Forecasting government expenditures in South Africa with a dynamic artificial neural networks: Does population aging play a role?

Manoochehr Ghiassi[1], Beatrice D. Simo-Kengne[2]*

[1] *Santa Clara University, Santa Clara, California, 95053, USA*
[2] University of Johannesburg, *Auckland Park, 2006, South Africa*
* *Corresponding author's e-mail: bdsimo-kengne@uj.ac.za*

Abstract

The government of South Africa spends a significant portion of its GDP in support of its public policy including heath care (8.79% in 2014) and social grants (3% in 2013/2014 of which 41% accounts for old age grants). Public policy strategies over a 5-year period from 2010/2011 to 2014/2015 has increased by 39%, moving from ZAR 33764billion ($2597 billion) to ZAR 50336billion ($3872 billion). The growth of the old age grants is expected to continue. Accurate forecasting of such expenditures enables policy makers and government planners for better assessment, planning, and the ultimate allocation of funds in support of their decisions. We address this specific objective by developing a set of time series forecasting models which consider governmental expenditure over time and accounts for the aging population in this process.

We offer two models: the first one based on ARIMAX and we introduce a second model that uses a Dynamic Architecture for Artificial Neural Network (DAN2). We assess the performance of these models by using RSME, MAPE and MAD statistics. The DAN2 model, using actual data, for quarterly public expenditure ratio of GDP from 1960-Q1 to 2016-Q4 along with the corresponding demographic values, resulted in 97% forecasting accuracy. Furthermore, the relatively high forecast precisions obtained across the two models, suggest that demographic changes is an important predictor of government expenditure in South Africa; implying that demographic monitoring is indispensable for efficient fiscal planning and management.

Keywords. Public expenditure, demographic trends, ARIMAX, ANN, DAN2.
JEL Classification. H5, J11, C45

1. Introduction

A sound governmental fiscal planning and management requires accurate expenditure predictions. This is particularly crucial for developing countries characterized by persistent budget imbalances. In addition, modifications in fiscal expenditures due to unanticipated changes in socio-economic and demographic conditions entail delayed effects which necessitate accurate forward-looking policy strategies and/or actions. To address such uncertainties, the economic and forecasting literature (Lassila et al., 2014; Kudrna et al., 2015) advocates incorporation of both demographic and macroeconomic factors in predicting future government expenditures.

While the number and/or quality of predictor matters for the forecasting accuracy, the prediction precision is also driven by the forecasting method employed. Some researchers, Timmermann (2006) have offered grouping of the forecast methods (combinations or ensembles) in order to provide a more accurate forecasts over the best single model; the principal attraction being the possibility to hedge against model uncertainty. They postulate that such a forecast combinations, makes use of various information sets, predictors, and modeling structures to better accommodate structural break than single model while alleviating the potential misspecification bias and measurement errors (Ghysels and Ozkan, 2015).

The most recently used class of forecast combinations is the mixed data sampling (MIDAS) models which consist of pooling different specifications to compare the forecasting performances of different predictors (Kuzin et al. (2013). An illustrative example is Andreou et al. (2013) who combine many cross sectional financial series to derive MIDAS predictors of output growth. Marcellino and Schumacher (2010) use the factor-MISDAS models based on the parsimonious principal component analysis whereas Andeou et al. (2011) build a MIDAS with autoregressive distributed lag (ADL-MISDAS) models to accommodates mixed frequencies indicators (daily and quarterly). ADL-MISDAS models have also been applied by Ghysels and Ozkan (2015) to forecast US federal government budget using root mean squared forecast errors (RMSE) as accuracy metric. Their combination includes fiscal and macroeconomic indicators at different frequencies.

Recently, researchers have offered Artificial Neural Networks (ANN) as an alternative approach to time series based forecasting (Zang et al., 1998; Zang, 2003; Ghiassi et al., 2006; Ghiassi et al., 2008 among others). For example, Ghiassi et al. (2006) introduce a dynamic architecture for artificial neural network (DAN2) to forecast electric consumption for Taiwan, and Ghiassi et al. (2008) to forecast urban water demand for the city of San Jose, California. They find that DAN2 delivers exceptional fit and forecasts using time series datasets. Similarly, Basaran et al. (2010) implement a feed forward neural network architecture to forecast public expenditures in Turkey and obtain high prediction precision.

This study aims to develop a model based on the performance of a traditional approach (ARIMAX) and compare and contrast its performance against a model based on DAN2 to predict government expenditures in South Africa on social welfare. The accurate forecast values from these models can be used to assist policy makers to better manage and/or keep fiscal imbalances at sustainable levels. Moreover, with the increasing pace of elderly cohort expected to trigger fiscal pressures, this study analyses the role of population aging in such forecast models.

In fact, in most developing countries, government is the primary finding support to old people whose vulnerability has been triggered by poverty and inequalities. Besides the free health care, this population group in South Africa is entitled to the financial assistance in the form of social grant (old age grant) with a significant eligibility increase since the political transition in 1994[1]. The improvement of the health system in South Africa has also increased life expectancy (Younger, 2016) while the retirement age has not been increased proportionally (due to contractual agreements, tendency of governments to use retirement as the vehicle for "job-creation," i.e. creating jobs for younger people). On the other hand, the government revenue has not kept pace with this rapid rate of growing proportionally to the rapid pace in population aging. Since the existing government revenue cannot keep up with demand, policy decisions ranging from "qualification age" must be considered and any such decision to "rationing of funds" will require accurate expenditure forecasts.

While fiscal policy and its macroeconomic impact have been studied in South Africa (Akanbi, 2013), it is surprising how public expenditure forecasting has received very little or virtually no attention. With the increasing pressure on government expenditures due to rapid changes in socio-economic and demographic structures, informed decision has become an imperative to achieve efficient budget allocation. Considering that aging population occurs as a result of rising survival and decline in fertility rate, all of which characterising the demographic transition in South Africa[2], this paper incorporates population aging as an important predictor of government expenditure in South Africa.

[1] According to Statistics South Africa, the number of beneficiaries has increased from under 3 million in 1997 to 11.8 million in 2007 and 17 million in 2017.
[2] Demographic stats.

The rest of the paper is organized as follows. Section 2 succinctly reviews the literature on expenditure modelling in relation to the fiscal environment. The methodology and empirical analysis are presented in sections 3 and 4, respectively. Concluding remarks and policy recommendations are then discussed in section 5.

2. Literature review

In the fiscal literature, government expenditure forecasting has evolved across time mainly due to the development of improved forecasting methods. From the regression analysis to the time series approaches, many studies predict that economic, social, demographic, institutional and political variables help explain the trend in public expenditures. Borcherding (1985) surveys the causes of government expenditure growth in the US and finds that familiar substitution, income, population/public good-tax sharing effects contribute to about 40% of the government spending besides institutional factors including rent-seeking political redistribution and bureaucracy. Peacock (2004) emphasizes that economic growth is a key driver of public expenditures in selected industrialised countries while Dizaji (2014) shows that oil price is an important predictor of public expenditures in oil exporting countries. In developing countries, the degree of openness, the level of economic development, the rent-seeking behaviour and the demographic factors are emphasized in explaining the growth in government spending (Fan et al., 2008). While higher proportion of young cohort affect the demand for education spending, population aging resulting from the demographic shift induces higher spending demand of health care, housing and social security (Murthy and Okunade, 2016; Braendle and Colombier, 2016, Jimeno et al., 2006).

Despite the wide range of predictors, government expenditure forecast has generally been characterised by inadequate accuracy; possibly due to the forecasting technique used. Though the omission bias might be an important source of poor accuracy, Vasconcelos de Deus et al. (forthcoming) explore various dimensions of government budget balance forecast in Brazil including, economic, political, intuitional, and find that the forecast is indeed of low quality and inefficient. They conclude that forecast errors are due to backward-looking effects and cyclical fluctuations. Similarly, Ericsson (2017) detects biases in the US government forecast of the federal debts and argues that these biases are closely linked to turning points in the business cycle. This suggest the inability of the existing techniques to better capture the nonlinearity induced by the cyclical movement in both predictors and predicted variables.

The time series forecasting toolbox comprises several approaches categorised into linear, nonlinear and hybrid. The linear forecasting literature is dominated by the popular autoregressive integrated moving average (ARIMA), applicable in both univariate and multivariate frameworks, which has found worthwhile applications in many disciplines such social, financial, economic, environmental and engineering. Given the restriction to linear problems, ARIMA was further extended to nonlinear frameworks which led the development of parametric nonlinear methods such as threshold autoregressive (TAR), autoregressive conditional heteroscedastic (ARCH) and general autoregressive conditional heteroscedastic (GARCH) among others.

However, due to misspecification bias and over-parametrisation (particularly severe in multivariate contexts) attributed to parametric nonlinear forecasting models (Khashei and Bijari, 2011), non-parametric approaches such as ANN, have recently received increased attention in time series forecasting. ANN methods offer the following advantages. (i) flexibility in nonlinear mapping enabling the approximation of any continuous function; (ii) imposition of a few restrictions to the data generation process of the observed time series which reduces the misspecification bias; (iii) potential to adapt and remain robust and accurate in nonstationary environment subject to changes across time; (iv) computational benefit by using less parameters than the alternative polynomial, spline and trigonometric expansions with similar approximation rate (Khashier and Bijari, 2011).

The overwhelming attraction to ANN has further materialised in empirical studies which confirm the evidence of accurate forecasting from ANN approaches (see for example Chakraborty et al., 1992; Poli and Jones, 1994; Cottrell et al., 1995; Balkin and Ord, 2000; Berardi and Zhang, 2003; Chen et al., 2005; Jain and Kumar, 2007; Giordano et al., 2007 among others). Although the comparative analysis between ANN and traditional forecasting methods, either linear or nonlinear, at time may provide mixed results (Foster et al., 1992; Taskaya and Casey, 2005), it seems that ANN may perform as good as linear alternative under ideal conditions or when using relatively high frequency data (Tang, 1991; Tang et al., 1993; Zang et al., 1998). According to Denton (1995), ANN outperform the linear models in the presence of misspecification, multicollinearity and outliers; suggesting that ANN solution is appropriate for complex dynamic systems.

Though Ghysels and Ozkan (2015) document the ability of real time forecasting models with mixed frequency data in improving government budget forecast, this multivariate framework is data-driven and therefore might not be appropriate for developing and emerging contexts due to lack of available data. Moreover, public expenditures are subject to multifaceted dynamics (economic, demographic, social, environmental, political…) which render its prediction a difficult task. Such complexities are indeed suitable to ANN models as illustrated by Basaran et al. (2010) who conclusively report a relatively high prediction precision of public expenditure in Turkey using the feed forward, back propagation (FFBP) architecture. Unlike the FFBP structure which is static with restricted adaptively, Ghiassi et al. (2006) develop the dynamic Architecture for Artificial Neural Network (DAN2) which overcomes the limitations of the traditional FFBP, provides more flexibility and a larger spectrum of nonlinear approximation while offering better accuracy.

Furthermore, because the data generation process of a time series is likely to comprise both linear and nonlinear components, component models for forecasting also known as hybrid forecasting techniques, have recently emerged as the optimal solution for accuracy improvement; hence the attention shift to hybrid ANN. Khashier and Bijari (2011) provide the evidence that hybrid ARIMA- ANN models overcome the inconsistency related to ANN forecast in comparison to traditional time series models. In line with the hybrid approach and considering that cyclical variabilities are often cited as a major source of error forecast of public expenditure (Ericsson, 2017 and Vasconcelos de Deus et al., forthcoming), this study proposes DAN2 as a promising solution to public expenditure forecast. To the best of our knowledge, this is the first such attempt for public expenditure forecasting in South Africa. In addition, our empirical strategy accounts for exogenous changes in population structures which represent an important driver of public expenditure growth in developing countries. We therefore offer two forecasting models for our analysis: ARIMAX and DAN2.

3. Methodology

While it is rational to expect government expenditures to be influenced by economic fundamentals, internal forces such as fiscal goals, policies and strategies are likely to also play an important role. Moreover, as indicated earlier, social security goal has been one of the key development strategies in South Africa, particularly in the post-apartheid administration. Not surprisingly, social benefits form an important components of the total expenditures. And given the subjection of this expenditure category to demographic development, government spending is also expected to be determined by demographic factors. Therefore, government expenditures are not only influenced by its own passed values but also by other exogenous predictors, namely economic and demographic. For modeling purposes, this can be reduced to presentation of government spending as a time series process based on GDP and demographics values, that is, public expenditure to GDP ratio as a time series function and one exogenous predictor: demographics.

The traditional solution forecast is the multiple regression model in which the total public expenditures are set to be explained by a set of predictors such as economic, financial, social, political and demographic variables (Borcherding, 1985; Fan et al., 2008). Following Box and Jenkins (1976) methodology and its popularity thereof, ARIMA has become a standard forecasting tool. ARIMA is based upon a simple principle that the history of a times series represents the information set required to efficiently predict its future behavior. Box-Jenkins specification can be viewed as a multiple regression model with a single independent variable (*x*), and at least one autoregressive (AR) and/or moving average (MA) terms. Accordingly, let y_t be our expenditure variable, the structure of ARIMA (*p, d, q*) is given by the following equation:

$$y_t = \alpha + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + ... + \gamma_p y_{t-p} + e_t - \phi_1 e_{t-1} - \phi_2 e_{t-2} - ... \phi_q e_{t-q} \qquad (1)$$

where p and q are the lag length of the AR and MA components, respectively and d is the order of integration; y_{t-i} ; $i = 1,..., p$ are the AR terms and e_{t-j} ; $j = 1,..., q$ are the MA terms.

Formulating this problem as an ARIMAX, the structure of ARMA (*p, q*) derives from the general ARMA (*p, q*) model in which there are p autoregressive and q moving average components, respectively.

$$y_t = \alpha + \beta x_t + \gamma_1 y_{t-1} + \gamma_2 y_{t-2} + ... + \gamma_p y_{t-p} + e_t - \phi_1 e_{t-1} - \phi_2 e_{t-2} - ... - \phi_q e_{t-q} \qquad (2)$$

ARIMAX *(p, d, q*) has a structure which is similar to ARIMA (*p, d, q*) where the time series is differenced d^{th} ; the differencing transformation helps to achieve stationarity for non-stationary variables. This specification emphasizes the importance of exogenous changes in improving the prediction performance.

Because this model imposes the normality assumption and models only linear relationships, we further consider a more realistic framework which is able to capture more complex relations including nonlinear associations. The machine learning algorithm tools, including ANN based systems, are shown to be a very effective modeling tool for solving complex, nonlinear, time series problems. The traditional ANN models are based on FFBP algorithm and are extensively suited in the literature (Zang et al., 1998; Zang, 2003; Tang and Fishwich, 1993; Basaran et al., 2010). Recent advances in ANN has introduced more effective architectures. One such model was introduced by Ghiassi and Saidane (2005) and is shown to be very effective for various modeling problems, including time series events (Ghiassi et al., 2006; Ghiassi et al., 2008; Wang et al., 2010). The prediction accuracy of DAN2 is reported to be better than the traditional ANN, ARIMA or various other modeling approaches (Ghiassi et al., 2006; Ghiassi et al., 2008; Wang et al., 2010; Velasquez and Franco, 2010; Guresen et al., 2011, amongst others). This motivates the use of DAN2 in this research, which is briefly summarized in the next section.

3.1 A dynamic architecture for artificial neural networks

Ghiassi and Saidane (2005) introduced a neural network model, DAN2 (A Dynamic Architecture for Artificial Neural Networks), which employs a different architecture than the traditional neural network (FFBP) models. The general philosophy of the DAN2 model is based upon the principle of learning and accumulating knowledge at each layer, propagating and adjusting this knowledge forward to the next layer, and repeating these steps until the desired network performance criteria are reached. Fig. 1 presents the overall DAN2 architecture. As in classical neural networks, the DAN2 architecture is composed of an input layer, hidden layers and an output layer. The input layer accepts external data to the model. In DAN2, unlike classical neural nets, the number of hidden layers is not fixed a priori. They are sequentially and dynamically generated until a level of performance accuracy is reached. Additionally, the proposed approach uses a fixed number of hidden nodes (four) in each hidden layer. This structure is not arbitrary, but justified by the estimation approach. At each hidden layer, the network is trained using all observations in the training set simultaneously, so as to minimize a stated training accuracy measure such as mean squared error (MSE) value or other accuracy measures such as F_1. As shown in Fig. 1, each hidden layer is composed of four nodes. The first node is the bias or constant (e.g. 1) input node, referred to as the C node. The second node is a function that encapsulates the ‘‘Current Accumulated Knowledge Element’’ (CAKE node) during the previous training step. The third and fourth nodes represent the current residual (remaining) nonlinear component of the process via a transfer function of a weighted and normalized sum of the input variables. Such nodes represent the ‘‘Current Residual Nonlinear Element’’ (CURNOLE nodes). In Fig. 1, the ‘‘I’’ node represents the input, the ''C'' nodes are the constant nodes, the ''G_k'' and ''H_k'' nodes represent CURNOLE nodes, and the ''F_k'' nodes are the CAKE nodes. The final CAKE node represents the dependent variable or the output. At each layer, the previous four nodes (C, G_k, H_k, and F_{k-1}) are used as the input to produce the next output value (F_k). The parameters on the arcs leading to the output nodes, (a_k, b_k, c_k, d_k), represent the weights of each input in the computation of the output for the next layer. The parameter connecting the CURNOLE nodes, μ_k, is used as part of the argument for the CURNOLE nodes and reflects the relative contribution of each input vector to the final output values at each layer. A detailed description of the architecture and its properties are fully presented in Ghiassi and Saidane (2005).

The training process begins with a special layer where the CAKE node captures the linear component of the input data. Thus, its input (content) is a linear combination (weighted sum) of the input variables and a constant input node. These weights are easily obtainable through classical linear regression. If the desired level of accuracy is reached, we can conclude that the relationship is linear and the training process stops. This step is used as the starting point. For nonlinear relations, additional hidden layers are required. At each subsequent layer, the input to the CAKE node is a weighted sum (linear combination) of the previous layer's CAKE, CURNOLE, and C nodes. Throughout training, the CAKE nodes carry an adequate portion of learning achieved in previous layers forward. This process ensures that the performance or knowledge gained so far is adjusted and improved but not lost. This property of DAN2 introduces knowledge memorization to the model. Ghiassi and Saidane (2005) show that the DAN2 algorithm ensures that during network training, the residual error is reduced in every iteration and the accumulated knowledge is monotonically increased. The training process defines creation of partitions among classes that could include linear and nonlinear components. The linear component of the input data is captured in the first CAKE node using ordinary least squares (OLS) or other simple and easy to compute approaches. The algorithm next transforms the input dataset to model the nonlinearity of the process in subsequent iterations. DAN2 uses a vector projection approach to perform data transformation. The transformation process defines a reference vector $R = \{rj; j = 1, 2, . . . , m\}$, where m represents the number of attributes of the observation records, and projects each observation record onto this vector to normalize the data as discussed in Ghiassi and Saidane (2005). This normalization defines an angle, α_i, between record i and the reference vector R. DAN2 uses the set of α_i's to train the network, and updates their values in every iteration. In

Ghiassi and Saidane (2005), the authors show that this normalization can be represented by the trigonometric function Cosine ($\mu_k\ \alpha_i + \theta_k$). In every hidden layer k of the architecture we vary ($\mu_k\ \alpha_i + \theta_k$) and measure the impact of this change on the output value. The modification of the angle ($\mu_k\ \alpha_i + \theta_k$) is equivalent to rotating μ_k and shifting θ_k, the reference vector, thus changing the impact of the projected input vectors and their contribution to the output for that iteration. The Cosine ($\mu_k\ \alpha_i + \theta_k$) uses two (nonlinear) parameters, μ_k and θ_k. The use of the latter can be avoided through the expansion of the cosine function in the form: A Cosine ($\mu_k\alpha_i$) + B Sine ($\mu_k\alpha_i$). We use this functional form as the transfer function in our model. The two CURNOLE nodes in Fig. 1 represent this formulation. At any given hidden layer k, if the Cosine ($\mu_k\ \alpha_i + \theta_k$) terms captured in previous layers do not adequately express the nonlinear behavior of the process, a new layer with an additional set of nodes is automatically generated, including a new Cosine ($\mu_k\ \alpha_i + \theta_k$) term. This process is analogous to how the Fourier series adds new terms to improve function approximation. Therefore, the number of layers in the DAN2 architecture is dynamically defined and depends upon the complexity of the underlying process and the desired level of accuracy. Thus, the output of this model is represented by the linear combination of the constant, CAKE, and CURNOLE nodes. Eq. (3) represents the functional form of this relationship at iteration (layer) k

$$F_k(X_i)= a_k + b_kF_{k-1}(X_i) + c_kG_k(X_i) + d_kH_k(X_i) \qquad (3)$$

where X_i represents the n independent input records, $F_k(X_i)$ represents the output value at layer k, $G_k(X_i)$ = Cosine ($\mu_k\alpha_i$), and $H_k(Xi)$ = Sine ($\mu_k\alpha_i$) represent the transferred nonlinear components, and a_k, b_k, c_k, d_k, and μ_k are parameter values at iteration k. The training process initially captures the linear component by using OLS or other simple and easy to compute approaches. If the desired level of accuracy is reached, the training terminates. Otherwise, the model generates additional layers to capture the nonlinear component of the process by minimizing a measure of total error as represented by $SSE_k = \Sigma[(X_i) - F\hat{}(X_i)]^2$. Substituting $F_k(X_i)$ from Eq. (3) results in:

$$SSE_k = \Sigma[a_k + b_kF_{k-1}(X_i) + c_kCos(\mu_k\alpha_i) + d_kSin(\mu_k\alpha_i) - \_\ F\hat{}(X_i)]^2$$

(4)

where $F\hat{}(X_i)$ are the observed output values. Minimizing Eq. (4) requires the estimation of five parameters. This formulation is linear in the parameter set A_k, where Ak = {a_k, b_k, c_k, d_k} and nonlinear in parameter μ_k. In Ghiassi and Saidane (2005), they present several nonlinear optimization strategies to estimate the nonlinear parameter μ_k. They also show that following this approach, at each layer the knowledge gained is monotonically increased, total error is reduced, and the network training improves. In Ghiassi et al. (2005), the authors compare DAN2 with traditional FFBP and recurrent neural network (RNN) models. The comparison spans both theoretical and computational perspectives using several benchmark datasets from the literature. Performance of DAN2 against these models as well as non-neural network alternatives is also presented. Their study shows that DAN2 outperforms all other alternatives and produces more accurate training and testing results in every case.

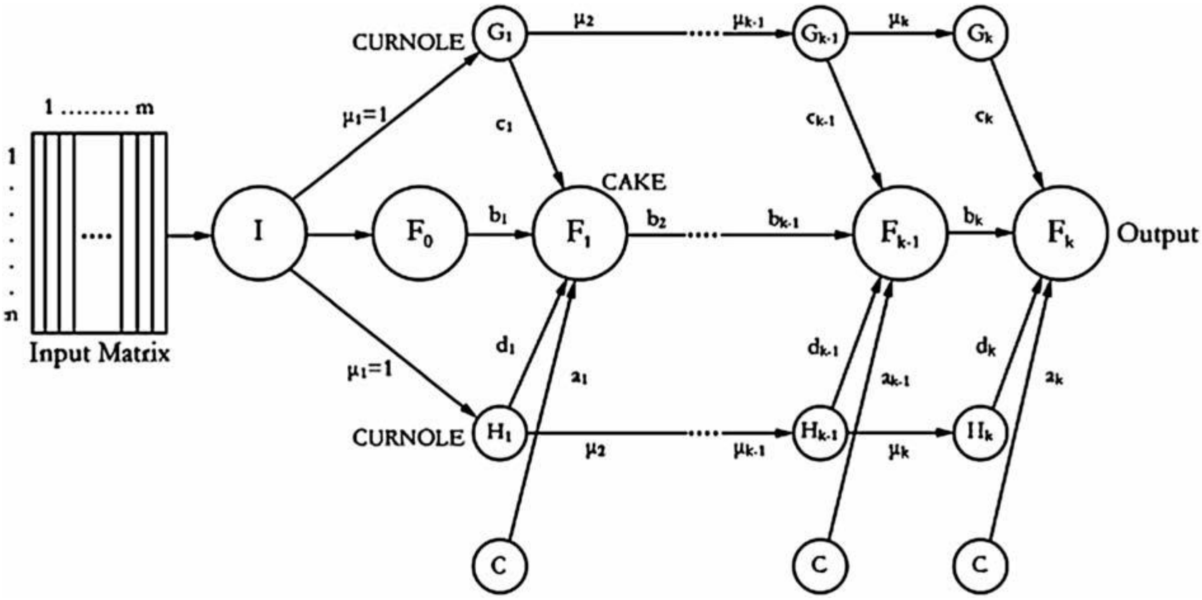

Figure 1: DAN2 Architecture (see Ghiassi and Saidane (2005))

In the forecasting literature, the Root Mean Squared Error (RMSE) has commonly been used to assess the forecast accuracy; the smaller the RMSE the better the accuracy. Accordingly, the recent forecast of the US government budget by Ghysels and Ozkan (2015) reports for the 1-step to 4-step forecast horizon, RMSEs ranging from 2.601 to 2.733 with the naïve model (AR); 2.386 to 2.640 with the Augmented distributed Lag (ADL) and 1.648 to 2.311 with the MIDAS. Some machine learning algorithm use some performance criteria (such as RMSE) as a stopping criterion. We use this guideline to assist us in reaching a low RMSE value, such as the RSME value of 1.648 obtained by Ghysels and Ozkan (2015) from the MIDAS models, in our forecasting models. As noted in Ghiassi and Saidane (2005), defining a minimum value for RMSE as a stopping rule may not always be attainable, in such cases, we report the lowest RMSE obtained for each model.

4. Empirical analysis

4.1. Data and preliminary analysis

This study uses quarterly data for South Africa from 1960 to 2016. Expenditure variable is measured as a percentage of GDP (EXP/GDP), and the demographic factor is proxied by the old age dependency ratio (ratio of old population to the working age population) (OADR). The data for these variables are obtained from the Federal Reserve of Saint Louis (FRED) Database, however, unlike the expenditure ratio of GDP, the aging variable is available at the annual frequency only. We transform the annual data into quarterly figures using quadratic polynomial interpolation. This data conversion method consists of fitting a local quadratic polynomial for each single observation of low frequency series, then use this polynomial to fill in all observations of the high frequency series associated with the period. The quadratic polynomial is formed by taking sets of three adjacent points from the source series and fitting a quadratic so that either the average or the sum of the high frequency points matches the low frequency data actually observed. For most points, one point before and one point after the period currently being interpolated are used to provide the three points. For end points, the two periods are both taken from the one side where data are available[3]. Figure 2 depicts both transformed and untransformed series which display similar trending behavior. Finally, we ensure that for each year the sum of the quarterly values match the corresponding actual yearly value.

The preliminary analysis in Table 1 (Panel A) displays a small probability of the Jarque-Bera test of normality for the demographic variable (OADR); indicating a rejection of the null hypothesis of normality. While this might be attributed to the potential nonlinearity in the data generation process, the normality assumption could not be rejected for the expenditure ratio of GDP; therefore suggesting the use of forecasting tools suitable for both linear and nonlinear linkages. Besides the linearity/nonlinearity, the choice of an appropriate forecasting model for time series is determined by the stationarity property of the variables as well as the optimum lag length. Panel B of Table 1 shows the unit root test results which indicate that both variables are non-stationary in levels, but of different order of integration. This justifies the decision to use ARIMAX as a benchmark, which implies further transformation to ensure stationarity before carrying out the forecasting analysis. The AIC information criteria points to the optimal lag length of 4; however, based on the correlation coefficient and partial correlation coefficient analysis recommended by Makridakis et al. (1998), we determined that using the last quarter and the corresponding quarter from last year, (that is, t-1 and t-4) for lags was more effective for this analysis. Additionally, these two

[3] This transformation is carried out in Eviews 10.

lags correspond with quarters of two subsequent years and the excellent model accuracy validates this choice, i.e Q1 1970, and Q1 1971, etc.

Table 1. Descriptive statistics and unit root test results

Panel A. Summary Statistics	OADR	EXP/GDP
Mean	7.75439	25.0903
Maximum	10.4750	33.2000
Minimum	6.3937	16.3000
Std. Dev.	1.1151	3.1926
Skewness	0.7660	-0.2936
Kurtosis	2.3864	3.0058
Jarque-Bera	25.8730***	3.2767
Probability	0.0000	0.1643
observation	228	228
Panel B. Unit root test results	**OADR**	**EXP/GDP**
Level	-1.8292	-3.0884
First Difference	-2.5431	-6.1966***
Conclusion	I(2)	I(1)

Note. *,** and *** indicate significance at the 10%, 5% and 1% level of significance, respectively. The unit root test results are based on trend and intercept specification using the popular augmented Dickey-Fuller (ADF) test (Dickey and Fuller, 1979 and 1981).

4.2. ARIMAX versus DAN2 results

For the sake of comparison, "identical" input values have been used for both models so as to ensure that changes, if any, can only be attributed to model specification. In addition, three main statistics are used to compare forecasting performance across models, namely the root mean squared error (RMSE), the mean absolute percentage error (MAPE) and the mean absolute error (MAE) referred to as accuracy. The first four observations are dropped on account of the lag length. The remaining 224 data points are divided into training and testing datasets. We use the traditional 80% (180 data points) for model training and the remaining 20% (44 data points) for model testing (hold out data).

Table 2. Forecasting output

Model	Training			Forecasting		
	RMSE	MAPE	MAE	RMSE	MAPE	MAE
ARIMAX	1.7337	5.85	94.15	3.1084	8.60	91.40
DAN2	0.7591	3.00	97.0	0.8483	3.14	96.9

Table 3. Forecast comparison

	Ye(DNA2)	*Ye(ARIMAX)*
Mean	25.2144	24.4315
Variance	8.39843	22.28216
Observations	224	224
Hypothesized Mean Difference	0	
df	370	
t Stat	**2.1154***	
P(T<=t) two-tail	**0.0351**	
t Critical two-tail	1.9664	

Note. * indicates significance at the 5% level of significance.

Table 2 displays the forecasting performance of the two models. DAN2's results show an accuracy (MAE) value of 97% for both the training and testing data sets. The balance between the training and testing accuracy values are an indication of excellent model fit and the absence of overfitting. DAN2 outperforms ARIMAX by 5.5% point for the hold out (testing) data set. The improvement is statistically significant at 95% as indicated by the mean comparison test run on predicted outputs from both models (Table 3). Irrespective of the performance criteria used, ANN appears to outperform the ARIMAX benchmark across both training and testing periods; thus substantiating the superiority of DAN2 over traditional time series tools.

In sum, with an accuracy of about 97%, DAN2 represents a promising ANN alternative for forecasting government expenditure. Moreover, the relatively good forecasting precision across models (ranging from 91% to 97%) confirms the casual role of population aging in predicting government expenditure in South Africa. From the policy perspective, our findings imply that demographic monitoring is indispensable for efficient fiscal planning and management in South Africa.

5. Conclusion

This study analyses the effectiveness of the machine learning algorithm DAN2 in forecasting government expenditures in South Africa. Results from quarterly data from 1960Q1 to 2016Q4 indicate that DAN2 outperform the ARIMAX benchmark, validating that DAN2 is a promising tool for government expenditure. Interestingly, all the models used in this research deliver relatively high predictive accuracy which confirms that demographic development is an important predictor of government spending in South Africa.

While these findings emphasize the importance of demographic monitoring in achieving effective fiscal projections, to what extent demographic changes drive government spending remains an open question which is beyond the scope of this study.

Acknowledgements

The authors wish to thank Dan Barkhorn for his contribution to this paper.

References

Andreou, E., Ghysels, E., & Kourtellos, A. (2011). Forecasting with Mixed-Frequency Data. In *Oxford Handbook of Economic and Forecasting* (M. Clement, & D. Hendry, Ed.). Oxford University Press.

Andreou, E., Ghysels, E., & Kourtellos, A. (2013). Should Macroeconomic Forecasters Use Daily Financial Data and How? *Journal of Business and Economic statistics* 31: 240-251.

Akanbi, O.A. (2013). Macroeconomic Effects of Fiscal Policy Changes: A Case of South Africa. *Economic Modelling*35: 771-785.

Balkin, S.D., & Ord, J.K. (2000). Automatic Neural Network Modeling for Univariate Time Series. *International Journal of Forecasting* 16:509–515.

Basaran, A.A., Aladag, C.H., Bagdadioglu, N., & Gunay, S. (2010). Public Expenditure Forecast with Artificial Neural Networks. 5th International Conference on Applied Statistics, Bucharest.

Berardi, V.L., & Zhang, G.P. (2003). An Empirical Investigation of Bias and Variance in Time Series Forecasting: Modeling Considerations and Error Evaluation. *IEEE Transactions on Neural Networks* 14 (3):668–679.

Berkman, M.,& Plutzer, E. (2004). Gray Peril or Loyal support? The Effects of the Elderly on Educational Expenditures. *Social Science Quaterly* 85: 1178–1192.

Borcherding, T.E. (1985). The Causes of Government Expenditure Growth: A Survey of the US Evidence. *Journal of Public Economics* 28(3): 359-382.

Borge, L.-E., & Rattso, J., (2008). Young and Old Competing for Public Welfare Services. Working paper no. 2223, CESifo, Munich.

Box, P., & Jenkins, G.M. (1976). Time Series Analysis: Forecasting and Control, Holden-day Inc., San Francisco, CA.

Braendle, T., & Colombier, C. (2016). What Drives Public Health Care Expenditure Growth? Evidence from Swiss Cantons, 1970–2012. *Health policy* 120: 1051-1060.

Braun, A.R., & Joines, D.H. (2015). The implications of a graying Japan for government policy. *Journal of Economic Dynamics and Control* 57: 1-23.

Brunner, E., & Balsdon, E. (2004). Intergeneraltional Conflict and the Political Economy of School Spending. *Journal of Urban Economics*56: 369–388.

Busemeyer, M.R. (2007). Determinants of Public Education Spending in 21 OECD Democracies, 1980–2001. *Journal of European Public Policy* 14: 582–610.

Cattaneo, M.A., & Wolter, S.C. (2009). Are the Elderly a Threat to Educational Expenditures? *European Journal of Political Economy* 25: 225-236.

Chakraborty, K., Mehrotra, K., Mohan, C.K., & Ranka, S. (1992). Forecasting the Behavior of Multivariate Time Series Using Neural networks. *Neural Networks* 5:961–970.

Chen, Y., Yang, B., Dong, J., & Abraham, A. (2005). Time-Series Forecasting Using Flexible Neural Tree Model. *Information Sciences* 174 (3–4):219–235.

Cottrell, M., Girard, B., Girard, Y., Mangeas, M., & Muller, C. (1995). Neural Modeling for Time Series: A Statistical Stepwise Method for Weight Elimination. *IEEE Transactions on Neural Networks* 6 (6):1355–1364.

Denton, J.W. (1995). How Good are Neural Networks for Causal Forecasting? *The Journalof Business Forecasting* 14 (2):17–20.

Dickey, D., & Fuller, W. (1979). Distribution of the Estimators for Autoregressive Time Series with a Unit Root. *Journal of American Statistical Association* 74(366a):427-431.

Dickey, D., &Fuller, W. (1981). Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root. *Econometrica* 49:1057-1072.

Dizaji, S.F. (2014). The Effects of Oil Shocks on Government Expenditures and Government Revenues Nexus (with an Application to Iran's Sanctions). *Economic Modelling* 40: 299-313.

Ericsson, N.R. (2017). How Biased are US Government Forecast of the Federal Debt? *International Journal of Forecasting*33: 543-559.

Fan, S., Yu, B., & Saurkar, A. (2008). Public Expenditure in Developing Countries: Trends, Determination and Impact. In *Public Expenditures, Growth and Poverty: Lesson for Developing Countries* (S. Fan, Ed.). Washington DC.

Fernandez, R., & Rogerson, R. (2001). The Determinants of Public Education Expenditures: Evidence from the States, 1950–1990. *Journal of Education Finance* 27: 567–584.

Foster, W.R., Collopy, F.,& Ungar, L.H. (1992). Neural Network Forecasting of Short, Noisy Time Series. *Computers and Chemical Engineering* 16 (4):293–297.

Ghiassi, M., & Saidane, H. (2005). A Dynamic Architecture for Artificial Neural Network. *Neurocomputing* 63: 397-413.

Ghiassi, M., Saidane, H., & Zimbra, D.K. (2005). A Dynamic Artificial Neural Network Model for Forecasting Time Series Events. *International Journal of Forecasting* 21: 341-362.

Ghiassi, M., Zimbra, D. K., & Saidane, H. (2006). Medium-Term System Load Forecasting with a Dynamic Artificial Neural Network Model. *Electric. Power Systems. Research* 76(5): 302–316.

Ghiassi, M., Zimbra, D.K., & Saidane, H. (2008). Urban Water Demand Forecasting with a Dynamic Artificial Neural Network Model. *Journal of Water resources Planning and Management* 134 (2): 138-146.

Ghysels, E., & Ozkan, N. (2015). Real-Time Forecasting of the US Federal Government Budget: A Simple Mixed Frequency Data Regression Approach. *International Journal of Forecasting* 31: 1009-1020.

Giordano, F., La Rocca, M., & Perna, C. (2007). Forecasting Nonlinear Time Series with Neural Network Sieve Bootstrap. *Computational Statistics and Data Analysis* 51:3871–3884.

Grob, U., & Wolter, S.C. (2007). Demographic Change and Public Education Spending a Conflict between Young and Old? *Education Economics* 15: 277–292.

Guresen, E., Kayakutlu, G., & Daim, T.U. (2011). Using Artificial Neural Network Modelsin Stock Market Index Prediction. *Expert Systems with Applications* 38: 10389-10397.

Harris, A.R., Evans, W.N., & Schwab, R. (2001). Education Spending in an Aging America. *Journal of Public Economics* 81: 449–472.

Isenah, M. G., & Olubusoye, O.E. (2014). Forecasting Nigeria Stock Market Returns using ARIMA and Artificial Neural Network Models. *CBN Journal of applied Statistics* 5(2): 25-48.

Jain, A., & Kumar, A.M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Soft Computing* 7:585–592.

Jimero, J.F, Rojas, J.A., & Puente, S. (2006). Modelling the Impact of Aging on Social Security Expenditures. Documentos Ocasionales. N.º 0601.

Khashei, M., & Bijari, M. (2011). A Novel Hybridization of Artificial Neural Networks and ARIMA Models for Time Series Forecasting. *Applied Soft Computing* 11:2664-2675.

Kudrna, G., Tran, C., & Woodland, A. (2015). The Dynamic Fiscal Effects of Demographic Shift: The Case of Australia. *Economic Modelling* 50: 106-122.

Kuzin, V., Marcellino, M., & Schumacher, C. (2013). Pooling versus Model Selection for Nowcasting GDP with Many Predictors: Empirical Evidence for Six Industrialised Countries. *Journal of Applied Econometrics* 28: 392-411.

Lassila. J., Valkonen, T., & Alho, J.M. (201). Demographic Forecast and Fiscal Policy Rules. *International Journal of Forecasting* 30: 1098-1109.

Makridakis, S. G., Wheelwright, S. C., & Hyndman, R. J. (1998). *Forecasting: Methods and Applications*. New York:John Wiley & Sons.

Marcellino, M., & Schumacher, C. (2010). Factor MISDAS for Nowcasting and Forecasting with Ragged-Edge Data: A Model Comparison for German GDP. *Oxford Bulletin of Economics and Statistics* 72: 518-550.

Murthy, V.N.R, & Okunade, A.A. (2016). Determinants of U.S. Health Expenditure: Evidence from Autoregressive Distributed Lag (ARDL) Approach to Cointegration. *Economic Modelling* 59: 67-73.
Nishiyama, S. (2013). Fiscal Policy Effects in a Heterogeneous-Agent Overlapping-Generations Economy with an Aging Population." Working Paper 2013-07, Working Paper Series Congressional Budget Office Washington, DC.

Peacock, A.T. (2004). The Growth of Public Expenditure" Schneider, C., (Eds.) the Encyclopaedia of Public Choice (pp. 594-597).

Plutzer, E., & Berkman, M. (2005). The Graying of America and Support for Funding the Nation's Schools." *Public Opinion Quaterly* 69: 66–86.

Poli, I., & Jones, R.D. (1994). A Neural Net Model for Prediction. *Journal of American Statistical Association* 89:117–121.

Taskaya, T., & M.C. Casey, M.C. (2005). A Comparative Study of Autoregressive Neural Network Hybrids. *Neural Networks* 18:781–789.

Tang, Z., Almeida, C., & Fishwick, P.A. (1991). Time Series Forecasting Using Neural Networks vs. Box-Jenkins Methodology. *Simulation* 57 (5):303–310.

Tang, Z., & Fishwick, P.A. (1993). Feedforward neural nets as models for time series forecasting. *ORSA Journal on Computing* 5 (4):374–385.

Timmermann, A. (2006). Forecast combinations." *Handbook of Economic Forecasting* 1: 135-196.

Vasconcelos de Deus, J.D.B., & de Mendonca, H.F. (forthcoming). Fiscal Forecasting Performance in an Emerging Economy: An Empirical Assessment of Brazil. *Economic Sytem*.

Velásquez, J. D., & Franco, C .J. (2010). Prediction of the Prices of Electricity Contracts Using a Neuronal Network with Dynamic Architecture. *Innovar* 20(36): 7-14.

Velásquez -Henao, J.D., Franco-Cardona, C.J., & Olaya-Morales, Y. (2012). A Review of DAN2 (Danamic Architecture for Artificial Neural Networks) Model in Time Series Forecasting. *Ingeniería y Universidad* 16(1): 135-146.

WANG, J., NIU, D., & LI, L. (2010). Middle-Long Term Load Forecasting Based on Dynamic Architecture for Artificial Neural Network. *Journal of Information and Computational Science* 7(8): 1711-1717.

Younger, D.S. (2016). Health Care in South Africa." *Neurologic Clinics* 34: 1127-1136.

Zhang, G. (2003). Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model." *Neurocomputing* 50: 159–175.

Zhang, G., Patuwo, E. B., & Hu, M. Y. (1998). Forecasting with Artificial Neural Network: The State of the Art. *International Journal of Forecasting* 14: 35– 62.

Figure 2. Aging trends