2014 International Joint Conference on Neural Networks (IJCNN)
July 6-11, 2014, Beijing, China

A Factor – Artificial Neural Network Model for Time Series Forecasting: The Case of South Africa

Ali Babikir
School of Math., Statistics and Computer Science
University of KwaZulu-Natal
Pietermaritzburg, South Africa South Africa
e-mail: ali.basher@gmail.com

Henry Mwambi
School of Math., Statistics and Computer Science
University of KwaZulu-Natal
Pietermaritzburg, South Africa South Africa
e-mail: MwambiH@ukzn.ac.za

***Abstract* - In this paper, the factor models (FMs) are integrated with the ANN model to produce a new hybrid method which we refer to as the Factor Artificial Neural Network (FANN) to improve the time series forecasting performance of the artificial neural networks. The empirical results of the in sample and out of sample forecasts indicate that the proposed FANN model is an effective way to improve forecasting accuracy over the dynamic factor Model (DFM), the ANN and the AR benchmark model. When we compare the FANN and ANN models the results confirm the usefulness of the factors that were extracted from a large set of related. On the other hand, as far as estimation is concerned the nonlinear FANN model is more suitable to capture nonlinearity and structural breaks compared to linear models. The Diebold-Mariano test results confirm the superiority of the FANN model forecasts performance over the AR benchmark model and the ANN model forecasts.**

Keywords— Artificial neural network; Dynamic factor model; Forecast accuracy; Root mean square error.

I. INTRODUCTION

In recent decades, considerable progress into handling large panels of time series data in forecasting using factor models has been made. The initial contributions in this area were the work of Geweke [18] and Sargent and Sims [31], who introduced the dynamic factor approach to macroeconomics. They exploited the dynamic interrelationships between the variables, and then reduced the number of common factors even further. However, the approach followed by [18, 31] is too restrictive, in that it assume orthogonality on the idiosyncratic components, while the work by Chamberlain [11] and [12] allow for the possibility of weakly cross-sectional correlation of the idiosyncratic components. In further improvements these large factor models have been improved by accounting for serial correlation and weakly cross-sectional correlation of idiosyncratic components, through advances in estimation techniques proposed [16], [27], [40]. This advance, in turn, has generated an increasing amount of interest in the usage of these models in academia, international organizations, central banks, and governments, simply because they can accommodate a large panel of time series when forecasting variables.

However, there is still a considerable degree of divergence in opinion as to whether or not factor models with large cross-sections of time series tend to outperform traditional econometric models with limited numbers of variables. On the one hand, [14, 16, 17, 19, 21, 32, 35-40, 42] found evidence of improvements in the forecasting performances of macroeconomic and financial variables using factor analysis, while on the other hand, other works found only minor or no improvements in forecasting ability [3, 20, 33, 34]. These conflicting results have led to attracting debate as to whether or not the victory claimed by the proponents of large models was precocious. Some attribute the success of large models to different circumstances pertaining to each study. For example, Banerjee et al. [8] found that small models forecast macroeconomic variables better than factor models. In addition, they also find that the performances of factor models differ between countries. Factor models are comparatively good at forecasting real variables in the US relative to the euro area, while the euro area nominal variables are easier to predict than the US nominal variables, using factor models. Furthermore, Boivin and Ng [10] claim that the composition of the data set and the dimensions of the cross-section are important in producing better forecasts from factor models.

Based on the success of the dynamic factor model many linear extensions were introduced such as factor augmented vector autoregressive (FAVAR) - Bernanke et al. [9] - and factor augmented error correction model (FECM) - Banerjee and Marcellino [7] - and their Bayesian applications. Our factor model extension brings together factor model and the nonlinear ANN model, the mixture that we believe can accommodate the structural breaks.

Against this backdrop, this paper exploits the information contained in the large-dimensional factor model framework developed by Forni et al. [16] (hereafter FHLR) for forecasting Johannesburg Stock Exchange (JSE) share prices and a measure of the short-term nominal interest rate (Treasury Bill Rate) for South Africa, over the out of sample period from 2007:01 to

2011:12[1], with an in-sample estimation period from 1992:01 to 2006:12. The forecasting performances of the Factor Models (FMs), estimated under linear dynamic factor model (DFM) and nonlinear Factor – Artificial Neural Network (FANN) assumptions with regard to the interaction between the factors and the variables of interest are investigated. The FMs are evaluated and compared with the performances of two other alternative models, namely Autoregressive (AR) and Artificial Neural Network (ANN) models, on the basis of the Root Mean Squared Error (RMSE) of the out-of-sample forecasts. In a related project Babikir and Mwambi (forthcoming) used the DFM and ANN to evaluate their individual and combined forecasts performance.

In this paper we introduce the FANN model, where we model the extracted factors using ANN nonlinear method to forecast the variables of interest. The nonlinear Factor-ANN results compare to the results of the DFM and ANN models. To the best of our knowledge this is the first attempt to use the FANN model to forecast variables in general and in South Africa in particular.

The empirical results show sizable gains in terms of the forecasting ability of the FANN compared to both the standard ANN and the DFM. Thus the FANN represents an improvement with respect to the standard ANN and the DFM.

The remainder of the paper is organized as follows: Section 2 describes the FMs and ANN forecasting models; Section 3 presents the data; and the results from forecasting models are discussed in Section 4. Finally, we close with Conclusions in Section 5.

II. THE MODELS

This paper uses the FM to extract common components from a large set of variables, after which these common components are used to forecast the variables of interest using the linear DFM and the nonlinear ANN methods.

A. *Estimation Of The Factors And The Dynamic Factor Model*

Let the panel of observations X_t be the N stationary time series variables with observations at times t = 1,......., T, where it is assumed that the series have zero mean. The idea behind the factor model is that most of the variance of the data set can be explained by a small number $q \ll N$ of factors contained in the vector f_t. In general the dynamic factor model representation is given by

$$X_t = \chi_t + \xi_t = \lambda(L)' f_t + \xi_t \tag{1}$$

where χ_t are the common components driven by factors f_t, and ξ_t are idiosyncratic components for each of the variables. In particular ξ_t is that part of X_t that cannot be explained by the common components. The common component is a function of the $q \times 1$ vectors of dynamic factors which are common to all variables in the set $f_t = (f_{1t} \,.....\, f_{qt})'$, the operator $\lambda(L) = 1 + \lambda_1 L + \cdots + \lambda_s L^s$ is a lag polynomial with positive powers on the lag operator L with $Lf_t = f_{t-1}$. This way the lags of the factors are allowed to affect the current movement of the variables. The model can be re-written in static representation as;

$$X_t = \Lambda' F_t + \xi_t \tag{2}$$

where F_t is a vector of $r \geq q$ static factors that comprise of the dynamic factors f_t and all lags of the factors. Basically there are three methods of estimating the factors in F_t from a large data set. These methods were developed by Forni et al. [16] (hereafter FHLR[2]), Stock and Watson [39] (hereafter SW) and [27]. In the current paper we employ the estimation method developed by FHLR. Below, we give a brief description of SW and FHLR methods and how they differ.

First we start with the SW model where the authors proposed estimating F_t with static principal component analysis (PCA) applied to X_t . The factor estimates are simply the first r principal components of X_t which according to SW are $F_t = \hat{\Lambda}' X_t$, where $\hat{\Lambda}$ is the $N \times r$ matrix of the eigenvectors corresponding to the r largest eigenvalues of the sample covariance matrix $\hat{\Sigma}$.

On the other hand, FHLR propose a weighted version of the principal components estimator suggested by SW, where the series are weighted according to their signal-to-noise ratio, which is estimated in the frequency domain. The estimation of common and idiosyncratic components is conducted using two steps. First, the covariance matrices of the common and idiosyncratic components of X_t are estimated via dynamic PCA. This involves estimating the spectral density matrix of $X_t, \Sigma(\omega)$, which has rank q. For each frequency ω, the largest q eigenvalues and the corresponding eigenvectors of $\Sigma(\omega)$ are computed, and the spectral density matrix of the common components $\sum_\chi(\omega)$ is estimated. Then it follows that the spectral density matrix of the idiosyncratic components is given by $\hat{\Sigma}_\xi(\omega) = \hat{\Sigma}(\omega) - \hat{\Sigma}_\chi(\omega)$. Inverse Fourier transform provides the time-domain autocovariances of the common and the idiosyncratic components given by $\hat{\Gamma}_\chi(k)$ and $\hat{\Gamma}_\xi(k)$ for lag k. Since dynamic PCA corresponds to a two-sided filter of the time series, this approach alone is not suited for forecasting. Second, a search is undertaken for the r linear combinations of X_t that maximize the contemporaneous covariance explained by the common factors $\hat{Z}_i{}'\hat{\Gamma}_\chi(0)\hat{Z}_i$, $i = 1, \ldots, r$. This optimization problem is subject to the normalization $\hat{Z}_i'\hat{\Gamma}_\xi(0)\hat{Z}_j = 1$ for $i = j$ and zero otherwise. This representation can be reformulated as the generalized eigenvalue problem such that $\hat{\Gamma}_\chi(0)\hat{Z}_i = \hat{\mu}_i \hat{\Gamma}_\xi(0)\hat{Z}_i$, where $\hat{\mu}_i$ denotes the *i*-th generalized eigenvalue and $\hat{Z}_i$ its $N \times 1$ corresponding eigenvector

[1] The choice of the out-of-sample span comes from the aim to investigate the performance of forecasting models during the period of financial crisis.

[2] For further technical details on this type of factor models, see Schumacher [33].

in their non-null spaces. The factor estimates according to FHLR are then obtained as $\hat{F}_t = \hat{Z}'X_t$ with $\hat{Z} = [\hat{Z}_1 \ldots \hat{Z}_r]$.

B. Dynamic Factor Model

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 [16, 39] which take the form:

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

where $\hat{f}_t$ are dynamic factors estimated using the method by FHLR while $\beta(L)$ and $\gamma(L)$ are the lag polynomials, which are determined by the Akaike Information Criterion (AIC) and the Schwarz Information Criterion (SIC). The u_{t+h} is an error term. The coefficient matrix for factors and autoregressive terms are estimated by ordinary least square (OLS) for each forecasting horizon h. To generate the estimate and forecast of the Autoregressive (AR) benchmark we impose a restriction to Eq. (3), where, we set $\beta(L) = 0$[3].

C. The Artificial Neural Network (ANN)

A neural network model can be described as a type of multiple regression in that it accepts inputs and processes them to predict some output. ANN can offer a valid approximation to the generating mechanism of a vast class of non-linear processes; see for example [24, 28, 41] for their use as forecasting tools. There are a number of properties that make the ANN model an attractive alternative to traditional forecasting models[4]. Most importantly ANN models control or are resistant to the limitations of traditional forecasting methods, including misspecification, biased outliers and assumption of linearity [23]. The most significant advantages of the ANN models over other classes of nonlinear models is that ANNs are universal approximators that can approximate a large class of functions with a high degree of accuracy, see [12, 43]. The network used in this paper is a single hidden layer feed-forward network with n nodes in the hidden layer and linear jump connection or linear neuron activation function (see Fig 1) specified as follows:

$$y_{t+h} = \alpha_0 + \sum_{j=1}^{n} w_j\, g\left(\alpha_{0,j} + \sum_{i=1}^{p} w_{i,j}\, y_{t-i}\right) + \sum_{i=1}^{p} \beta_i\, y_{t-i} + \varepsilon_{t+h} \quad (4)$$

where inputs y_{t-i} represent the lagged values of the variable of interest and the output y_{t+h} is the variable being forecast, h indicates the forecast horizon, where $w_{i,j}(i = 1,2, \ldots, p, j = 1,2, \ldots, n)$ and $w_j(j = 1,2, \ldots, n)$are the weights that connect the inputs to the hidden layer and the hidden layer to output respectively, α_0 is the bias. The function g is a logistic function given by $g(x) = \frac{1}{1+e^{-x}}$. The ε_{t+h} is an error term. The third summation in Equation (4) shows the jump connection or skip-layer network that directly links the inputs y_{t-i} to the output y_{t+h} through β coefficients. The most important feature about this model is the combination of the pure linear model and feed-forward neural network. Therefore, if the relationship between inputs and output is pure linear, then only the skip-layer given by coefficient set β should be significant, and if the relationship is nonlinear one expects the coefficients set w and α to be highly significant, while the jump connections coefficient β will be relatively insignificant. Finally however, if the underlying relationship between input and output is mixed, then we expect all types of coefficient sets to be significant. The model is estimated by recursive least square using the Broyden, Fletcher, Goldfarb and Shanno (BFGS) algorithm [26]. The selection of the lag lengths and the number of nodes in the hidden layer are chosen on the basis of the training set or the in-sample RMSE, where n=5.

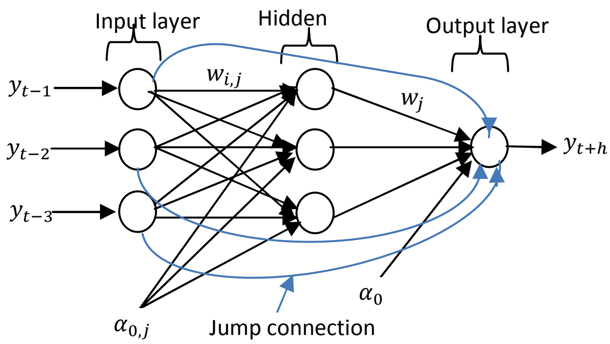

Fig. 1: Structure of the best fitted network, $\mathbf{N}^{(3,5,1)}$

D. Proposed Factor – Artificial Neural Network (FANN) Model

Previous researches argued that combined models improve the predictive performance of time series forecasting. The combined models reduce the risk of using an inappropriate model as the underlying process cannot easily be determined; thus the hybrid model can reduce this risk failure and obtain more accurate results. In this paper we propose a hybrid model of artificial neural network and factor model in order to yield an enhanced predictive and forecast performance. The factor model (FM) extract components that are common between the 228 time series variables. The factor model expresses individual time series as the sum of two unobserved components, a common component, which is driven by a small number of common factors, and an idiosyncratic component which is specific to each variable. The FM is able to extract a few factors that explain the co-movement of all variables. Our proposed model used the FHLR approach explained above to extract these factors at the first step.

In the second step, a neural network is used in order to model the nonlinear and linear relationships existing in

[3] We use the autoregressive model as our benchmark.

[4] For more details about the strengths and drawbacks of ANN, see Ramlall [25].

the factors f_t and y_t the variable we need to forecast (see Fig 2), as follows:

$$y_{t+h} = \alpha_0 + \sum_{j=1}^{n} w_j \, g\left(\alpha_{0,j} + \sum_{i=1}^{p} w_{i,j} \, f_{t,i}\right) + \sum_{i=1}^{p} \beta_i \, f_{t,i} + \varepsilon_{t+h} \qquad (5)$$

where $w_{i,j}(i = 1,2,\ldots,p, j = 1,2,\ldots,n)$ and $w_j(j = 1,2,\ldots,n$ are the weights that connect the inputs to the hidden layer and the hidden layer to output respectively, p is the number of factors. In our application we arrive at $p = 5$ as determined by Bai and Ng [5] approach and also supported by Onatski [29, 30] test, n is the number of nodes in the hidden layer, α_0 is the bias. The function g is a logistic function, where $g(u) = \frac{1}{1+e^{-u}}$. The coefficients β represent the linear part of the equation (5) which directly links the inputs f_i to the output y_{t+h}. The ε_{t+h} is an error term. The number of nodes in the hidden layer are determined on the basis of the training set or in-sample RMSE, where n=3.

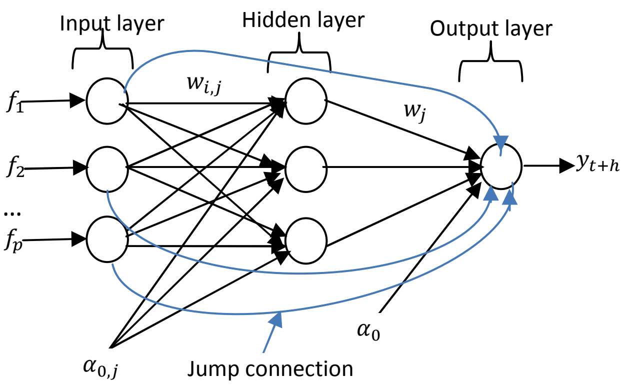

Fig. 2: The best fitted network structure, $\mathbf{N}^{(5,3,1)}$

III. DATA

The data set contains 228[5] monthly series, 203 from South Africa, covering the financial, real and nominal sectors, two global variables and 23 series of major trading partners and global financial markets. Thus besides the national variables, the paper uses a set of global variables such as gold and crude oil prices. In addition the data also includes series from financial markets of major trading partners namely the United Kingdom, the United States, China and Japan. The in-sample period contains data from 1992:01 to 2006:12, while the out-of-sample set spans from 2007:01 to 2011:12. The Augmented Dickey-Fuller (ADF) test is used to assess the degree of integration of all series. All non-stationary series are made stationary through differencing. 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. All series are standardized to have a mean of zero and a constant variance.

Recently the determination of the number of the factors has been developed for both the case of the static factor model [1] and [5] and for the dynamic factor model [2, 6, 22, 29, 30]. To specify the number of static factors, [1] and [5] use information criterion, based on AIC and BIC, to help guide the selection of the optimal number of factors r in a large data set. We apply the Bai and Ng [5] approach which proposes five static factors. Onatski [29] developed a statistical test to test and determine the number of dynamic factors under the null hypothesis that the number of factors is equal to k_0 against the alternative $k_1 > k_0$ (for details see [29]). In our case the test suggests two dynamic factors, which both explain more than 87% of variation of the entire data panel.

IV. RESULTS

A. *In-Sample Results*

In this subsection we evaluate the in-sample predictive power of the fitted models. We estimate the forecasting models using the full sample, in order to check the robustness of our in-sample results. In-sample forecasting is most useful when it comes to examining the true relationship between the set of predictors and the future predictions of the variable of interest. Table 1 below reports the RMSE[6] of the in-sample forecasting results. The table reports the RMSE statistics for the AR benchmark model and the ratio of the RMSE for the other models to the RMSE for the AR benchmark model. Thus, the ratio that is higher than one indicates that the method under analysis is worse than the benchmark, so the model with a lowest RMSE ratio is deemed to perform better than the other models. Our proposed FANN model out preformed all other models with a large reduction in RMSE relative to the AR benchmark model for both variables. The reason is potentially because we merge the factors that efficiently handle large amounts of information that include external variables that influence South African economy with ANN nonlinear estimation model. The ANN model also provides fairly better in-sample forecasts compared to the AR benchmark model and DFM model. In general the FANN and ANN nonlinear models preform much better than DFM and AR linear models.

TABLE 1: THE RMSE OF THE IN-SAMPLE FORECASTS

Forecasting model	Treasury Bill Rate	JSE all Share prices
AR (benchmark model)	0.8860	0.9747
DFM	0.9369	0.9511
FANN	**0.6868**	**0.6536**
ANN	0.7731	0.8431

Note: the first 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.

[5] The data sources are South Africa Reserve Bank, ABSA Bank, Stats South Africa, National Association of Automobile Manufacturers of South Africa (NAAMSA), South African Revenue Service (SARS), Quantec and World Bank.

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

B. Out-Of-Sample Forecasting Results

In this subsection, we evaluate out of sample forecasts of the Treasury Bill Rate and JSE all share prices over the period 2007:01 to 2011:12. This period includes the global financial crisis that impacted the South African economy at the end of 2008 and 2009. We consider short forecast horizon of 3 months and long forecast horizon of 12 months. Table 2 below reports the RMSE statistics for the AR benchmark model in the first row and the ratio of the RMSE of other models to the RMSE for the AR benchmark model. The result of the AR benchmark model shows that the RMSE increases as horizon increases, and indicates that more accurate forecasts for the AR are available at shorter horizons. Note; in this paper we choose iterated forecast instead of direct forecast, on the other hand, the forecasts constructed recursively, using all available data to estimate parameters. The results of the two variables can be summarized as follows:

Treasury Bill Rate: the proposed FANN model outperforms all other models for short horizon producing the lowest RMSE followed by the AR benchmark model. For the long horizon, the ANN outperforms all other models followed by the FANN model. The FANN result shows that the RMSE increases as the forecast horizon increases. Compared to the DFM[7] model the FANN model performs better, thus the estimation method used to model the factors matters.

***JSE all Share prices**:* the FANN model stands out in forecasting the JSE all share prices for both short and long horizons with a sizable reduction in RMSE relative to the AR benchmark model of 8 percent to 19 percent. The DFM outperforms the ANN and AR benchmark model, thus the derived factor models FANN and the DFM outperform univariate linear and nonlinear models AR and ANN respectively. These results clearly indicate the importance of the information contained in the common factors, which in turn, are derived from 228 monthly series. The performance of the FANN model over the DFM model indicates the role of the estimation method that captures the nonlinearity associated to the variables of interest. Babikir et al. [4] found evidence of structural breaks in the JSE share return index in the end of 2008 and mid of 2009. These events are included in our out-of-sample period, thus it shows that the FANN model captures well the structural breaks compared to the DFM and the other models.

We attribute the forecast performance of derived factor models the FANN and the DFM for the JSE all Share prices over the Treasury Bill Rate to the data set used to extract the factors which contains more financial than macroeconomic variables.

[7] Gupta R. and Kabundi A. [21] found that the DFM model outperforms the other models they used to forecast Treasury Bill Rate for South Africa.

TABLE 2: THE RMSE OF OUT-OF-SAMPLE (2007:01 – 2011:12) FOR 3 AND 12 MONTH HORIZONS

Forecasting model	h = 3	h = 12
Treasury Bill Rate		
AR benchmark	0.5208	0.6919
DFM	1.1334	0.9829
FANN	**0.9453**	0.9364
ANN	1.0620	**0.7524**
JSE all Share prices		
AR benchmark	1.7743	1.8187
DFM	0.9655	0.9532
FANN	**0.8150**	**0.9273**
ANN	1.0325	1.0947

Note: the first 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.

In order to assess the FANN model forecast accuracy, we perform the cross model test of the FANN against other models, namely AR, DFM and ANN. The cross-model test is based on the statistic proposed by Diebold and Mariano [15], which 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 FANN model and e_{2t}^2 denotes the forecast errors from the AR benchmark model, the DFM and ANN. The S statistic follows a standard normal distribution asymptotically. Note, a negative and significant value of S indicate that the FANN model outperforms the other model in out-of-sample forecasting. Table 3 below shows the test results. In general the FANN model outperforms the AR and ANN in predicting the two variables of interest and for each of the short and long horizon forecasts. In other words, based on RMSE and on the Diebold and Mariano test statistics, we have relatively strong evidence that there is a significant statistical gain from using the FANN model over other models. We note that there is no significant statistical difference between the forecasts of factors derived models namely the FANN and the DFM in most cases. For the JSE all share prices variable the forecast of the FANN model out performs linear and nonlinear univariate models for 12 month horizons with at least 5% level of significant, and outperforms all other models for the 3 month horizons in particular.

TABLE 3: DIEBOLD – MARIANO TEST (2006:01 – 2011:12)

Model	Forecasting Horizons	
	3 month	*12 month*
Treasury Bill Rate		
FANN vs. AR	-3.3174 ***	-2.5733**
FANN vs. DFM	-1.2825	0.0503
FANN vs. ANN	-0.5206	3.040**
JSE all share prices		
FANN vs. AR	-3.0829**	-2.0972**
FANN vs. DFM	-3.7276***	0.7960
FANN vs. ANN	-2.7126**	-2.3940**

Note: ***, ** and * indicate significant at the 1%, 5% and 10% levels respectively.

V. CONCLUSION

In this paper, the Factor Models (FMs) are applied to introduce a new hybrid method for improving the time series forecasting performance of the artificial neural networks. The model used the factors that were extracted from 228 monthly series. Five static factors and two dynamic factors were extracted which explain more than 87% of the variation in the data panel. These factors are then used as independent variables or inputs to the ANN in a model we call the factor ANN model (FANN) and to estimate the common linear DFM. Besides the FANN and the DFM, we estimate standard ANN and AR benchmark model. The four models were then used to forecast the Johannesburg Stock Exchange (JSE) share prices and the Treasury Bill Rate over the estimation period 1992:01 to 2006:12. The models were evaluated based on the RMSE for 3 and 12 month ahead forecasts over an out-of-sample horizon of 2007:01 to 2011:12.

The in-sample results showed the superiority of the FANN over the other models. The FANN outperformed the AR benchmark model with a large reduction in RMSE of around 31 percent to 35 percent. The model outperformed the standard ANN model but the ANN model outperformed the DFM, which in turn, performed better than AR benchmark model.

In general the out-of-sample results revealed that the best performed model appears to be our proposed FANN model, followed by the DFM model. These results confirmed the usefulness of the factors that were extracted from large related variables. On the other hand, as far as estimation is concerned the nonlinear FANN model was suitable to capture nonlinearity and structural breaks compared to linear models. Thus the structural breaks associated to the financial crisis that affected the economy can explain the failure of the linear DFM compared to the nonlinear FANN model. The results of Diebold-Mariano test suggested that the FANN model produced forecasts that were significantly better than the AR benchmark model forecasts, and the standard ANN model forecasts.

Further research can evaluate the FANN forecasting performance in small and large simulated samples and compare it to FAVAR model.

ACKNOWLEDGMENT

The authors would like to thank the Editor and the reviewers for their insightful and constructive comments.

REFERENCES

[1] Alessi L., Barigozzi M. and Capasso M., “A robust criterion for determining the number of static factors in approximate factor models,” Working paper No 903, European Central Bank, 2008.

[2] Amengual D. and Watson M.W., “Consistent estimation of the number of dynamic factors in a large N and T panel,” Journal of Business and Economic Statistics, vol. 25, pp. 91-96, 2007.

[3] Angelini, E., Henry, J., & Mestre, R., “Diffusion index-based inflation forecasts for the euro area,” Working Paper No. 61. European Central Bank, 2001.

[4] Babikir A., Gupta R., Mwabutwa C. and Owusu-Sekyere E., “Structural breaks and GARCH models of stock return volatility: The case of South Africa. Economic Modelling,” vol. 29, pp. 2435–2443, 2012.

[5] Bai, J., & Ng, S., “Determining the number of factors in approximate factor models,” Econometrica, vol. 70, pp. 191–221, 2002.

[6] Bai, J., & Ng, S., “Determining the number of primitive shocks in factor models,” Journal of Business and Economic Statistics, vol. 25, pp. 52–60, 2007.

[7] Banerjee, A. and Marcellino, M., “Factor augmented error correction models,” CEPR Discussion Paper, 6707, 2008.

[8] Banerjee, A., Marcellino, M., & Masten, I., “Leading indicators for euro-area inflation and GDP growth,” Oxford Bulletin of Economics and Statistics, vol. 67, pp. 785–814, 2005.

[9] Bernanke, B., Boivin, J. and Eliasz, P., “Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach,” Quarterly Journal of Economic, vol. 120, pp. 387-422, 2005.

[10] Boivin, J., & Ng, S., “Are more data always better for factor analysis?,” Journal of Econometrics, vol. 132, pp. 169–194, 2006.

[11] Chamberlain, G., “Funds, factors, and diversification in arbitrage pricing models,” Econometrica, vol. 51, pp. 1281–1304, 1983.

[12] Chamberlain, G., & Rothschild, M., “Arbitrage, factor structure and mean-variance analysis in large markets,” Econometrica, vol. 51, pp. 1305–1324, 1983.

[13] Chen, A., Leung, M. T., and Hazem, D., “Application of neural networks to an emerging financial market: Forecasting and trading the Taiwan Stock Index,” Computers and Operations Research, vol. 30, pp. 901–923, 2003.

[14] Cristadoro, R., Forni, M., Reichlin, L., & Veronese, G., “A core inflation indicator for the euro area,” Journal of Money, Credit and Banking, vol. 37, pp. 539–560, 2005.

[15] Diebold, F. X., & Mariano, R. S., “Comparing predictive accuracy,” Journal of Business and Economic Statistics, vol. 13, pp. 253– 263, 1995.

[16] Forni M., Hallin M., Lippi M. and Reichlin L., “The generalized dynamic factor model: One sided estimation and forecasting,” Journal of the American Statistical Association, vol. 100, pp. 830-840, 2005.

[17] Forni, M., Hallin, M., Lippi, M., & Reichlin, L., “Coincident and leading indicators for the euro area,” The Economic Journal, vol. 111, pp. 62–85, 2001.

[18] Geweke, J., “The dynamic factor analysis of economic time series”. In D. J. Aigner, & A. S. Goldberger (Eds.), Latent variables in socio-economic models (pp. 365–383). Amsterdam: North Holland, 1977.

[19] Giannone, D., & Matheson, T. D., “A new core inflation indicator for New Zealand,” International Journal of Central Banking, vol. 3, pp. 145–180, 2007.

[20] Gosselin, M.-A., & Tkacz, G., “Evaluating factor models: an application to forecasting inflation in Canada,” Working Paper No. 18. Bank of Canada, 2001.

[21] Gupta R. and Kabundi A, “A large factor model for forecasting macroeconomic variables in South Africa,” vol. 27(4), pp. 1076–1088, 2011.

[22] Hallin M. and Liska R., “Determining the number of factors in the general dynamic factor model,” Journal of the American Statistical Association, vol. 102, pp. 603-617, 2007.

[23] Hill, T., O'connor, M. and Remus, W., “Neural network models for time series forecasts,” Management Science, vol. 42, pp. 1082-1092, 1996.

[24] Hornik, K., Stinchcombe, M. and White, H., “Multi-layer feed forward networks are universal approximators,” Neural Network, vol. 2, pp. 359-66, 1989.

[25] Indranarain Ramlall, “Artificial Intelligence: Neural Networks Simplified,” International Research Journal of Finance and Economics, vol. 39, 2010.

[26] Jorge Nocedal and Stephen J. Wright, “Numerical Optimization, 2nd ed., Berlin, New Yor. Springer, 2006.

[27] Kapetanios, G., & Marcellino, M., “A parametric estimation method for dynamic factor models of large dimensions,” Journal of Time Series Analysis, vol. 30(2), pp. 208–238, 2009.

[28] Omidi, A., Nourani, E. and Jalili M., “Forecasting stock prices using financial data mining and Neural Network. In the Proceedings of 3rd IEEE International Conference on Computer Research and Development, Shanghai, pp. 242-246, 2011.
[29] Onatski A., “Testing hypotheses about the number of factors in large factor models,” Econometrica. Vol. 77, pp. 1447-1479, 2009.
[30] Onatski A., “Determining the number of factors form empirical distribution of eigenvalues,” The Review of Economics and Statistics, vol. 92, pp. 1004-1016, 2010.
[31] Sargent, T. J., & Sims, C. A., “Business cycle modeling without pretending to have too much a priori economic theory,” In C. Sims (Ed.), New methods in business research. Federal Reserve Bank of Minneapolis, 1977.
[32] Schneider, M., & Spitzer, M., “Forecasting Austrian GDP using the generalized dynamic factor model,” Working Paper No. 89. Oesterreichische National Bank, 2004.
[33] Schumacher, C., “Forecasting German GDP using alternative factor models based on large dataset,” Journal of Forecasting. Vol. 26(4), pp. 271–302, 2007.
[34] Schumacher, C., and Dreger, C., “Estimating large-scale factor models for economic activity in Germany: do they outperform simpler models?,” Jahrbücher für Nationalökonomie und Statistik, vol. 224, pp. 731–750, 2004.
[35] Stock, J. H., & Watson, M. W., “New indexes of coincident and leading economic indicators,” NBER Macroeconomics Annual, pp. 351–393, 1989.
[36] Stock, J. H., & Watson, M. W., “A probability model of the coincident indicators. In K. Lahiri, & G. Moore (Eds.), Leading economic indicators: new approaches and forecasting record,” Cambridge: Cambridge University Press, 1991. pp. 63–95.
[37] Stock, J. H., & Watson, M. W., “A comparison of linear and non-linear univariate models for forecasting macroeconomic time series.” In R. Engle, & R. White (Eds.), Cointegration, causality, and forecasting: A festschrift in honor of Clive W.J. Granger, Oxford: Oxford University Press. 1999, pp. 1 – 44.
[38] Stock, J. H., &Watson, M. W., Forecasting inflation,” Journal of Monetary Economics, vol. 44, pp. 293–335, 1999.
[39] Stock, J. H., & Watson, M. W., “Forecasting using principal components from a large number of predictors,” Journal of the American Statistical Association, vol. 97, pp. 147–162, 2002a.
[40] Stock, J. H., &Watson, M. W., “Macroeconomic forecasting using diffusion indexes,” Journal of Business and Economic Statistics, vol. 20, pp. 147–162, 2002b.
[41] Swanson, Norman R. and Halbert White, “A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks,” The Review of Economics and Statistics, vol. 79 (4), pp. 540–550, 1997.
[42] van Nieuwenhuyze, C., “A generalized dynamic factor model for the Belgian economy: identification of the business cycle and GDP growth forecasts,” Journal of Business Cycle Measurement and Analysis, vol. 2, pp. 213–248, 2005.
[43] Zhang, G. P., and Qi, G. M., “Neural network forecasting for seasonal and trend time series.” European Journal of Operational Research, vol. 160, pp. 501–514, 2005.