. This means that the presence of external shock in huge magnitude can alter a series significantly from a given period to another in a remarkable way.

A notable point of view is the shift in institutional arrangement in the market for financial assets before the onslaught of the global financial crisis. The changes recorded in the market meant an upsurge in highly risky exotic mortgage assets and this generally changed the dynamics of the housing price series data. Chow (1960) is regarded as the pioneer for structure break laid out the vital point in his test procedures. i_ the absence of a multiple break points, ii_ an exact estimate for the break period, iii_ sameness for the estimate of error variance given the periods for considerations.

The points however limit the efficiency of the test result and range of studies following this sought to overlook these points. In about the same period as chow (1960), Quandt (1960) in his seminal work assumed the break date were not known and a changing nature of the variance of error given the break date. Thus, Quandt (1960) unlike Chow (1960) mainly focused on just a break point in his series. The shortcoming of this approach is the characterization of the distributional test statistic based on a break point date that is not known.

Andrews (1993) provided a remarkable for the structural break test when he came up with his derivation of for the Quandt test statistic distribution alongside its critical values. Although Quandt (1960) and Andrews (1960) both considered just the possibility of one break point, their approach is still considered insufficient in estimating structural break.

In the period following, Bai (1997a) came up with his new estimators which are a distributional statistic with asymptotic properties. For the purposes of estimating multiple break points, Bia and Perron (1998 and 2003) are the leading lights in this regards. The method they employed is considered more robust and a much more generic methodology for time series structural break modelling.

2.1 Neural Networks Based Forecasting

With the advancement in computer technology, it is easier to handle large data set for multivariate analysis for the estimation of a complexity of relation in the given data set. This data set can be linear or non-linear in nature. An example is the demand and supply for housing and a wide range of asset prices.

The application of computer technology in the estimation of historical data set in pervasive with the advantage of ease of analyzing and predicting price movement in the market.

The notion underlying the above holds that with the ability and flexibility of predicting price movement using data from historical analysis, we are better able to understand the market reaction to some fundamental variables. This follows that with such an understanding of the daily data relationship we are better able to deduce the likely value in the coming days ahead.

Thus, we can extrapolate from the future market behavior and make informed deduction about the market movements. The mostly employed computerized machine with a reliable ability for the prediction of market movement is known simply as ANN. This is done with historical market data.

ANN makes use of an approach which is completely computational using the human biological makeup and functions of the brain. It has a semblance of the man’s brain given the manner in which it operates and attracts knowhow such as from illustrations. The knowledge and information component generate by the ANN machine utilizes a set of weight that is essential for its operations. This technique has been employed for a number of mental assignments with a cognitive component; these include a speech rebirth/synthesis, the identification of signs and patterns, medical diagnosis, character building etc.

These days there is an increased drive at commercializing the usage of ANN as a reliable technique for the forecast of price movement in the market for the financial asset. They're evidences provided that supports the potential advantages of ANN in terms of its forecasting ability (Dhar and Stein, 1996; Ward and Sherald, 1995). There is however structural limitations to this system in term of its network design.

Kaastra and Boyd (1996) pioneered a network system capable of forecasting both financial and economic series. Theirs is a clearer, practical and not technical framework for setting up the neural network system and its salient point for a starters detail explanation for the setup variables and parameters.

They employed an ANN system based on straight forward back propagation system in their buildup and they noted that this system has the capacity to estimate any class of function and with accuracy of forecasting. This point is also supported other studies particularly Window et.al (1994). The evidence showed that the ANN has the capacity to estimate any pattern of movement whether irregular or illogical.

The general deductions hold for this method. But the numerous variable and parameters as a numerous repetition of the process makes it cumbersome.

Research evidence shows that for any system with non-linear instability patterns such as the market for housing, the utilization of the ANN methodology serve properly (Do and Grudnitiski, 1990). This view is also echoed by Lan Guoliang (2003). His study was based on the utilization of the ANN technique in forecasting sales rate for housing in the housing market.

In the same vein the ANN model has also been applied in the approximation of the indexation of prices more recently. The ANN technique came in handy for predicting the performance of the housing market by Khalafallah (2008). An historical data set relating to the performances of the housing market was utilized in this study for multiple ANN analysis. The conclusion from this studies the evidences that justify the forecasting ability of ANN. Khalafallah (2008) allow for an error prediction of -2 and +2 % respectively. Rather than the more known method, this study that uses the mean housing price, Khalafallah (2008), employed a different technique for predicting the price movement in the housing sector. This method uses the ratio of the bid and ask price

Although the OLS techniques have been utilized housing market research, a list of factors can however reduce the adequacy of the technique. This range from co-linearity problems, bias in model specification, a non-linear pattern and present of outliers.

The ANN technique provides an easier approach in the context of the problems highlighted when data tend towards a non-linear pattern (Lenk et.al 1997; Owen and Howard, 1998).

Tay and Ho (1992) employed the Back propagation ANN technique in their analysis of home price movement in Singapore in comparison with the OLS approach. They reported an error mean of 3.9% compared to7.5% for the OLS method. This study used a data of 1055 variable (properties) and conclusion provides evidences for the ANN based technique in term of its accuracy. This conclusion is also in line with other studies such as of McCluskey (1996) and Do and Grudnitiski (1992).

On the other hand, some studies employed an approach closely related to the ANN method using a fewer data. For on a study in England and Wales Evans et.al (1992), employed a data set containing 34 variable (homes).

However, some the literature surveyed, an ANN based method with a larger data set to perform well for accuracy and reliability than a fewer data set approach. In this regards McCluskey et.al (1996) and Rossini (1997) reported a mixed results. Using a method drawn from Evans et.al (1992) - an approach based on various ANN designs, McCluskey et.al (1992) employed a 416 sets of data from the Irish housing market. They compared this approach with the OLS method. Their evidence provides support for the OLS rather than the ANN methods.

In a different study of southern Australia market using the ANN method it is found that, this approach perform better with robustness under some conditions than the OLS based method Rossini (1997). This view is also held by the duo of Rossini and

Kershaw (1999) using a fixed index for home prices employing both the OLS and ANN method.

The method was also applied to the US housing market using 4 separate ANN methodologies by Borst (1992). The employed a data set without breaks or outliers to analyze the effect of data dynamics for a multiplicity of non-independent observations of ranges of prices. He reported an error value for the absolute mean of 8.7% to 12.4% with 22 and 20 sample sets of data for each.

Tay and Ho (1992) and Do and Gudnitski (1992) employed a data set with the following inputs-numbers of bedrooms, garage, fireplace, floor and the home age. They also include the square footage of the home etc. This study used a two node layer for input prices and a 10 neuron in a hidden compartment. The evidence from the study shows a smaller mean absolute error for the ANN which is 6.9% compared the OLS which was 11.3%

A recent research by Nguyen and Cripps (2001), studied the application of ANN approach and OLS method for only a given home sale. This study found evidence for the ANN based method as against the OLS. Their result was based on the size of the data set and functionality of the method employed in the model. If fewer set of data is employed with a simple function model, the OLS perform better than ANN but with a larger set of data ANN is more reliable and yield better outcomes.

The general conclusion is that with an ANN based approach as the data set gets larger with more input variable, the predictive accuracy of the model is enhanced. In another study by Worzala et.al (1995), the approach followed by other authors

(Evans, 1992; Borst 1991; and Do and Grudnitski, 1992) was re-analyzed and examined. Using triple data set cases, they analyzed a 288 sample of variable as an input factor in their model.

In the first case, they choose the whole data, for the next they employed homes for a given interval of prices and for the third an equal number of homes were used.

The goal of this study was two folds – to determine if the ANN method produced better outcomes than OLS and if it yields a similar for the two giants in the software industry (NeuroShell and @Brain).

Unlike the previous studies, Worzala (1995) evidenced runs contrary. He reported varying results for the two groups of software he had employed in his estimation. A contrary view also exists in the literature on the usage of the ANN model for valuation of real estate. James and Lam (1996) evidences call for more effort at validating the real data set to ensure the validation of property valuation and appraisal technique.

A close look at the empirical literature shows that more of the previous studies concentrated more on housing data set and only a little attempt to extract data from other market source with close association with the housing market such as the job market.

2.2 Genetic Algorithms

This method is used when the goal is to find a robust solution to our research questions. This may include finding a suitable parametric variable and the best optimal forecasting structure of our time series data set.

The examination of this system has been applied to solving the challenges of home price movement and forecasting (Wilson et.al, 2004). The employed the GAs system mainly based on a non-linear approach to select and design dependence structure in the system. The nonlinear system chosen was known as the GT (Gamma test). This was used in 2 ways with a subset drawn from the data set using the GT for a selection of 8 economic time series variables and then followed thereafter by the estimation of the ANN prediction model using the GT estimation results.

2.3 ANFIS

Adaptive Neuro-fuzzy inference system (ANFIS) model has been implemented in many scientific fields including energy, stock market, robotic applications and many others. Loads of scientific articles applied this method to maximize their forecasting accuracy. ANFIS controller has been developed by Atsalakis and Valavanis (2009) with the purpose of short-term stock market prediction. Later on in 2011 Atsalakis introduced a model that enables prediction of trend of stock prices through applying Elliot waves theory and neural fuzzy systems. On 2010 Atsalakis applied Neuro-fuzzy models to time-series model in order to forecast wind energy production, the result of which compared with traditional prediction models.

Chapter 3

DATA AND METHODOLOGY

3.1 Data

The National association of realtors (NAR) calculated median and mean housing prices for the nation beside four census region on a monthly basis. Based on their findings the mean sale price doesn't balance with the median and it usually exceeds the median, the reason of which can be the nature of home price distribution. Despite the existence of some slight seasonal patterns of the sales price data, NAR found seasonal patterns difficult to match the model, the consequence of which was unadjusted seasonal data. To reduce home price instability (since the home price is non stationary) the only way was finding annual natural logarithmic differences in house price indexes to approximate growth rate which is,

$$r_t = \Delta_{12}\ lnP_t = \ln P_t\ - lnP_{t-12}$$

Where P_t is the median home price

Using census x-12 method, the data has been seasonally adjusted with levels. Figures 1 show the seasonally adjusted level of the median home sale prices and annual growth rate r_t in the four census regions in the US respectively. Our four census regions and states are include the Northeast comprising Connecticut, Maine, Massachusetts, New Hampshire, New Jersey, New York, Pennsylvania, Rhode

Island, and Vermont; the Midwest which covers Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin; the South covering Alabama, Arkansas, Delaware, District of Columbia, Florida, Georgia, Kentucky, Louisiana, Maryland, Mississippi, North Carolina, Oklahoma, South Carolina, Tennessee, Texas, Virginia, and the West encompassing Virginia; and West: Alaska, Arizona, California, Colorado, Hawaii, Idaho, Montana, Nevada, New Mexico, Oregon, Utah, Washington, and Wyoming. The analysis covers monthly data for the period 1968:01 to 2000:12 in the sample period (384 observations), and forecast the time period of 2001:01 to 2010:05 out of sample period (138 Observations) beside that, I compared ex-ante forecasts from 2010:06 to 2012:06. (24 steps ahead). The data correspond to an annual growth rate of median home prices, which were actually analyzed in the thesis.

3.2 Methodology

This thesis attempts to forecast house price in the US more specifically in four states by using the overall price of houses in America through the use of NAR, MLP, ANFIS, and Genetic Algorithm. These are used to do a 24 steps ahead out of sample forecast with the inside samples are from 1968:01 to 2000:12 and our out of sample forecasting period will fall between 2001:01 and 2010:05.

Using the four methods separately for the individual time period (24 steps ahead) will give different results. After computing the errors for each method, each result will be compared to the other in order to assess its level of accuracy. Thereafter, the simple combination method will be used to check the results as a combination by using the simple mean. Each of them contains a matrix with 138 rows and 24 columns also the mean for each is calculated to be able to use simple combination method.

3.2.1 Time Series Forecasting.

Every day there's concern about future behavior of surrounding phenomena rises that can be solved through understanding and knowing their structures and functioning. Forecasting the weather, the price of stock, the price of oil are all our favorite concerns. In a different approach study of each of this phenomenon in a numerical and quantitative sequence can help finding future values of them. These sequences regardless of their nature and structure can be analyzed through time- series. However a lot of data and information about the phenomenon can be used in time series analysis. In a time-series with "n" sample: $X_1,\ X_2,\ \ X_3,\ \ldots X_i,\ \ldots\ X_n,$ Future values are dependent on previous value. X_k=f ($X_{k-1},\ \ X_{k-2},\ \ldots X_{k-p},$) In linear time-series models such as ARMAX, AR, MA, AR there are different classical approaches that can forecast future values of time-series based on their past value. Therefore, forecasting nonlinear time-series needs intelligent tools such as neural network because the regression models are limited by their linearity. Regression models fall short in modeling non-linear relationships; therefore non-linear models have been proved to be better than regression models in their ability of modeling any i.e. both linear and non-linear relationships via logistic functions (Bishop, 1995). The latter models are expected to perform better than simple regression as far as regression is a linear technique operating via bringing non-linear problems in a linear format Albeit the mentioned better performance of non-linear systems their application is very complicated since the latter tries to fit all data and present noise encountered. Therefore, after making sure that any noise within the data absorbed and internalization of all necessary information is completed the processing must be stopped.

3.2.2 Neural Networks

This section introduces the neural networks, its history and application. Although the first Artificial neural nets-ANN has been applied to computational methods less than 50years ago, having sagacity and flexibility made the method an important factor in pattern recognition, clustering, modeling estimation, identification and forecasting as well as in neural training integral observations have made it a more efficient method. Samples given to neural network should be as refined and uniformed as possible. Neural network saves the observations in an inner parameter category. In fact changes in each observation causes inner parameter changes toward maintaining relationships between observations, so mostly neural networks react in case of Encountering with training samples even if they contain some errors. The neural network has some special characteristics illustrating its capacity and talent like the brain is sometimes so talented, progressive and successful and sometimes processes slow and not successful, neural networks in smaller dimension, depend on its inner structure can process differently. Therefore it might process successfully or weak so selecting a suitable structure based on field of problem is highly important. As success is based on effective training, learning the process of neural network is also dependent on the basic training situation and selecting the initial value of basic parameters is effective in getting the result of training. The main reason for using neural network in forecasting time series is non-linear, and ANN has the capacity of modeling the non-linear relationship without any knowledge about the relationship between output and input variables.

3.2.2.1 The Simplest Form of Neural Network

The system has a semblance with the Brain. It is made up of the interconnectivity of neurons packed in a layer that disseminate information from one to another.

A network in its simple nature consists of an input and output layer, operating in a manner similar to the input and output system. This system utilized the estimate of the neuron have inputted into the system to estimate the output value.

A graphical depiction of this network system is shown in figure 2 and each is depicted by a circle and the connectivity flow by an arrow link.

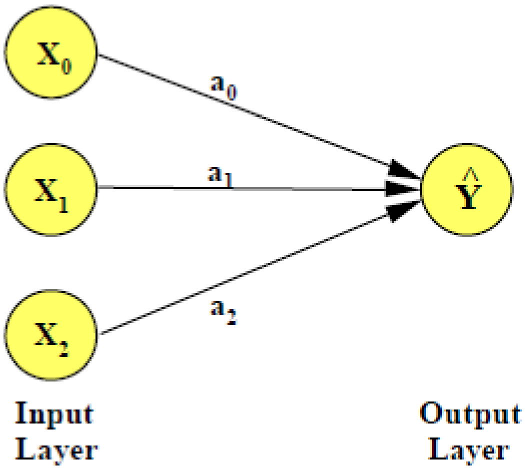


Figure 2. A Basic Feed Forward Neural Network

The output Y and the inputs X_0, X_1 and X_2 are n x 1 vectors and the observations are represented by n.

This illustration is defined as a feed-forward approach in the sense that the transmission of information is forward only the input to the output variables. An expression for the weights in the input-output connection is represented by a_i

This expression stresses the advantages of a given input in the estimation of the output variables. In the estimation of the output variable *t*, the value of each observation is multiplied by its weight for each connection. The total is obtained as;

$$a_0X_{0t} + a_1X_{1t} + a_2X_{2t} \qquad (1)$$

An activation key is thus utilized in the processing of the function denoted as *f (x)*.

In this process the identity in the feed-forward network based neuron system is the activating function f (x) = x.

In the regards, the estimate from (1) makes up the result of the observed t value of the network for the output value.

$$Y_t = a_0X_{0t} + a_1X_{1t} + a_2X_{2t} \qquad (2)$$

Naturally, a single input variable known as the bias has the same value for the total observations.

The network output is described as equation as given below, for a bias X_0

$$Yt = a0 + a1X1t + a2X2t \qquad (3)$$

Generally, various studies use a targeted value for output (Yt) for which the system attempt to estimate from the estimation procedure for the corresponding input value.

The forecast error obtains for the individual observation is estimated as yet less it.

From the foregoing, given the algorithm iterative process the popular of which is called the back-propagation, the assigned weight of the network system is restructured until it yield a low value for the absolute or sum squared error in the whole sample.

This process yields alternative weight as it is repeated over time and the learning process in the network is simulated. Describe the process above can be liken to the OLS system given the identification activation key in the dual strata feed-forward system. In this analysis, the exogenous variables are the input neurons while the endogenous variables are the output neuron respectively.

The coefficient of the estimated OLS regression is the assigned weight given to the networks while the bias is the OLS intercept.

The bias term in this model for the network output is specified as below;

$$Y1 = a01 + a11X1 + a21X2 + a31X3$$

$$Y2 = a02 + a12X1 + a22X2 + a32X3 \qquad (4)$$

A system of equations, linear in expression is derived from our irregular regression equation (*à la* Zellner). Given our sequences of series data a neural equivalent of a VAR model is obtained from the network system with lagged values. In a similar fashion given the linkage between Y1 and Y2, a simultaneous equation system is derived.

3.2.2.2 Nonlinear Activation Functions

The foregoing illustration is based on identification of the neuron output activating function. To fully appreciate the merits of the neural network systems is it imperative to employ an activating function with nonlinear relationship. Wide ranges of network system that are neutral make good use of the functional relationship that is nonlinear. This allows for the reproduction of data set with a nonlinear organization. Thus, it is the case that in order to facilitate the optimal utilization of the algorithm system for the determination of the correct weight, an activating functional relation that is continuous, with differential and monotonic relation is necessary.

$$F(x) = \frac{1}{1+e^{-x}} \qquad (5)$$

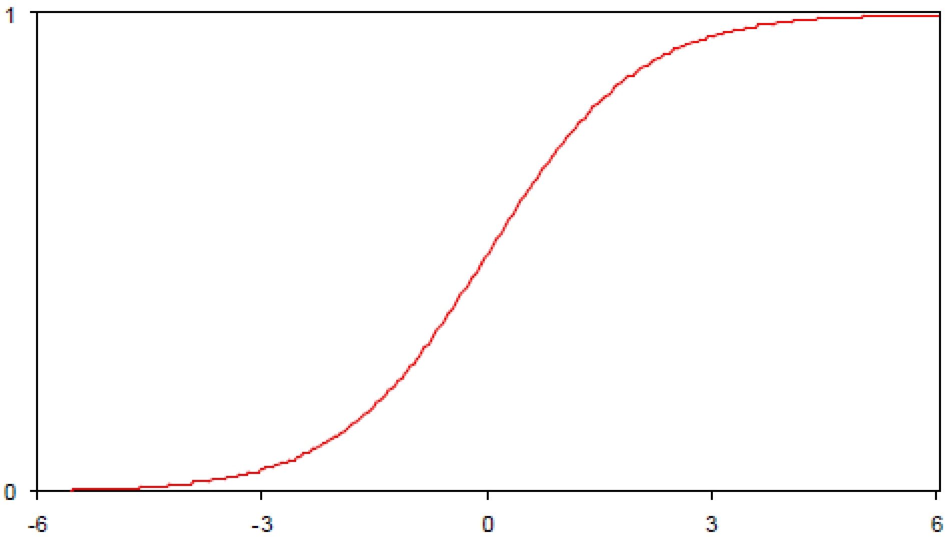


Figure 3. The Logistic Function

A mostly applied function in this regards is a logistic function with a cumulative distribution of the neural network system. This function has a boundary with ranges 0 and 1 as depicted in Figure 2. Thus, by using a boundary with ranges of values, the aim is to recreate real neurons using the activation function. With a value of about 1, the neuron generates a higher activity level given the transmitted sign conveyed to it and with 0 value it simply react to the conveyed signal to it.

It is important to note that for a non positive value of the forecast variable a hyperbolic tangential function is to be used the activating function. This class of function has a semblance with earlier describe function i.e. the logistic function. Unlike the former, it has boundary ranges of -1 and 1

From the analysis of the network system in Figure 1 Described as the feed-forward networking system, an activation system with logistic function is essential. This gives rise to network system with semblance as the Binary Logit probability model.

It is also possible to derive a binary probit model if our cumulative distribution functions are normal for the activating function. The usage of other boundary function is capable of yielding network systems that can easily address the problem of nonlinearity for a bounded dependent variable. Given an unbound variable, we may decide to use an activation system that is equally unbounded .E.g. $f(x) = x^3$.

Although, leading studies in the field favored to maintain an activating function is bounded rather than unbounded using the network structure with layers hidden.

3.2.2.3 Neural Networks with Hidden Layers

The system that had been mentioned previously has a dual strata linkage of input and output structure. This system is described as a complex system for application to practical analysis. This structure is built to include a single or double stratum as shown in Fig 3.

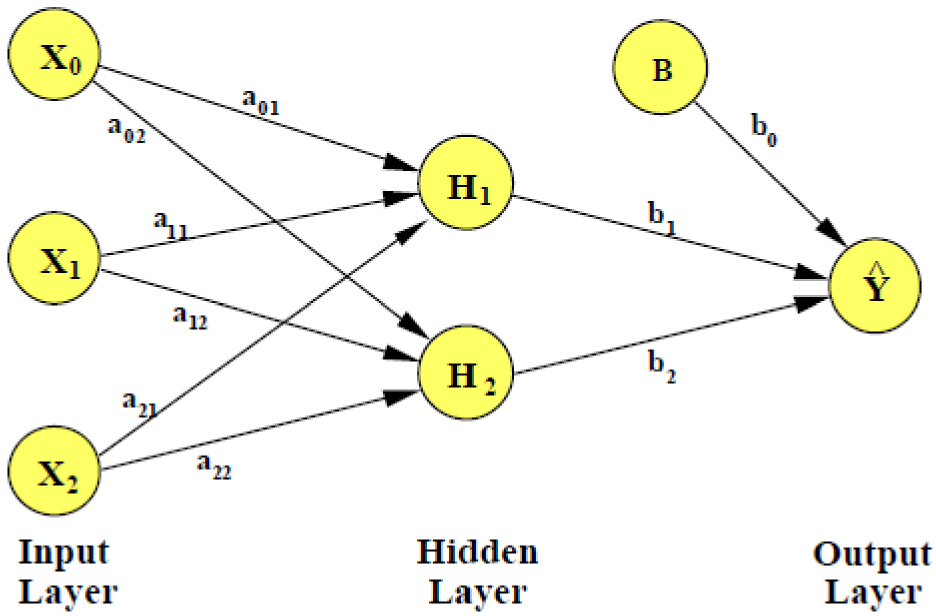


Figure 4. A Feed Forward Neural Network with One Hidden Layer

Here the connecting weight of the input and hidden factor is denoted by a_{ij} . The term which serves as the intercept-the bias term is assumed to be X_0 while that that captures the hidden term unit for the output is assumed to be B. Unlike the unit in terms of the input and output, the expression of the hidden term does not apply to any apparent concept and to direct implication is attributable to them. They only represent the product of the estimation of the values for the output units. Thus they do not have any implication in econometric analysis.

The hidden term display similar result as the output term for the units. That is, they both provide the estimate for the weighted total outcome for the variable in the input unit, processing this outcome with the utilizing the activating functions. Here the case to note is that the outcome of the network is not restricted to the hidden unit as against the output given that the activating function from the logistic function is applied only to the hidden units. This means then that the network will not be restricted to only the estimation of the variables assumed to be bounded.

For an unbound variable which is not dependent, an identification function is used as an activating function. This means that this output will be the estimate of the weights of the sum total of the value of the hidden units, denoted the weighted coefficient b_j our hidden unit term is assumed to have a very essential impact on this analysis. Many studies indicate that a system of network with 3 neural features in its layers having an activating function of the hidden unit and at the same time, a logistic function can be safely used as a general approximating function. Thus, with a given amount of units for the hidden term, it is possible for the system to yield level of estimation for both a nonlinear and linear functions. This is an advantage for a sequence of data generation activity that involves a complexity of nonlinear function. Thus, this complex non-linear function for system of neural networks is capable of being applied in economic research covering a wide range of variables as exchange rate relations, growth in employment levels and growth rate of a country real GDP.

Additionally, an understanding of the data generated is not essential for such estimation as is the situation for the traditional regression for nonlinear function It is crucial to note that no exact number of layers for the units in terms of the hidden term in a network is theoretically mentioned. However, from our analysis a larger amount for the hidden term unit logically suits the feature of the universally adopted approximation. But an addition of too many a term makes the system threatened by an over fitness syndrome in our data. The implication here is that system may be accurately suitable for forecasting for the estimated window; however it may perform badly for forecasting out of sample, Furthermore the addition more hidden terms for the unit variable increase the model estimation period.

Finally, the system design often involves a series of rigorous testing until the appropriate model is determined with a good capability for accurate forecasting.

The next section will provide an explanation regarding the criteria for error selection and processes.

3.2.2.4 Estimation of Network Weights

A variety of processes called the iterative algorithms are very carefully used in the estimation of the weight in a network system. This process is described as the back propagation algorithm. It is the case that this system of the algorithm performs poorly due to their slow performance rate (Sarle , 1994)

However, an optimal outcome can be easily obtained via much faster system utilizing a standard algorithmic system. An e.g. include the algorithm system used for a nonlinear regression analysis. It is also the case that for every day usage, the algorithm system for neural networking are least important and do not require tedious training for applying to the back propagation method (Sarle, 1994).

In the light of the above, a detailed description of the algorithm neural network is out of the scope of this thesis. A network analyst for the neural system often has his date set in two folds; a training set and a test set. The first data set, use the algorithm system for the estimation of the weight assigned to the network and the second data set does the forecasting evaluations for the system. This result yields an out of sample forecast ex post. The mean square error term minimizing the forecast error is utilized here as an error selection criterion.

3.2.2.5 Early Stopping

Empirical test indicates that neural system is frequently susceptible to over fit as regards the training data set leading to the dismal performance out of sample. To arrest this shortfall, a wide range of processes is followed. The Early stopping is one essential process followed in this regards. This process includes a three step procedure; the first part involves the set of training steps and validity set for the forecast accuracy. The training set estimates the weight of the system to be used for the forecast out of sample. The validating set consists of the data set not employed in the training but ensure the accuracy of the forecast out of sample for the system.

Given a series of estimation, a forecast is attempted out of sample via the aid of the validating set and mean square error is estimated. This process is described succinctly in Fig.4.

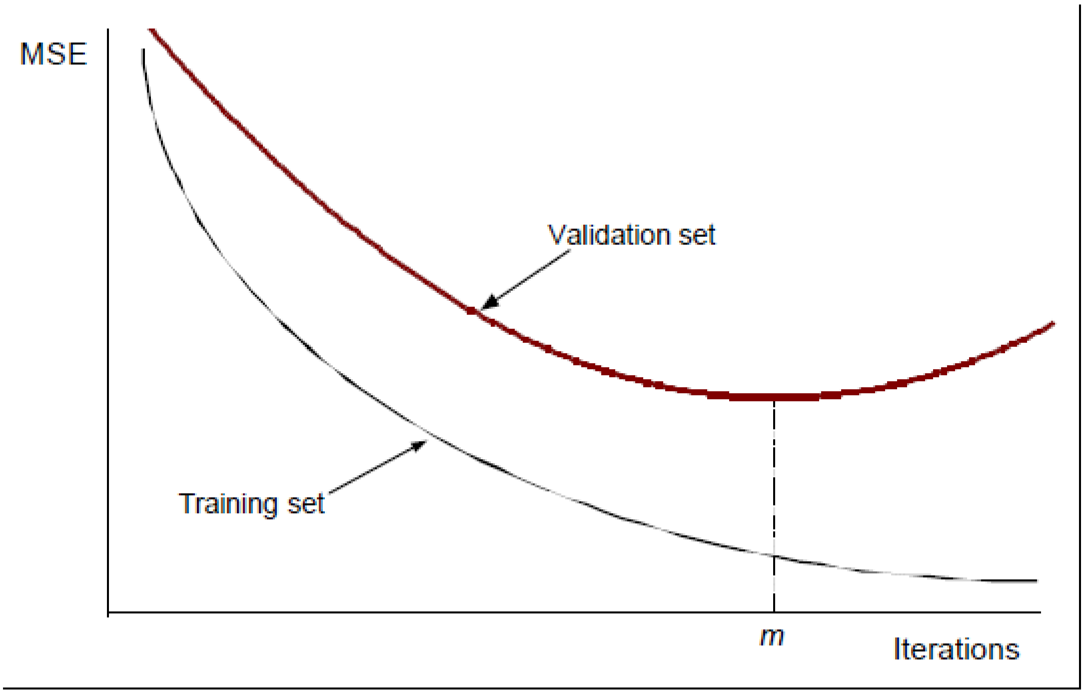


Figure 5. Early Stopping Estimation

With a constant repetition of this process the estimated error in both the training data set and validating data set reduces. Empirical evidences show that after a whole series of trail the estimated error tends to increase as the system become familiar with observation data set and thus fails in its generalization to another data set. This process is generally completed when the forecast error is at its minimum in the validating as against the training set. This is the case after a series of **m** repeated processes (Fig.5).

This serves to prevent the system from familiarization with the data set of training data and enough it potentiality to for forecasting out of sample. This approach is a new development in as econometric analysis does not utilize information for estimating coefficient out of sample for forecasting. Rather such forecasting is done at another level after coefficient estimation.

The advantage of this model is its orientation towards accurate forecasting results. Thus the early stopping processes yield an optimal biased estimate which is not an unbiased estimate for the forecast validity of the model in population

For unbiasedness in the system, the test set data must be employed in an out of sample forecasting. Thus, to ensure the accuracy of forecasting out of sample in comparison with the OLS the sample utilized such a task be excluded from both the training and validating set data of the system.

In spite of its advantages, this system is regarded as less efficient as it does not utilize all the data set contain in a sample rather it uses just the training data set to the estimate the weight for the sample observations. Further, for a few sample set of data, the subdivision of the data into various sets such as the training, validating and testing data set respectively results in fewer elements in each data set to ensure the reliability and accuracy of the estimate statistic.

In the final analysis, the estimated statistic may be prone to arbitrariness in the selection of the data from the observations. Finally, despite its potential shortfall, the early stop approach is utilized in a plethora of empirical literature and as help the development of network systems with reliability and accuracy.

A very general type of ANN which is called MLP is stipulated below.

3.2.3 The Multi-Layered Perceptron (MLP)

Application of Artificial Neural Networks covers a wide range of practical problems, such as, classification, noise reduction and prediction (Masters, 1993). Two kinds of

ANN are available one of them being supervised and the other one unsupervised network, differing for the methods of training. As it is obvious from the name the unsupervised network needs no supervisor during training i.e. sample inputs are present without any relevant outputs. Mostly classification problems are addressed via the latter networks. Such networks are often used in classification problems (Masters, 1993), in the supervised networks however, inputs and relevant to the latter outputs are used together. The multi-layered perceptron (MLP) network might be a good example of the supervised learning network. The main advantage of MLP networks is their capacity of management of nonlinear functions. The success of MLP networks proves through a large number of studies of forecasting house price indices. The following figure illustrates an example of a typical MLP.

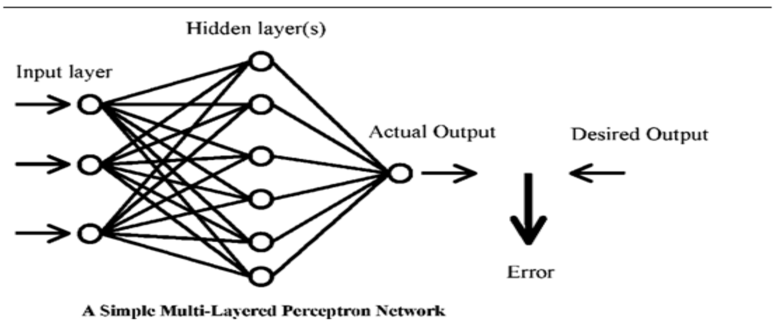


Figure 6. Typical MLP

Interconnected layers of neurons or nodes are the main particles of the MLP in figure 5 Depending on the characteristics being modeled we determine the number of the input layer nodes. The error got through the comparison of actual and desired output is the main measure for us for defining the weights of each processing node. The aforementioned is what is called a feed-forward back-propagation ANN. It is a complicated task to establish such a neural network albeit good forecasting performance. The following issues should be taken into account while modeling: first of all, the number of layers and the number of nodes in those layers, and that of arcs which grants interconnection should be determined. Moreover, activation function for the processing nodes, the training algorithm, data normalization methods, training data and performance measures are issues that should be considered very attentively. (Zhang et al, 1998). An ANN consists of an input and an output layer, and also one or more hidden nodes layers. ANN model building and predictive ability are significantly affected by the number of inputs and hidden nodes. However, one should consider that the number of input nodes has greater effect than that of hidden nodes on forecasting (Zhang et al, 2001). The number of variables in the input vector is a means of determining the number of in out layers nodes. The method used for univariate time series forecasting, is a series of lagged observations operating under the assumption that future values are related in some way to the series that precedes them. The number of consecutive data points used defines the window size. In multivariate forecasting, however, the number of indicators determines the number of inputs (leading indicators or factors believed to influence the outcome)." An input vector may contain windows from more than one time series and a number of indicators, with the number of input nodes corresponding to the total number of observations used. In order to approximate any complex non-linear function with any desired accuracy that a single hidden layer has been proved to be sufficient for an ANN " (Hornik, 1991), in spite of the higher efficiency and more compact of architecture of two hidden layers than a single layer (Chester, 1990; Srinivasan et al, 1994; Zhang, 1994). Trial and error determined the number of hidden nodes in each layer. As I mentioned the Multilayer Perceptron neural network with back propagation is one of common practical neural network. In theory it's been proven that MLP, given right internal structure, can simulate any non-linear systems.

Structure of neural network contains numbers of activity function Perceptron in various layers. Each Perceptron using their weight coefficients gets previous layers outputs and send it to the next layer. Neural network contains an output layer, an input layer and minimum one hidden layer and numbers of Perceptron in each layer depending on structure of network is different. Computational algorithms in back propagation have high diversity and performance. In simplest algorithm, weight coefficients minimize the objective function of network, which is slop of output error. Therefore in each training process weight coefficients are as below:

$$w_{k+1} = w_k - a_k g_k$$

Where w_k stands for weight coefficient of network, g_k is gradient of network's output error and a_k is learning coefficient of network. This method that also known as a gradient descent algorithm is applied in two ways: incremental mode and batch mode. Incremental mode performs after complete observation of training samples and computing weight coefficient while in the Batch mode functioning process starts after observation of training samples and computing total gradient. In most cases the incremental model performs better than Batch mode. Beside methods introduced earlier, several modern computational methods have been offered objectives of which is to reduce and speed up calculations. These methods are also called faster training methods and they divide into two groups: heuristic technique and numerical optimization techniques. With Heuristic technique training coefficient changes in the teaching process, this coefficient in different stage of teaching become smaller till we get minimum gradient in less time. In Numerical optimization techniques through applying computational methods such as Conjugate Gradient, Quasi Newton or modern method of Levenberg-Marquardt, that increase the speed of computation to reach gradient error and at the same time reduce the mass of computation.

3.2.4 Recurrent Neural Model for Multi-Step Prediction

The neural model we employed in this work is an alternative of classical neural model the aim of which is to forecast the future behavior of some prediction horizons. Throughout applying this model we became able to constrain particular learning stages that had long-term forecasting aim. The recurrent model is based on a partially recurrent neural network. The network includes sending feedback connections to a multilayer feed forward neural network by the output neuron to the input layer. The value of the prediction horizon defines the number of recurrent connections. If the horizon is (h), then a bunch of (h) neurons will make the input layer of the network. These groups of (h) neuron's (also called context neuron) duty are memorizing previous network output. Capturing the basic or measured time series data is a duty of remaining neurons in the input layer. Figure 1

On the number forecasting horizon (h) become more than the amount of external input neurons (d+1), whole input neurons of the network come to be context neurons, therefore no measured time series value is going to inject into the network.

Based on multi-step time series prediction, training method of partially recurrent neural network found, as what you can see below: At each instant k+1 outset with k=d, STEP 1:

The neurons in the input layer obtain the measured series x (k)…., x (k-d). Therefore the amount of context neurons, those that memorized the former network outputs, is zero. And network output is given by:

$$\tilde{X}(k+1)=\tilde{F}(x\ (k),.\ \ldots..,\ x(k-d\ ),w) \qquad (5)$$

STEP 2:

Increase in amount of context neuron by one unit, x (k +1). So the prediction at moment k+2 is given through:

$$\tilde{X}(k+2)=\tilde{F}(\tilde{x}(k+1),x(k),\ldots.X(k-d+1),w) \qquad (6)$$

STEP 3:

Assuming step2 continued till (h) context neurons achieved. When we reach instant k+h+1, the output of the recurrent model will be:

$$\tilde{X}(k+h+1\ )=\tilde{F}(\tilde{x}(k+h),\ldots.\tilde{x}(k+1),\ x(k),\ldots.x(k-d+h),w) \qquad (7)$$

STEP 4:

In this stage the parameter set of the recurrent model, w3 is updated to constrain a training stage aiming long-term forecasting, the learning is based on the summation of local errors and prediction horizon. For instance: along the interval [k+1, k+h+1]. Therefore the parameter set w3 is following the negative slope of error function given by:

$$e(k+1) = \frac{1}{2}\sum_{i=1}^{h}(x(k+i+1) - x^{\sim}(\ \ k+i+1))^2 \qquad (8)$$

We can find out training throughout the traditional back propagation algorithm because the internal structure of the partially recurrent network is like a feed forward neural. However other propagations of the algorithm should be conceivable.

STEP 5:

This step is where the process goes back to stage one, by increasing one unit time variable (k). This process continues till our instant reaches k=N-h, where N stands for number of patterns.

Structure of recurrent model and feed forward models is almost the same, with just one distinctive feature that is: the approaching way of parameter sets in these models. This is the process of learning of the system. The parameter set **w** of the feed forward model obtained training a multilayer feed forward network and it doesn't change during the forecasting procedure. Meaning parameter **w** is updated via measuring local error in each moment.

When we use recurrent model measured error during prediction interval [k+1, k+h+1] is responsible for keeping parameter updated at each moment. Likewise the set of parameters w has been specified to minimize the error in the future (eqs.8). As a conclusion, we can say recurrent model is built in a way that its function is multi-step forecasting, while feed forward model is trained particularly for one-step forecasting, meaning it’s only able to predict next sampling time. Recurrent model distributes errors that occur at moment to next sampling time, but in recurrent model (distributed error decrease during training steps). Since learning is done via predicted output at very early time steps, so correction will be done in each stage and we can expect better forecasting for the future. MLP is inherently feed forward as it was mentioned however through special adjustment of the loop it was made available to work like a recurrent neural network. In this type of network one hidden layer is used (10 neurons), and because of the need of comparison of the results of this network with the results of other methods (GA) 2 nodes (two lags) are used for input and 1 node is used for output. Moreover, Levenberg-Marquardt is used for back propagation (trainlm) as training function. The MLP has been adjusted in the loop for forecasting 24 steps ahead and a recursive network has been built for forecasting since MLP is not recursive. The schematic of MLP has shown in the figure 6.

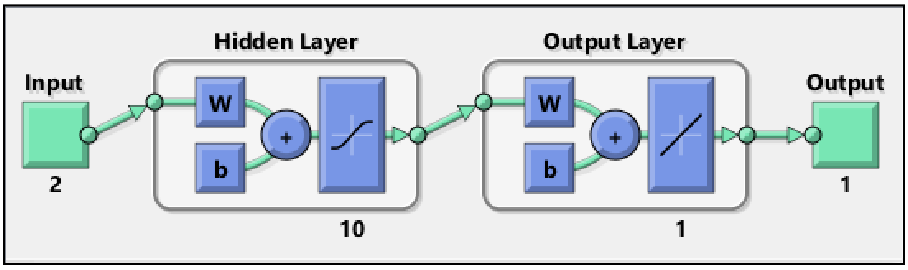


Figure 7. The Schematic of MLP

3.2.5 Non-Linear Autoregressive Model (NAR)

A non-linear autoregressive model of order P will be given by the following input-output relation

$$y(t) = f((y(t-1), \dots . y(t-b), d(t-1), \dots . . d(t-b))$$

Where d are the targets for the time series that we want to predict, y shows the past predicted values by the model, b is the output order, and f represents a nonlinear function. NAR (p) stands for Non-linear autoregressive model of order (p) while f stands for an unknown function of applying a NAR model. It is possible to use values from “t – 1” to “t –b since the intension is to produce predicted values at the current time. It is shown that b shows the number of past predictions fed into the model. In this case since two lags are used, it implies that two past predictions are used. The real values of the time series that are desired to be forecasted and fed into the system are represented by the targets d. For the past predicted values the same order is used

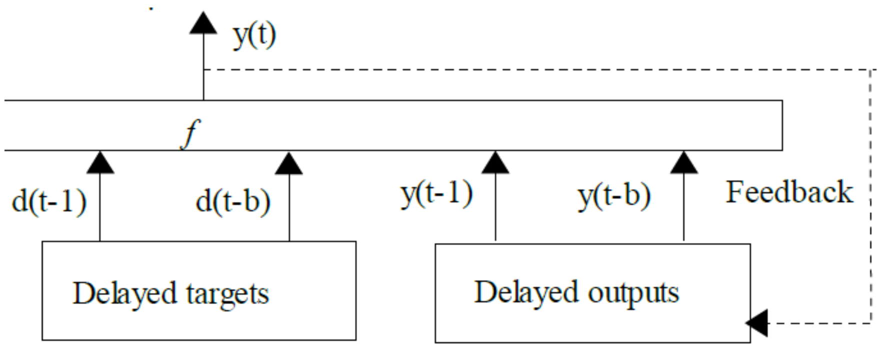


Figure 8. NAR Model

Given the network’s output at time t, the forecast is y(t), and since we have the targets d(t) we can figure out the error e(t) as the difference between d(t) and y(t).v an interesting fact of the NAR networks is that in any given problem can be modeled and solved using a NAR network. A NAR network theoretically can substitute any persistent network with no power loss of the level of calculation accuracy. A key issue with recurrent neural network is their shortfall in representing long run dependencies since as “n” step increases in some case the quantity of errors diminishes. Although this issue can be surpassed using NAR, it is worth noting the NAR differs from the AR model since it predict the next step depending on its past value rather than the actual data used in ARIMA. In addition, a nonlinear structure can be possibly used regardless of the simplicity of the AR model used in case of linear structures.

Many solutions have been suggested to solve the problem of vanishing gradient in training RNN's with most including embedding memory in neuron networks while others proposed the improved learning algorithms Newton's type algorithms, annealing algorithms. The embedded memory can help to speed up the propagation of gradient information which in turn reduces the effects of vanishing gradient which includes the formation of spatial representation of temporal pattern, installing time delays into the neurons or their connections, use of recurrent connections, use of neurons with activation that summed up inputs over time. The topology of NAR is different from that of the MLP, and this type of network does not need any adjustment for putting in a loop because it has a loop during the training.

3.2.5.1 Learning Algorithms

A dynamic propagation algorithm is used to calculate the gradient. It is usually more intensively computationally than static back propagation and takes more time. Dynamic network errors can be more complex than those for static network. In some training method there are some advantages of availability at training time of the true output set. By using the true real output set and not estimating output to train the network with a decoupled network has a feed forward architecture which can be trained with a classic static back propagation algorithm in training feedback connection decoupled. Some of the hardships faced by training programs are related to number parameters, which how many connections or weight are contained in the network. The number is large and can lead to over training the data and producing a false fitness that may lead to better forecasts. This leads to the use of an algorithm including the regularization technique, which involves modifying the performance function for reducing the value. With the new performance function leads to the network having a smaller weight and biases therefore leading to the network respond to be smoother and not over fit. The network training function which usually updates the weight and bias values are usually modified to include regularization technique, this is according to Levenberg Marquardt optimization. "With this method there is a reduction in the combination of squared errors and weight and determines the correct combination so as to produce a network which generalizes well. This process is referred to as Bayesian regularization. The function approximation problem for networks that contain up to few hundred weights the Levenberg Marquardt algorithm will have the fastest convergence" (Andrzej 2002). It is an advantage as very accurate training is needed. As number of weight increase, the advantage of algorithm decreases. The neutral training is more efficient it some pre-process steps are performed to the networks input and targets are performed. By normalizing the input and target value means we map them in the interval (-1, 1). By this, it simplifies the problem of the outliers for the network. Normalization input and target that are returned will fall in the interval (-1, 1). NAR has a loop by itself, and it is a recursive network which uses output as an input in the next step. The architecture of NAR has shown in the figure. 8. In this study NAR with 2 lag as feedback delays is employed and with one hidden layer with 10 neurons and after that the result of each part of the data put into the loop for finding the 24 steps ahead forecasting, in the form of univariate and recursively

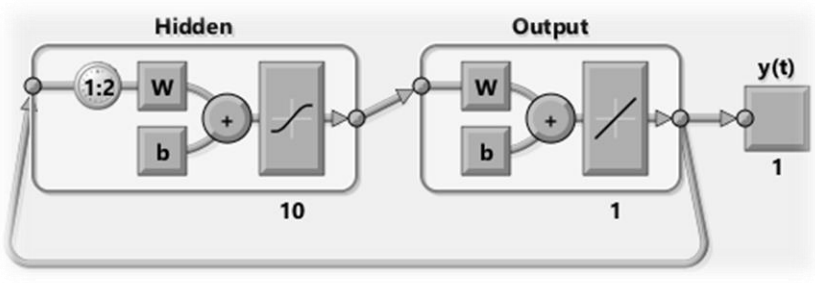


Figure 9. NAR Feed Back (Loop).

3.2.6 Adaptive Neuro Fuzzy Inference Model (ANFIS)

This model expanded by Yang in 1996 and it enables fuzzy systems with training parameters to use the training algorithm for back propagation error (Morgan, 1998). In the structure of ANFIS they used IF-THEN, TSK type of fuzzy rule that can be used for modeling and mapping input and output data. Normal description of the model is recognizing dependent f^ so that it can be used almost instead of actual dependent function.

3.2.6.1 Learning Algorithm in Neural Fuzzy Systems

Training of neural fuzzy systems can be divided into two main groups: stipulation of the structure and stipulation of (estimation) parameters. Stipulation of structure refers to finding suitable amount of fuzzy rules and a right partitioning through input output space. But the stipulation of parameters depends on adjustment and regulation of system's parameters such as member functions and other parameters. There are some different training algorithms which we now elaborate.

3.2.6.1.1 Training Algorithm after Linear Propagation

This method is the most common training way in neural networks. Normally primary fuzzy set and the rules will be defined by the scientist. Then the rules of the game will be improved via a gradient descent algorithm. Having slow convergence near minimum areas is the biggest weakness of the method after propagation of error. It also can be trapped in local minimum

3.2.6.1.2 Adaptive Vector Quantification

This method was introduced by Kosko and Kong where at the beginning input and output data divided into overlapping fuzzy sets using clustering scheme are sorted. Then based on the best relation between fuzzy sets, input and output variables of the fuzzy rule set will be developed. Fuzzy sets in this model are not adjustable and this is one of the weaknesses of the method.

3.2.6.1.3 Consolidated Training Algorithm

ANFIS is a sample of consolidated training algorithm that uses modulation of two methods: least squares - minimizing gradient; to modify and adjust the structure of fuzzy neural system.

3.2.6.1.4 Orthogonal Least Squares Algorithm

Wang and Mendel used this method to determine the parameters of a fuzzy system with fuzzy function. In neural fuzzy systems or fuzzy function, first a fuzzy system based on various fuzzy functions will create pairs of inputs and outputs. Then the most significant function will be defined through least square. Even though training in neural fuzzy systems is different in function, there is no detailed comparison about the function and advantage of each one in case of training speed and approximate non-linear function.

One of the most general types of the Neuro _fuzzy system is ANFIS that the training algorithm for this type of system is CONSOLIDATED TRAINING ALGORITHM pioneered by (Jang1996)

ANFIS is a multilayer neural network-based fuzzy including layers where the training & forecasted values exposed through the input and output nodes that operating as a membership function (MFs) and rules are presented in the hidden layers. You can find the topology of it in Figure 10

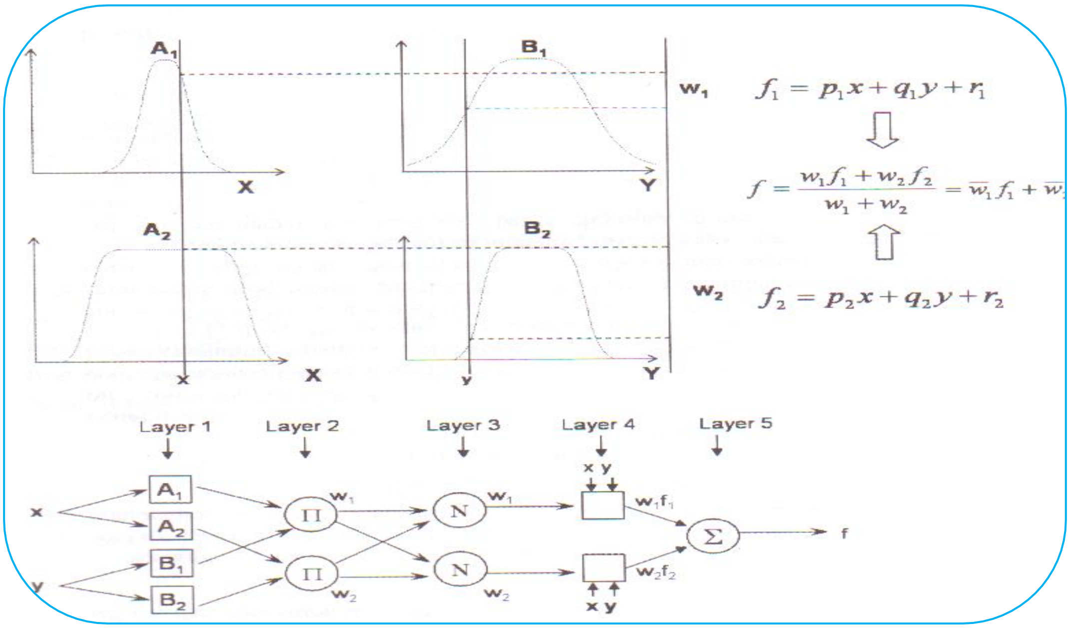


Figure 10. ANFIS Architecture and Sugeno _Type Model (Jang,1997)

The ANFIS model utilizes a hybrid learning algorithm. It uses a combination of the least square method and the back-diffusion gradient. It uses the fuzzy inference system (FIS) membership function parameter in a bid to associate it with certain training data set. To optimize the fuzzy sets of the premises, a back-propagation algorithm is used to obtain linear coefficients in consequent term; a least square process has been employed. To check if the data fits in the model, a data testing set is applied and to reduce error between the target output and computed values, the parameters involving the (MFs) membership functions must be dynamic. The fuzzy inference system is assumed to comprise two inputs and a single output for easy comprehension. In this thesis, the specifications that are used in ANFIS are as follow

Table 1. ANFIS Specifications

Number of membership functions	**Type of membership functions**	**Function**
3	**Gaussian curve built-in membership function (gaussmf)**	**Genfis1**

Three membership functions are used therefore we have 9 rules that are shown in figure 11.

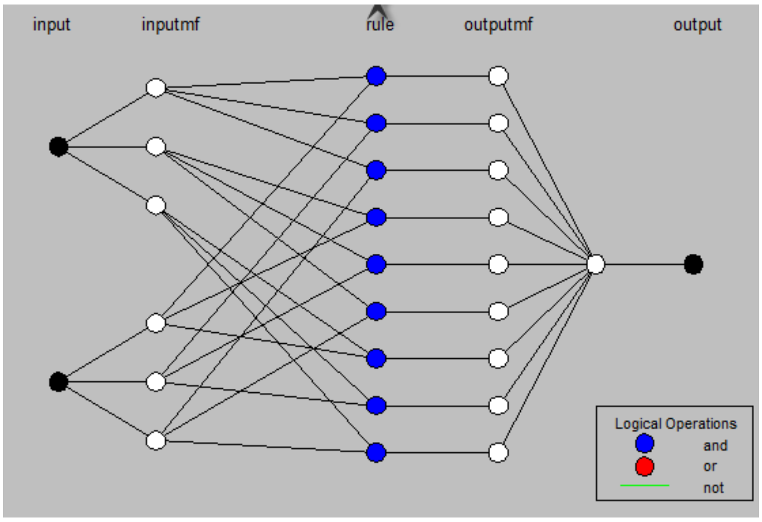


Figure 11. ANFIS Structure

For generating ANFIS, I used genfis1 (Generate Fuzzy Inference System structure from data using grid partition). The genfis1 generates a Sugeno-type FIS structure which is imposed as an initial condition (initialization of the membership function parameters) for ANFIS. traininggenfis1 (data) generates a single-output Sugeno-type fuzzy inference system using a grid partition of the data (no clustering). The genfis1(data, numMFs, inmftype, outmftype) generates a FIS structure from a training data set, data, with the number and type of input membership functions and the type of output membership functions are explicitly specified. The logics for genfis1 are as follows: the training data matrix should be entered together but the last columns representing input data, and the last column representing the single output. numMFs is a vector that determines the number of membership functions related to each input. If the aim is to associate the same number of membership functions with each input, then the numMFs is specified as a single number. The inmftype signifies a string array in which each row specifies the membership function type associated with each input. The outmftype is a string specifying the membership function type associated with the output. There can only be one output, because this is a Sugeno-type system and it must either be a constant or linear function. The number of membership functions associated with the output and the number of rules generated by genfis1 are same. The results of trials relating to the type of output functions show that having a Gaussian function is preferable. The two types of ANFIS are mamdani and sugeno but this research uses.

“Sugeno, or Takagi-Sugeno-Kang, method of fuzzy inference introduced in 1985 is similar to the Mamdani method in many respects. The initial two parts of the fuzzy inference process i.e. fuzzifying the inputs and applying the fuzzy operator, are exactly the same. The major difference between Mamdani and Sugeno is that the Sugeno output membership functions are either linear or constant. On the other hand an evident comparison between the two is that Sugeno is a more compact and computationally efficient representation than a Mamdani system. The Sugeno system lends itself to the use of adaptive techniques for constructing fuzzy models .These adaptive techniques can be used to customize the membership functions so that the fuzzy system best models the data" (www.Mathwork.com).

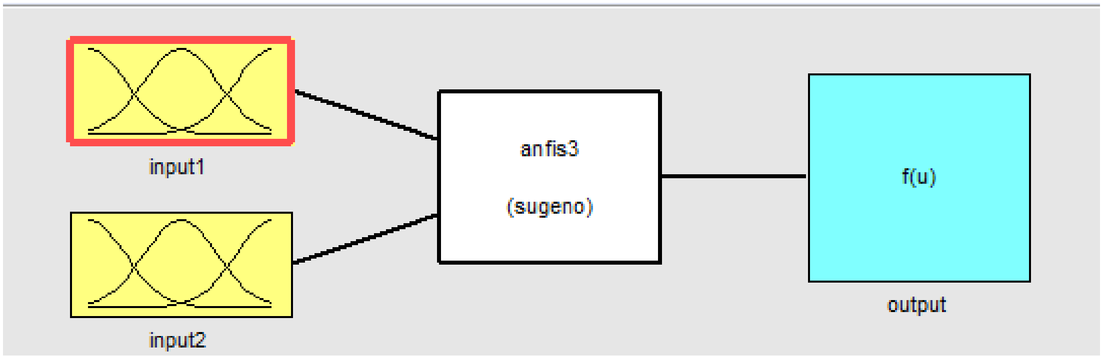


Figure 12. ANFIS Scheme

3.2.7 Genetic Algorithm (GA)

Genetic algorithm defined as a feature of human brain and has been used in forecasting processes as a tool of numerical optimization, the general Algorithm (G.A) which was pioneered by Holand (1975) and elucidated by GoldBerg (1987). According to Dorsen and Mayer (1995) compared G.A with other conventional algorithms to estimate models with the known specification model. Also Schmertmann (1996) is illustrated in his work the use of GA to estimate a non linear model parametrically and non- parametrically, from the synthetic data when there is a known covariate and the function is usually known to the researcher. According to Farley and Jones (1994), they use the GA to select covariance as a leading indicator model of economic activities in the U.S while Varetho (1998) was able to use G.A to predict bankruptcy in Italy.

GA has been used to estimate time series models that are the main focus of this article. Accordingly, Koza (1997) used time series data for money output and the price level in U.S. Szpino (1997) was able to use GA to estimate the dynamic models of unemployment and personal income from the time series data and the estimation model of wages from the cross sectional data. In order to estimate of personal income, he used one explanatory variable with fourth lag while in the unemployment, there are several explanatory variables but the lag is assumed to be the first order.

The econometric application as a whole is not being related to classical themes such as model misspecification, spurious correlation and nested and non-nested hypothesis testing. There has been a high concentration as a non Paramedic tool for describing data as a whole. According to Leamer (1978), he shows the difference between the several different aspects of specification search. The latter includes post data model construction in which the additional covariate are more considered interpretive research which has no restrictions imposed and reduces the number of parameters in the simplification search. According to Leamer he proposes the Bayesian decision making framework for the selecting models, in which also GA covers similar ground by using a Darwinian framework that maximum fitness is achieved when parsimony is set at a premium. The main focus of this paper is to use algorithms that generalize to nonlinear models estimated from the non stationary data. With this paper it has agreed and disagreed with other works with regards to specification search algorithms proposed by Hoover and Perez (1999). The main aim is to ensure that the algorithm doesn’t get stuck at optimum. This thesis looks at the GA specification as an optimal function form and the search of parameter which is globally optimal is looking at these issues and treating them simultaneously showing the curse of dimensionality. GA not only must it estimate non_ parametrically the interaction between the covariates, but also must choose the relevant covariates as well as their lags at the same time selecting between different functional forms to add to that. It looks out for the global optimum of the relevant fitness criterion, hence because the data generation process is generally unknown and has to be discovered through specification search. With GA, it tends in probability to select stronger dynasties in proportion to their fitness and leaves weaker ones to disappear in probability GA also selects a superior dynasty the best is not always selected and worst is not always excluded. GA uses the principle of selection, crossover and random mutation check whether new and better strains and variety models specification and crossovers it fitters the strains to either float or surface, while mutation prevent GA from involving with inferior solutions. It carries out this task efficiently. It carries out this task efficiently because GA doesn't precede model by models in an exhaustive search. According to Mitchell, who summarized GA's efficiency in numerical optimization found that at any level of generation, when GA is explicitly measuring fitness of the "n" strings in the population, it implicitly estimates the average fitness of a much larger number of schemes.

The fifth and last objective is that of comparing results obtained blindly from GA with those from painstaking applications of conventional specification research. The main suggestion is that GA is that GA serves a useful and efficient technique for exploring the specification space and as a practical aid to build better econometric model.

3.2.7.1 Genetic Algorithms as Optimization Procedure

If we consider maximization problem in which target function F of N arguments θ_n is not available to reach a simple analytical solution, we need to apply the numerical optimization procedure. Traditional maximization algorithms such as one for Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm, repeats an argument of optimizing the target function (F) by (1) estimating the curvature around the candidate and (2) using this curvature to find the optimal direction and length of the change of the candidate solution. This type of algorithm is also called “hill-climbing” algorithm and it’s very effective in its use of the shape of the target function. The algorithm fails where we have ill-behaved target function. By ill-behaved we mean they’re in non-constant or almost flat around optima or they are having breaks and kinks. Another weakness of BFGS is that it can’t perform well where the problem has many dimensions. While BFGS is unable to function well in large dimension problems, Genetic algorithm uses totally diverse approaches which can help them handle different types of problems with bigger diversity. The most important concept is that we should pay more attention to the population of the arguments that only compete in their relative functional value. This competition is known to be evolutionary where mutation operators allow for blind search experimentation but the functional value refers to the out last of a particular candidate during a time. So the target function can be as general as necessary and argument can cover any kind such as members, integers, probabilities or binary variables. The only limitation for arguments is that they must fall into certain defined bounded interval a_n, b_n

3.2.7.2 Binary Strings

Genetic algorithm (GA) uses H chromosomes $g_{h,t} \in H$, that are binary strings divided into N genes $g^{n}_{h,t}$ each of them encodes a candidate parameter, $\theta^{n}_{h,t}$ for the argument θ^{n} when a chromosome h ∈ {1,…H} is at time t ∈ {1,..., T} it can be determined as:

$$g_{h,t} = \{g^{1}_{h,t}, \ldots\ldots, g^{N}_{h,t}\} \qquad (9)$$

As a result each gene n ∈ {1,…, N} has its length equal to integer L_n and is a string of binary entries (bites).

$$g^{n}_{h,t} = \{g^{n,l}_{h,t}, \ldots\ldots, g^{n,Ln}_{h,t}\},\ g^{n,l}_{h,t} \in \{0,1\}\ for\ each\ j \in \{1, \ldots \ldots L_n\} \quad (10)$$

Equation 10, encodes θ^{n} an integer, notice that argument θ^{n} is now probability. Keep in mind that. A $\sum_{l=0}^{Ln-1} 2^{j} = 2^{Ln} - 1$, normalized sum as you can find in equation 11 can easily decode a specific gene of. $g^{n}_{h,t}$.

$$\theta^{n}_{h,t} = \sum_{l=1}^{Ln} \frac{g^{n,l}_{h,t}\, 2^{l-1}}{2^{Ln} - 1} \quad (11)$$

A gen of zeros and ones associated with θ^{n} =0 and θ^{n} =1 respectively, and other plausible binary strings conceal the [0, 1] interval with an $\frac{1}{2^{Ln-1}}$ increment. A real variable θ^{n} from an $[a_n, b_n]$ interval can also ended by linear transformations of probability:

$$\theta^{n}_{h,t} = a_n + (b_n - a_n) \sum_{l=1}^{Ln} \frac{g^{n,l}_{h,t}\, 2^{l-1}}{2^{Ln} - 1} \qquad (12)$$

The accuracy of the representation is given through $\frac{b_n - a_n}{2^{Ln} - 1}$. Two advantages of binary representation are: enabling efficient search through the parameter space and translating any type of well-defined argument to strings of logical values.

3.2.7.3 Evolutionary Operators

Evolutionary operators are the core of genetic algorithm. GA repeats the population of chromosomes for T period, and T can be either large or predefined or dependencies of some convergence standard. First at each period t ∈ {1,..,T} each chromosome's fitness equals to a non-declining transformation of function value F

$$V(\mathcal{F}(\theta_{h,t}^{n})) \equiv V(h_{k.t}) \rightarrow \mathbb{R}^{+} \cap \{0\} \qquad (13)$$

Chromosomes at each period can go through some evolutionary operators such as: procreation, mutation, crossover and election. The duty of these operators is generating an offspring population of chromosomes through parent population t and transforming both groups of populations into a new generation of chromosomes t+1.

3.2.7.4 Procreation

Procreation model based on natural selection process, where original population subset is subject of consideration and out of each subset few chromosomes will be selected. And they will, given more advantages perform better. And the process will be continued till number of old and new generations become equal, however we expect the new generation to come out better. For the population at time t, GA picks $\prod \subseteq \mathbb{H}$ subset of π chromosomes and takes k<π in set $\mathbb{K}$ and power function is what can help us finding the probability of chromosome h ∈ $\prod$ being picked into the $\mathbb{K}$ as its zth element, where z∈ $\{1, \dots \dots k\}$. The power functions are as follows:

$$\text{Prob}(g_z = g_{nh.t}) = \frac{V(g_{h.t})}{\sum_{j \in l} V(g_{h.t})} \qquad (14)$$

By repeating the process with different $\prod$s, number of chromosomes in all set $\mathbb{K}$ will be equal to H. For example in **roulette** (as a procreation) GA picks one chromosome from the whole and the probability of those chromosomes which picked is equal to its function value relative to the function value of all other chromosomes and it is repeated H times .

3.2.7.5 Mutation

After procreation, the next step for each generation is that each binary entry in each chromosome should have a chosen δ_m the probability to be swapped: turning into a zero and vice versa, expecting chromosomes to show the diversity of numbers and attain better fit. And this stage is where the binary representation reaches its maximum efficiency. Mutation of bites that are close to the end of the gene, will lead to new different arguments which are substantially different from the original ones. Also obtaining slight changes from the beginning of the gene are possible through mutating bites. In this case GA can easily evaluate arguments which are far away and at the same time close to current encoded chromosomes. This process helps maximum coverage of GA, however it might not fixate on a local maximum.

3.2.7.6 Crossover

Consider $0 \leq$ CL, C H $\leq \sum_{n=1}^{N} Ln$ Are two predefined integers crossover functioning through the division of chromosome population into pair and exchanging 1 St C_L and last CH bites among chromosomes in each pair with specific probability δ_c In this operator chromosome experiment different combinations of individual arguments that are already successful by themselves.

3.2.7.7 Election

Since experimentation done by the mutation and crossover operators does not require efficient usage of binary sequences, there might be some negative effect on results. As an example a chromosome which actually decodes argument should not mutate at all. In order to avoid such effect, it’s normal to divide the creation of a new generation into two stages: first, the chromosomes that can withstand mutation and crossover in some predefined order and Second, to compare chromosomes in terms of fitness with parent population. And if offspring generation performed weakly, they will pass to a new generation in order to make them perform at least as good as old generation.

3.2.7.8 Behavioral Interpretation

Brain can use numerical procedures very well, because there is simply biological for them. As an example we can consider a dog trying to intercept a moving object like human trying to catch a ball (Chardenon, Montagne, Buekers, and Laurent, 2002). In this scene the brain of the animal needs to translate one number – the number of retina photoreceptor cells, stimulating special fashion into another number that refers to a number of special nervous signals sent to particular muscles, which will define the exact muscle contraction. The surrounding environment is an important factor in increasing efficiency of special numerical responses. In case of dogs mentioned before, interception can be dependent on wind strength or type of ground. All activities of the brain causes from a reason, for instance if the wind is strong, it may force someone to run faster. Meaning a slight change in the retina can cause a stronger muscle construction. We use this model to optimize human brain through GA. GA is applicable to many types of problem and beside its very simple and efficient. Through GA we can discover an elegant and diverse solution. And we should provide the brain to access numerical tools directly not just via conscious control.

3.2.7.9 Experimental Validity

The ideal way of using econometric in forecasting purpose is estimating ARMA or GARCH representation of time-series. Econometric approach can be hard to put into practice. One needs to have the skill and software to work in this area. This increases in the case of the structural estimations, for an agent to get the basic price of an asset, knowing the concept and the way of applying it into specific market is necessary. Another thing to be considered is that short time-series increase the identification and precision issues. The agent never directly will be asked to use equilibrium concepts such as fundamental price the Nash equilibrium of the model. They do not know whether the short-run dynamics of price can have an effect on long-run attraction. They will be asked to forecast the number or price in the next period, not predicting long-run behavior.

3.2.7.10 GA and other Learning Models

GA is a generalized version of all forecasting methods and first order rule plus heuristic switching model. Beside it is also a generalized version of other models including experience-weighted attraction model (EWA), (Camerer and Huatlo, 1999,). In EWA, the chromosomes are responsible for encoding the number of a strategy or probability of a mixed strategy. Therefore they will fight some predefined transformation of the past experience. And in the classical EWA model evaluation of each strategy that might be involved in case of large games, is duty of the agent.

In this research, the GA for forecasting is used and also for investigating the linear relationships in the data. Two lags are considered besides intercept as cost function and after that the result of the GA are included in the loop in order to investigate 24 steps ahead forecasting i.e. sets the maximum number of iterations equal to 100 and for procreation of parents the **Roulette-wheel** is used**.** I also optimize our cost function and find the best answer for minimum cost that has mentioned in the next chapter by using GA.

3.2.8 Forecast Combination Method

A special characteristic that makes forecast combination method different from others is the way it uses historical data and information to get the combination of prediction to the point the weight given to the each individual could change during the time. We have some forecast sampling methods like (the simple forecasts MEAN, the discounted DISC, principle component PC method) some of them need a holdout period to compute the weights used to combine the individual model forecasts, therefore we have to use the first P_0 observations from out of sample time period as the primary holdout period.(Rapach and Strauss ,2010).

Mixture of forecasts such as y_{t+h}^{h} created at time t and $y^{\wedge h}_{CB,t+h} \mid$ t are linear combination of each forecast models

$$y^{\wedge h}_{CB,t+h} \mid t = \sum_{i=1}^{n} w_{i.t} \; y^{\wedge h}_{i,t+h} \mid t \qquad (15)$$

Where $\sum_{i=1}^{n} w_{i.t}$ = 1. And the weights $\{wi.t\}_{i=1}^{n}$ are estimated and individual out of sample forecast and there are y_{t+h}^{h} observations available from the beginning of holdout out of sample time t.

I used the simple combination forecasts to calculate combination forecast regardless of historical performance of individual forecasts. According to Stock and Watson (1999, 2003, 2004) simple combining methods perform well in forecasting inflation and output growth through applying a number of potential predictors. Later in (2004) they noticed that there's a slight difference in the performance of the mean and trimmed mean forecast while especially median has higher relative MSFE than both of them. The mean combination forecast includes $W_{i,t} = \frac{1}{n}$ (i=1… n) in Equation 15 So the simple combining methods do not require a holdout out-of- sample period.

Chapter 4

EMPIRICAL FINDINGS

In this section I have examined and combined the results which were obtained from the four methods of forecasting. Since it was discussed in section three, in order to out of sample forecasting for US house price index; l used four methods of " MLP, NAR, ANFIS, GA " and also the combination of them by ''simple mean'' method.

With regard to the nature of the tested data, the existence of the linear or non-linear relationship is investigated by the above mentioned methods which the results can be observed through the tables 2-6.

Table 2 by using the RMSE criteria, is introducing the obtained error which is related to US data. Since it can be observed in the table 2, in 24 steps for out of sample forecasting the RMSE criteria are specified separately. With attention to the table's data the existence of the non-linear relationship that is due to the forecasting method through the neural network of MLP and NAR, are better than the two above methods. The RMSE of MLP is 0.07 and RMSE of NAR is 0.1. The RMSE for ANFIS is 0.26 that is relatively a good answer.

The reason ANFIS did not represent better answer relative to two previous methods is the range of data, because it started introducing the fuzzy rules according to the

Table 2. US Results

STEPS	RMSE GA	RMSE ANFIS	RMSE NAR	RMSE MLP	RMSE COMBINATION
1	0.05	0.03	0.10	0.03	0.04
2	0.05	0.04	0.10	0.03	0.04
3	0.06	0.09	0.10	0.04	0.05
4	0.07	0.85	0.10	0.05	0.28
5	0.10	0.50	0.10	0.06	0.17
6	0.12	0.94	0.10	0.08	0.31
7	0.15	0.70	0.10	0.08	0.23
8	0.19	0.33	0.10	0.07	0.11
9	0.24	0.37	0.10	0.09	0.13
10	0.30	0.21	0.10	0.08	0.09
11	0.38	0.39	0.10	0.07	0.14
12	0.48	0.05	0.10	0.07	0.05
13	0.63	0.05	0.10	0.08	0.05
14	0.80	0.18	0.10	0.08	0.08
15	1.04	0.92	0.10	0.08	0.31
16	1.34	0.05	0.10	0.09	0.06
17	1.75	0.05	0.10	0.06	0.05
18	2.26	0.05	0.10	0.08	0.06
19	2.95	0.05	0.10	0.06	0.05
20	3.93	0.05	0.10	0.08	0.06
21	3.74	0.05	0.11	0.07	0.05
22	4.77	0.05	0.11	0.11	0.06
23	7.16	0.06	0.11	0.06	0.05
24	7.93	0.09	0.11	0.09	0.07
MEAN	3.37	0.26	0.10	0.07	0.11

range of data in training part and as soon as the range and trend of data changes in testing part ; then ANFIS reacts strongly to this case but when the ANFIS encounter some steps which their errors are large for next steps, its results became favorable for instance in US in step 15 the error was 0.92 but after that the errors declined until step 24 , but totally considering the range and trend of the applied data related to US, the results of data is acceptable.

Concerning RMSE of GA, l should say that as I mentioned before, I have used two lags in cost function and after few times running of GA (about 70 times), the relative acceptable results were not observed. This happens for two reasons: one. Because in every iteration, GA chooses the data accidently and as a result the convergence for fitting the data is accomplished with delay, and two, on the other side the relationship between the two previous days 's data and today's data is not suitable. In fact GA rejects the existence of linear relation while forecasting, which with attention to obtain results concluded that there is no linear relation. Therefore in combined forecasting method. GA and its results were ignored.

As for this point can be observed in the table 2, the obtained results from the combination of three methods (NAR, MLP, and ANFIS) are acceptable. As a whole on this table the results of MLP are the best answer (0.07)

Balcilar et al (2012) declared that obtained results through the AR and STAR methods in their article following which it can be perceived that neural network and intelligent methods have more potency in forecasting and modeling the data. For example, RMSE for AR method in step 1 is 0.02 which can be seen that the obtained results from intelligent methods are better in every five methods. The table --- shows the results of Balcilar et al (2012).

The author with regards to the nature of GA as an optimizer and research data has calculated the optimized values of cost function f and four census regions as follows, the results can be observed in the figure 13-17 The criteria that was used for cost function minimization is SSE (Sum of squared for error).

Table 3. Out-Of-Sample Point Forecast Evaluation, Linear AR And STAR Models Source Balcilar et al(2012)

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

Table 4. Out-Of-Sample Point Forecast Evaluation, Linear AR And STAR Models Source Balcilar et al(2012)

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

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

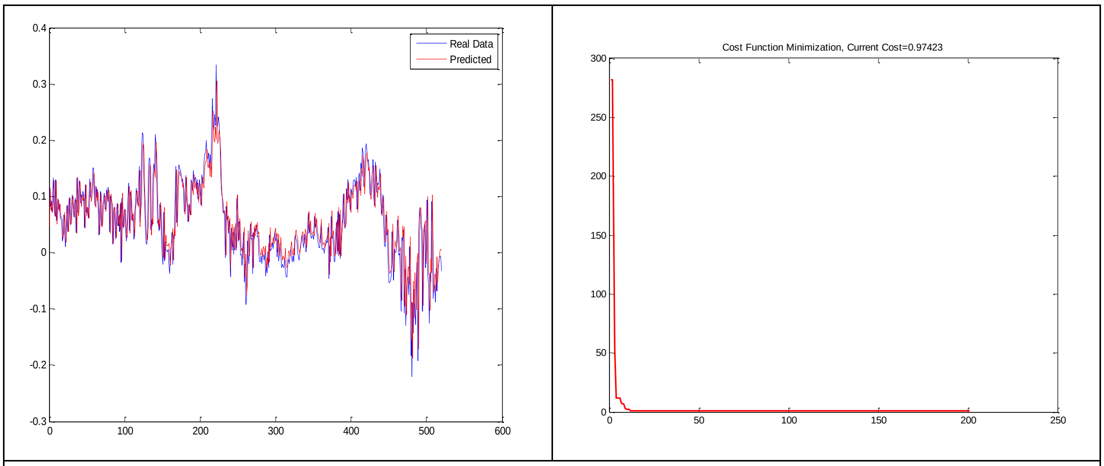
a Forecast horizon (in months).

b Linear AR model RMSFE.

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

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

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

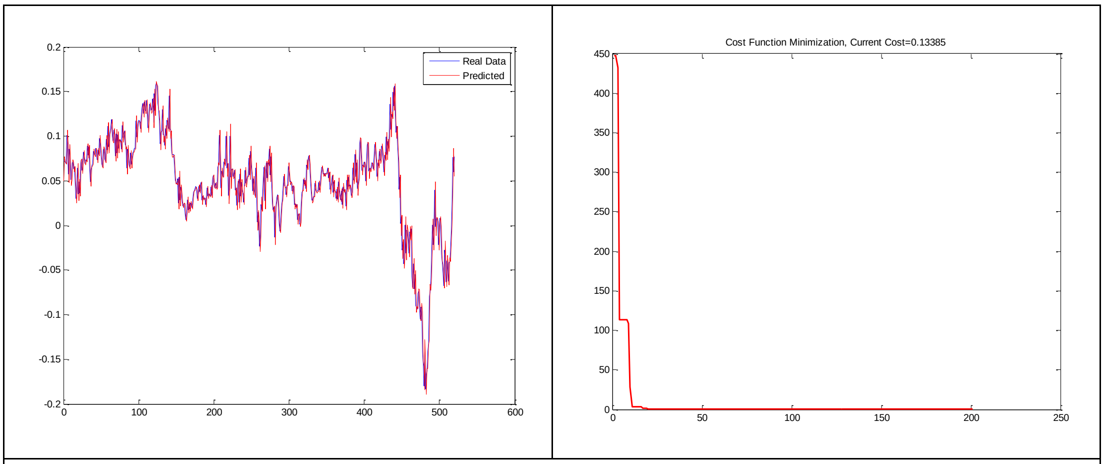


Cost function: $Y_t^{\wedge}$=a*Y_{t-1}+ b*Y_{t-2}+c

Coefficient	Optimal value
a	0.9563
b	-0.1060
c	0.0108

SSE=0.97

Figure 13. NorthEast _GA Optimization Results.

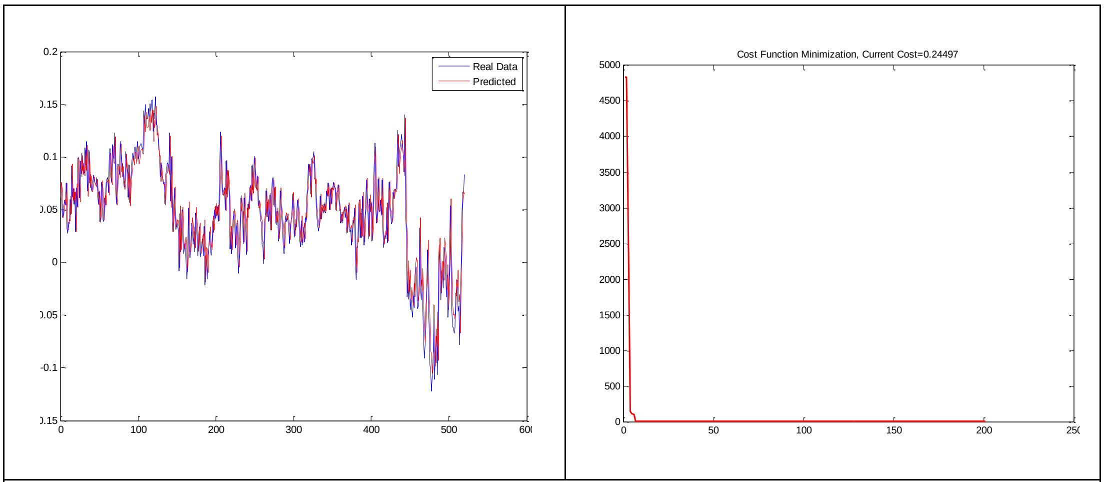


Cost function: $Y_t^{\wedge}$=a*Y_{t-1}+ b*Y_{t-2}+c

Coefficient	Optimal value
a	0 .2318
b	-0.2434
c	-0.0002

SSE = 0.13

Figure 14. US _GA Optimization Results.

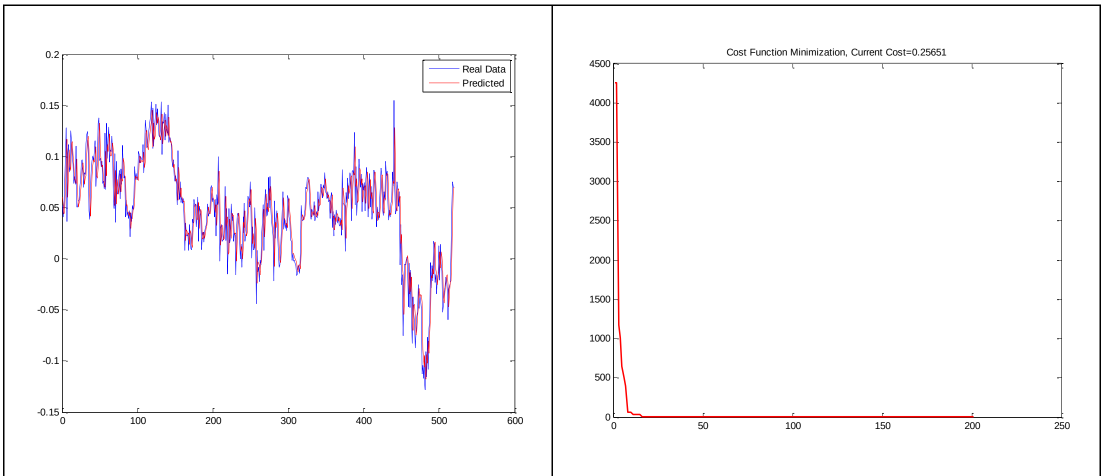


Cost function: $\hat{Y}_t = a*Y_{t-1} + b*Y_{t-2} + c$

Coefficient	Optimal value
a	0.9939
b	-0.0989
c	0.0065

SSE = 0.24

Figure 15. MidWest_GA Optimization Results

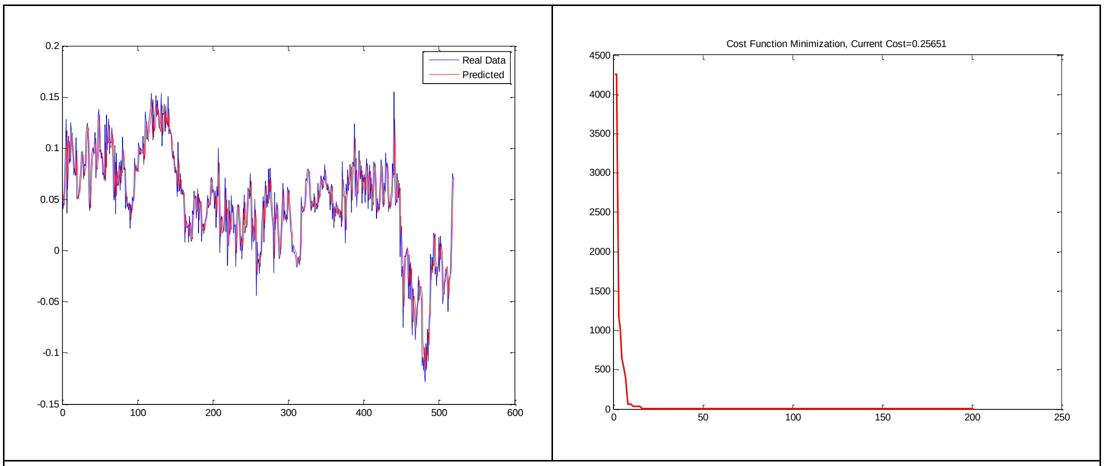


Cost function: $\hat{Y}_t = a*Y_{t-1} + b*Y_{t-2} + c$

Coefficient	Optimal value
a	0.6966
b	0.2571
c	0.0022

SSE = 0.25

Figure 16. South _GA Optimization Results



Cost function: $\hat{Y}_t = a*Y_{t-1} + b*Y_{t-2} + c$

Coefficient	Optimal value
a	0.7512
b	0.2148
c	0.0019

SSE = 0.50

Figure 17.West_GA Optimization Results

The results related to ‘’ West- Northeast- Midwest and South ‘’ can be observed in following tables orderly

Since it has been observed, GA did not have acceptable results in all of the tables which the reasons have been explained before

In table 2 for the Northeast, RMSE related to neural network is better relative to the other results; on the other side I excluded GA in the calculation of combination method like the previous section. Also the results of those tables can be compared with the results of Balcilar et al (2012). As we can see again the results of intelligent methods are better than the other methods (AR…) that was attained in the article of Balcilar.et al (2012).

Table 5. MidWest Results

STEPS	RMSE GA	RMSE ANFIS	RMSE NAR	RMSE MLP	RMSE COMBINATION
1	0.05	0.03	0.08	0.04	0.0395
2	0.05	0.04	0.08	0.05	0.0430
3	0.06	0.04	0.08	0.08	0.0455
4	0.07	0.06	0.08	0.08	0.0542
5	0.10	0.14	0.08	0.08	0.0689
6	0.12	0.39	0.08	0.11	0.1461
7	0.15	0.17	0.08	0.09	0.0803
8	0.19	0.19	0.08	0.08	0.0807
9	0.24	0.78	0.08	0.08	0.2742
10	0.30	0.19	0.08	0.08	0.0809
11	0.38	0.08	0.08	0.08	0.0580
12	0.48	0.27	0.08	0.09	0.1023
13	0.63	0.09	0.08	0.08	0.0615
14	0.80	0.23	0.08	0.08	0.0928
15	1.04	0.06	0.08	0.08	0.0584
16	1.34	0.06	0.08	0.08	0.0593
17	1.75	0.06	0.08	0.09	0.0595
18	2.26	0.06	0.08	0.09	0.0600
19	2.95	0.06	0.08	0.09	0.0606
20	3.83	0.06	0.08	0.09	0.0610
21	3.74	0.07	0.08	0.09	0.0614
22	4.77	0.07	0.08	0.09	0.0623
23	6.16	0.07	0.08	0.09	0.0620
24	7.93	0.07	0.08	0.10	0.0622
MEAN	2.69	0.14	0.08	0.08	0.0764

Table 6. West Results

STEPS	RMSE GA	RMSE ANFIS	RMSE NAR	RMSE MLP	RMSE COMBINATION
1	0.07	0.05	0.14	0.05	0.0647
2	0.07	0.08	0.14	0.08	0.0702
3	0.09	0.17	0.14	0.14	0.1048
4	0.09	1.85	0.15	0.12	0.6256
5	0.10	0.64	0.15	0.18	0.2502
6	0.11	0.65	0.15	0.12	0.2221
7	0.14	2.09	0.15	0.18	0.7176
8	0.16	3.97	0.15	0.16	1.3330
9	0.20	3.21	0.15	0.21	1.0494
10	0.23	0.93	0.15	0.18	0.3183
11	0.29	0.11	0.15	0.23	0.1283
12	0.36	0.25	0.15	0.19	0.1339
13	0.45	1.49	0.15	0.23	0.5091
14	0.58	0.11	0.15	0.20	0.1150
15	0.74	0.11	0.15	0.23	0.1308
16	0.96	0.12	0.15	0.19	0.1157
17	1.24	0.12	0.15	0.23	0.1331
18	1.64	0.12	0.15	0.20	0.1174
19	2.02	0.12	0.15	0.24	0.1357
20	2.70	0.12	0.15	0.21	0.1219
21	3.55	0.12	0.15	0.25	0.1410
22	4.81	0.12	0.15	0.22	0.1264
23	6.37	0.12	0.15	0.26	0.1431
24	8.64	0.12	0.16	0.22	0.1275
MEAN	2.74	0.70	0.15	0.19	0.2890

Table 7. South Results

STEPS	RMSE GA	RMSE ANFIS	RMSE NAR	RMSE MLP	RMSE COMBINATION
1	0.05	0.03	0.08	0.03	0.0386
2	0.05	0.04	0.08	0.05	0.0426
3	0.06	0.08	0.08	0.05	0.0506
4	0.07	0.13	0.08	0.05	0.0588
5	0.10	0.42	0.08	0.04	0.1514
6	0.12	0.04	0.08	0.05	0.0448
7	0.15	0.14	0.08	0.06	0.0555
8	0.19	0.87	0.08	0.06	0.2815
9	0.24	0.06	0.08	0.05	0.0456
10	0.30	0.09	0.09	0.05	0.0501
11	0.38	0.05	0.09	0.06	0.0470
12	0.48	1.11	0.09	0.06	0.3685
13	0.63	0.46	0.09	0.07	0.1658
14	0.80	0.06	0.09	0.05	0.0528
15	1.04	0.09	0.09	0.05	0.0489
16	1.34	0.92	0.09	0.06	0.3178
17	1.75	0.05	0.09	0.06	0.0497
18	2.26	0.05	0.09	0.07	0.0520
19	2.95	0.05	0.09	0.05	0.0497
20	3.83	0.05	0.09	0.06	0.0506
21	3.64	0.05	0.09	0.06	0.0514
22	4.87	0.05	0.09	0.07	0.0516
23	6.16	0.05	0.09	0.07	0.0533
24	6.93	0.05	0.09	0.05	0.0511
MEAN	2.29	0.21	0.09	0.05	0.0929

Table 8. North East Results

STEPS	RMSE GA	RMSE ANFIS	RMSE NAR	RMSE MLP	RMSE COMBINATION
1	0.07	0.05	0.10	0.10	0.0704
2	0.07	0.09	0.10	0.07	0.0735
3	0.09	0.86	0.10	0.07	0.2833
4	0.09	0.14	0.10	0.07	0.0864
5	0.10	0.67	0.10	0.14	0.2442
6	0.11	5.84	0.10	0.08	1.9536
7	0.14	0.07	0.10	0.07	0.0697
8	0.16	0.07	0.11	0.07	0.0762
9	0.20	0.07	0.11	0.14	0.0928
10	0.23	0.07	0.11	0.08	0.0791
11	0.29	0.07	0.11	0.08	0.0734
12	0.36	0.07	0.11	0.08	0.0796
13	0.45	0.07	0.11	0.15	0.0959
14	0.58	0.07	0.11	0.09	0.0822
15	0.74	0.08	0.11	0.08	0.0763
16	0.96	0.08	0.11	0.08	0.0824
17	1.24	0.08	0.11	0.15	0.0986
18	1.64	0.08	0.11	0.09	0.0851
19	2.02	0.08	0.11	0.08	0.0791
20	2.70	0.08	0.11	0.09	0.0852
21	3.55	0.08	0.11	0.15	0.1013
22	4.79	0.08	0.11	0.09	0.0876
23	6.37	0.08	0.11	0.09	0.0812
24	8.64	0.08	0.11	0.09	0.0872
MEAN	2.65	0.38	0.11	0.09	0.1760

The interpretation of these tables is exactly the same as table 2 for US in all of them the results of combination method are good and the neural networks (MLP _NAR) had strong power to model and predict. ANFIS was also an appropriate way for forecasting of applied data

It should be noted that the diagrams for 24 steps of forecasting for each part can be observed. I mention just a few important samples. That has been shown in Appendix. In these diagrams MLP and NAR indicates better results and perform better forecasting

4.1 The Methods of Evaluation of Accuracy of Forecasting Models

Different methods on how to evaluate the accuracy of forecasting models have attracted attention of many economists in recent decades. In current dissertation RMSE and DM-test was used for the comparison of the forecast error. Whatever the error, less the forecast is better

The DM - test has a few provisions limiting what can be calculated as follows:

Suppose that the forecasting error of model j^{th} is $e_{t,j}$ and the function of this case is $g(e_{t,j})$ then the null hypothesis would be:

$$E[g(e_{t,j})] = E[g(e_{t,i})]$$

$$\text{Or} \quad E[d_t] = 0 \quad \text{since that} \quad d_t = E[g(e_{t,j})] - E[g(e_{t,i})]$$

That means the power of accuracy in both forecast models should be equal (null hypothesis) otherwise one model outperforms ($\mathbb{H}_1$ alternative hypothesis). Suppose that $\{d_t\}_{t=1}^{T}$ is the mean of difference series of the error function. Therefore we have

$$\text{DM}= \frac{\bar{d}}{\sqrt{\frac{2\pi f^{\wedge}{}_{d\,(0)}}{T}}}$$

Where $\frac{2\pi f^{\wedge}{}_{d\,(0)}}{T}$ is the variance of $\{d_t\}_{t=1}^{T}$ and $f^{\wedge}{}_{d\,(0)}$ is spectral density of the loss differential at frequency 0

As I mentioned when the null hypothesis was rejected, that means two forecast models are not equal so in this case the sign of DM_ test identify which model is better (according to subtractive equation ($E[g(e_{t\,,j}\,)]$ - $E[g(e_{t\,,i}\,)]$) , if the sign of DM_ test will be positive so the second model will be better than the first model, otherwise the first model will be better if the sign of DM_ test will be negative. The obtained results of the DM - test displayed as follows

As an example in table 7 the DM_ANFIS_NAR for step one is -6.7687 (The star values show the rejection in null hypothesis) so the null hypothesis was rejected at 5% because it is bigger than the critical values ($\pm$ 1.96) and as can be seen the sign of this value is negative therefore the accuracy for ANFIS is better than the NAR in step 1 for US.

Table 9. US_DM Test Critical Values=± 1.96 at 5% level

STEP	DM_ANFIS_MLP	DM_ANFIS_NAR	DM_MLP_NAR
1	-1.4523	-6.7687*	-6.6095*
2	1.2339	-4.1782*	-4.9842*
3	1.1879	-0.1130	-4.2784*
4	1.4638	1.4496	-4.1494*
5	1.3196	1.2880	-3.2607*
6	1.0553	1.0510	-1.6262
7	1.4318	1.4171	-1.9503
8	1.2591	1.2042	-3.1717*
9	1.2730	1.2493	-0.7685
10	0.9772	0.8934	-2.2624*
11	1.0421	1.0064	-3.1841*
12	-1.0666	-2.4656*	-3.2516*
13	-1.0677	-2.4043*	-1.2698
14	0.9533	0.8105	-2.9730*
15	1.0654	1.0603	-2.6636*
16	-1.4354	-2.3303*	-2.3868*
17	-1.5405	-2.3411*	-2.5104*
18	-1.2852	-2.3330*	-2.9425*
19	-1.4192	-2.2985*	-2.5031*
20	-1.2092	-2.2785*	-2.9285*
21	-1.5156	-2.2651*	-2.3013*
22	-1.2967	-2.2760*	0.1643
23	-0.0104	-2.2756*	-2.4720*
24	0.3297	-0.5850	-2.6351*

Table 10. North East_DM Test Critical Values=± 1.96 at 5% level

STEP	DM_ANFIS_MLP	DM_ANFIS_NAR	DM_MLP_NAR
1	-1.0118	-5.2511*	-0.2211
2	1.4564	-0.5302	-3.8993*
3	1.1280	1.1184	-4.0280*
4	0.9697	0.6409	-3.0241*
5	1.0205	1.0141	0.5817
6	1.0202	1.0201	-2.7884*
7	-0.3429	-3.8237*	-3.4427*
8	-0.4837	-3.8243*	-3.4357*
9	-1.0527	-3.8997*	0.6623
10	-1.2972	-3.8923*	-3.5136*
11	-0.8960	-3.8951*	-3.3512*
12	-1.1653	-3.8420*	-3.2715*
13	-1.1078	-3.8592*	0.7488
14	-1.9454	-3.7535*	-3.4009*
15	-1.8087	-3.6900*	-2.9237*
16	-1.9841*	-3.5830*	-2.8735*
17	-1.1489	-3.5491*	0.8169
18	-2.0966*	-3.5082*	-3.3095*
19	-2.6491*	-3.5058*	-2.8038*
20	-2.9550*	-3.4767*	-2.8088*
21	-1.1963	-3.4720*	0.8816
22	-2.3268*	-3.4766*	-3.3994*
23	-3.2576*	-3.3759*	-2.7084*
24	-3.2014*	-3.3690*	-2.7999*

Table 11. MidWest _DM Test Critical Values=± 1.96 at 5% level

STEP	DM_ANFIS_MLP	DM_ANFIS_NAR	DM_MLP_NAR
1	-2.2641*	-7.1719*	-7.0125*
2	-1.9120	-5.4814*	-4.5183*
3	-1.6748	-4.9424*	0.0681
4	-1.0850	-1.3739	-0.0290
5	1.1132	1.0426	0.1781
6	1.1850	1.1924	1.1012
7	0.9788	0.9545	0.5393
8	1.0894	1.0268	0.1291
9	1.0488	1.0452	0.1190
10	1.0192	0.9653	0.0901
11	0.0685	0.1938	0.2402
12	1.0264	1.0212	0.6674
13	0.4526	0.4728	0.3417
14	1.0317	1.0015	0.2935
15	-1.3702	-3.3244*	0.2642
16	-1.3218	-3.3823*	0.2729
17	-1.4025	-3.4341*	0.3604
18	-1.5865	-3.5062*	0.5753
19	-1.3095	-3.5673*	0.3337
20	-1.4910	-3.6140*	0.5175
21	-1.4197	-3.6921*	0.4520
22	-1.3385	-3.6987*	0.4064
23	-1.4977	-3.7315*	0.5342
24	-1.8288	-3.6529*	0.8693

Table 12. South_DM Test Critical Values=± 1.96 at 5% level

STEPS	DM_ANFIS_MLP	DM_ANFIS_NAR	DM_MLP_NAR
1	-1.0840	-7.7053*	-7.6571*
2	-0.3105	-3.8808*	-2.8695*
3	0.6892	-9.4354e-04 ≅ 0	-2.2118*
4	0.9673	0.6497	-3.6250*
5	1.0242	0.9943	-3.8451*
6	-0.6529	-3.1768*	-3.4682*
7	1.0745	0.7605	-1.8780
8	1.0293	1.0237	-1.7808
9	0.8619	-2.3815	-2.8369*
10	1.0175	0.2807	-2.7711*
11	-0.8508	-2.5068*	-2.6533*
12	1.0537	1.0519	-1.6793
13	1.0474	1.0386	-1.7424
14	0.9009	-2.0498*	-2.3555*
15	1.0620	0.0620	-2.2996*
16	1.0744	1.0713	-2.0638*
17	-0.9689	-2.1179*	-1.6090
18	-1.2034	-2.0857*	-1.6609
19	-0.0856	-2.0858*	-2.0618*
20	-0.9889	-2.0452*	-2.2812*
21	-0.8150	-2.0203*	-1.8541
22	-1.0724	-1.9980*	-1.5093
23	-1.2940	-1.9675	-1.5889
24	-0.2643	-1.9648	-1.9314

Table 13. West_DM Test Critical Values=± 1.96 at 5% level

STEP	DM_ANFIS_MLP	DM_ANFIS_NAR	DM_MLP_NAR
1	-0.3650	-6.3754*	-6.3871*
2	0.0448	-4.7054*	-4.4886*
3	0.4533	0.6333	0.0025
4	1.4410	1.4401	-0.5964
5	1.0110	1.0259	0.7178
6	1.1386	1.1336	-0.6065
7	1.0721	1.0749	0.8990
8	1.0332	1.0332	0.2921
9	1.2630	1.2635	1.1830
10	1.1234	1.1292	0.8378
11	-1.4565	-2.5676*	1.1947
12	0.6037	0.8968	0.9581
13	1.0480	1.0530	1.1843
14	-1.5330	-2.5588*	1.1143
15	-1.3828	-2.5818*	1.1591
16	-1.3895	-2.4723*	0.9702
17	-1.3359	-2.5825*	1.1220
18	-1.3524	-2.5344*	0.9486
19	-1.3240	-2.6112*	1.1212
20	-1.3878	-2.6289*	1.0382
21	-1.3575	-2.6685*	1.1835
22	-1.4007	-2.8069*	1.1067
23	-1.3616	-2.8735*	1.1873
24	-1.3968	-2.9689*	1.1054

Chapter 5

CONCLUSION

Many recent researches showed that there is a link between the housing market and economic activity. Besides they mentioned that house-price fluctuations lead real activity, inflation, or both. Therefore the existence of good model to forecast is very crucial for policy makers. Good policy requires that first identification of relationship for data (linear or non _linear) because it can affect not only housing prices rather all the economy.

In this thesis I predicted housing price in the US and fourth census regains (Northeast, West, Midwest and South). I have taken the overall price of house in America and used four methods for: NAR, MLP, ANFIS, and Genetic Algorithm to do out of sample forecasts for US and other fourth regions. Our inside samples are from 1968:01 to 2000:12 (384 observations) and our out of sample forecasting period will fall between 2001:01 and 2010:05 (138 observations). Finally I used ex-ante forecast design and predicted 24 steps ahead (over 2010:6 to 2012:6) with four models and also with a combination model to compare the results of each model. Using four methods separately for the individual time period (24 steps ahead) will give different results

After computing the errors for each method, I compared the results in each method with other methods to see the level of accuracy of each method. And then I used simple combination method to check the results as a combination of all three method (I have dropped GA in combination form since the results of this method are not good) .The result showed that the MLP and NAR altogether had better answer in all parts of the data (US and four census regions) and they could have a better forecast accuracy while the results of ANFIS specially in the initial steps showed that the ability of it for forecasting are better than other methods. According to the results of this research l concluded that the neural network (MLP and NAR) and in a sense ANFIS has a suitable ability to model and forecast especially when there is non-linear relationships between the data

GA follows the linear relationship between the data with respect to its cost function (with two lags and intercept) therefore the results of this method can be evident justification based on the existence of linear relationship between the data. The results of GA in all parts of the data were not desirable and can be concluded that the neural networks (MLP _NAR) and also Neuro_Fuzzy systems (ANFIS) dominate the GA model.

The combination method had a good and suitable answer in each part except in the west because the results of ANFIS in this section were not very good toward the other sections (is 0.70 as a mean of RMSE) and it was because of the range of the data. In that case when the range of data in training is different from the range of data in testing part the membership function could not work correctly therefore the accuracy of forecasting is not ideal. But in other sections the results of combination method were very agreeable.

I also computed the Diebold _Mariano test for identifying which model has better fitness in each steps of forecast. DM test at each step showed the power of each model for forecasting. In many steps there are no difference between the ANFIS and MLP since the null hypothesis could not be rejected, but in the case of ANFIS and NAR and MLP and NAR there are significant differences between these methods with respect to the power of accuracy.

REFERENCES

Aiken, & Milam. (1999). Using a neural network to forcast inflation. *Industrail Managment and Data Systems* .

Balcilar, M., Gupta, R., & Miller, S. M. (2012). The Out-of-Sample Forecasting Performance of Non-Linear Models of Regiona Housing Prices in the USl. *University of Pretoria Department of Economics Working Paper Series* .

Balcilar, M., Gupta, R., & Shah, Z. B. (2011). An in_sample and out_of _sample emprical investigation of the nonlinearity in house prices of South Africa. *Economic Modelling* , 28.891-899.

Baure, R. J. (1994). *Genetic Algorithem and investment strategies* . New York : John wiley .

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

Booker, L. B. (1985). Improving the Performance of Genetic Algorithem . *International Conference on Genetic Algorithem And Their Applications*, (pp. 80-92).

Booker, L. B., Goldberg, D. E., & Holland, J. H. (1989). Classifier Systems and Genetic Algorithem. *Artificial Intelligence* , 40,235-282.

De Jong, K. A. (1975). An Analysis of Behavior of a Genetic Algorithem. *Machine Learning* , 161-188.

Dzielinski, A. (2002). Neural Network-Based NARX Models in Non-Linear AdaptiveE Control. *Int. J. Appl. Math. Comput. Sci* , Vol.12, No.2, 235–240.

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

Fonseca, C. M., & Fleming, P. J. (1995). AN Overview of Evolutionary Algorithems in Multiobjective Optimization. *Evolutionary Computation* , 3(1),1-16.

Fu, J. (1998). A Neural Network Forecast of Economic Growth and Recession . *The Journal of Economics* , NO. 1,51-66.

Goldberg, D. E. (1989). Genetic Algoeithem in Search,Optimization and Machine Learning. *Evolutionary Computation* , 25-36.

Greene, D. P., & Smith, S. F. (1994). Using Coverage as a Model Bulding Constraint in Learning Classifier Systems . *IEEE* , 67-91.

Grefenstette, J. J. (1989). A System for Learning ControL Sategies with Genetic Algorithem. *International Confrence on Genetic Algorithems*, (pp. 183-190).

Hansen, J. V., McDonald, J. B., & Nelson, R. D. (1999). Time series prediction with genetic algorithem designed neural networks. *Computational intelligence* , 15(3),171-184.

Hill, Tim, Marquez, Leory, Marcus, & Remus. (1994). Artificial Neural Network Models for Forecasting and Decision Making. *Jouranl of Forecasting* , 10,5-15.

Holland, J. h. (1986). The Possibility of General Purpose Learning Algorithems . *Artificail Intelligence* .

Hornik, K., Stinchcombe, M., & White. (1989). Multi Layer Feed Forward Networks Are Universal Approximators. *Neuaral Network* , 2,551-560.

http://www.mathworks.com/help/fuzzy/genfis1. (2012).

Jang, J. S. (1993). ANFIS :Adaptive network_based fuzzy inference system. *IEEE Trans* , 23,665-685.

Stock, J. H., & Watson, M. V. (2004). Combination forecasts of output growth in aseven country data set . *Journal of Forecasting* , 23:405-430.

Tal, Benny, Nazareth, & Linda. (1995). Artificial Intelligence And Economic Forecasting. *Canadian Business Economics* , 69-74.

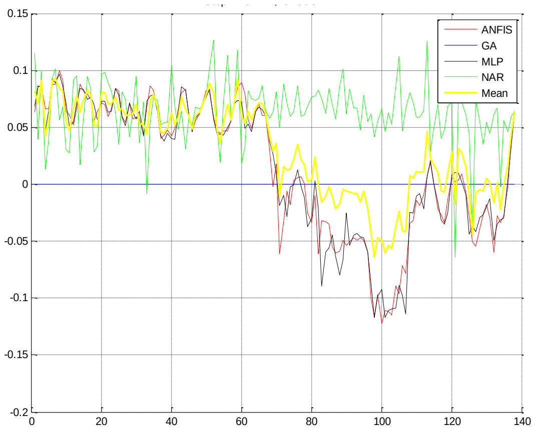
Wan, L., Luo, B, J., & Yan. (2012). A type of combination forecasting method. *American Journal of Operation Research* , 2,476-472.

Zhang, G. P., & Kline, D. M. (2007). Quarterly time series forecasting with neural networks. *IEEE Transacaction on Neural Network* , 18(6),1800-1814.

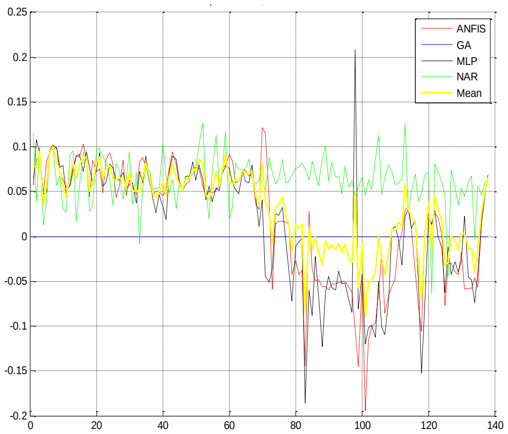
APPENDICES

Appendix A. South

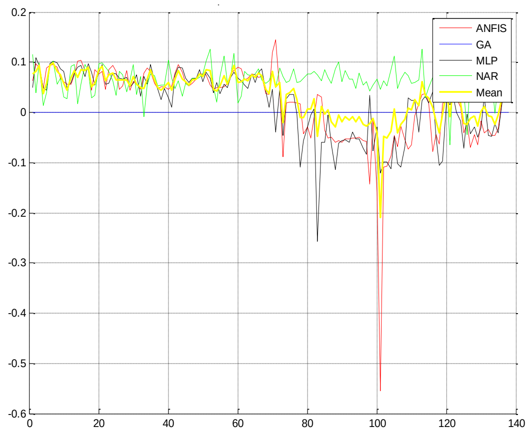
Step 1 for South



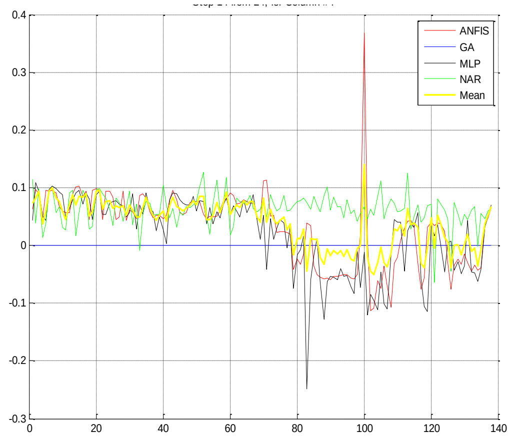
Step 6 for South



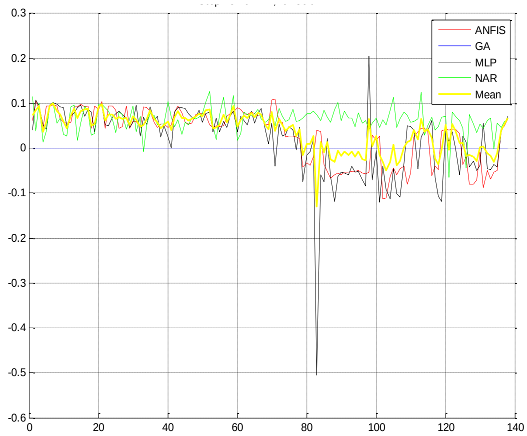
Step 9 for South



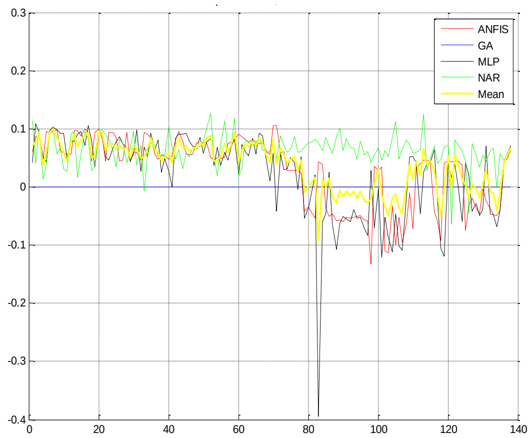
Step 14 for South



Step 18 for South

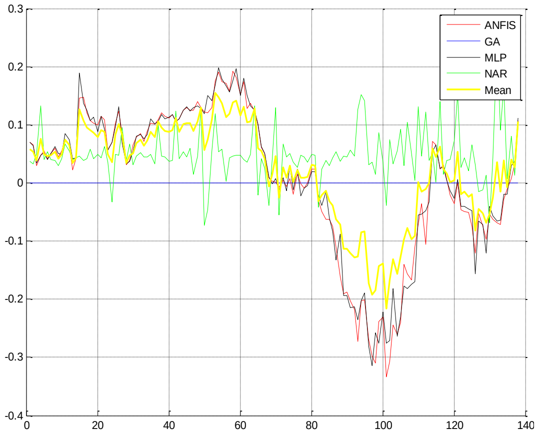


Step 21 for South

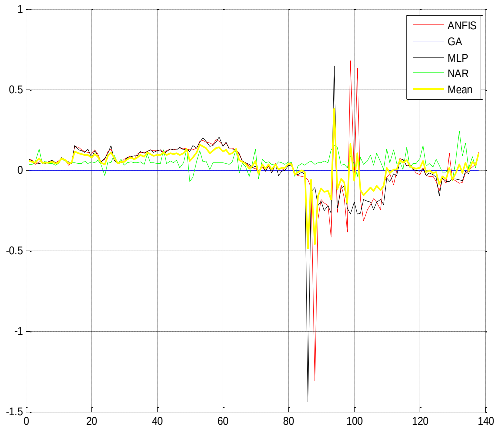


Appendix B. West

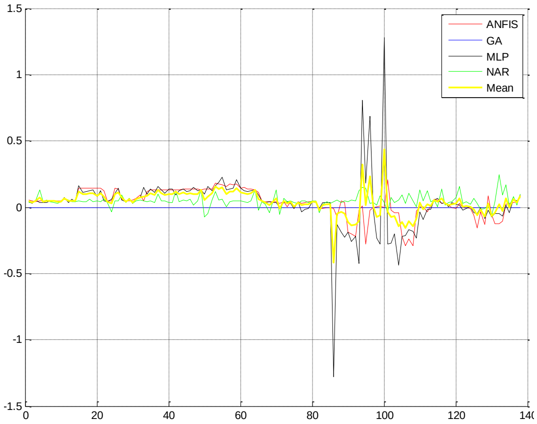
Step 1 for West



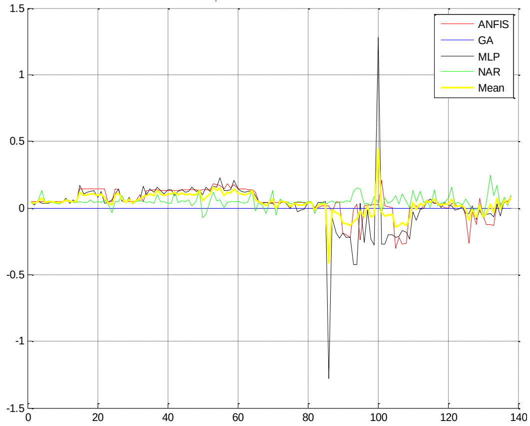
Step 3 for West



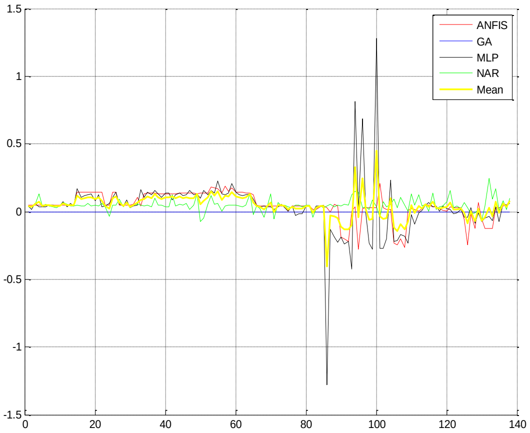
Step 11 for West



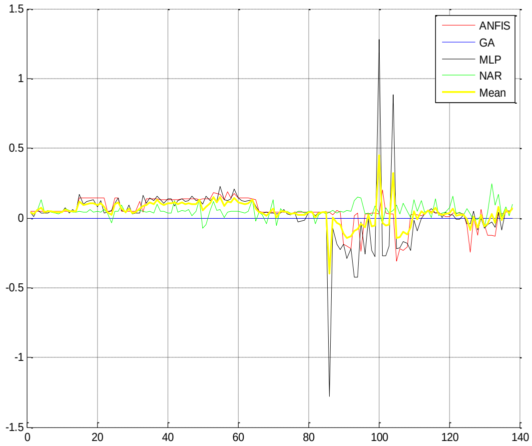
Step 16 for West



Step 19 for West

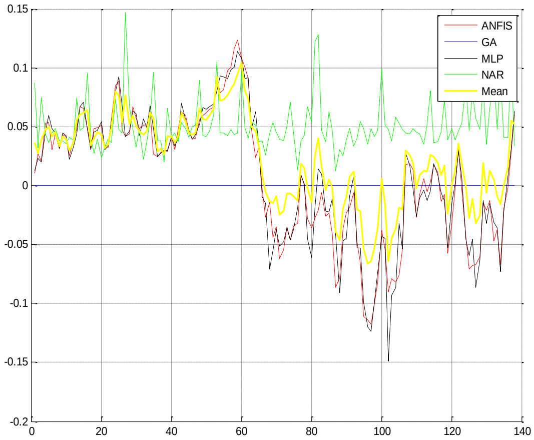


Step 24 for West

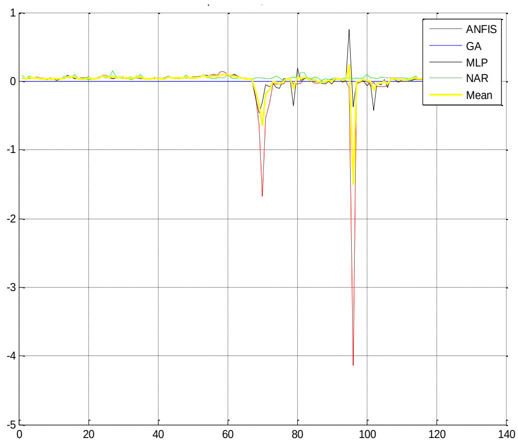


Appendix C. Midwest

Step 1 for Midwest

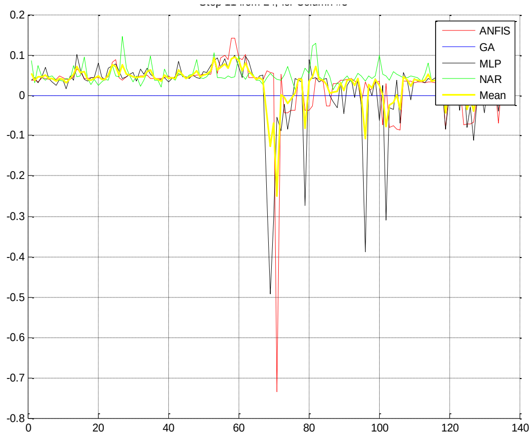


Step 6 for Midwest

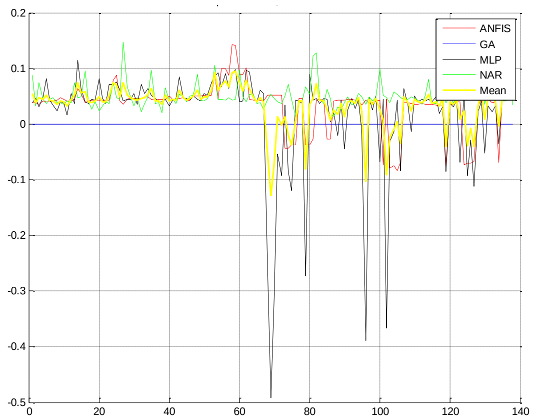


Step 8 for Midwest

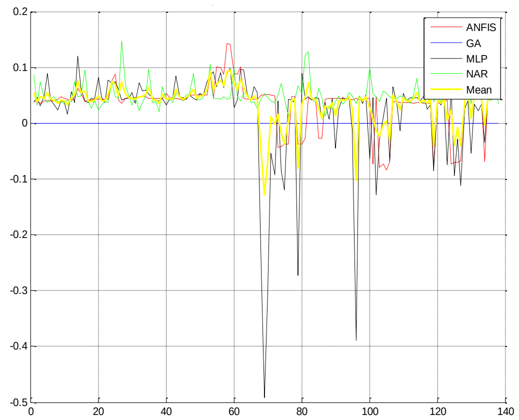
Step 11 for Midwest



Step 18 for Midwest

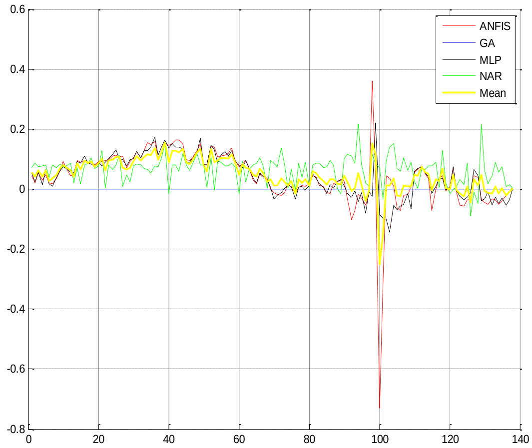


Step 22 for Midwest

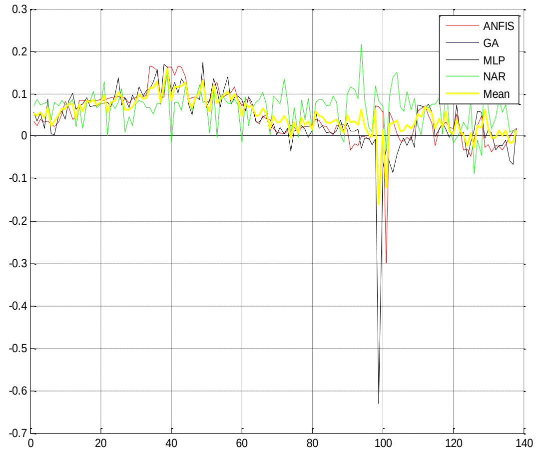


Appendix D. Northeast

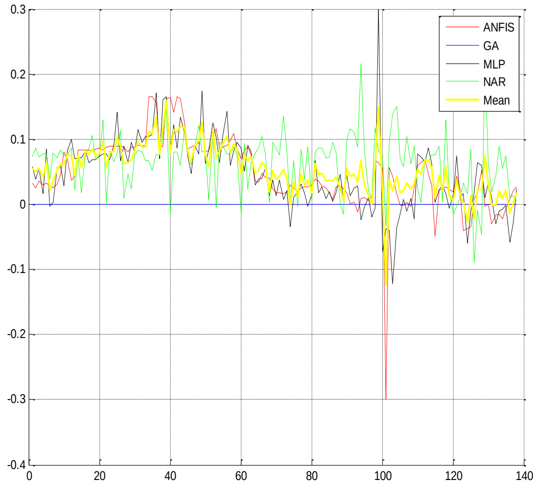
Step 3 for Northeast



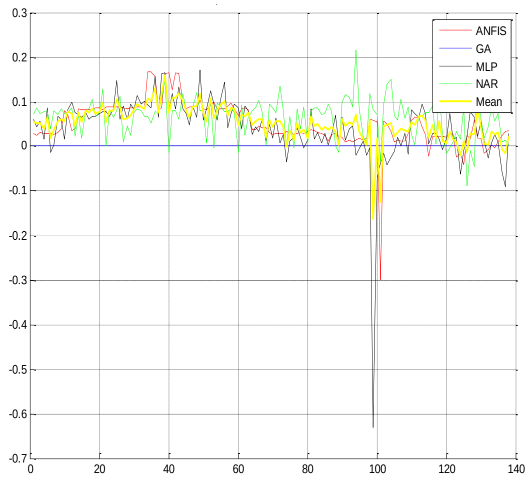
Step 7 for Northeast



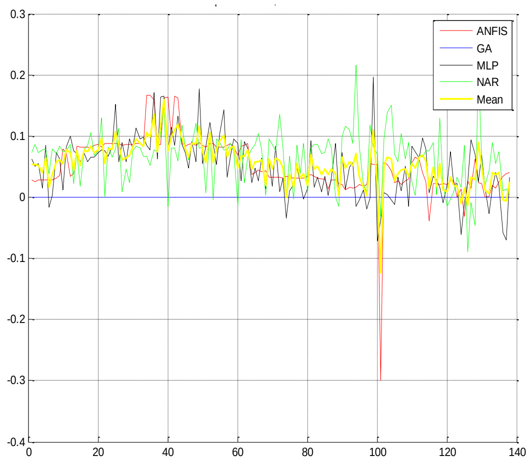
Step 10 for Northeast



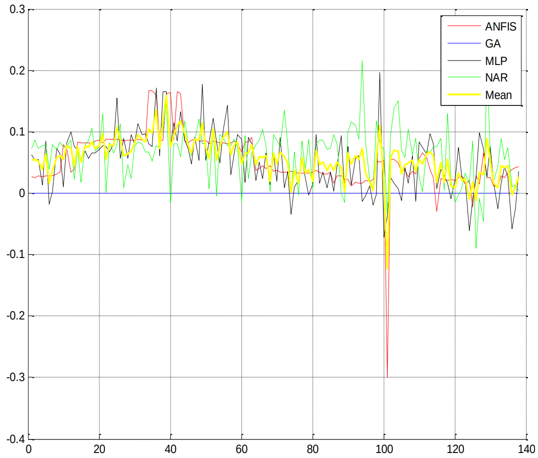
Step 15 for Northeast



Step 20 for Northeast

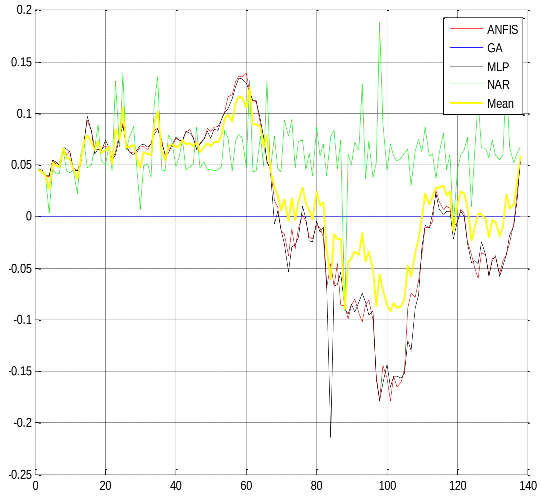


Step 24 for Northeast

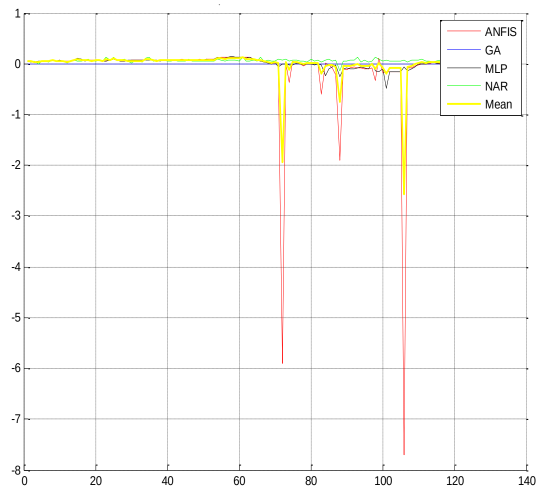


Appendix E. US

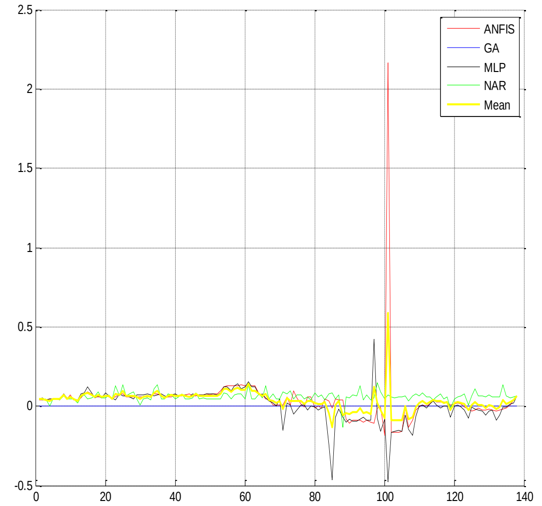
Step 1 for US



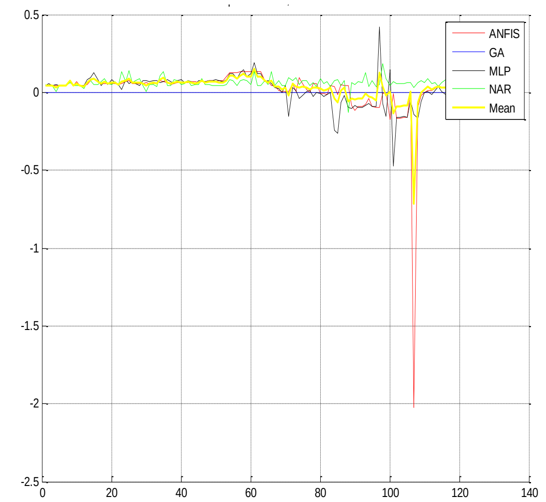
Step 4 for US



Step 10 for US



Step 14 for US



Step 18 for US

Step 22 for US