SBP Working Paper Series

No. 99 October, 2018

A Thick ANN Model for Forecasting Inflation

Muhammad Nadim Hanif

Khurrum S. Mughal

Javed Iqbal

STATE BANK OF PAKISTAN

SBP Working Paper Series

Editor: Sajawal Khan

The objective of the SBP Working Paper Series is to stimulate and generate discussions on different aspects of macroeconomic issues among the staff members of the State Bank of Pakistan. Papers published in this series are subject to intense internal review process. The views expressed in these papers are those of the author(s) not State Bank of Pakistan.

Price per Working Paper (print form)

Pakistan:	Rs 50 (inclusive of postage)
Foreign:	US$ 20 (inclusive of postage)

Purchase orders, accompanied with cheques/drafts drawn in favor of State Bank of Pakistan, should be sent to:

Chief Spokesperson,
External Relations Department,
State Bank of Pakistan,
I.I. Chundrigar Road, P.O. Box No. 4456,
Karachi 74000. Pakistan.

Soft copy is downloadable for free from SBP website: http://www.sbp.org.pk

For all other correspondence:

Postal: Editor,
SBP Working Paper Series,
Research Department,
State Bank of Pakistan,
I.I. Chundrigar Road, P.O. Box No. 4456,
Karachi 74000. Pakistan.

Email: wps@sbp.org.pk

ISSN 1997-3802 (Print)
ISSN 1997-3810 (Online)

Published by State Bank of Pakistan, Karachi, Pakistan.
Printed at the SBP BSC (Bank) – Printing Press, Karachi, Pakistan

A Thick ANN Model for Forecasting Inflation

Muhammad Nadim Hanif [1], Khurrum S. Mughal [2] & Javed Iqbal [3,*]

Abstract

Inflation forecasting is an essential activity at central banks to formulate forward looking monetary policy stance. Like in other fields, machine learning is finding its way to forecasting; inflation forecasting is not any exception. In machine learning, most popular tool for forecasting is artificial neural network (ANN). Researchers have used different performance measures (including RMSE) to optimize set of characteristics - architecture, training algorithm and activation function - of an ANN model. However, any chosen 'optimal' set may not remain reliable on realization of new data. We suggest use of 'mode' or most appearing set from a simulation based distribution of optimum 'set of characteristics of ANN model'; selected from a large number of different sets. Here again, we may have a different trained network in case we re-run this 'modal' optimal set since initial weights in training process are assigned randomly. To overcome this issue, we suggest use of 'thickness' to produce stable and reliable forecasts using modal optimal set. Using January 1958 to December 2017 year on year (YoY) inflation data of Pakistan, we found that our YoY inflation forecasts (based on aforementioned multistage forecasting scheme) outperform those from a number of inflation forecasting models of Pakistan economy.

JEL Classification: C45, E31, E37

Key Words: Artificial Neural Networks, Inflation Forecasting

Acknowledgment

Authors would like to thank an anonymous reviewer for his valuable comments, which led to significant improvement in the forecast evaluation approach.

Contact for correspondence:
Muhammad Nadim Hanif
Lead Economist
Research Department
State Bank of Pakistan
I.I. Chundrigar Road
Karachi 74000.
Email: nadeem.hanif@sbp.org.pk

[1] Lead Economist, Research Department, State Bank of Pakistan, Karachi (nadeem.hanif@sbp.org.pk)
[2] Research Economist, Research Department, State Bank of Pakistan, Karachi (khurrum.mughal@sbp.org.pk)
[3] Joint Director, Research Department, State Bank of Pakistan, Karachi (javed.iqbal6@sbp.org.pk)
[*] All authors contributed equally in this research study.

Non-technical Summary

Central banks use monetary policy as a tool to tame inflation without being prejudice to economic growth and without compromising financial stability. While taking monetary policy stance, policy makers are forward looking. Inflation forecasting play a crucial role in such forward guidance to policy makers. Central banks use different models, including theoretic and atheoretic, to forecast inflation. Uncertainties associated with each model's inflation forecast may result from structural break(s) in inflation series and/or from unaccounted for nonlinearities in structural relationship(s) used in the modeling.

With advancements in computing power and in the field of computer science, artificial neural network (ANN) has emerged as a tool for forecasting. It identifies patterns in large data sets and accounts for nonlinearities and structural changes in the data generating process of the variable of interest. An ANN - consisting of a combination of certain network specific characteristics - is 'trained' on the available data and the trained network is used to produce forecast based upon its 'learning' during the training process. Deciding upon an optimal combination of network specific characteristics is difficult because of an indefinite set of choices to start with. Moreover, an optimal network may produce altogether different forecast if trained again.

We propose a three-step process in order to address aforementioned issues with ANN based forecasting. At first stage, an optimal combination of chosen characteristics is selected using minimum root mean square error (RMSE) criterion. At second stage, a simulation based frequency distribution of 'minimum RMSE' is used to find the most likely optimal combination of characteristics. Finally, in order to have a reliable forecast, a large number of forecasts generated from the 'most likely' optimal network are averaged.

In this paper, we have used our thick ANN approach to obtain 24 months ahead YoY inflation rate forecasts for Pakistan using its data from January 1958 to December 2017. We found forecasts from the thick ANN model better than forecasts from a suite of inflation forecasting models of Pakistan.

1. Introduction

Objective of macroeconomic policy making is to achieve sustainable economic growth in low inflation environment. In addition to many other relevant data and survey reports, central banks also consider short to medium term inflation forecasts, to decide their monetary policy stance, particularly the policy interest rate. The reliability of these inflation forecasts matter a lot in effectiveness of monetary policy. During the last couple of decades, central bankers have developed many inflation forecasting models including theoretic and atheoretic. These models fail, at least occasionally, due to structural break(s) and unaccounted for nonlinearities in structural relationships.

With availability of large magnitude of high frequency data and advancements in computing, forecasters have started using Machine Learning for better forecasting, particularly to account for structural break(s) in the series (of variable of interest) and/or nonlinearities in the underlying relationships. Machine Learning (ML) has different philosophy behind it compared to Traditional Statistics (TS). There is a significant difference in approach and applications of the two: ML is a branch of Artificial Intelligence (AI) which aims to discover regularities in data through pattern recognition so that these regularities can be generalized (with purpose). ML heavily relies on computing power whereas TS techniques were mostly developed when currently available computing powers were not available and statisticians had to rely on small samples and relatively slow computations that forced them to make heavy assumptions about data and its distributions (Hassibi, 2016).

Central banks are not behind in utilizing ML for policy related analysis yet there is a little evidence for use of these techniques by central banks for inflation forecasting as regular tool. Only 22% central banks use ML tools according to a survey from 50 central banks[1]. These tools are data driven and are able to identify the underlying data generating process of a time series making them very effective in forecasting. One of the tools to forecast inflation under the broader umbrella of ML is Artificial Neural Networks (ANN). It keeps a part of the data for testing its own ability to identify the pattern(s) for prediction, and uses backpropagation to reach a trained network. The trained network is later used for forecasting. Even though ANN can be a very useful tool for macroeconomic forecasting yet it is not very widely used due to complexity in choice of an optimal architecture and has often been termed as a black box [Olden & Jackson (2002)].

Forecasting using ANN is not new and has been under academic investigation along with its use by practitioners. Zhang, Patuwo & Hu (1998) provided a detailed review of research in the field of forecasting using ANN. The authors highlighted major issues while using ANN models, one of which is selecting the network architecture. A network consists of parameters like input nodes, output nodes, and hidden layers & nodes. Values of these parameters define a network's architecture and depends upon the underlying pattern detected in the given time series and desired period for forecast. Each node is assigned a random weight which is optimized based on the gradient descent using a training algorithm. Various approaches have been designed to determine the optimal architecture. These approaches are either based on rules of thumb or statistical techniques, but both approaches have been criticized. Rules of thumb might be very subjective for a forecaster and might not guarantee an optimal architecture once the structure of a series is changed. In case of statistical techniques, the weights are unknown and are

[1] Glass, E. (2017, Nov 06). *Big Data in Central Banks: 2017 Survey*. Retrieved from https://www.centralbanking.com/central-banks/economics/data/3315546/big-data-in-central-banks-2017-survey

identified using a training algorithm. Initial weights are chosen randomly, in the start of each training, and thus the final weights of a trained network differ. Curry & Morgan (2006) highlighted the issue(s) arising out of random initialization of weights at the start of training process. Further, this difference (of weights) between two trained networks on same data and same architecture leads to reluctant use of ANN by practitioners.

Another important aspect, which has not received much attention under the literature addressing the optimality of a network, is network characteristics. The forecasts may also differ from various other network characteristics including the choice of training algorithm and activation function, even if the most commonly used training algorithm (Levenberg-Marquardt backpropagation) and activation functions (hyperbolic tangent and log sigmoid) are applied. If ANN is to be considered as one of the tools for (inflation) forecasting, it is imperative to (a) move ahead of optimization of architecture alone and consider an optimum combination of a training algorithm and activation function along with an architecture, and (b) address the reliability and stability issue of forecasts from an optimum combination. This paper aims to develop a scheme where reliable ANN based forecasts can be obtained through multistage simulations after reaching an *optimal set of architecture, training algorithm and an activation function of an ANN model*.

We contributed to literature in at least four dimensions. First, for forecasting a given time series (Pakistan's year on year inflation in this study), we propose considering a combination of a training algorithm and an activation function along with an architecture at optimization stage. Second, we propose a simulation based frequency distribution of optimal set of architecture, training algorithm and an activation function, to identify 'modal set'. Third, solution for the problem of instability in forecasts from same architecture owing to randomly initialized initial weights has been proposed using a simulation based thick ANN model. Fourth, we use a new measure of forecast evaluation in this study as conventional measures (like RMSE) fail to account for systematic bias in forecasts.

The paper is organized as follows. Section 2 briefly summarizes the textbook knowledge on basic procedure of ANN. Section 3 presents critical review of literature on use of ANN for forecasting. Section 4 briefly discusses the data used for inflation forecasting of Pakistan. Section 5 presents the proposed scheme – consisting of optimization and multistage simulations. Section 6 presents the results while section 7 concludes the paper.

2. Recapitulation of ANN

The human brain is comprised of multiple neurons which are interconnected with each other making it billions of connections which process information. ANN is a basic level attempt to process nonlinear learning like human brain. It is comprised of input nodes, hidden layers consisting of hidden nodes, and output nodes. It is trained on a part of given set of data, based on which its predictions are compared with actual available data. The error in prediction is fed back to the network for necessary adjustments. There are three distinct functional operations that take place in a mathematical neuron. These are: the weight function, the net input function and the transfer function. Assuming inputs denoted by $x_1, x_2\ldots\ldots, x_n$ are fed to the neurons where each one is multiplied by a random initial weight w_i to form the sum $\Sigma w_i x_i$. This is the most commonly weighting function used although there are other types of weighting functions. This leads to the input function where a bias "b_i" is added: $b_i+\Sigma w_i x_i$. The w_i is like a parameter in liner function, x_i is a variable while the b_i is similar to a constant. To get the output the net input is processed through a transfer function and activation function. Hence, the three process can be represented as:

$$y = f\left(b_i + \sum_{i=1}^{n} w_i x_i\right) \qquad (I)$$

Although liner transfer functions are also used but the most commonly applied transfer functions are log-sigmoid (having values between 0 and 1) and hyperbolic tangent sigmoid function (having values between -1 and +1). Reason for popularity of these functions is the simplicity in calculation of their first derivative which is required for weight adjustment during back-propagation. Hence in each case equation I would become:

$$y = tansig\left(b_i + \sum_{i=1}^{n} w_i x_i\right) \qquad (II)$$

$$y = logsig\left(b_i + \sum_{i=1}^{n} w_i x_i\right) \qquad (III)$$

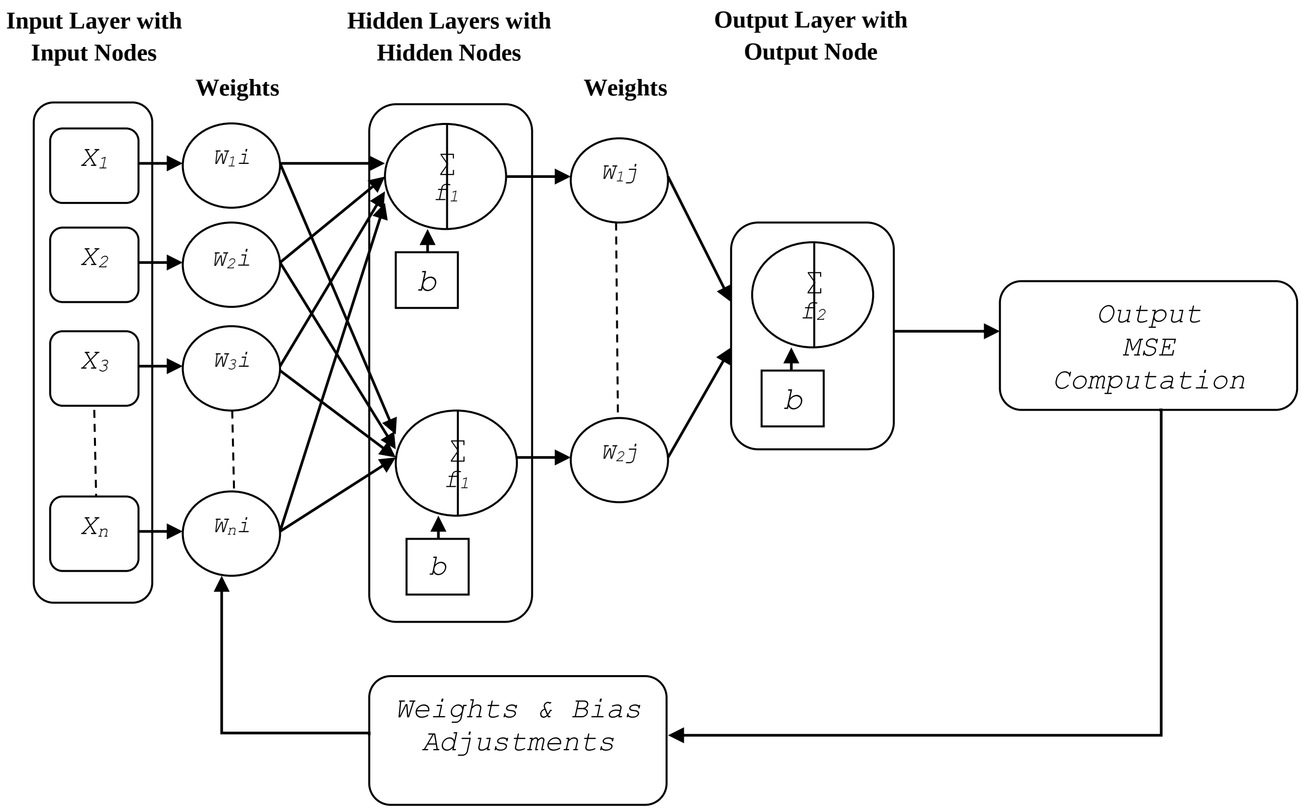
Once the output is generated it is compared with the actual output for calculation of error. The objective is to minimize the error through adjusting w_i in the direction where error is minimized. This description is of a single layer ANN. In order to learn various patterns multilayer neural networks are used which have either one or few hidden layers in between the input and output layers (Figure 1).

Each hidden layer has certain number of pre-decided neurons. The process is similar where the hidden unit also attaches a weight to the input received and then delivers output which becomes the input of the final output layer. The process that takes place in hidden unit can be represented as:

$$G_j = g\left(b_j + \sum_{i=1}^{n} \gamma_{ij} x_i\right) \qquad (IV)$$

where subscript i and j represent input and hidden units respectively. x_i is the input to hidden layer and γ_{ji} is a weight connecting the input and hidden unit while b_j represents bias attached. The function g is the activation function which gives the output G_j.

Figure 1 A: General Layout of an ANN with Back-propagation



As already mentioned, this output becomes input of the output layers and similar process initiates which has the following form:

$$F_h = f\left(b_h + \sum_{j=1}^{n} \beta_{hj} G_j\right) \qquad (V)$$

Where, the subscripts h and j represent hidden and output units respectively. β_{hj} is the weight while b_h is the bias added at this layer. The final output of output layer is F_h. After transfer function *f* is applied we obtain:

$$F_h = f\left(b_h + \sum_{j=1}^{n} \beta_{hj} g\left(b_j + \sum_{i=1}^{n} \gamma_{ij} x_i\right)\right) \qquad (VI)$$

This is the forecast based on ANN model.

3. Literature Review

Artificial neural networks have been extensively used for forecasting in many fields including but not limited to rainfall (Chang, Rapiraju, Whiteside & Hwang, 1991), airborne pollen (Arizmendi, Sanchez, Ramos & Ramos, 1993), industrial production (Aiken, Krosp, Vanjani & Govindarajulu, 1995), wind pressure profile (Turkkan & Srivastava, 1995), and international airline passenger traffic (Nam and Schaefer, 1995). It has also been applied to financial data and macroeconomic variables like commodity prices (Kohzadi, Boyd, Kermanshahi & Kaastra, 1996), bankruptcy and business failure (Odom & Sharda, 1990; Coleman, Graettinger & Lawrence, 1991; Salchenberger, Cinar & Lash, 1992; Tam & Kiang, 1992; Fletcher & Goss, 1993; Wilson & Sharda, 1994), and credit scoring (Blanco, Pino-Mejías, Lara & Rayo, 2013; Khashman, 2011). There is an extensive list of research studies where authors have used various techniques within ANN domain to forecast variables of their interest. Some models are univariate while others include exogenous variables. However, the focus of this study is on a univariate model for forecasting inflation.

ANN based forecasting research has received contributions from researchers in academia as well as from practitioners including central bankers. Although, little is known which central bank(s) use macroeconomic forecasts from ANN models as a policy guide, yet there is published research by central bankers using such tools: Andreou & Zombanakis (2006) forecasted Euro exchange rate versus the U.S. Dollar and the Japanese Yen. Hall, Muljawan, Suprayogi & Moorena (2008) applied ANN to assess bank credit risk for Indonesia. Monge (2009), Haider & Hanif (2009), and Choudhary & Haider (2012) forecasted inflation of Costa Rica, Pakistan, and 28 OECD countries respectively. Chakraborty & Joseph (2017) used ML techniques for forecasting UK CPI inflation for two years as well as for detection of alerts on the balance sheets of financial institutions in the context of banking supervision. León, Moreno & Cely (2017) used pattern recognition algorithm to classify Columbian commercial banks based on 25 indicators of banks' financial position.

The research in ANN based forecasting has resulted into two broad streams of literature. First is where the ANN forecasts of macroeconomic series are compared with other econometric model based forecasts using various performance criterion. For example: ANN with Autoregressive Integrated Moving Average (ARIMA) based models - Haider & Hanif (2009); Binner, Bissoondeeal, Elger, Gazely & Mullineux (2005); Lee, Sehwan & Jongdae (2007); Merh, Saxena & Pardasani (2010); Adebiyi, Adewumi & Ayo (2014); Yao, Tan & Poh (1999); Wijaya, Kom & Napitupulu (2010); Hansen, McDonald & Nelson (1999)- ANN with various Vector Autoregressive (VAR) models - Binner et. Al. (2005), Aydin & Cavdar (2015), Mirmirani & Cheng (2004), Pendar & Haji (2017) - and ANN with various types of Autoregressive Conditional Heteroscedasticity (ARCH) models - Binner et al. (2005); Dhamija & Bhalla (2010); Maciel & Ballini (2010); Charef & Ayachi (2016); Pendar & Haji (2017). Such studies tend to evaluate whether the forecasts from ANN outperform the econometric models or not. There have been mixed results across type of models, the series being forecasted, as well as time periods. Some have found ANN to perform better over short horizons while other have concluded that ANN is a better approximation for the time series undergoing structural changes [McAdam & McNelis (2005)].

The second strand of literature has attempted to answer the issue of optimal choice of model but mostly revolves around network architecture. The input nodes can automatically be determined by number of variables as well as the lag structure of a particular series while the output nodes are based on forecast horizon. *The most thought over issue is identifying the number of hidden layers and nodes as well as the choice of final model (trained network).* Some have used certain rules of thumb [Shirakawa, Shimizu, Okubo & Daido (1998)] while other have used statistical methods [Qi & Zhang (2001); BuHamra, Smaoui & Gabr (2003)] or various algorithms [Siestema and Dow (1988); Reed (1993); Roy, Kim & Mukhopadhyay (1993)], while others have employed hit and trial sort of mechanisms [Kumar, Aggarwal & Sharma (2013)]. Hansen, McDonald & Nelson (1999) argued that several rules of thumb exist to identify number of nodes in hidden layers like using the average of number of input and output units. Using such rules, however, does not guarantee the effectiveness of an architecture. Hence this leads to testing several network architectures to find out the most optimal one. Murata, Yoshizawa & Amari (1994) suggested the network information criterion which is a generalized Akaike's Information Criterion (AIC) applicable to unrealized models under a general loss function including a regularization term. Their proposed criteria measures the relative merits of two models and determines whether or not more neurons should be added to a network. However, Eğrioğlu, Aladağ & Günay (2008) proposed a new strategy based on AIC, BIC, RMSE, MAPE, DA, and modified direction accuracy (MDA) while Wang, Massimo, Tham & Morris (1994) introduced the canonical decomposition technique. Kursa & Rudnicki (2010) preferred using statistical significance of features to remove the less significant ones over the random trials.

Bredahl Kock & Teräsvirta (2016) comprehensively investigated three automated modeling techniques for selection of an optimal model namely; QuickNet (QN) [White (2006)], Autometrics [Doornik (2009)], and Marginal Bridge Estimator (MBE) [Huang, Horowitz, & Ma (2008)]. The authors aimed to focus on the best model selection technique out of these three, and not the most optimal non-liner model. They compared direct and recursive forecasts as well as forecasts with differenced and level variables. They preferred direct forecasts over recursive and level variables over differenced ones; still they also argued that results differ across countries: In terms of model selection technique, the authors concluded that MBE delivers the best model among the three. Autometrics selects models very well when there is a non-linear model which fits the data, otherwise it does not. Moreover, it also performs poorly when the series is a simple first order AR model. The QN's performance lies somewhere in between the other two techniques.

Curry & Morgan (2006) selected the best fitting model using R^2 as a criteria of comparison among networks; based on two training algorithms; polytope and backpropagation, having 1-2 hidden layers, and two different activation functions: logistic activation and hyperbolic tangent. While changing various hidden nodes in 1-2 layers the authors carried out five simmulations for each. The weights were initialized randomly in each case. Using three datasets the authors concluded that model's level of fit can be replicated but they were unable to replicate the precise values of weights. Based on this the authors concluded that there is an identification problem in these models and associated this issue with their difficulty in choosing a statistical technique to identify the optimal structure. McAdam & McNelis (2005) used thick modeling strategy; combining forecasts of several ANNs, based on different numbers of neurons in the hidden layer, and different network architectures using genetic algorithm. The combination forecast is the trimmed mean forecast at each period, coming from an ensemble of networks. The authors used RMSE as one of the criteria to evaluate the thick model's performance with a liner model. In short both the studies discussed above, attempted to find the single or few most suitable model(s).

In summary, there have been attempts to identify the optimal network architecture. However, it is evident that no single approach has yet emerged which can be generalized. Once new data is available, which is a routine in macroeconomics, the previous training might not be as much applicable to the new scenario as was with the previous dataset. Hence, with new information, a new network has to be trained which might have different architecture from the previously trained network. Most importantly, the least attempted issue in literature (despite being highlighted by Curry & Morgan (2006)) is the complication arising from randomly generated weights as well as the capability of gradient descent process to end up with different weights after every training session. Further, to the best of our knowledge, there has been no study to address the choice of optimal ANN considering multiple network characteristics (set of architecture, training algorithm and an activation function) and not just the network architecture. This paper proposes a three optimization levels scheme to obtain a reliable forecast while addressing all the aforementioned gaps in the existing literature. Moreover, forecast evaluation in this study is made using a newly proposed formula (by Hanif, Iqbal & Mughal (2018)) for forecast evaluation.

4. Data

We used monthly year on year (YoY) inflation rate of Pakistan from Jan 1958 to Dec 2017. Data for last 24 months (Jan 2016 to Dec 2017) was kept out of the net to test its out of sample forecast performance. Hence, the data used for ANN was from Jan 1958 to Dec 2015. Out of which, we used 70% of the data for training and remaining 30% was divided equally for testing and validation. We have used Neural Network Toolbox from MATLAB 2017R. We employed Nonlinear Autoregressive (NAR) formulation which trains a network for univariate time series and is a feed-forward network with backpropagation. Our output nodes were defined by the forecast horizon, which we kept at 24 months (Jan 16 – Dec 17).

5. Proposed Scheme

Before going into the steps involved in our proposed scheme, it will be useful if we select values for a couple of parameters that will be required in the simulation. These include output nodes (or forecast horizon) and the number of input nodes (lags). Forecast horizon generally useful for monetary policy makers in case of monthly inflation rate is 24 months. After going through the literature on inflation forecasting in Pakistan (for example Hanif & Malik (2015)) we used 13 lags in this study. For hidden nodes and layers; we tested for 10 to 50 hidden nodes (with intervals of 10) and in 1 to 2 hidden layers. In case of 2 layers the numbers of nodes were equally divided among each layer and same activation function was used in both hidden layers.

In order to train the network we test the performance of two more available training algorithms in the (MATLAB) toolbox i.e. scaled conjugate gradient backpropagation (SCG) and resilient backpropagation (RP) in addition to the most widely used algorithm - Levenberg-Marquardt backpropagation (LM). Similarly, we tested two most widely practiced activation functions i.e. log-sigmoid and hyperbolic tangent sigmoid[2]. We used same activation function in all layers and also varied the activation function by creating hybrid combinations in each case, where the first and second layer activation functions were different. For example, the Hybrid 1 for LM training algorithm with 1 layer consists of log-sigmoid activation function in the first layer while hyperbolic tangent sigmoid in the output layer. All these

[2] Log-sigmoid function has the form ${}^{1}\!/_{1\,+\,e^{-x}}$ and gives the output in the range [0,1] while tan sigmoid of hyperbolic tangent function has the form ${}^{e^{x}\,-\,e^{-x}}\!/_{e^{x}\,+\,e^{-x}}$ and has [-1,1] output range.

choices create multiple sets of neural networks characteristics[3] (based on different architectures, training algorithms, and activation functions) to be trained and tested.

Now we formulate our proposed scheme where the issue of identifying most optimal architecture and the stability in forecast with new trained network owing to randomly initialized weights can be addressed. Our scheme is described as in the following:

i. *The first stage is identifying the **most optimal set** of architecture, training algorithm and activation function. For this a set of these parameters and characteristics is run with various values, and forecasts are collected from each. Comparison is made using the minimum RMSE.*

ii. *At second stage, a frequency distribution is generated for all the sets with 1000 simulations, where minimum RMSE is identified among all the sets in each simulation. Hence, we find **optimal modal ANN** (most frequent set based on minimum RMSE) of this distribution.*

iii. *At third stage the optimal modal ANN is rerun enough number of times to train multiple networks, whose forecast is then averaged to formulate a thick ANN model based forecast. The purpose of this averaging after having 1000 forecasts for creating a **thick model** is to ensure that the random weight initialization issue is addressed.*

6. Results and Discussion

We used minimum (out of sample) RMSE as criteria for model selection. We went through three levels of optimization proposed in this study to find the best possible model from among all exhaustive options within the set of choices we have used (which can be different for different studies).

At the first stage, we conducted multiple simulations from 10 to 50 hidden nodes with a gap of 10 nodes in one layer while holding a training algorithm and activation function fixed, hence having multiple network architectures. This makes a total of five 24 month forecasts from a single algorithm and activation function when we have one hidden layer. Once we include another hidden layer, the nodes are divided equally among the two layers but the activation function is fixed and we have five more 24 months out of sample forecasts at the same algorithm and activation function. With hybrid layers from each training algorithm in 1 and 2 layers we have a total of 36 combinations to be tested.

[3] Although there can be indefinite number of ANN models with different characteristics based on various network architectures (input & output nodes, and hidden layers & nodes) training algorithm and activation functions but the study has been limited to 180 ANN models for the sake of computational ease.

Table 1: Out of Sample RMSE of 180 Forecasts with Multiple Hidden Nodes & Layers, Training Algorithms, and Activation Functions								
Layers and Activation Function Combination			**Training Algorithm**	**Number of Nodes**				
				10	20	30	40	50
Same Activation Function	**1 layer**	tansig	LM	0.64	0.71	0.64	0.60	1.20
			SCG	1.30	0.91	0.83	1.96	1.65
			RP	0.62	0.85	0.60	1.72	6.96
		logsig	LM	9.84	9.84	9.84	9.84	23.19
			SCG	10.02	9.97	9.90	9.84	9.84
			RP	10.42	22.08	10.26	25.32	9.86
	2 layers	tansig	LM	1.06	0.53	0.71	0.52	0.67
			SCG	0.60	0.64	14.25	0.73	1.97
			RP	2.24	0.71	1.92	0.74	2.35
		logsig	LM	9.84	9.84	9.90	9.84	9.84
			SCG	23.66	21.10	23.72	9.84	9.92
			RP	9.88	24.59	9.97	9.84	9.86
Hybrid	**1 Layer**	Hybrid 1 (tan-log)	LM	9.84	9.84	9.84	9.84	9.84
			SCG	10.03	9.93	10.05	23.06	10.22
			RP	10.35	10.08	9.87	9.99	9.84
		Hybrid 2 (log-tan)	*LM*	0.85	0.89	0.83	0.57	0.79
			SCG	3.69	0.72	0.76	1.84	5.89
			RP	0.72	1.07	1.17	1.10	0.92
	2 Layer	Hybrid 3 (log-log-tan)	LM	1.17	0.79	0.99	0.59	*0.51*
			SCG	0.64	0.94	3.73	0.78	0.77
			RP	2.37	1.68	3.20	0.78	0.56
		Hybrid 4 (log-tan-log)	LM	26.05	23.03	19.31	21.89	9.84
			SCG	9.84	9.95	10.48	21.87	9.84
			RP	9.85	9.88	9.85	9.84	9.85
		Hybrid 5 (tan-log-log)	LM	9.84	9.84	9.84	9.92	9.84
			SCG	9.89	21.91	23.45	23.10	22.70
			RP	24.73	9.84	9.84	22.56	23.46
		Hybrid 6 (tan-tan-log)	LM	23.40	10.82	9.84	9.84	9.84
			SCG	24.66	21.71	10.10	24.22	9.84
			RP	18.98	9.84	9.85	24.69	9.84
		Hybrid 7 (tan-log-tan)	LM	1.18	0.92	1.79	0.54	2.16
			SCG	1.00	2.12	0.98	0.93	1.31
			RP	1.90	2.88	2.54	1.59	0.68
		Hybrid 6 (log-tan-tan)	LM	14.25	0.54	1.49	0.86	14.25
			SCG	0.65	1.09	2.13	1.04	0.89
			RP	1.76	1.51	0.88	1.04	1.14

Table 2: Frequency Distribution of Minimum RMSE from 1000 Simulations

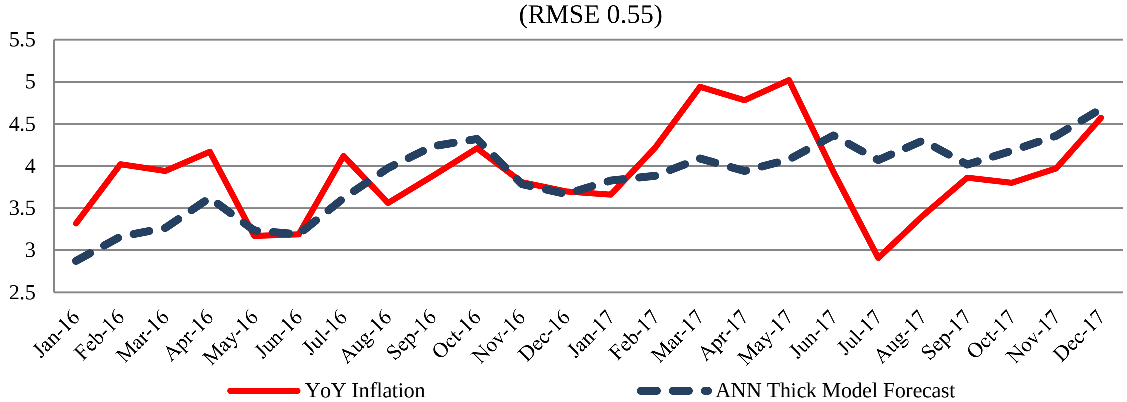
Layers and Activation Function Combination			Training Algorithm	Number of Nodes				
				10	20	30	40	50
Same Activation Function	1 layer	tansig	LM	10	24	25	18	20
			SCG	16	13	14	5	4
			RP	16	8	22	14	16
		logsig	LM	20	27	23	19	19
			SCG	0	0	0	0	0
			RP	0	0	0	0	0
	2 layers	tansig	LM	4	15	14	11	13
			SCG	11	8	5	4	8
			RP	7	9	11	5	12
		logsig	LM	0	0	0	0	0
			SCG	0	0	0	0	0
			RP	0	0	0	0	0
Hybrid	1 Layer	Hybrid 1 (tan-log)	LM	0	0	0	0	0
			SCG	0	0	0	0	0
			RP	0	0	0	0	0
		Hybrid 2 (log-tan)	**LM**	8	26	25	24	**29**
			SCG	4	5	3	2	4
			RP	2	8	6	3	11
	2 Layer	Hybrid 3 (log-log-tan)	LM	3	21	19	18	22
			SCG	9	7	4	7	0
			RP	6	5	6	13	14
		Hybrid 4 (log-tan-log)	LM	0	0	0	0	0
			SCG	0	0	0	0	0
			RP	0	0	0	0	0
		Hybrid 5 (tan-log-log)	LM	0	0	0	0	0
			SCG	0	0	0	0	0
			RP	0	0	0	0	0
		Hybrid 6 (tan-tan-log)	LM	0	0	0	0	0
			SCG	0	0	0	0	0
			RP	0	0	0	0	0
		Hybrid 7 (tan-log-tan)	LM	5	16	11	20	14
			SCG	11	6	11	2	4
			RP	6	12	15	12	11
		Hybrid 6 (log-tan-tan)	LM	5	11	19	22	23
			SCG	6	3	5	5	8
			RP	4	12	13	17	15

Table 1 presents all the RMSEs for each forecast calculated against out of sample inflation rate for the period Jan 16 to Dec 17. Once all the results are obtained we select forecast with lowest RMSE. The minimum RMSE of 0.51 – shaded and italics in Table 1 – is with the network architecture that has LM algorithm having 2 hidden layers and 50 hidden nodes and hybrid 3 as combination of activation functions. 50 hidden nodes in total means there are 25 nodes in each hidden layer. Table 1 is showing results of just one exercise.

In order to achieve stage 2 optimization we repeat this exercise 1000 times which gives us a frequency distribution of lowest RMSE. These results are presented in Table 2. The results show that for the given dataset and time period the probability of lowest RMSE in a given simulations is highest for hybrid 2 combination – shaded & bold in Table 2 – trained through LM algorithm with 1 layers having 50 nodes with log-sigmoid activation function in the first layer and tan-sigmoid activation function in the output layer. This exercise has resulted into identification of most optimal modal ANN[4].

In order to obtain reliable and stable forecasts, we move to third stage. Once the architecture with minimum RMSE is identified we obtain 1000 out of sample 24 months' forecasts with hybrid combination 2. Figure 2 presents the final forecast which is the average of these 1000 simulations.

Figure 2: ANN Thick Model Forecast
(RMSE 0.55)



The exercise with repeated simulations and creating a thick model results into stability of forecasts. To show the robustness of our approach we developed 10 thick ANN models each consisting of 1000 forecasts. Their RMSE varied from 0.54 to 0.59 (Table 3) and we consider this stable. Following the practice in usual ANN forecast based literature; we also compared the forecasting performance of our thick ANN model forecast with a suite of inflation forecasting models published in Hanif & Malik (2015). The authors investigated the performance of their 16 inflation forecasting models for Pakistan with a specific focus on performance across various inflation regimes (low, moderate, and high). The authors concluded that ARDL type models were the most preferred ones among their estimated models.

[4] The code can easily be expanded to include more layers and test other algorithms as well as various combinations of activation functions.

Table 3: Stability of ANN Thick Model	
ANN Thick Model	**RMSE**
1	0.54
2	0.57
3	0.56
4	0.54
5	**0.53**
6	**0.59**
7	0.55
8	0.56
9	0.54
10	0.55

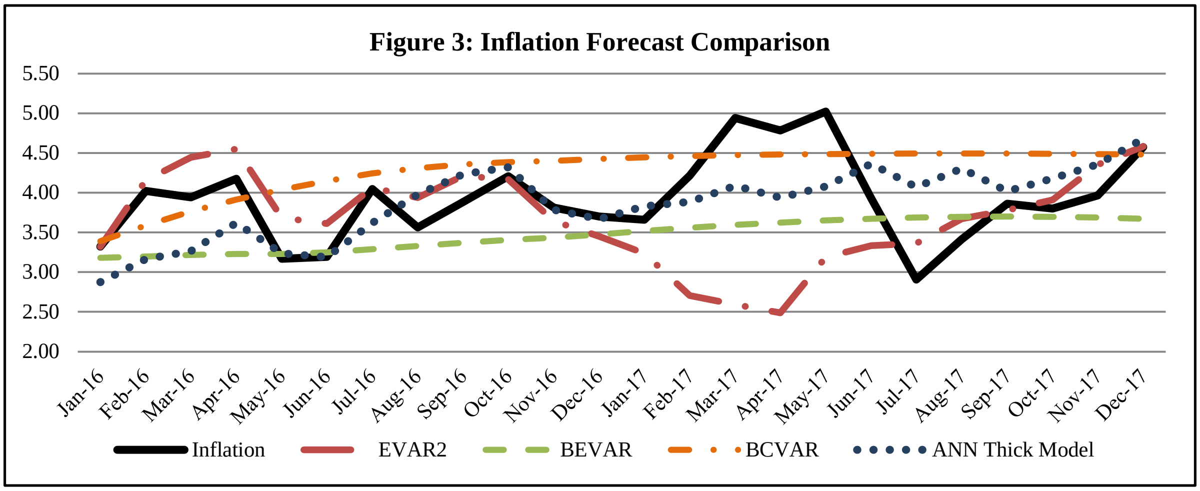
Considering new data, we re-estimated Hanif & Malik (2015) models to forecast inflation for 24 months from Jan 2016-Dec 2017 (using the data from Jan 1992 to Dec 2015). While comparing inflation forecasts from various type of models , it is conventional to use popular criterion of forecast evaluation from closely knitted family of criteria, like RMSE. However, these criteria fail to account for systematic bias in a forecast, i.e. persistent over/under forecast, since they only consider the distance between actual and forecast (in one manner or other). Hanif, Iqbal and Mughal (2018), referred to as HIM (2018) thereafter, has introduced a new out-of-sample forecast evaluation criterion which considers not only the RMSE but also accounts for systematic bias, if any, in the forecast by using 'number of times' the forecast series crosses the actual series . HIM (2018) forecast evaluation criteria is:

$$\frac{1}{2} * (\frac{1}{e^{RMSE}} + \frac{c}{h-1})$$

where, c is the number of times forecast series crosses the actual series and h is the number of forecast horizons. This formula is applied to cases where forecast is evaluated for at least two periods ahead. In ideal (worst) situation when forecasts perfectly match (off track) with (from) the observed underlying series this sum will be 1 (0). Higher the value for HIM criterion, better is the forecast.

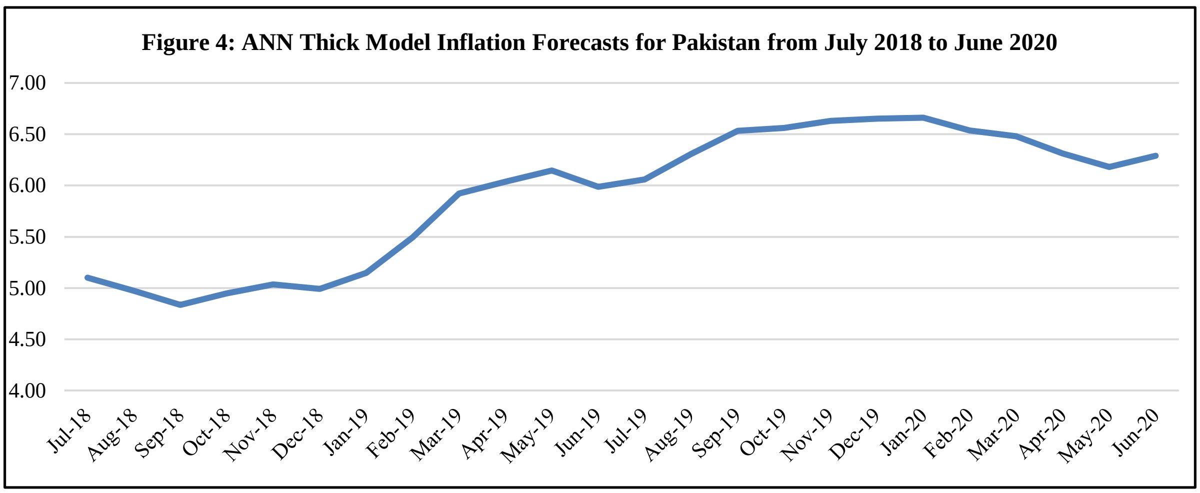
Table 4: Comparison of Inflation Forecasting Models of Hanif & Malik (2015)		
Model	**Criteria**	
	RMSE	**HIM (2018)**
EVAR2	0.88	0.31
BEVAR	0.68	0.34
BCVAR	0.65	0.37
ANN Thick Model	0.55	0.44

Based on RMSE and HIM (2018) criteria - results in Table 4 - the three best models from Hanif and Malik (2015) are a) External Structural VAR (EVAR2), b) Bayesian External VAR model (BEVAR), and c) Bayesian Credit VAR model (BCVAR) for inflation forecasts from Jan 2016 – Dec 2017. ANN thick model has the lowest RMSE and the highest value for HIM (2018) measure when compared with the best performing model from Hanif and Malik (2015). A comparative depiction of 24 months forecasts based on this study with the best 3 from Hanif & Malik (2018) is also made in Figure 3.



The Bayesian models have too little variation across 24 months horizon (when compared with ANN and EVAR2). The dispersion between realized inflation and forecasts from EVAR2 increases from Oct 16 to Jun 17. The inflation forecasts from ANN-thick model are very close to realized inflation.

Based on our three stage ANN scheme, the following figure presents out of sample inflation forecasts for Pakistan for the future - from July, 2018 to June, 2020 (Figure 4)[5].



[5] These forecasts are not official forecasts of the institutions where authors are employed.

7. Summary and Conclusion

The main purpose behind this study was using artificial neural networks (ANN) to obtain better (inflation) forecasts (for Pakistan) while improving reliability of forecasts for practical purpose. The literature on ANN based forecasting highlights issues like choice of an optimal network architecture and different forecast with every retrained model (emanating from random initialization of weights). Moreover, from a practitioner's perspective, it is necessary to go beyond an optimal architecture and consider the complete network characteristics (including training algorithm, and activation function). There is lack of consensus in the literature on approach to identify an optimal network architecture let alone the issues pertaining to random weight initialization and complete network characteristics.

We developed and implemented a three-stage scheme, using RMSE of out of sample 24 period ahead forecasts of (inflation). At stage I, we reach an optimum set of characteristics - architecture, training algorithm and activation function - with the lowest RMSE from 180 different possible sets. At stage II, we run 1000 simulations of stage I and generate a frequency distribution for 180 different sets. 'Mode' or 'most appearing set' is selected as an optimal modal combination of ANN characteristics. However, owing to random weight initialization, this optimal modal combination may result into a different network on each training and possibly different forecasts. Hence, the stability in forecasts could still be at stake. At stage III, we run 1000 simulations using optimal modal combination of ANN characteristics to generate 1000 out of sample 24 period ahead (inflation) forecasts. Average of these 1000 forecasts are then used as thick and reliable ANN based 24 period ahead (inflation) forecasts (for Pakistan). Robustness of the proposed scheme is also ensured by doing 10000 simulations.

This thick ANN model is to found to outperform all the 15 econometric models of Pakistan economy [in Hanif & Malik (2015)] for forecasting 24 months ahead headline inflation based upon a newly introduced approach to forecast evaluation.

An ANN thick model for forecasting scheme, designed in this paper, can be applied to forecast economic indicators (like stock prices, exchange rate, industrial production etc.) and variables from other fields (like marketing, environment, psychology etc.).

References

Adebiyi, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Comparison of ARIMA and artificial neural networks models for stock price prediction. *Journal of Applied Mathematics*, 1-7.

Aiken, M., Krosp, J., Vanjani, M., & Govindarajulu, C. (1995). A neural network for predicting total industrial production. *Journal of End User Computing, 7*, 19-19.

Andreou, A. S., & Zombanakis, G. A. (2006). *Computational intelligence in exchange-rate forecasting*. Bank of Greece, Working Paper No. 49.

Arizmendi, C. M., Sanchez, J. R., Ramos, N. E., & Ramos, G. I. (1993). Time series predictions with neural nets: application to airborne pollen forecasting. *International journal of biometeorology*, *37*(3), 139-144.

Aydin, A. D., & Cavdar, S. C. (2015). Comparison of prediction performances of artificial neural network (ANN) and vector autoregressive (VAR) Models by using the macroeconomic variables of gold prices, Borsa Istanbul (BIST) 100 index and US Dollar-Turkish Lira (USD/TRY) exchange rates. *Procedia Economics and Finance, 30*, 3-14.

Blanco, A., Pino-Mejías, R., Lara, J., & Rayo, S. (2013). Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with applications*, *40*(1), 356-364.

Binner, J. M., Bissoondeeal, R. K., Elger, T., Gazely, A. M., & Mullineux, A. W. (2005). A comparison of linear forecasting models and neural networks: an application to Euro inflation and Euro Divisia. *Applied economics*, *37*(6), 665-680.

Bredahl Kock, A., & Teräsvirta, T. (2016). Forecasting macroeconomic variables using neural network models and three automated model selection techniques. *Econometric Reviews*, *35*(8-10), 1753-1779.

BuHamra, S., Smaoui, N., & Gabr, M. (2003). The Box–Jenkins analysis and neural networks: prediction and time series modelling. *Applied Mathematical Modelling*, *27*(10), 805-815.

Chakraborty, C., & Joseph, A. (2017). *Machine learning at central banks*. Bank of England, Working Paper No. 674.

Chang, I., Rapiraju, S., Whiteside, M., & Hwang, G. (1991). A neural network to time series forecasting. *Proceedings of the Decision Science Institute* (Vol. 3, pp. 1716-1718).

Charef, F., & Ayachi, F. (2016). A Comparison between neural networks and GARCH models in exchange rate forecasting. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, *6*(1), 94-99.

Choudhary, M. A., & Haider, A. (2012). Neural network models for inflation forecasting: an appraisal. *Applied Economics*, *44*(20), 2631-2635.

Coleman, K. G., Graettinger, T. J., & Lawrence, W. F. (1991). Neural networks for bankruptcy prediction: The power to solve financial problems. *AI Review*, *4*, 48-50.

Curry, B., & Morgan, P. H. (2006). Model selection in neural networks: some difficulties. *European Journal of Operational Research*, *170*(2), 567-577.

Dhamija, A. K., & Bhalla, V. K. (2010). Financial time series forecasting: comparison of neural networks and ARCH models. *International Research Journal of Finance and Economics*, *49*, 185-202.

Doornik, J. A. (2009). Autometrics. In Castle, J. L., Shephard, N., eds. *The Methodology and Practice of Econometrics*. Oxford University Press: United Kingdom.

Eğrioğlu, E., Aladağ, Ç. H., & Günay, S. (2008). A new model selection strategy in artificial neural networks. *Applied Mathematics and Computation, 195*(2), 591-597.

Fletcher, D., & Goss, E. (1993). Forecasting with neural networks: an application using bankruptcy data. *Information & Management, 24*(3), 159-167.

Haider, A., & Hanif, M. N. (2009). Inflation forecasting in Pakistan using artificial neural networks. *Pakistan Economic and Social Review, 47*(1), 123-138.

Hall, M. J., Muljawan, D., Suprayogi, & Moorena, L. (2009). Using the artificial neural network to assess bank credit risk: a case study of Indonesia. *Applied Financial Economics, 19*(22), 1825-1846.

Hanif, M., Iqbal, J., & Mughal, K. S., (2018). *A new approach to forecast evaluation*. Unpublished, State Bank of Pakistan.

Hanif, M., & Malik, M. J. (2015). Evaluating the Performance of Inflation Forecasting Models of Pakistan. *SBP Research Bulletin*, *11*, 43-78.

Hansen, J. V., McDonald, J. B., & Nelson, R. D. (1999). Time Series Prediction With Genetic Algorithm Designed Neural Networks: An Empirical Comparison With Modern Statistical Models. *Computational Intelligence*, *15*(3), 171-184.

Hassibi, K. (2016). *Machine Learning vs. Traditional Statistics: Different philosophies, Different Approaches*. Retrieved from: https://www.datasciencecentral.com/profiles/blogs/machine-learning-vs-traditional-statistics-different-philosophi-1

Huang, J., Horowitz, J. L., Ma, S. (2008). Asymptotic properties of bridge estimators in sparse high dimensional regression models. *Annals of Statistics*, *36*, 587–613.

Khashman, A. (2011). Credit risk evaluation using neural networks: Emotional versus conventional models. *Applied Soft Computing, 11*(8), 5477-5484.

Kohzadi, N., Boyd, M. S., Kermanshahi, B., & Kaastra, I. (1996). A comparison of artificial neural network and time series models for forecasting commodity prices. *Neurocomputing*, *10*(2), 169-181.

Kumar, R., Aggarwal, R. K., & Sharma, J. D. (2013). Energy analysis of a building using artificial neural network: A review. *Energy and Buildings, 65,* 352-358.

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software*, *36*(11), 1-13.

Lee, C. K., Sehwan, Y., & Jongdae, J. (2007). Neural network model versus SARIMA model in forecasting Korean stock price index (KOSPI). *Issues in Information System*, *8*(2), 372-378.

León, Carlos and Moreno, Jose and Cely, Jorge, (2017). Whose balance sheet is this? Neural networks for banks' pattern recognition. CentER Discussion Paper Series No. 2017-009. Available: https://ssrn.com/abstract=2922516

Maciel, L. S., & Ballini, R. (2010). Neural networks applied to stock market forecasting: An empirical analysis. *Journal of the Brazilian Neural Network Society*, *8*(1), 3-22.

McAdam, P., & McNelis, P. (2005). Forecasting inflation with thick models and neural networks. *Economic Modelling*, *22*(5), 848-867.

Merh, N., Saxena, V. P., & Pardasani, K. R. (2010). A comparison between hybrid approaches of ANN and ARIMA for Indian stock trend forecasting. *Business Intelligence Journal*, *3*(2), 23-43.

Mirmirani, S., & Cheng Li, H. (2004). A comparison of VAR and neural networks with genetic algorithm in forecasting price of oil. In *Applications of Artificial Intelligence in Finance and Economics* (pp. 203-223). Emerald Group Publishing Limited.

Monge, M. E. (2009). *Performance of Artificial Neural Networks in Forecasting Costa Rican Inflation*. Available: http://www.cemla.org/red/papers2009/COSTARICA-Esquivel.pdf

Murata, N., Yoshizawa, S., & Amari, S. I. (1994). Network information criterion-determining the number of hidden units for an artificial neural network model. *IEEE Transactions on Neural Networks*, *5*(6), 865-872.

Nam, K., & Schaefer, T. (1995). Forecasting international airline passenger traffic using neural networks. *Logistics and Transportation Review*, *31*(3), 239.

Odom, M. and R. Sharda, 1990. A neural network model for bankruptcy prediction. *Proceedings of the IEEE International Conference on Neural Networks* (San Diego, CA) 163-168.

Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": a randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling*, *154*(1-2), 135-150.

Pendar, M., & Haji, M. (2017). Comparison of Neural Network Models, Vector Auto Regression (VAR), Bayesian Vector-Autoregressive (BVAR), Generalized Auto Regressive Conditional Heteroskedasticity (GARCH) Process and Time Series in Forecasting Inflation in Iran. *International Journal of Industrial Mathematics*, *9*(2), 119-128.

Qi, M., & Zhang, G. P. (2001). An investigation of model selection criteria for neural network time series forecasting. *European Journal of Operational Research*, *132*(3), 666-680.

Reed, R. (1993). Pruning algorithms-a survey. *IEEE Transactions on Neural Networks*, *4*(5), 740-747.

Roy, A., Kim, L. S., & Mukhopadhyay, S. (1993). A polynomial time algorithm for the construction and training of a class of multilayer perceptrons. *Neural Networks*, *6*(4), 535-545.

Salchenberger, L. M., Cinar, E., & Lash, N. A. (1992). Neural networks: A new tool for predicting thrift failures. *Decision Sciences, 23*(4), 899-916.

Sietsma, J., & Dow, R. J. (1988, July). Neural net pruning-why and how. In *IEEE International Conference on Neural Networks* (Vol. 1, pp. 325-333). IEEE San Diego.

Shirakawa, K., Shimizu, M., Okubo, N., & Daido, Y. (1998). Structural determination of multilayered large-signal neural-network HEMT model. *IEEE Transactions on Microwave Theory and Techniques*, 46(10), 1367-1375.

Tam, K. Y., & Kiang, M. Y. (1992). Managerial applications of neural networks: the case of bank failure predictions. *Management Science, 38*(7), 926-947.

Turkkan, N., & Srivastava, N. K. (1995). Prediction of wind load distribution for air-supported structures using neural networks. *Canadian Journal of Civil Engineering, 22*(3), 453-461.

White, H. (2006). Approximate nonlinear forecasting methods. *Handbook of Economic Forecasting*, *1*, 459-512.

Wijaya, Y. B., Kom, S., & Napitupulu, T. A. (2010, December). Stock Price Prediction: Comparison of Arima and Artificial Neural Network Methods-An Indonesia Stock's Case. In *Advances in Computing, Control and Telecommunication Technologies (ACT), 2010 Second International Conference on* (pp. 176-179). IEEE.

Wilson, R. L., & Sharda, R. (1994). Bankruptcy prediction using neural networks. *Decision Support Systems, 11*(5), 545-557.

Yao, J., Tan, C. L., & Poh, H. L. (1999). Neural networks for technical analysis: a study on KLCI. *International Journal of Theoretical and Applied Finance, 2*(2), 221-241.

Wang, Z., Di Massimo, C., Tham, M. T., & Morris, A. J. (1994). A procedure for determining the topology of multilayer feedforward neural networks. *Neural Networks*, *7*(2), 291-300.

Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35-62.