Connectionism ARCH modelprocessing Area of research:Engineering and Information Science and Technology

RAKOTONIAINA RABENORO Barry[1], ANDRIAMANOHISOA Hery Zo[2], ROBINSON Matio[3]

[1]RAKOTONIAINA RABENORO Barry, Ph D student, University of Antananarivo in the STII, Madagasikara
[2]ANDRIAMANOHISOA Hery Zo, Professor, University of Antananarivo in ESPA, Madagasikara
[3]ROBINSON Matio, Ph D, University of Antananarivo in ESPA, Madagasikara

ABSTRACT

With the existence of volatility, several experts try to explain the reason for volatility. The reason for volatility is not only just existing information, but also several other factors. The ARCH and GARCH models have the advantage of allowing complex time series to be modeled with sufficient parameters, and are for this reason in particular to predict volatility. It is increasingly recognized, however, that the inclusion of non-stationarities in the series is inevitable in order to apply such models to long-term real data.
The purpose of this paper was to highlight the utility of non-linear models and conditional heteroscedasticity, which now has powerful analytical and modeling tools based on sound theoretical bases to model stationary time series with nonlinear dynamics. The concept of conditional variance characterizes the models that come to broaden the class of classical models based essentially on a linear dependence structure between a variable at moment t and its past values and those of white noise and its past values. However, ARCH models are problematic when the number of historical data becomes very large in which case conditional variances tend to become negative. Indeed, the problem with ARCH models is that volatility is predicted. But variables tend to be negatively correlated, a phenomenon that ARCH models cannot incorporate because they restrict volatility to be affected only.
Connectionism emerged as the beginning of a new paradigm in cognitive science that could bridge the gap between the study of behaviors and the study of underlying neurophysiological processes. A connected network consists of units, also called formal neurons, connected to each other by connections. Thus, mathematically speaking, a network isused to transform a set of values, which form a vector in the input space, into another set of values, which form a vector in the output space. One of the main interests of the connected networks is their learning capabilities that the values of the output units of a network depend not only on the values of the input units, but also on the values of the connection weights and unit thresholds. This means that you can change the correspondence between the inputs and outputs of network by changing these weights and thresholds.

Keyword*: volatility, ARCH, non-linear model, connectivity, formal networks*

Introduction

Statistics are among the modern techniques being sought to implement in developing countries. Statistical information, indeed, comes from any development technique. It is rarely possible to predict the future when past situations and trends are unknown.
It should be noted, however, that thestatistical data are both inaccurate,uncertain and difficult to compare over time. In these circumstances, there are sufficient reasons for theim perfection of these data. First, they address the inherent barriers to the environment, such as the vast size of the territories, the difficulty of entry due to lack of channels of communication, wide spread illiteracy, etc.
Yet a strong, transparent and efficient statistical system must be in place to provide reliable, consistent,timely and up to date data in as diverse arange of areas as possible for decision makers.
First, in developing countries such as Madagascar, the data obtained and presented are insufficient. They did not fail to draw attention to the seriousness of the lack of information. Second, a system has always a state and

dynamics. As a matter of fact, a system is a set forming a coherent and autonomous unit of real or conceptual objects, organized according to a goal by means of a set of relationships (mutual interactions, dynamic interactions...), immersed in an environment. And finally, despite a perfect knowledge of the basic components of a system, it's impossible to predictits behavior

Is the network connection an effective modeling tool for the representation of complex reality?

On the one hand, the objective is to develop a strong model that allows predicting the evolution in the course of a system such as economics indicators ; on the other hand, to find acompromise between two opposing objectives of the system by :

- The relationship must be complex enough to represent the system as accurately as possible;
- The relationship must be sufficiently simple so that the parametric estimate is not very costly in time of calculation and the variance of the estimated parameters

First, a three layer network of formal neurons, including a hidden layer, ensures good as an approximation as possible of any function of several variables. Second, neural networks are improved by Bayesian learning methods.

1 Critical summary of statistical templates

Researchers can base their ideas that many relationships of elements in the model in complex reality are nonlinear. A common aspect of non-linearity is to abandon the the linear model principle. Certainly, for structural reasons, the values of certain parameters change c ontinuously in the course of time discreetly.

The model comes first from ideas about the relationship between y and x....These ideas may have a very close connection with any theory. And a linear model is called a statistical model that can be written as:

$$Y = \sum_{j=1}^{k} \theta_j X^j + \varepsilon \qquad (1.01)$$

With Y is a random variable in witch we observe and want to explain and/or predict ;

k variables $X^1, X^2, \dots, X^k$ are real variables ;

θ_j model settings ;

ε is the term of error in the model.

The non-linearity of the model is due to the fact that the rupture dates are assumed to be unknown and endogenously estimated. As a result, the standard linear model becomes inadequate. The correct model can then contain exogenous factors whose influence changes from one period to another.

A model with regime change therefore describes the set states formally distinct from one another. Moreover, regime models change are particularly suited to studying the asymmetric dynamics presented by multiple variables.

In particular, a time series is characterized by the concept of asymmetry. The two types of asymmetrie presented by[5]:

- Deepness that is defined by asymmetry in the distribution of the affected series ;
- Steepness, which is defined by asymmetry in the distribution of the first difference of the chosen series.

Once asymmetry is detected, there is a significant difference between the two phases, which must be taken into account. The non-linear effect of series must be addressed.

Remember that a time series $\{y_t\}$ is strictly static if its joint distribution of $\left(y_{t_1}, \dots, y_{t_k}\right)$ is the same as $\left(y_{t_{1+t}}, \dots, y_{t_{k+t}}\right)$, regardless of the number of momentsc onsidered $(t_1, \dots, t_k)$[1]. This condition is difficult to verify and low or second order stationarity is used. Thus, a time series $\{y_t\}$ is low stationary if its mean is not dependent on t and if the covariance between $\{y_t\}$ and $\{y_{t-l}\}$ depends only on *l* and not ont.

There are three commonly used specifications for linear modeling of stochastic mean processes:

- The Mobile Average specification (MA) :

$$Y_i = c + \sum_{i=0}^{p} \theta_i u_{t-i} \qquad (1.02)$$

With $\theta_t = \Psi_i, \theta_q \neq 0$

- Self-Regressive specification (AR)

$$Y_i = \mu + \sum_{i=0}^{p} \phi_i Y_{t-i} + u_t \qquad (1.03)$$

- Moving Average Self-Regressive specification (ARMA)

$$Y_i = \mu + \phi Y_{t-1} + \phi^2 Y_{t-2} + \dots + \phi^p Y_{t-p} + u_t + \theta_1 u_{t-1} + \dots + \theta_q u_{t-q} \qquad (1.04)$$

ARMA models have the following disadvantages:

- They do not allow for the consideration of asymmetry phenomena or large scale breaks;
- Incomplete exploitation of the information contained in the series.

But other series are not suitable for such modeling; a stationary series is obtained only after differentiation: they are called "difference stationary". ARIMA and SARIMA models are stationary series in difference that have stochastic trend (not deterministic) or stochastic (not deterministic) seasonality.

A non-stationary process is called the ARIMA (p,d,q) process X_t for which the differentiated process of order d, $Y_t = (1-B)^d X_t, t \in \mathbb{Z}$ is stationary, and checks an ARMA recurrence relationship (p,q):

$$Y_t = \sum_{i=1}^{p} \varphi_i Y_{t-i} + \sum_{i=0}^{q} \theta_i \varepsilon_{t-i}, \forall\, t \in \mathbb{Z} \tag{1.05}$$

In which p is the number of self-limiting terms, d is the number of differences, and q is the number of moving averages.

The ARIMA model estimate assumes that a stationary series is being developed. This means that the average of the series is constant overtime, as well as variance. The best way to eliminate any tendency is to differentiate, i.e. to replace the original series with the adjacent series of differences.

An order 1 differentiation assumes that the difference between two successive y values is constant:

$$Y_t - Y_{t-1} = \mu + \varepsilon_t \tag{1.06}$$

The ARIMA method is appropriate only when the time series is stationary (i.e. averages, variances, and autocorrelations must be substantially constant over time)[3]. Sichel and alii pointed out that cycling asymmetry is an important feature that cannot be addressed by linear models.

So, ARCH models have been introduced. This new statistical tool is developed for various elements such as the underlying dynamic model, inference procedures, etc. ARCH models are applicable to the analysis of various series which are sufficiently irregular.

Compared to conventional self-regressive models, there is a linear formulation of the forecast of Y_t, assumed to be equal to $a_0 + a_1 Y_{t-1}$

$$Y_t = a_0 + a_1 Y_{t-1} + u_t \tag{1.07}$$

On the other hand, the error is allowed to have an order of magnitude based on past values:

$$h_t = E_{t-1}(u_t) = V_{t-1}(u_t) = b_0 + b_1 u_{t-1}^2$$

An endogenous evolution of these volatility is directly taken into account. Indeed, this type of formulation quickly appeared too restrictive to achieve the correct fit between the statistical model and the actual series[4].

Thus, it was necessary to introduce ARCH-M (ARCH in Mean) formulations:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 h_t + u_t \tag{1.08}$$

$$E_{t-1}(u_t) = 0, V_{t-1}(u_t) = h_t = b_0 + b_1 u_{t-1}^2$$

The quadratic form of volatility has often appeared to be inadequate for the actual series [4]. This leads to the introduction of formulations in which mean and conditional variance have more flexible forms.

$$Y_t = g_0(Y_{t-1}) + h_0(Y_{t-1})\varepsilon_t \tag{1.09}$$

$$E_{t-1}(\varepsilon_t) = 0, V_{t-1}(\varepsilon_t) = 1$$

In practice the form of g_0 and h_0 functions is not specified, so non-parametric approaches to estimation of g_0 and h_0 are applied directly.

2 Artificial Neurones Networks

Inspired by the functioning of the human brain, the neural approach is a relatively recent technique that provides a mathematical model for approximating relationships between variables. Unlike conventional approaches, this is one without imposing a specific functional form on the data, and without making assumptions a riori about the distribution of the variables in question.

Neuron networks are a data processing technique that will soon be part of the toolbox of any engineer concerned with making the most of the relevant information from the data he owns: make forecasts, develop models, etc. The main objective of neural network research was to increase our knowledge of the brain mechanism through the development of artificial systems able of replicating complex calculations similar to those of the human brain.

The main advantage of neural networks is that they are universal sparse approximators[3]. This means that they require fewer adjustable parameters to provide accuracy comparable to that obtained with conventional techniques.

Moreover, the advantage of neural networks over conventional regression methods is that they generally require a smaller number of adjustable parameters to obtain a given nonlinear precision model [3].

21 Neuron model

The neuron can be represented by a cell with multiple inputs and one output, and can be modeled by two operators:

- A summation operator developing a "post-synaptic potential":

$$p = \sum_i (w_i . x_i) \tag{1.10}$$

With w_i the weight and x_i the input.

- A decision operator that calculates the state of the neuron's output based on its potential p: this operator is called "activation function": s = F(p)

In a network, there are three types of neurons: incoming neurons, also called perceptual cells because of their property to acquire data whose origin is outside the network; outgoing neurons, which define the output of the network; and hidden neurons, which have no relation to the world outside the network, just to the other neurons in the network.

22 Topology

The topology of a neuron network is defined by its architecture (or structure)and the nature of its connections.

221 Architecture

The architecture of the network is described by the number of layers and the number of neurons in each layer.

222 Connections

The connection model defines how neurons in a network are interconnected. In general, the meaning of transferring information to a network is defined by the nature of connections: direct or recurring. Direct connections are those directed from a lower layer of index to a higher layer of index. Connections are said to be recurring when neuron outputs from one layer are connected to inputs from a lower index layer.

23 Learning and use

231 Implementation phase

The neuron network is used to perform a particular function. This function will be developed during a learning phase. The result of this function is obtained during a phase of network use (or propagation). Propagation through the network occurs by changing the state of the neurons, from the first hidden layer to the exit of the network.

In a network of neurons, information is encoded by the weights associated with connections. Learning is done by computation algorithms designed to adapt these weights to the excitations presented at the entrance of the network. Once the learning is complete, the weights are no longer changed. In fact, the learning is done by presenting data with some redundancy to a stand alone network. Hence, the objective of the network is to establish regularities.

232 Error gradient rule

The objective of this algorithm is to minimize an E cost function. The equation expresses this cost function from the quadratic error, for an input-output torque

$$E = \sum_i (d_k - s_k)^2 \tag{1.11}$$

With d_k the desired output for the index neuron k and s_k the output obtained by the network.

Learning involves a first phase of calculation in the direct direction where each neuron performs the weighted sum of its inputs and then applies the activation function f(derivative function) to obtain the update of the neuron. The equation (1.11) corresponds to this update with the p_i post-synaptic potential of neuron i, x_j the state of the neuron of the previous layer and w_{ij} the weight of the connection between the two neurons:

$$s_i = f(p_i) = f\left(\sum_{j=0}^{n} \left(w_{ij} x_j\right)\right) \tag{1.12}$$

This phase, known as propagation, allows to calculate the output of the network according to the input. The back propagation algorithm is to perform a gradient descent on criterion E. The E gradient is calculated for all weights as follows:

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial p_i} \frac{\partial p_i}{\partial w_{ij}} = \frac{\partial E}{\partial p_i} x_j \tag{1.13}$$

The gradient will then be noted C_j , and $C_j = -\frac{\partial E}{\partial p_i}$

The gradient attached to the output cells is then obtained by the equation:

$$C_j = -\frac{\partial E}{\partial p_i} = -\frac{\partial}{\partial p_i}\left(\sum_k (d_k - s_k)^2\right) = 2(d_i - s_i) f'(p_i) \tag{1.14}$$

For neurons in the hidden layers, the order of gradient calculation is the reverse of the order used to update states in the network. It takes place from the exit layer to the entrance; we are talking about backpropagation.

In the case of the total gradient algorithm, examples of the learning base are presented successively to the network, gradients accumulated over time, and weight change occurs only after all examples are presented (as opposed to the stochastic gradient where weight change is performed for each example presented). Weight change is obtained sthrough equation:

$$w_{ij}^{t+1} = w_{ij}^{t} + \alpha C_i s_j \quad (1.15)$$

The learning phase is often stopped when the error calculated on the whole learning base is below a user-defined threshold.

3 Results on experimentation

31 Exchange rate

The variable we carried out the experiment is the exchange rate. Thus, the exchange rate corresponds to the price of a national currency (in our case, Ariary) as expressed in another currency (for example, theUS dollar or the euro) or according to a basket of currencies.

Madagascar has a floating exchange rate regime. Since we are pursuing an inflation target in order to preserve the domestic value of Ariary, we cannot at the same time define an objective in terms of its external value. herefore, there is no established (fixed) value between Ariary and other currencies. The Ariary's exchange rate against the reference currency, and, moreover, against any other currency, floats; it depends on the demand and supply of Ariary on the Interbank Market of Currencies (MID).

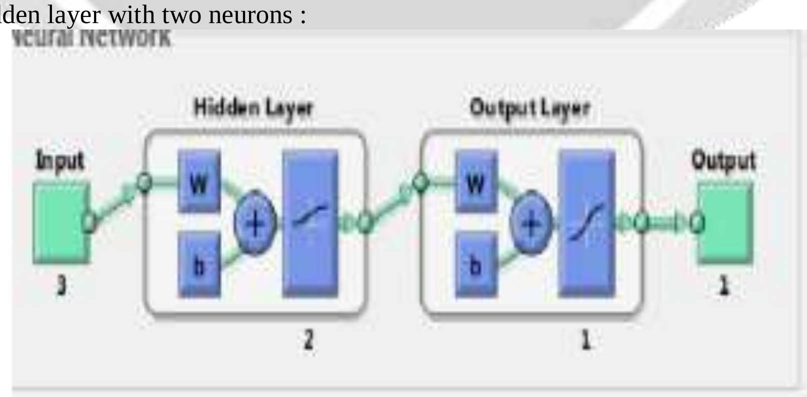
For an economy as open and trade-dependent as ours, the external value of the currency is a particularly relevant factor, as it affects, inter alia, the prices and volume of goods and services exported and imported. More precisely, the rise or fall in the external value of Ariary will make goods and services from Madagascar more or less expensive for foreign buyers, which will also tend to encourage or reduce external demand for our products. The upward or downward movement of Ariary in relation to other currencies will also affect imported goods, making them more or less accessible, and will therefore lead to a growth or decrease in the volume of our imports.

32 Network and learning

321 Network Architecture

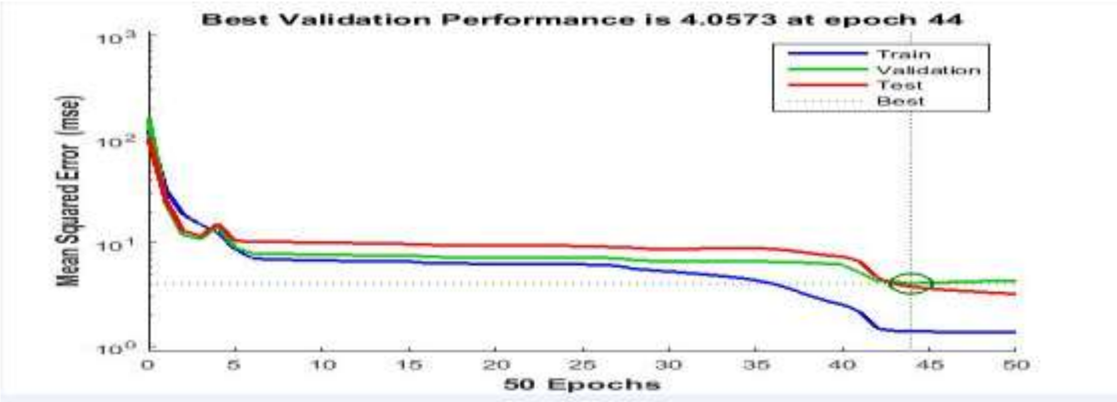
Our network is of type perceptron multi-layer architecture feedforward type, uses the learning supervised by the method of backpropagation. At first we started with three-layer perceptron, a hidden layer and an output layer as shown in the figure below, we varied the number of neurons per layer until we improved the results. In order to achieve good results we have used several means, one of them is to increase the number of layers in the network, because an architecture with several hiddenlayers will increase the accuracy of theestimates.

For the model of a hidden layer with two neurons :

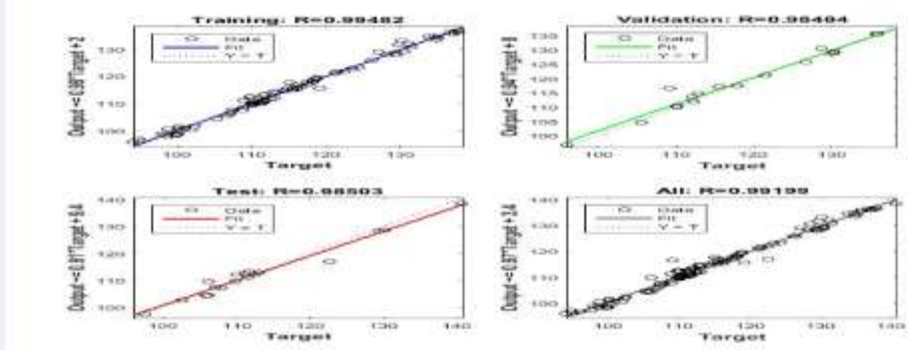


We have the following results:

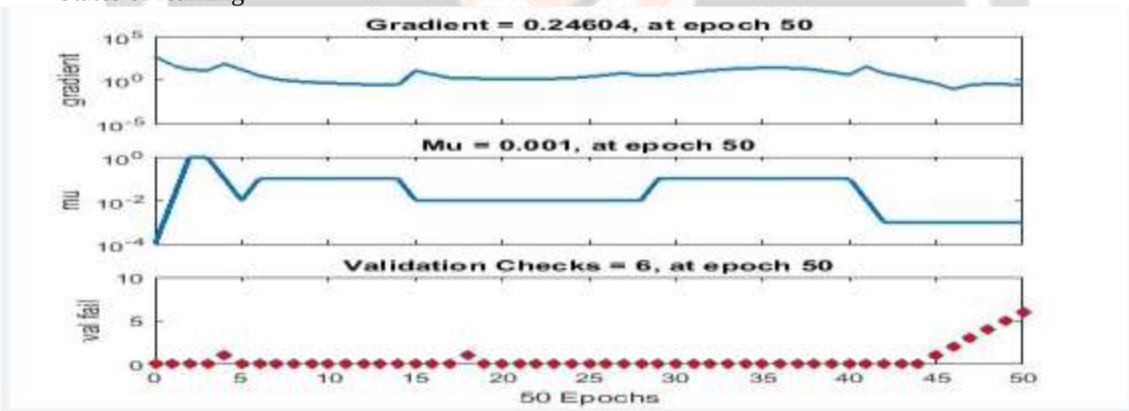
- Performance



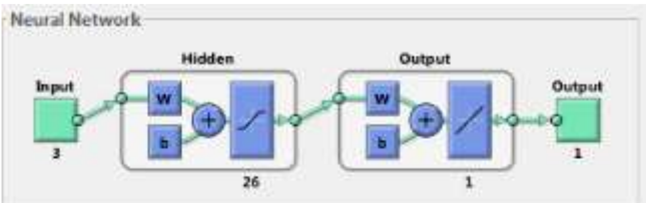
- Regression



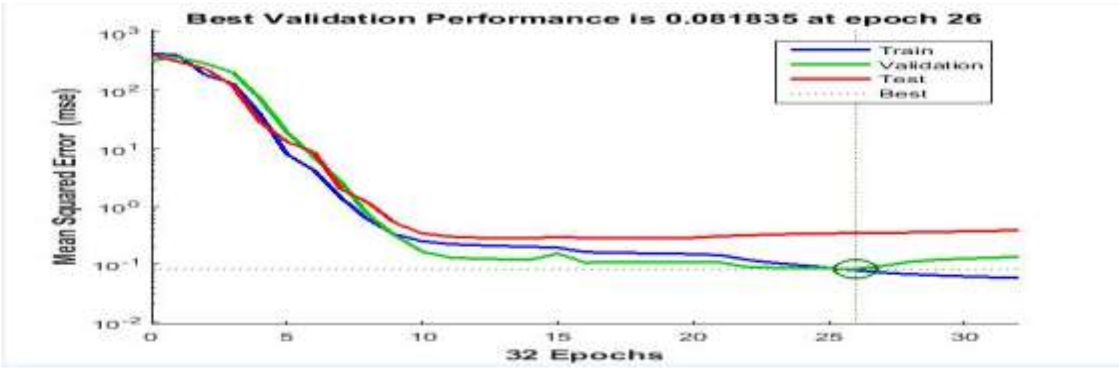
- States of learning



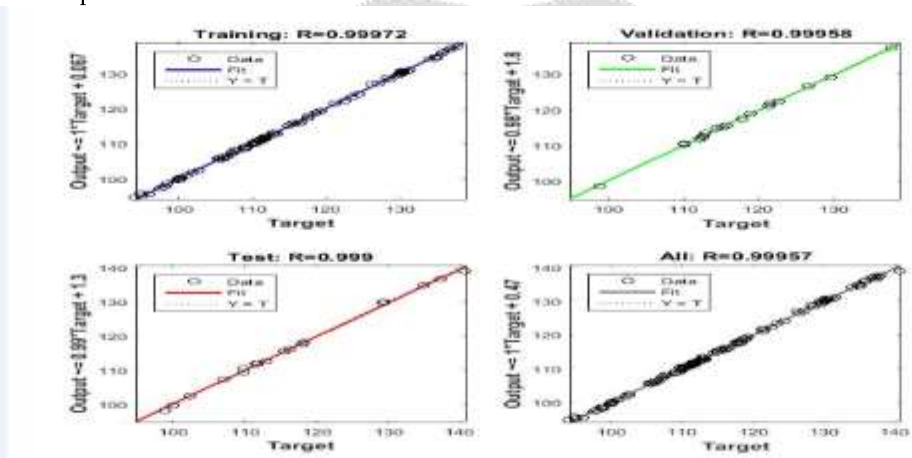
We used learning with a single hidden layer with different number of neurons like five (5), then ten (10), and the results remain unstable.
We found that with a number of neurons equal to twenty-six (26) in the hidden layer, we get the lowest average Quadratic Error.



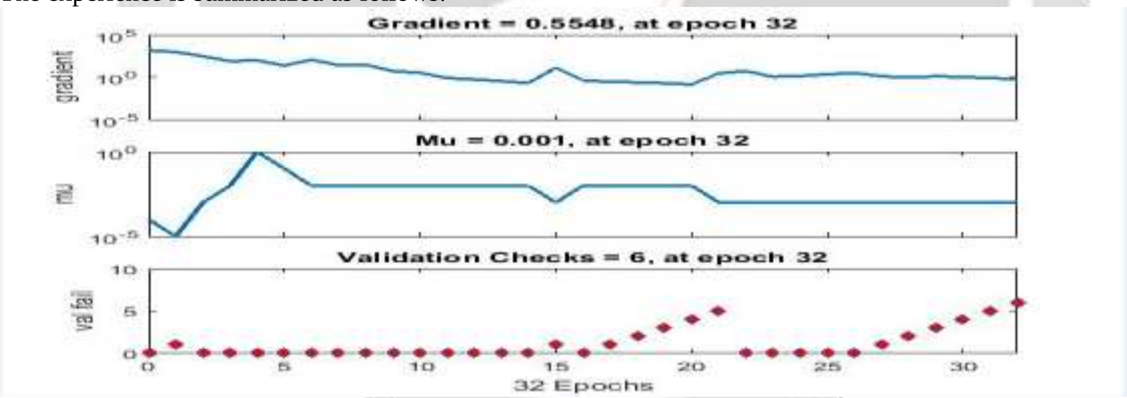
Performance is given by the following plot:



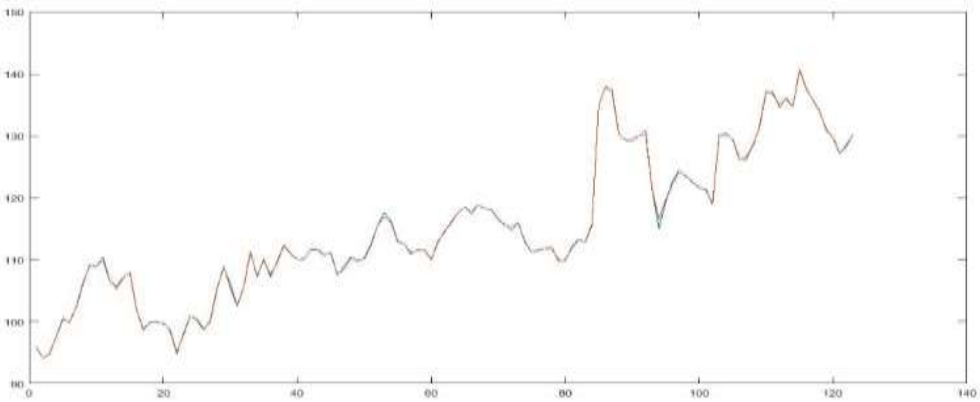
The regression is presented as follows:



The experience is summarized as follows:



And finally, the presentation of the results of the simulation with reality can be seen in the plot below:



Weight matrix :

$$
\begin{bmatrix}
-4{,}6269 & 5{,}7756 & 3{,}7895 \\
7{,}5594 & -6{,}2233 & 6{,}3176 \\
-4{,}2768 & -6{,}5645 & 3{,}2583 \\
-7{,}9106 & -1{,}1283 & -1{,}7466 \\
-2{,}9362 & 7{,}9392 & 2{,}9842 \\
-47674 & 7{,}102 & -4{,}5136 \\
-1{,}6096 & -5{,}1804 & 4{,}9262 \\
-2{,}4298 & -2{,}4907 & 6{,}8553 \\
6{,}785 & -2{,}3216 & 3{,}0187 \\
4{,}9074 & 5{,}8297 & -3{,}2525 \\
-7{,}887 & -7{,}6063 & 0{,}51827 \\
4{,}6541 & 4{,}4296 & -3{,}3699 \\
-6{,}2341 & 3{,}112 & -1{,}8457 \\
-4{,}4572 & -2{,}6481 & -6{,}6086 \\
4{,}6648 & -2{,}5807 & -7{,}7332 \\
5{,}6819 & -1{,}7381 & -3{,}0775 \\
-5{,}1528 & 0{,}49836 & -4{,}9179 \\
5{,}2627 & -3{,}4235 & -4{,}8119 \\
-6{,}8622 & 3{,}1541 & 3{,}6248 \\
3{,}6195 & -6{,}8304 & -19104 \\
5{,}7765 & -5{,}0598 & 5{,}6261 \\
0{,}78668 & 3{,}804 & 7{,}1198 \\
4{,}268 & -5{,}2222 & 3{,}4685 \\
2{,}7411 & -3{,}4032 & -6{,}9899 \\
-3{,}928 & 0{,}94174 & -5{,}5072 \\
-7{,}3002 & -4{,}5225 & -3{,}3455
\end{bmatrix}
$$

322 Bayesian inference algorithm integrated into the connection network

Sufficient efforts have been made to improve neural networks. In these perfections, we can reveal the incorporation of Bayesian methods into the learning of neural networks.

- Principles

Normal learning was accomplished by discerning a value from the weight vector that minimizes a cost or error function.

According to Bayesian methods, all parameters, especially network weights, are considered random variables from a probability distribution. Learning a network of neurons thus allows us to determine the probability distribution of weights knowing the learning data. Indeed,weights are assigned a priori probability, and once the learning data has been observed, this probability a priori is converted to a posteriori probability thanks to the Bayes theorem.

- Integration of MCMC methods into learning

The Mont Carlo methods by Markov chains (MCMC), are marked in the framework of Bayesian formalism.

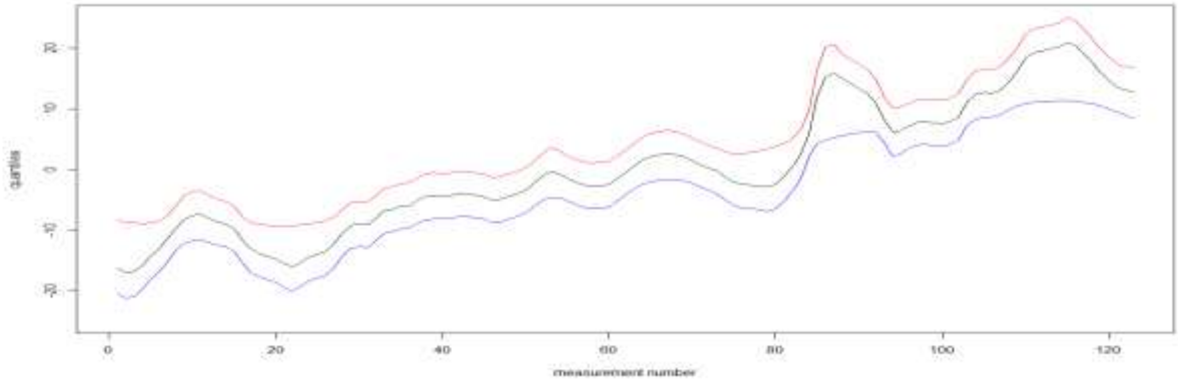
The Bayesian learning process begins with the definition of the model, and the distribution a priori $p(\theta)$ for the parametric model. A priori distribution expresses knowledge of parameter values before the data is observed. After observing new data D, the distribution a priori is transformed into a posteriori distribution via the Bayes theorem.

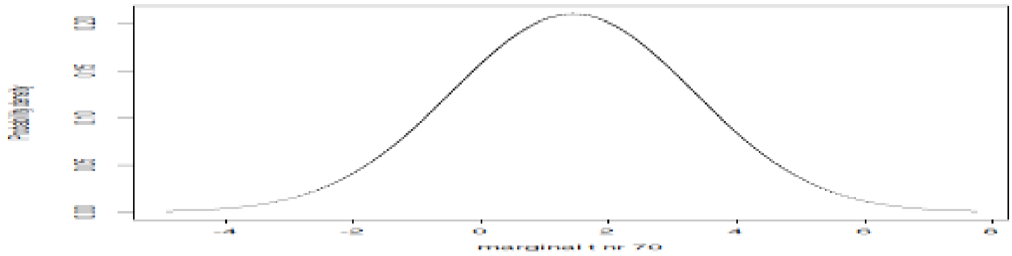
- Correction of adjustment

Adjusting a model to a particular dataset (examples) is all the better because the number of adjustable parameters is large. However, an excessive number of parameters may compromise the generalization (interpolation) capacity of the model.

If the network of neurons has an excessive number of parameters, due to an excessive number of hidden neurons, its output can pass with great precision through all learning points, but provide results lacking significance between these points; if it has too few parameters, the model is not rich enough to account for the complexity of the unknown regression function.

To resolve the adjustment, we define θ_1 as a fixed value, which is an effective way. But the problem is that by doing so, the data would no longer affect the distribution of θ_1. So let's look at an alternative approach that can eliminate this problem.





The above plot presents the earlier and later distribution of ρ for u_t .And, we can see that ρ is closed to 1,which means that u_t depends largely on u_{t-1}.

Neuron network have the ability to deal with diverse and varied problems. The result can be a prediction, classification or analysis of data. They are used to deal with unstructured problems, i.e. problems that do not have prior information.
Moreover, neura networks can learn from incomplete or noisy data. This data imperfection can be overcome by adding additional neurons to the hidden layer.
Bayesian methods for learning neural networks can bring severa ladvantages, because it is no tnecessary to define the size of the network to avoid over-fitting, and thenumber of hidden neurons can move towards infinity. However, the only cause that must limit network size is the capacity of the machines used and the free time to perform the necessary calculations, but since the parameters used are derived from a probability distribution, it is essential to learn a parameter to calculate integrals involving distributions of the other parameters. To calculate these integrals, we use the Monte Carlo Markov Chain (MCMC) method. The integration of this method with the learning of neural networks has led us to satisfactory results compared to other traditional learning algorithms.

4 Conclusion

A system is an organized collection of objects that interact to form a whole.This system is an open system. This naturally led us to the general theory of systems and the techniques of modeling complex systems. We cal lthese systems "complex" because they appear to be unpredictable. It is not possible to predict directly bycalculation the outcome of the processes involved or even to prejudge any behavior based on a simple composition of the individual behaviors of its constituents. However, it is possible to develop plausible scenarios (emergenc ephenomena, heavy trends, etc.),based on a model.
Modeling and simulation of complex systems is a difficult challenge. Our ability to build tools to study and manipulate models is a key issue .However, studying a system is all the more difficult because it is very heterogeneous, dynamic and open.When talking about systems, it is fundamental to understand that one is placed in a modeling logic, which immediately leads to a distinction between the concepts of real and formal systems.
The possibility is to present a modeling tool that allows the construction of a representation of acomplex system. The aim is to build a model of heterogeneous phenomena of interacting entities and allow the different phenomena to be represented separately by coexisting in a single model.
An econometric model is an equation whose role is to explain a phenomen on through variables that are initially considered determinative. The aim is to capture the most important fact(s) of the reality he seeks to represent. An econometric model is therefore a surely simplified representation of a phenomenon. A question then arises as to the reliability of an econometric model and whether it can be an acceptable representation of reality. A model is

necessarily something simple because it's easier to understand, understand and test. However, this simplification has risks: on the one hand, over-simplification; andon the other hand, unrealistic assumptions.
Connectionism is potentially relevant to cognitive research. In fact, the cognitive system would consist essentially of a non-symbolic network of artificial neurons that maintain activation and inhibition ratios between them. Connectionism will bring benefits to this complex reality modeling.

References

[1] Aragon Y., Séries temporelles avec R: méthodes et cas, Springer, Verlag France, 284p ;
[2] Delignières D., Séries temporelles: modèles ARIMA, Séminaire EA "Sport – Performance – Santé", 19p
[3] Dreyfus G, Les réseaux de neurones, Mécanique industrielle et matériaux, 51p
[4] Franchot A., L'économétrie des modèles dynamiques: avantages et limites des modèles ARCH, Journal de la société statistique de Paris, tome 133, n°4, p 53-54
[5] Sichel, Business Cycle Asymmetry: A Deeper Look, Economic Inquiry, 224-236