Bureau d'économie théorique et appliquée (BETA)
UMR 7522

Documents de travail

« Forecasting with Neural Networks Models »

Auteurs

Francis Bismans, Igor N. Litvine

Document de Travail n° 2016 – 28

Mai 2016

Faculté des sciences économiques et de gestion
Pôle européen de gestion et d'économie (PEGE)
61 avenue de la Forêt Noire
F-67085 Strasbourg Cedex

Secrétariat du BETA
Géraldine Del Fabbro
Tél. : (33) 03 68 85 20 69
Fax : (33) 03 68 85 20 70
g.delfabbro @unistra.fr
www.beta-umr7522.fr

cnrs
UNIVERSITÉ DE LORRAINE
UNIVERSITÉ DE STRASBOURG

Forecasting with Neural Networks Models

Francis Bismans

BETA, University of Lorraine

COEF, Nelson Mandela Metropolitan University, Port Elizabeth

and

Igor N. Litvine,

COEF, Nelson Mandela Metropolitan University, Port Elizabeth

Summary. This paper deals with so-called feedforward neural network model which we consider from a statistical and econometric viewpoint. It was shown how this model can be estimated by maximum likelihood. Finally, we apply the ANN methodology to model demand for electricity in South Africa. The comparison of forecasts based on a linear and ANN model respectively shows the usefulness of the latter.

Keywords. Artificial neural networks (ANN), electricity consumption, forecasting, linear and non-linear models, recessions

JEL classification. C45, C53, E17, E27, Q43, Q47

1. Introduction

Artificial neural networks (ANN) subsume a set of models which have been developed in the cognitive sciences to understand functioning of human brain. These models were originated in the publications by McCulloch and Pitts (1943) and in the study of perceptron by Rosenblatt (1958). However, at that time the capabilities of computing were very limited, so these early models were too simple to explain the complexities of the actual operation of the brain.

Consequently, with growth of computing powers, more complex ANN structures and network learning methods were designed, peculiarly in the investigations of Rumelhart et al. (1986) and McClelland et al. (1986). Since then, the research in neural networks is growing progressively.

In this study we are interested to apply econometric approach to ANN models. From this viewpoint, the inspiring and path breaking contribution is that of Kuan and White (1994). More recently, Kuan (2008) gave a review of the matter from an econometric perspective. In brief, ANN models for an econometrician constitute a specific set of non-linear models and “learning” is understood as an estimation of model parameters.

Applications of econometric ANN models are been numerous in the field of market finance. For a presentation of main results, see e.g. Franses and van Dijk (2000, chapter 5). Forecasting, especially macroeconomic, was also an area to prospect well, as evidenced by the studies of Swanson and White (1997), McMenamin (1997), Zhang, G. et al. (1998), Rech (2002), White (2006), Medeiros et al. (2006), and Ozdemir et al. (2010), a short list in the vast literature on the subject.

The structure of this paper is the following. Section 2 presents the standard ANN model from a twofold viewpoint – heuristic and econometric-statistical. Section 3 deals with the estimation and forecasting method of the considered model. As far as the fourth section is concerned, it develops an application focused on the electricity production in South Africa. Finally, conclusions and research prospects are discussed in the section 5.

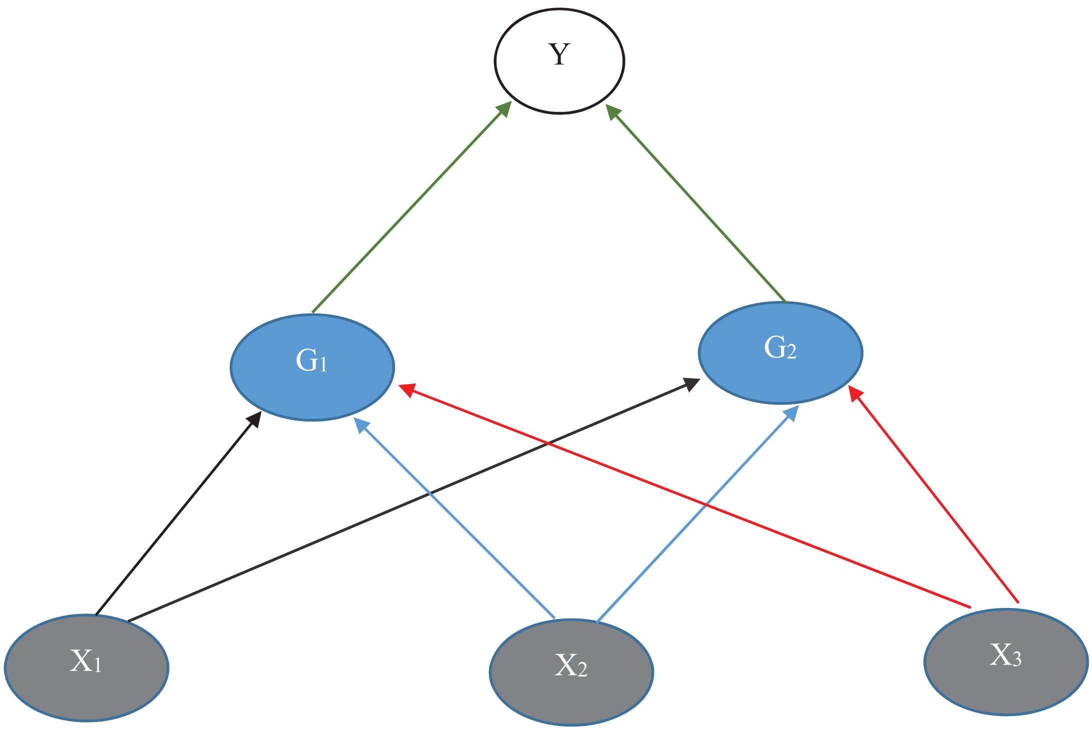
2 The model

1.1 A first heuristic approach

Neural networks models mimics the brain operation. They use therefore a specific terminology such as layer, nodes, output, input and so one. For example consider the simple “feedforward neural network”. The output is denoted by y and the three inputs are x_1, x_2 et x_3. The basic idea is that the inputs feed into the nodes and in a second step they feed onward to the output layer. Thus there is no feedback.

The following diagram furnishes a representation of the links between output and inputs structured in three layers.

Figure 1. A simple feedforward network



The Xs represent the **inputs layer**, and Y is the **output layer**. The nodes G_1 and G_2 constitute the **hidden layer**. These Gs are called **activation functions**. The analogy with biological neural systems is traightforward: the input units are neurons which convey outside information to the neurons in the intermediate hidden layer; these latter generate nervous signals, forwarded to the neurons in the output layer. One last point is timely: the activations functions are possibly nonlinear, such as e.g. a logistic mapping; the parameters of these functions are called the **connection strengths** or **connection weights**.

Evidently, in the reality, i.e. in biological neural systems, the number of processing units is enormous: they amount to billions. Such numbers are usually not considered in the economic models.

1.2 The **ANN(*k,q*)** *model*

Translating the previous diagram in a well-defined statistical model yields the following nonlinear equation:

$$Y = \sum_{i=1}^{2} G_i\left(\boldsymbol{\gamma}'\mathbf{x}\right), \tag{1}$$

where G is an activation function, $\mathbf{x}$ a column vector of inputs and the γs are the parameters of the model.

Naturally, for practical purposes, the structure of eq. (1) must be more complex. Therefore, from an econometric perspective, the "single hidden-layer feedforward" model will be written as follows:

$$y_t = \boldsymbol{\alpha}'\mathbf{z}_t + \sum_{j=1}^{q} \beta_j G(\boldsymbol{\gamma}'_j \mathbf{z}_t) + \varepsilon_t, \quad t = 1, \cdots, T, \tag{2}$$

where y_t is the dependent variable (= output), $\mathbf{z}_t = (1, y_{t-1}, \cdots, y_{t-p}, x_{1t}, \cdots, x_{kt})'$ is the vector of the explaining variables, including the constant and the delayed values of y_t, $\boldsymbol{\gamma}$ and $\boldsymbol{\alpha}$ are $(p+k+1)$ vectors of parameters and the Gs are the activation functions. Consequently, the relationships between y_t and $\mathbf{x}_t$ are possibly nonlinear. Moreover, as usual, ε_t is a Gaussian white noise with null mathematical expectation and constant variance.

Some discussion on activation functions is now required. These functions are usually restricted to those which have values between zero and one. From this viewpoint, the logistic function

$$G(x) = \frac{1}{1+e^{-x}}, \quad x \in \mathbb{R}, \tag{3}$$

is the leading choice. Mathematically, this kind of function is differentiable everywhere and its derivative is easily computed:

$$\frac{dG(x_0)}{dx} = G(x_0)\left[1 - G(x_0)\right].$$

However, many other choices are possible such as smooth cumulative distribution functions, sine and cosine functions, hyperbolic tangent functions, etc. – see Kuan (2008). Furthermore, McMenamin (1997) has even proposed to use the π-based activation function, say π^x, which the derivative is given by

$$\frac{d\pi^x}{dx} = \pi^x \cdot \ln(\pi) = 1.145 \ln(\pi).$$

Nevertheless, all these mappings must be also bounded and should be asymptotically constant.

In total, eq. (2) and the relevant specific activation functions constitute the ANN (k, q) model. Fundamentally, this model belongs to the class of nonlinear models and it subsumes many other models, well known in the econometric literature – see e.g. Franses and van Dijk (2000) – such as the switching-regression model or the Smooth Transition Autoregressive (STAR) model. Better, as noted by Kock and Teräsvirta (2011), ANN-model is also a so-called "universal approximator", just like the Kolmogorov-Gabor polynomial, all used to determine unknown nonlinear functional forms.

2. Estimation and forecasting method

The problem will be drastically simplified if ANN (k, 1) model is considered:

$$y_t = \boldsymbol{\alpha}'\mathbf{z}_t + \beta G(\boldsymbol{\gamma}'\mathbf{z}_t) + \varepsilon_t, \quad t = 1, \cdots, T, \tag{4}$$

where $\varepsilon_t \sim N(0, \sigma^2)$ and G is the logistic function (3).

3.1 Estimation

The estimation strategy consists of two steps. In the first stage we shall estimate the linear part of the model, obtaining the vector of parameters $\hat{\boldsymbol{\alpha}}$ by OLS. The second step of estimation yields the estimates of the vector of parameters $\hat{\boldsymbol{\gamma}}$. Of course, in this latter case, the estimation procedure is more intricate due to the non-linearity of the model.

Practically, by using estimated parameters in the first step, Eq. (4) can be written:

$$w_t = y_t - \hat{\boldsymbol{\alpha}}'\mathbf{z}_t = \beta G(\boldsymbol{\gamma}'\mathbf{z}_t) + \varepsilon_t, \qquad \varepsilon_t \sim N(0, \sigma^2).$$

Consequently, the likelihood is given by

$$L_c(\beta, \boldsymbol{\gamma}) = \prod_{t=1}^{T} \frac{2}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(w_t - \beta G(\boldsymbol{\gamma}'\mathbf{z}_t))^2}.$$

Finally, the log-likelihood function is

$$\ln L(\beta, \boldsymbol{\gamma}) = -\frac{T}{2}\ln(2\pi) - \frac{T}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{t=1}^{T}\left(w_t - \beta G(\boldsymbol{\gamma}'\mathbf{z}_t)\right)^2, \tag{5}$$

with $G(\boldsymbol{\gamma}'\mathbf{z}_t) = \left(1 + e^{-\gamma'\mathbf{z}_t}\right)^{-1}$.

So it remains to maximize this log-likelihood (5) using a usual algorithm such as that of Newton-Raphson or better, of Broyden-Fletcher-Goldfarb-Shannon (BFGS). For a discussion of the more specific algorithms relevant to neural feedforward networks, see Fine (1999, chapter 5). In our study, we used built-in optimization algorithm of Wolfram Mathematica®.

3.2 Forecasting methodology

Here the procedure involves two stages as well:

1) In a first step, the linear dynamic explicative model

$$y_t = \sum_{r=1}^{R} \delta_r y_{t-r} + \sum_{k=0}^{K} \sum_{p=0}^{p_K} \boldsymbol{\alpha}'_k \mathbf{x}_{k,t-p} + u_t$$

is estimated using (T-h) observations and we retain h data points for prediction purposes. Forecasting can be then realized for these h future periods.

2) The second step, based on the estimation of the complete ANN model (4), yields the forecasts for the same forward periods and allows a direct comparison with the set of predicted values in the precedent stage.

Recall, this prediction trial is a first attempt and our main objective is merely to show the utility of the ANN models.

4 An exercise in prediction

4.1 The data

The series to be "explained" and to be used for prediction, denoted Elec, is that of the monthly electricity production in South Africa for the period 2002-M1 to 2010-M6. It furnishes the production in thousands of Mwh.

The explicative variables are three in number:

- The consumer prices of services index, denoted CPI, which gives a picture of the prices variation for all urban areas (2012-M12 = 100);
- The total volume of manufactured production, denoted Prod, which constitutes an index with 2010 = 100;
- A binary variable, Rec, which takes the value one during the recessions and the value zero during the expansion phases of the economy. This dummy series was constructed based on the dating of the South African business cycle by Bismans and Majetti (2012) and Bismans and Le Roux (2013).

All the series are monthly and have been downloaded from the data base of the South African Reserve Bank.

4.2 Estimation of the linear model

The search for an adequate model applies the general-to-specific (Gets) methodology vindicated by David Hendry (see for a recent and path-breaking reference, Hendry and Doornik (2014), especially chapter 1).

Following this approach, we start the process of discovery with a general unconstrained model and after some reduction operations, we get a dynamic, parsimonious and well-specified model. Table 1 presents the final results of the process.

Table 1. The final linear model

	Coefficient	Std. Error	p-value
Elec (-2)	0.754	0.0875	0.000
Elec (-4)	-0.453	0.0834	0.000
Rec	-1608,9	494.1	0.002
Rec (-1)	1654	524.8	0.002
Rec(-6)	1568	297.9	0.000
Prod (-1)	128,9	14.38	0.000

Some comments are due at this point. Firstly, the variable CPI is not included among the regressors and thus has no statistical effect on the electricity production. Secondly, the variable "Rec" is peculiarly significant. Logically, the entry in recession lowers immediately the electricity output, but after one month and especially six months, the relationship becomes positive. Thirdly, the manufactured production contributes positively, but with one lag, to the electricity production.

4.3 The enlarged ANN-model

Given Eq. (4), it remains to estimate the part

$$\beta G(\boldsymbol{\gamma}'\mathbf{z}_t) + \varepsilon_t,$$

where $\beta = 1$, G is the logistic CDF and $\mathbf{z}_t$ is the vector compound by Elec(-2), Elec(-4), Rec, Rec(-1), Rec(-6) and Prod(-1).

The estimation process is conducted along the lines defined in subsection 3.2. However, the likelihood function is peculiarly difficult to maximize. The following graph gives a picture of the likelihood's profile in the last step of optimisation algorithm.

Graph. Profile of the likelihood function

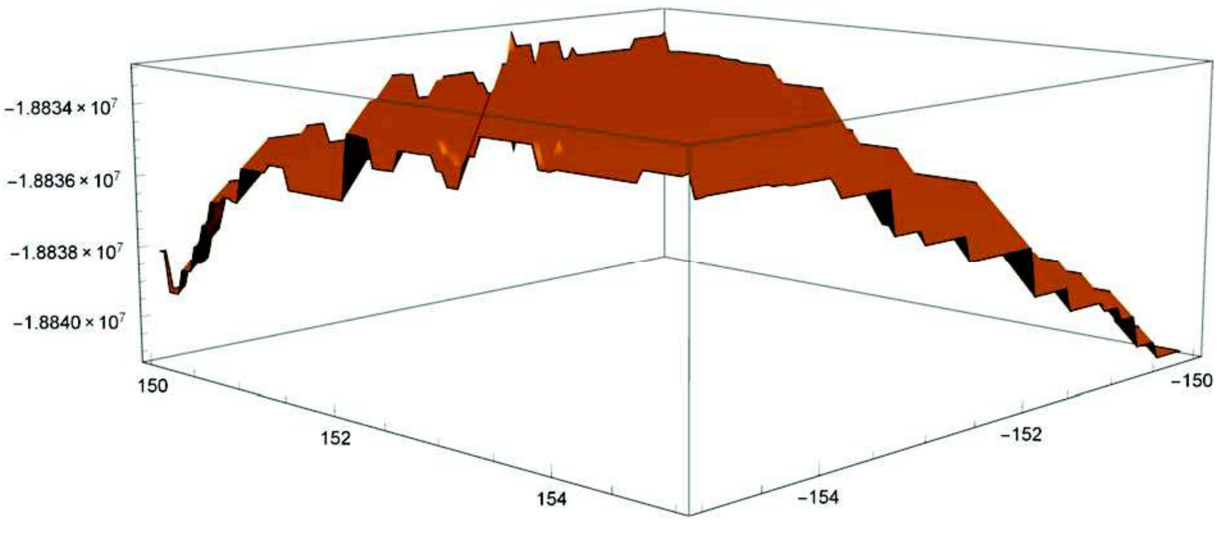


Table 2 shows the ML estimators of the gamma parameters

Table 2. The ANN-model (non-linear part)

	Elec (-2)	Elec (-4)	Rec	Rec (-1)	Rec (-6)	Prod (-1)
Coefficients	150	-151.7	-0.0022	-0.0002	0.0008	0.0399

As a general rule the estimated coefficients must not be interpreted in the same way as we do for a linear model. They are only used to get a better prediction of the electricity consumption.

4.4 Forecasts

The out-of-sample forecasts are presented in the following table for a horizon of six months.

Table 3. Dynamic forecasts

Horizon	Actual values	Linear model	ANN-model
2010-01	20124.4	18900.0	18748.4
2010-02	18861.7	18124.1	17972.6
2010-03	20914.6	17818.5	18590.3
2010-04	19844.7	17782.6	18187.1
2010-05	21149.2	17741.2	19675.9
2010-06	21352.6	18284.8	19660.4

The comparison of point predictions depicts that globally, the ANN-model is a better tool for forecasting in comparison to the linear model. To refine the analysis, two additional evaluation indicators are computed: the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE). Finally, the Theil's coefficient – a statistic which compares the forecasts of the proposed model with a naïve prediction given by a random walk process – is equally implemented.

Table 5. Some evaluation statistics

	Linear model	ANN-model
RMSE	2480.2	1605.9
MAPE	10.945	7.526
Theil's U	1.97	1.24

All the statistics deliver the same information: the predictive performance of the neural network model is superior to that of its linear counterparty. However, it is possible to push the interpretation further by considering the coefficient of Theil.

Theil's U takes a value of one with the naïve model. Values lesser than 1 indicate an improvement comparatively to this naïve benchmark and values higher than the unity translate a deterioration of the forecasts, always with respect to this benchmark.

From this viewpoint, both used models - linear and feedforward neural network - do less well than a simple random walk, for which the best prediction of a variable in *t* is the observed value of this variable during the immediately preceding period.

Explaining this apparent paradox is not difficult: it follows from this that the strong seasonality in the series was not taken into consideration. Nevertheless, one must recall that the only objective of this contribution was to compare linear and ANN models from a forecasting perspective.

5 Conclusion

Undoubtedly, the study has proved empirically that the ANN-model demonstrated superior predictive properties than the linear one. However, two limitations of the canonical ANN benchmark structure will be exceeded in the future: on one side, considering one node in the hidden layer is a simplification, at best temporary, to be abandoned. On the other side, the seasonality should be modeled explicitly. It's our prospective way!

References

Bismans, F., and R. Majetti (2012), Dating the South African Business Cycle, *Journal for Development and Leadership*, 1, pp. 1-12.

Bismans, F., and P. Le Roux (2013), Dating the Business Cycle in South Africa by Using a Markov-switching Model, *Studies in Economics and Econometrics*, 37, pp. 25-39.

Fine, T.L. (1999), *Feedforward Neural Network Methodology*, New York: Springer-Verlag.

Franses, P.H, and D. van Dijk (2000), *Non-linear Time Series Models in Empirical Finance*, Cambridge: Cambridge University Press.

Hendry, D.F, and J.A. Doornik (2014), *Empirical Model Discovery and Theory Evaluation. Automatic Selection Methods in Econometrics*, Cambridge (MA)-London : The MIT Press.

Kock, A.B, and T.Teräsvirta (2011), Forecasting with Nonlinear Time Series Models, *The Oxford Handbook of Economic Forecasting*, Eds M.P. Clements, and D. Hendry, Oxford-New York: Oxford University Press, pp. 61-87.

Kuan, C-M (2008), Artificial Neural Networks, *The New Palgrave Dictionary of Economics*, 2nd edition, Eds Durlauf, S.N., and L.E. Blume, London-New York, Palgrave Macmillan.

Kuan, C-M, and H. White (1994), Artificial Neuronal Networks: An Econometric Perspective, *Econometric Reviews*, 13, pp. 1-91.

McClelland, J.L., D.E. Rummelhart, and the PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 2, Cambridge, MA: MIT Press.

McCulloch, W.S., and W. Pitts (1943) A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bulletin of Mathematical Biophysics*, 5, pp. 115-133.

McMenamin, J.S. (1997), What not Pi? A Primer on Neural Networks for Forecasting, *The Journal of Business Forecasting Methods & Systems*, 16 (3), pp. 1-20.

Medeiros, M.C., T. Teräsvirta, and G. Rech (2006), Building Neural Networks Models for Time Series: A Statistical Approach, *Journal of Forecasting*, 25, pp. 49-75.

Ozdemir, O., A. Aslanargun, and S. Asma (2010), ANN Forecasting Models for ISE National-100 Index, *Journal of Modern Applied Statistical Methods*, 9, pp. 579-583.

Rech, G. (2002), Forecasting with Artificial Neural Network Models, *Working Paper* Series in Economics and Finance, n° 491, Stockholm School of Economics.

Rosenblatt, F. (1958), The Perceptron: A Probabilistic Model for Information Storage and Organisation in the Brain, *Psychological Reviews*, 62, pp. 386-408.

Rummelhart, D.E., G.E. Hinton, and the PDP Research Group (1986), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, Cambridge, MA: MIT Press.

Swanson, N.R. and H. White (1997), A Model Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks, *Review of Economics and Statistics*, 79, pp. 540-550.

White, H. (2006), Approximate Nonlinear Forecasting Methods, in *Handbook of Economic Forecasting*, vol.1, Eds Elliot, G., C.W.J. Granger and A. Timmerman, Amsterdam: Elsevier, pp. 459-512.

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