

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

186,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Dynamic Factor Model and Artificial Neural Network Models: To Combine Forecasts or Combine Models?

Ali Babikir, Mustafa Mohammed and
Henry Mwambi

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/intechopen.71536>

Abstract

In this chapter, we evaluate the forecasting performance of the model combination and forecast combination of the dynamic factor model (DFM) and the artificial neural networks (ANNs). For the model combination, the factors that are extracted from a large dataset are used as additional input to the ANN model that produces the factor-augmented artificial neural network (FAANN). Linear and nonlinear forecasts combining methods are used to combine the DFM and the ANN forecasts. The results of the best combining method are compared to the forecasts result of the FAANN model. The models are applied to forecast three time series variables using large South African monthly data. The out-of-sample root-mean-square error (RMSE) results show that the FAANN model yields substantial improvement over the individual and best combined forecasts from the DFM and ANN forecasting models and the autoregressive AR benchmark model. Further, the Diebold-Mariano test results also confirm the superiority of the FAANN model forecast's performance over the AR benchmark model and the combined forecasts.

Keywords: artificial neural network, dynamic factor model, factor-augmented artificial neural network model, forecasts combination, forecasting

1. Introduction

Prediction of economic or financial variable using related independent variables could be done by either using a super model which contains all the available independent variables or using the forecast combination methodology. Generally, it is admitted in the literature of econometrics that the forecast obtained by all the information integrated in one step is much better than the combination of forecast from individual models. For example, [17] argued that "The best forecast is obtained by combining information sets, not forecasts from information sets. If both

models are known, one should combine the information that goes into the models, not the forecasts that come out of the models.” Authors of Refs. [13, 23, 25] expressed similar opinions. As it seems the investigators in this field lean more to prefer the combination of information in one model.

The main questions that arise in researchers’ minds are “To combine or not to combine” and “how to combine.” In this chapter, we are concerned with the question of “combining forecasts from different models or combining information in one model.” This is an area that has been discussed by many researchers but not in detail (see [9, 11, 12, 29, 35, 40]).

Huang [29] state that “the common belief that combination of information is better than combination of forecasts might be based on the in-sample analysis.” On the contrary, from out-of-sample analysis, they found out that combination of forecasts performs better than combination of information. Many articles typically account for the out-of-sample success of combination of forecasts over combination of information by pointing out various disadvantages that combination of information may possibly possess. For example, (a) in many forecasting situations, particularly in real time, combination of information by pooling all information sets is either impossible or too expensive (see [12, 13, 42]); (b) in a data substantial medium where there are much closed input variables in hand, the super combination of information model may bear from exclusion problem [42]; and (c) in the absence of linearity and, simple dynamics, building an excellent model using combination of information is more likely to be misspecified [26]. We believe that the above-mentioned points can be maintained through the precise selection of the model that is used to estimate the combined information. In our case we used the artificial neural networks to overcome the nonlinearity problem that can be inherent in the series. On the other hand, the factor model is used to tame the problem of the dimensionality, where a large dataset can be summarized in few numbers of factors.

The seminal work of [7] opened the door to examine the prediction combination in different fields of studies in economics and finance. Consequently, a new scope in forecasting study has been to combine the forecasts generated by individual models, using different combinations of techniques. This lets the ultimate forecast result to extract strength from the individual forecasting techniques that cannot be carried out by a single method. Empirically, forecast combinations have been used successfully in diverse areas such as forecasting gross national product, currency market volatility, inflation, money supply, stock prices, interest rates, meteorological data, city populations, and outcomes of football games.

Factor models were introduced in macroeconomics and finance by [22, 36]. The literature on the large factor models starts with [19, 37]. Further theoretical advances were made among others [4, 5, 20]. Upon the successive performance of the DFMs in forecasting, factors augmented to other models are introduced. For example, Bernanke et al. [8] proposed a forecasting model which they called the factor-augmented vector autoregressive (FAVAR) model, a model which merges a factor model with a vector autoregressive component. A factor-augmented vector autoregressive moving average (VARMA) model is suggested by Dufour and Pelletier [16]. Factor-augmented error correction model (FECM) was introduced by Banerjee and Marcellino [6]; Ng and Stevanovic [38] proposed a factor-augmented autoregressive distributed lag (FADL) framework for analyzing the dynamic effects of common and idiosyncratic

shocks. Babikir and Mwambi [2] introduced a factor-augmented artificial neural network (FAANN) that showed improved forecasts compared to DFM and AR models.

On the contrary, artificial neural networks (ANNs) have become one of the most scientific projection methods and have been extensively used in different fields of projection goal. Artificial neural networks have several aspects that make them interesting and authentic for projection work. First, ANNs are common functional approximators. Second, ANNs are data-induced self-flexible approach in that there are less a priori presumptions to be stated about the models for the problem under examination; thus, ANN modeling is not similar to classical model-based approaches. Third, an ANN model is a nonlinear model which is in contrast to the conventional time series forecasting models, which postulate linearity of the series under consideration. [45] demonstrated that systems of the real world are often nonlinear. These advantages of ANNs have attracted attention in time series forecasting and have become a competitive method to traditional time series forecasting methods, and the literature is very vast in this area. The hybrid approach or combining models represent the most important developments in ANNs over the last decade. More hybrid models of ANNs with different forecasting models have been introduced in the recent time, which successfully improve the forecasting performance. [44] proposed the integration of the generalized linear autoregression (GLAR) model with artificial neural networks in order to obtain accurate forecasts for foreign exchange market. [43] proposed a hybrid model called SARIMABP that combines the seasonal autoregressive integrated moving average (SARIMA) model and the back-propagation neural network model to predict seasonal time series data. [34] introduced a hybrid model of ANNs and ARIMA models for forecasting purpose. [1] introduced a hybrid model where the factors were used as input to the ANN model. The model produced more accurate forecasts compared to ANN and DFM.

In this chapter, through the artificial neural networks framework and factor model, for in-sample and out-of-sample forecasting, we show analytically that combination of forecasts—of dynamic factor model and artificial neural networks—can be outclassed by combination of models (information)—of the factors to be used as additional input variables to the artificial neural networks.

To the best of our knowledge, the evaluation of the forecasting performance of the combination of information or models of factors and ANN—the FAANN—and combination of forecasts of ANN and DFM using different linear and nonlinear combinations is new, and this is the first attempt in general and in South Africa in particular. The empirical results show sizable gains in terms of the forecasting ability of the FAANN compared to both the standard ANN and the DFM and their forecasts combination; in other words it seems that combination of models outperforms combination of forecasts meaning that combination of information could be better than the combination of forecasts.

The remaining of the chapter is formulated as follow: Section 2 in brief expresses the DFM, the ANN, and the FAANN projection models and the combination techniques; Section 3 introduces the data; the results obtained from forecasting models and their combinations are presented in Section 4; finally, Section 5 gives a concise conclusion of the study and some suggestions for future researches.

2. Individual forecasting models and combination methods

In this section, we introduce briefly the symbols, formation, and estimation methods in forecasting models; also, we introduce and discuss the various combining methods.

2.1. Individual forecasting models

2.1.1. The dynamic factor model and the estimation of factors

This subsection handles DFM to get common elements from a large group of variables; then, these common components are used to predict the variables of interest.

Suppose that we have a group of observations, X_t be the N stationary time series variables having observations at times $t = 1, \dots, T$, where it is considered that the series have zero mean. Factor model assumes that most of the variation in the dataset can be explained by a small number $q \ll N$ of factors involved in the vector f_t . We can express the dynamic factor model representation as follows:

$$X_t = \chi_t + \xi_t = \lambda(L)'f_t + \xi_t \quad (1)$$

where χ_t is the common components driven by factor f_t and ξ_t is the idiosyncratic components for each of the variables. ξ_t is the portion of X_t that cannot be explained by the common components. χ_t is a function of the $q \times 1$ vectors of $\lambda(L)'f_t$; the operator $\lambda(L) = 1 + \lambda_1 L + \dots + \lambda_s L^s$ is a lag polynomial with positive powers on the lag operator L with $L f_t = f_{t-1}$. The static representation of the model can be rewritten in as

$$X_t = \Lambda' F_t + \xi_t \quad (2)$$

where F_t is a vector of $r \geq q$ static factors that compose of the dynamic factors f_t and all lags of the factors. From a set of data, there are three different methods of estimating the factors in F_t . These methods were developed by Stock and Watson [39] hereafter SW [30] and Forni, Hallin, Lippi, and Reichlin [20] hereafter FHRLR¹. In the current chapter, we employ the estimation method developed by FHRLR. For more details of the dynamic factor model estimation, see Babikir and Mwambi [2]. Thus, the estimated factors will be used to forecast the variables of interest. The forecasting model is specified and estimated as a linear projection of an h -step ahead transformed variable y_{t+h} into t -dated dynamic factors. The forecasting model follows the setup in [3, 21, 41] with the form

$$y_{t+h} = \beta(L)\hat{f}_t + \gamma(L)y_t + u_{t+h} \quad (3)$$

where \hat{f}_t represents the dynamic factors that estimated using the method by FHRLR, while $\beta(L)$ and $\gamma(L)$ are the lag polynomials, which are determined by the Schwarz information criterion (SIC). The u_{t+h}^h is an error term. The coefficient matrix for factors and autoregressive terms are estimated by ordinary least squares (OLS) for each forecasting horizon h . To find

¹For further technical details on this type of factor models, see [35].

the estimate and forecast of the AR benchmark, we enforce a condition to Eq. (3), where we set $\beta(L) = 0$.

2.1.2. The artificial neural network model

The ANN is one of the most popular and successful biological-inspired forecasting methods, which emulate the framework of the human brain; thus, ANNs have gradually achieved immense importance in forecasting among other fields. The ANN model is one of the generalized nonlinear nonparametric models (GNLNPMs). Compared to the traditional econometric models, the advantage of ANNs is that they can handle complex, nonlinear relationships without any prior assumptions about the underlying data-generating process (see [28]; **Figure 1**).

The properties of the ANN model made the method an attractive alternative to traditional forecasting models. Most importantly, ANN models deal with the limitations of traditional forecasting methods, including misspecification, biased outliers, and assumption of linearity [27]. One of the most recognized ANN structures in time series forecasting problems is the multilayer perceptron (MLP). An MLP is basically a feedforward architecture of an input, one or more hidden, and an output layer. The network structure illustrated in this chapter gives forward network connected with linear neuron activation function. Basically, the input nodes are connected forward to all nodes in the hidden layer, and these latent nodes are joined to the single node in the output layer, as shown in **Figure 1**. The inputs in this model serve as the independent variables in the multiple regression model and are joined to the output node—which is similar to the dependent variable—through the latent layer. We follow [33], in describing the network model. Thus, the model can be specified as follows:

$$n_{k,t} = w_0 + \sum_{i=1}^p w_i y_{t-i} + \sum_{j=1}^J \phi_j N_{t-1,j} \quad (4)$$

$$N_{k,t} = f(n_{k,t}) \quad (5)$$

$$y_t = \alpha_{i,0} + \sum_{k=1}^K \alpha_{i,k} N_{k,t} + \sum_{i=1}^p \beta_i y_{t-i} \quad (6)$$

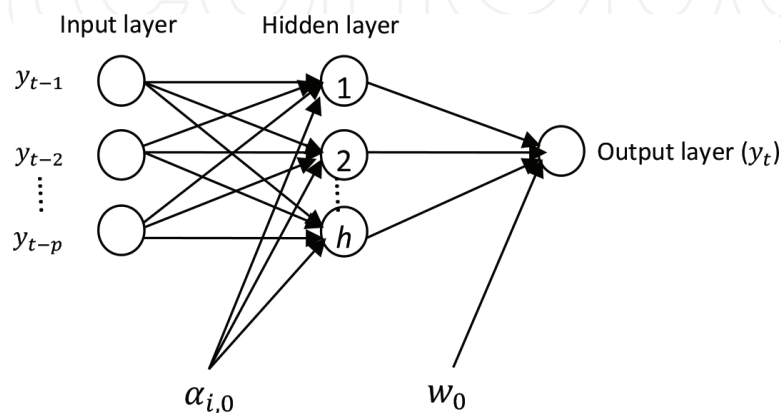


Figure 1. A $p \times h \times 1$ structure of a feed forward neural network.

where inputs y_{t-i} represent the lagged values of the variable of interest and the output y_t is their forecasts. The w_0 and $\alpha_{i,0}$ are the bias, and w_i and $\alpha_{i,k}$ denote the weights that link the inputs to the latent layer and the latent layer to output, respectively. The \varnothing_j and β_i connect the input to the output via the latent layer. The p -independent variables are connected linearly to form K neurons which then are combined linearly to produce the prediction or output. Eqs. (4)–(6) link inputs y_{t-i} to outputs y through the hidden layer. The function f is a logistic function meaning that $N_{k,t} = f(n_{k,t}) = \frac{1}{1+e^{-n_{k,t}}}$. The second summation in Eq. (6) shows that we also have a jump connection or skip-layer network that directly links the inputs y_{t-i} to the output y_t . The beauty of this ANN structure is that the model combines the true linear model and nonlinear supply-forward neural network. So, if the association between inputs and output is true linear, in this case, the coefficient set β , which is skip layer should be significant, in contrast if the association is a nonlinear in nature the jump connections coefficient β to be insignificant, while the coefficients set w and α be highly significant. Certainly, if the association between input and output is mixed, then we watch for all coefficient sets to be significant. For the best network selection in this chapter, beside the minimum error, we use Bayesian information criterion (BIC), which is usually preferred more than the other three criteria, because it has the ability to penalize the extra parameters more severely; mathematically, BIC is given by the following as described in [31]

$$BIC = N_{p,h} + N_{p,h} \ln(n) + n \ln\left(\frac{S(W)}{n}\right) \quad (7)$$

where $N_{p,h} = h(p+2) + 1$ is total number of parameters in the network, $n = N_{train} - p$ is the number of effective observations, N_{train} is the in-sample observation, $S(W)$ is the network misfit function, and W is the space of all weights and biases in the network. The in-sample sum of squared error (SSE) is usually used to determine the function $S(W)$. Eventually, the optimal model is the model with minimum BIC value.

2.1.3. Factor-augmented artificial neural networks (FAANN)

The FAANN model is a hybrid model of artificial neural network and factor model in order to combine information of factors and lagged values of interested variable to be forecasted for more accurate forecasts in hand. The nonlinear function uses the series, its lag, and factors to formulate the FAANN model that defines as follows:

$$y_t = f\left[\left(y_{t-1}, y_{t-2}, \dots, y_{t-p}\right), (F_1, F_2, F_3, F_4, F_5)\right] \quad (8)$$

where f is the nonlinear functional form determined via ANN. In the first stage, the factor model is used to extract factors from a large related dataset. In the second stage, a neural network model is used to model the nonlinear and linear relationships existing in factors and original data. Thus, based on the model structure depicted on **Figure 2**,

$$y_{t+h} = \alpha_0 + \sum_{j=1}^h \alpha_j g\left(\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i} + \sum_{i=p+1}^{p+5} \beta_{ij} F_{t,i}\right) + \varepsilon_t \quad (9)$$

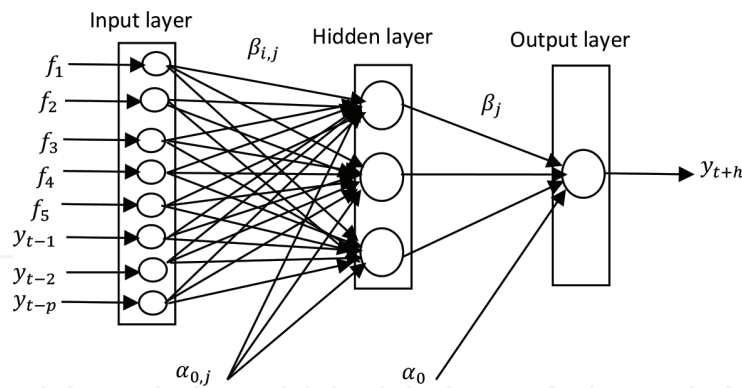


Figure 2. The FAANN model architecture ($N^{(p+5,h,1)}$).

As previously noted, the α_j ($j = 0, 1, \dots, h$) and β_{ij} ($i = 0, 1, \dots, p; j = 1, 2, \dots, h$) are the parameters of the model that called the connection weights. As we have stated earlier, p and h are the numbers of input and hidden nodes, respectively, and ε_t is the error term. **Figure 2** shows the FAANN model structure used.

2.2. Forecast combining methods

To combine individual forecasts composed by the DFM and ANN models, we used four combination methods. The combining methods involve three linear combining methods (the mean, VACO, and discount MSFE-based methods) and one nonlinear combining method (ANN). Just as some of the combining methods need a holdout period to calculate the weights used to combine individual forecasts, we use the first 24 months of the out of sample as holdout observations. For all combining methods, we form combination forecasts over the post holdout out-of-sample period. Brief details about the above combining methods are given below.

2.2.1. Mean combination method

The mean serves as a convenient criterion as has been shown to achieve better results compared to other fancy methods. For instance, see [10, 21, 32]. Compared to single forecasts, the performance of the simple average combination method is found to be superior (see [18]). The simple average combination method can be expressed as

$$\hat{y}_t^c = \sum_{i=1}^m w_i \hat{y}_t^i \quad (10)$$

where \hat{y}_t^c is the combined forecast at time t , \hat{y}_t^i is the forecast from i th individual forecasting model, $w_i = \frac{1}{m}$ is the individual forecast weight for model i , and m is the number of individual models. There are different forms of weights, but generally the weights have to satisfy the condition $\sum_{i=1}^m w_i = 1$.

2.2.2. Variance-covariance (VACO) combination method

The method uses the historical achievement of the individual forecasts to compute the weights. Thus, according to the VACO method, the weights determined as follows:

$$w_i = \frac{\left[\sum_{j=1}^T (y_j - \hat{y}_j^i)^2 \right]^{-1}}{\sum_{i=1}^m \left[\sum_{j=1}^T (y_j - \hat{y}_j^i)^2 \right]^{-1}} \quad (11)$$

Then, the combined forecast is given by $\hat{y}_t^c = \sum_{i=1}^m w_i \hat{y}_t^i$ where y_j is the j th actual value, \hat{y}_j^i is the j th forecasting value from i th individual forecasting model, and T is the total number of out-of-sample points. The weight in Eq. (11) is based on the inverse sum of squared deviation for model i as the numerator, and the denominator is the sum of these inverse contributions from all models. This guarantees that $\sum_{i=1}^m w_i = 1$.

2.2.3. Discounted mean square forecast error (DMSFE) combination method

The DMSFE method weights recent forecasts more heavily than distant ones. [32] suggest that the weights can be calculated as

$$w_i = \frac{\left[\sum_{j=1}^T \delta^{T-j+1} (y_j - \hat{y}_j^i)^2 \right]^{-1}}{\sum_{i=1}^m \left[\sum_{j=1}^T \delta^{T-j+1} (y_j - \hat{y}_j^i)^2 \right]^{-1}} \quad (12)$$

where δ is the discount factor with $0 < \delta \leq 1$, if $\delta = 1$ and then the DMSFE and VACO methods become one method, which means that the VACO is a special case of the DMSFE. Note that as mentioned above the sum of all weights is equal to one.

2.2.4. Artificial neural network (ANN) combination method

Linearity of combinations of the individual forecasts is the corner stone of linear combination method, but if the individual forecasts are based on nonlinear methods, the combinations are defined to be insufficient or if the true relationship is nonlinear. For the success of the ANN as a combination method over the linear methods, among others, see [15, 25]. Here, we use the same setup used in subsection (2.1.2); the output \hat{y}_t^c of combined forecasts can be given by

$$\hat{y}_t^c = \alpha_{i,0} + \sum_{k=1}^K \alpha_{i,k} N_{k,t} + \sum_{i=1}^m \beta_i \hat{y}_t^i \quad (13)$$

where \hat{y}_t^i is the forecast from i th individual forecasting model.

3. Data

For FAANN and DFM models, data are gathered that include 228 monthly time series² of which 203 are collected from South Africa, including the financial, real, nominal sectors, and confidence indices, 2 global variables, and 23 series of major trading partners and global financial markets. The AR criterion model will be used for the data which composed only the variable of interest, namely, deposit rate or share prices for gold mining or long-term interest rate. Thus, besides the national variables, the chapter uses a set of global variables such as gold and crude oil prices. Also, the data incorporate series from financial markets of major trading partners, namely, the United Kingdom, the United States, China, and Japan. For estimation data cover the period January 1992 through December 2006, while the period from January 2007 through December 2011 will be used for goodness of fit for the extracted model. For the degree of integration of all series, the augmented Dickey-Fuller (ADF) test will be used. Difference of the series is used for all nonstationary series in this study. The Schwarz information criterion (SIC) is used in selecting the appropriate lag length in such a way that no serial correlation is left in the stochastic error term. Finally, all series are standardized to have a mean of zero and a constant variance.

4. Evaluation of forecast accuracy

To evaluate the forecast accuracy of model combination or information combination, we used three datasets from South Africa, namely, deposit rate, gold mining share prices, and long-term interest rate, in order to demonstrate the in-sample and out-of-sample appropriateness and effectiveness of the combination of models or information of the DFM and ANN models.

4.1. In-sample forecast evaluation

In this subsection, we evaluate the in-sample predictive power of the combined model forecast—the FAANN model—and other fitted models which include AR (benchmark model), DFM, and ANN and best combined forecasts of the DFM and ANN models. To achieve this, a full sample from January 1992 to December 2011, giving a total of 240 observations of the three datasets—deposit rate, gold mining share prices, and long-term interest rate—is used to estimate the forecasting models in order to check the robustness of in-sample results of competed models and compare it to the AR benchmark model. In-sample forecasting is most useful when it comes to investigate the true relationship between the independent variables and the forecast of dependent variable. **Table 1** reports the root-mean-square error (RMSE)³ of the in-sample forecasting results. The FAANN model outperformed all other models. The maximum reduction in RMSE over the AR benchmark model is around 24%, while the

²The data sources are the South Africa Reserve Bank, ABSA Bank, Statistics South Africa, National Association of Automobile Manufacturers of South Africa (NAAMSA), South African Revenue Service (SARS), Quantec, and World Bank.

³The RMSE statistic can be defined as $\sqrt{\frac{1}{N} \sum (Y_{t+n} - \hat{Y}_{t+n})^2}$, where Y_{t+n} denotes the actual value of a specific variable in period $t+n$ and \hat{Y}_{t+n} is the forecast made in period t for $t+n$.

Variable	Model				
	FAANN	DFM	ANN	AR	Best combined forecasts of DFM and ANN
Deposit rate	0.1687	0.1849	0.1793	0.1954	0.1694
Share prices for gold mining	1.5922	1.7782	1.7787	1.8187	1.6215
Long-term interest rate	0.1253	0.1537	0.1546	0.1640	0.1438

Table 1. The RMSE of the in-sample forecasts.

minimum reduction is around 14% considering all variables. Regarding the in-sample forecasting, the FAANN model provides lower RMSE with a reduction of between 9 and 19% for all variables compared to the DFM. Despite that the same factors are augmented to AR and ANN to produce the DFM and the FAANN models, the in-sample results provide significant differences between estimation methods which favor the nonlinear method over the linear one. This is potentially due to the flexibility and property of the ANN models as universal approximators that can be used to different time series in order to obtain accurate forecasts. Comparing the forecasting performance of the FAANN and standard ANN model, the FAANN model produced lower RMSE of 6–19% for all variables. These results indicate the importance of the factors—which summarized 228 related series into five factors—that are used as input to the ANN to produce the FAANN model. Regarding the in-sample forecasting performance of the forecasts of combined models or information—the FAANN model—compared to the best forecast combination of the DFM and ANN models, the FAANN model outperforms the best forecast combination with reduction in the RMSE around 0.01–13% for all variables. These results confirm the superiority of the combination of information or models when a precise estimation method is used to estimate the combined information over the combined forecasts of individual models.

4.2. Out-of-sample forecast evaluation of individual models

In this subsection, we estimate the individual forecasts of the AR, DFM, and ANN and the best combined forecasts of the DFM and ANN models and the FAANN model that combine information of the factors and ANN for the three variables of interest, namely, deposit rate, gold mining share prices, and long-term interest rate, over the in-sample period January 1992 to December 2006 using monthly data, and then compute the out of sample for 3-, 6-, and 12-month-ahead forecasts for the period of January 2007 to December 2011. We employ iterative forecast technique to compute the RMSE for the three forecasting horizons used for the three variables across all of the different models in order to compare the forecast accuracy generated by the models. The starting date of the in-sample period depends on data availability of some important financial series. The out-of-sample period includes the occurrence of the financial crisis that affected economies and financial sectors in particular. Thus, we used this period as out of sample in order to show the suitability and efficiency of the combination of information—FAANN model—to produce accurate forecasts for such data that exhibits inherent nonlinearity or the data that faced fluctuations during the financial crisis. The result of each single variable can be summarized as follows:

- *Deposit rate forecasting results:* for the FAANN model estimation firstly, MATLAB package is used to estimate the factors. Secondly, R software using Broyden, Fletcher, Goldfarb, and Shanno (BFGS) algorithm is used to find and estimate the optimum network architecture. The network with the lowest in-sample RMSE and the Bayesian information criterion (BIC) is selected as best-fitted network, which is composed of eight inputs, five neurons in the hidden layer, and one output (in abbreviated form $N^{(8-5-1)}$). **Table 2** reports the RMSEs of the 3-, 6-, and 12-month-ahead and the average of the 3-, 6-, and 12-month-ahead RMSEs. The benchmark for all forecast evaluations is the AR model forecast RMSEs. For both long and short horizons, the FAANN model outperforms all other models followed by the DFM for the short horizons and the ANN in long horizon. The RMSE of the FAANN model decreases as the forecast horizon increases which in turn agreed with [24] who found that the ANNs significantly forecast better in long horizon. Results reveal that the FAANN performed better with large reductions in RMSE of around 25–46% of the RMSE compared to the AR benchmark model and the reduction on the average RMSE around 37%.
- *Gold mining share prices:* we used the same steps where software and algorism were implemented to the previous variable to estimate the FAANN model. The optimum network is composed of eight inputs, seven neurons in the hidden layer, and one output (in abbreviated form $N^{(8-7-1)}$). **Table 3** presents the RMSE results of the FAANN, the DFM, the ANN, and the AR benchmark. As expected based on the in-sample results, the FAANN model stands out in forecasting both short and long horizons with a sizable reduction in

Model	3 months	6 months	12 months	Average
FAANN	0.7465	0.6373	0.5359	0.6399
DFM	0.9501	0.9153	0.9438	0.9364
ANN	0.9693	0.9160	0.8869	0.9241
AR	0.1862	0.1949	0.2314	0.2041

Note: The last row reports the RMSE for the AR benchmark model; the remaining rows represent the ratio of the RMSE for the forecasting model to the RMSE for the AR. Bold entries indicate the forecasting model with the lowest RMSE.

Table 2. Out-of-sample (January 2007–December 2011) RMSE for deposit rate.

Model	3 months	6 months	12 months	Average
FAANN	0.9053	0.9121	0.8227	0.8800
DFM	0.9655	0.9661	0.9532	0.9616
ANN	0.9566	0.9556	0.9215	0.9446
AR	1.7743	1.7924	1.8187	1.7951

Note: See note to **Table 2**.

Table 3. Out-of-sample (January 2007–December 2011) RMSE for gold mining share prices.

Model	3 months	6 months	12 months	Average
FAANN	0.7281	0.6051	0.5498	0.6277
DFM	0.9834	0.9042	0.9584	0.9487
ANN	0.9893	0.8981	0.8306	0.9060
AR	0.2052	0.2140	0.2308	0.2167

Note: See note to **Table 2**.

Table 4. Out-of-sample (January 2007–December 2011) RMSE for long-term interest rate.

RMSE relative to the AR benchmark model of 10–18%. The average of the RMSE reduction over the forecast horizons is 12%. On average the FAANN outperforms the ANN and DFM models with reduction in RMSE of 6 and 8%, respectively.

- *Long-term interest rate:* for estimation purpose the same package and algorithm that are used with previous variables are implemented. Thus, the optimal network in abbreviated form is $N^{(8-3-1)}$. **Table 4** results show the performance of the FAANN model where the model produces more accurate forecasts compared to all competing model on both the single-level forecast horizons and the average of these horizons. Compared to the AR benchmark, the FAANN provides a reduction in the RMSE range from 45–27%, while the average RMSE reduction is around 38%. The performance of the FAANN model stands out followed by the ANN and the DFM with average reduction in RMSE of 9 and 5%, respectively, relative to the AR benchmark model. Comparing the FAANN performance to the ANN and the DFM, the FAANN model RMSE reduction is around 28 and 32%, respectively.

4.3. Out-of-sample forecast evaluation of the combined forecasts of the DFM and ANN models

Table 5 reports the results of combining forecasts of the DFM and ANN models. We aim of using the DFM and ANN models in particular to merge their advantages where the ANN model with its flexibility to account for potentially complex nonlinear relationships that is not easily captured by traditional linear models, and the DFM model can accommodate a large number of variables. Similar to **Table 2**, **Table 5** shows the ratio of the RMSE for a given combining method to the RMSE for the AR benchmark model. We found that the AR benchmark model poorly performs compared to all combining methods. Generally, the nonlinear ANN combining method outperforms all other combining methods for all variables at all forecasting horizons; hence, it offers a more reliable method for generating forecasts of the variables of interest. Compared to the AR, the nonlinear ANN combining method provides a large reduction in RMSE of around 7–20% relative to the AR model overall forecasting horizons and variables. The nonlinear ANN combining method also beats the best individual forecasting of the DFM and the ANN models for all variables and overall forecasting horizons with sizable reductions in RMSE of around 1–15% of the RMSE of the best individual forecasts. We note in addition that the discount MSFE with $\delta = 0.9$ as a combining method performs

Combination method	h = 3	h = 6	h = 12
Deposit rate			
AR	0.1862	0.1949	0.2314
Mean	0.915	0.890	0.851
VACO	0.921	0.892	0.846
DMSFE, $\delta = 0.95$	0.923	0.903	0.848
DMSFE, $\delta = 0.90$	0.905	0.884	0.837
ANN	0.907	0.882	0.835
Gold mining share prices			
AR	1.7743	1.7924	1.8187
Mean	0.946	0.942	0.937
VACO	0.943	0.946	0.951
DMSFE, $\delta = 0.95$	0.945	0.941	0.937
DMSFE, $\delta = 0.90$	0.945	0.942	0.936
ANN	0.921	0.929	0.911
Long-term interest rate			
AR	0.2052	0.2140	0.2308
Mean	0.951	0.923	0.902
VACO	0.952	0.942	0.922
DMSFE, $\delta = 0.95$	0.956	0.953	0.954
DMSFE, $\delta = 0.90$	0.951	0.952	0.935
ANN	0.827	0.815	0.804

Note: See note to Table 2.

Table 5. Forecast combining results of the DFM and ANN-RMSE for variables (January 2007–December 2011).

nearly as well as the best individual model for all variables and forecasting horizons. The combining method of variance–covariance (VACO), on average, performs less accurate compared to other combining methods' overall forecasting horizons and variables. We note that the combined forecasts produce more accurate forecasts for long horizons which we attributed to the contribution of the nonlinear model in the combination as nonlinear models produce more accurate forecast in the long horizon.

4.4. Comparison of forecasting performance of combination of models or information and combination of forecasts

Here, we compare the forecasting performance of the combination of models (information)—the FAANN model—with the best forecast combinations of the ANN and DFM models.

Forecasting model	h = 3	h = 6	h = 12
Deposit rate			
AR (benchmark model)	0.1862	0.1949	0.2314
FAANN	0.7465	0.6373	0.5359
Combined forecasts of DFM and ANN	0.907	0.882	0.835
Gold mining share prices			
AR (benchmark model)	1.7743	1.7924	1.8187
FAANN	0.9053	0.9121	0.8227
Best combined forecasts of DFM and ANN	0.921	0.929	0.911
Long-term interest rate			
AR (benchmark model)	0.2052	0.2140	0.2308
FAANN	0.7281	0.6051	0.5498
Best combined forecasts of DFM and ANN	0.827	0.815	0.804

Note: See note to **Table 2**.

Table 6. Forecast results of the best combination of DFM and ANN model and FAANN-RMSE for variables (January 2007–December 2011).

Table 6 presents the RMSE ratios of the FAANN model and the best forecast combination to the AR benchmark model over the out-of-sample period. Compared to the DFM, the results indicate that the FAANN model generates accurate forecasts for all variables and with all forecast horizons. The improvement of the FAANN model is compared to the DFM between 2 and 10% reduction in RMSE for all variables and horizons. Thus, these results indicate the superiority of augmentation of factors to nonlinear method (FAANN) over the linear one (DFM) across the three different series and three different time horizons.

To confirm the RMSE results, the test of equal forecast accuracy of Diebold and Mariano [14] is used to evaluate forecasts. The test of equal forecast accuracy employed here is given by $S = \frac{\bar{d}}{\sqrt{\hat{V}(\bar{d})}}$, where $\bar{d} = \frac{1}{T} \sum_{t=1}^T (e_{1t}^2 - e_{2t}^2)$ is the mean difference of the squared prediction error and $\hat{V}(\bar{d})$ is the estimated variance. Here, e_{1t}^2 denotes the forecast errors from the FAANN model, and e_{2t}^2 denotes the forecast errors from the AR benchmark model or the best combined forecasts of DFM and ANN. The S statistic follows a standard normal distribution asymptotically. Note, a significant negative value of S means that the FAANN model outperforms the other model in out-of-sample forecasting. **Table 7** shows the result of the Diebold and Mariano test between the FAANN and the AR benchmark and between the FAANN and the best combined forecasts of DFM and ANN. The test results confirm that the FAANN models provide the lowest RMSEs. In summary the FAANN models provide significantly better forecasts at the 5% and 10% level compared to the AR and the best combined forecasts of DFM and ANN models.

Model/variable	Forecasting horizons		
	3 months	6 months	12 months
Deposit rate	−2.095**	−2.108**	−3.159**
FAANN vs. AR	−1.944*	−1.799*	−2.064**
FAANN vs. best combined forecast from DFM and ANN			
Share prices for gold mining	−2.420**	−2.527**	−2.753**
FAANN vs. AR	−1.812*	−1.673*	−1.961**
FAANN vs. best combined forecast from DFM and ANN			
Long-term interest rate	−2.402**	−2.339**	−2.429**
FAANN vs. AR	−1.741*	−2.138**	−1.861**
FAANN vs. best combined forecast from DFM and ANN			

Note: ** and * indicate significant value at the 5 and 10% levels, respectively.

Table 7. Diebold-Mariano test (January 2007–December 2011).

5. Conclusion

In this chapter we aim to evaluate the forecasting performance of the model combination and forecast combination for the ANN and DFM models. In the model combination, we merge the factors that were extracted from a large dataset—288 series in our case—with ANN which produces the FAANN model. For the forecast combination, we used different linear and nonlinear combination methods to combine the individual forecasts of the DFM and the ANN models. Using the period of January 1992 to December 2006 as in-sample period and January 2007 to December 2011 as out-of-sample period, we compare the forecast performance of the FAANN with DFM, ANN, and AR benchmark model for 3-, 6-, and 12-month-ahead forecast horizons for three variables, namely, for deposit rate, gold mining share prices, and long-term interest rate. The study has provided evidence using both the RMSE and Diebold and Mariano test as the comparison criteria that FAANN models best fit the three considered variables over the 3-, 6-, and 12-month-ahead forecast horizons.

Tables 2–4 show the ability of the model combination—FAANN model—to produce accurate forecast that outperforms DFM and ANN and their best forecast combination results. The results seem not contradicted with in-sample model forecast performance as in **Table 1**. The FAANN model outperformed the AR benchmark model with large reduction in RMSE of around 25–46% considering all variables and forecast horizons. Compared to the DFM and ANN models, the FAANN model produces more accurate forecasts that yielded a decrease in RMSE of around 6–43% and 5–40%, respectively. We attribute the superiority of the FAANN to the flexibility of ANN to account for potentially complex nonlinear relationships that are not easily captured by linear models and the contribution of the factors to the model. On the other hand, the ANN and the DFM outperformed the AR benchmark with a reduction in the RMSEs of around 1–17% and 2–10%, respectively, for all variables and across all forecast horizons. **Table 6** shows comparison results of the forecasting performance of the combined models—the FAANN—and the best forecast combination of the DFM and the ANN models. The results

indicate that the combined models or information produced forecasts that are better than the best combined forecasts of the DFM and the ANN models. In other words, the nonlinear model that uses large dataset of economic and financial variables in addition to the lags of the interested variable improves the forecasting performance over models that are estimated separately—the DFM and the ANN. We also observed that the FAANN residual decreases as the forecast horizon increases.

Author details

Ali Babikir^{1*}, Mustafa Mohammed² and Henry Mwambi¹

*Address all correspondence to: ali.basher@gmail.com

1 School of Mathematics, Statistics and Computer Science, University of KwaZulu-Natal, Pietermaritzburg, South Africa

2 Faculty of Science and Arts, University of Jeddah, Saudi Arabia

References

- [1] Babikir A, Mwambi H. A Factor—Artificial Neural Network Model for Time Series Forecasting: The Case of South Africa. *Proceeding of IEEE International Joint Conference on Neural Networks (IJCNN)*; Beijing, China. 2014. p. 838–844
- [2] Babikir A, Mwambi H. Factor augmented artificial neural network model. *Neural Processing Letters*. 2016;**45**:507-521. DOI: 10.1007/s11063-016-9538-6
- [3] Aruoba S, Diebold F, Scotti C. Real-time measurement of business conditions. *Journal of Business & Economic Statistics*. 2009;**27**:417-427
- [4] Bai J. Inferential theory for factor models of large dimensions. *Econometrica*. 2003;**71**(1): 135-171
- [5] Bai J, Ng S. Determining the number of factors in approximate factor models. *Econometrica*. 2002;**70**(1):191-221
- [6] Banerjee A., Marcellino, M.. Factor augmented error correction models. *CEPR Discussion Paper*, 6707; 2008
- [7] Bates JM, Granger CWJ. The combination of forecasts. *Operations Research Quarterly*. 1969;**20**:451-468
- [8] Bernanke B, Boivin J, Elias P. Measuring the effects of monetary policy: A factor-augmented vector autoregressive (FAVAR) approach. *Quarterly Journal of Economics*. 2005; **120**:387-422

- [9] Chong YY, Hendry DF. Econometric evaluation of linear macro-economic models. *The Review of Economic Studies*. 1986;671-690
- [10] Clemen RT, Winkler RL. Combining economic forecasts. *Journal of Business & Economic Statistics*. 1986;4:39-46
- [11] Clements MP, Galvao AB. Combining predictors and combining information in Modeling: Forecasting US recession probabilities and output growth. University of Warwick; 2005
- [12] Diebold FX. Forecast combination and encompassing: Reconciling two divergent literatures. *International Journal of Forecasting*. 1989;5:589-592
- [13] Diebold FX, Pauly P. The use of prior information in forecast combination. *International Journal of Forecasting*. 1990;6:503-508
- [14] Diebold FX, Mariano RS. Comparing predictive accuracy. *Journal of Business & Economic Statistics*. 1995;13:253-263
- [15] Donaldson RG, Kamstra M. Forecast combining with neural networks. *Journal of Forecasting*. 1996;15:49-61
- [16] Dufour J-M, Pelletier D. Practical Methods for Modelling Weak VARMA Processes: Identification, Estimation and Specification with a Macroeconomic Application. Discussion paper; 2013
- [17] Engle RF, Granger CWJ, Kraft DF. Combining competing forecasts of inflation using a bivariate ARCH model. *Journal of Economic Dynamics and Control*. 1984;8:151-165
- [18] Fang Y. Forecasting combination and encompassing tests. *International Journal of Forecasting*. 2003;19:87-94
- [19] Forni M, Hallin M, Lippi M, Reichlin L. The generalized factor model: Identification and estimation. *The Review of Economics and Statistics*. 2000;82:540-554
- [20] Forni M, Hallin M, Lippi M, Reichlin L. The generalized dynamic factor model, one sided estimation and forecasting. *Journal of the American Statistical Association*. 2005;100(471): 830-840
- [21] Forni M, Hallin M, Lippi M, Reichlin L. The generalized factor model: Consistency and rates. *Journal of Econometrics*. 2004;119:231-255
- [22] Geweke J. The dynamic factor analysis of economic time series. In: Aigner DJ, Goldberger AS, editors. *Latent variables in socio-economic models*. North Holland: Amsterdam; 1977. pp. 365-383
- [23] Granger CWJ. Invited review: Combining forecasts - twenty years later. *Journal of Forecasting*. 1989;8:167-173
- [24] Tkacz G, Hu S. Forecasting GDP Growth Using Artificial Neural Networks. Working Paper 3. Bank of Canada; 1999

- [25] Harrauld PG, Kamstra M. Evolving artificial neural network to combine financial forecasts. *IEEE Transactions on Evolutionary Computation*. 1997;1:40-52
- [26] Hendry DF, Clements MP. Pooling of forecasts. *The Econometrics Journal*. 2004;7:1-31
- [27] Hill T, O'connor M, Remus W. Neural network models for time series forecasts. *Management Science*. 1996;42:1082-1092
- [28] Hornik K, Stinchcombe M, White H. Multi-layer feed forward networks are universal approximators. *Neural Networks*. 1989;2:359-366
- [29] Huang H, Lee T-H. To combine forecasts or to combine information? *Econometric Reviews*. 2010;29(5-6):534-570
- [30] Kapetanios G, Marcellino M. A parametric estimation method for dynamic factor models of large dimensions. *Journal of Time Series Analysis*. 2009;30:208-238
- [31] Kihoro JM, Otieno RO, Wafula C. Seasonal time series forecasting: A comparative study of ARIMA and ANN models. *African Journal of Science and Technology*. 2004;5(2):41-49
- [32] Makridakis S, Winkler R. Averages of forecasts: Some empirical results. *Management Science*. 1983;29:987-996
- [33] McAdam P, Hughes Hallett AJ. Non linearity, computational complexity and macro economic modeling. *Journal of Economic Surveys*. 1999;13(5):577-618
- [34] Khashei M, Bijari M. An artificial neural network (p, d,q) model for time series forecasting. *Expert Systems with Applications*. 2010;37:479-489
- [35] Newbold P, Harvey DI. Forecast combination and encompassing. In: Clements MP, Hendry DF, editors. *A Companion to Economic Forecasting*. Blackwell Publishers; 2001
- [36] Sargent TJ, Sims CA. Business cycle modeling without pretending to have too much a priori economic theory. In: Sims C, editor. *New Methods in Business Research*. Federal Reserve Bank of Minneapolis; 1977
- [37] Schumacher C. Forecasting German GDP using alternative factor models based on large datasets. *Journal of Forecasting*. 2007;26:271-302
- [38] Serena Ng and Dalibor Stevanovic: Factor Augmented Autoregressive Distributed Lag Models. working paper; 2012
- [39] Stock JH, Watson MW. Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*. 2002a;97:147-162
- [40] Stock JH, Watson M. Combination forecasts of output growth in a seven country data set. *Journal of Forecasting*. 2004;23:405-430
- [41] Stock JH, Watson MW. Macroeconomic forecasting using diffusion indexes. *Journal of Business & Economic Statistics*. 2002b;20:147-162

- [42] Timmermann A. Forecast combinations. Elliott G, Granger CWJ, Timmermann A, ed. Forthcoming in Handbook of Economic Forecasting. North Holland. 2005
- [43] Tseng FM, Yu HC, Tzeng GH. Combining neural network model with seasonal time series ARIMA model. Technological Forecasting and Social Change. 2002;**69**:71-87
- [44] Yu L, Wang S, Lai KK. A novel nonlinear ensemble forecasting model incorporating GLAR and ANN for foreign exchange rates. Computers and Operations Research. 2005;**32**:2523-2541
- [45] Zhang G, Patuwo BE, Hu MY. Forecasting with artificial neural networks: The state of the art. International Journal of Forecasting. 1998;**14**:35-62

