Forecasting High Frequency Data: An ARMA-Soft RBF Network Model for Time Series

Dušan Marček

Department of Applied Informatics, Faculty of Economic, VŠB-Technical University of Ostrava

Outlines

- Basic principles of identifying input-output functions of systems and forecasting
- Statistical ARCH-GARCH models
- A classification table of statistical and NN methods for time series modelling and forecasting
- ARMA-GARCH MODEL– Data
- RBF neural netvorks models
- Empirical Comparison, Conclusions

Basic principles of identifying input-output functions of systems and forecasting

- There are three major approaches to forecasting – explanatory, time series, machine learning
- Explanatory forecasting assumes a cause and effect relationship between the inputs into the system and its output. In this system any change in inputs will affect the output of the system in a predicable way.
- Time series
- In a special case machine learning such as SV regression or ANN offers relatively new ways for improving forecast method.

Explanatory approach

The econometric approach adopted from early days of econometrics is referred to as "AER" or Average Economic Regression and is concerned with the functional form of the multiple regression model in the form

$$y_t = \beta_0 + \beta_1 x_{1t} + ... + \beta_p x_{pt} + u_t$$

The formulation of an econometric forecasting model requires the following steps:

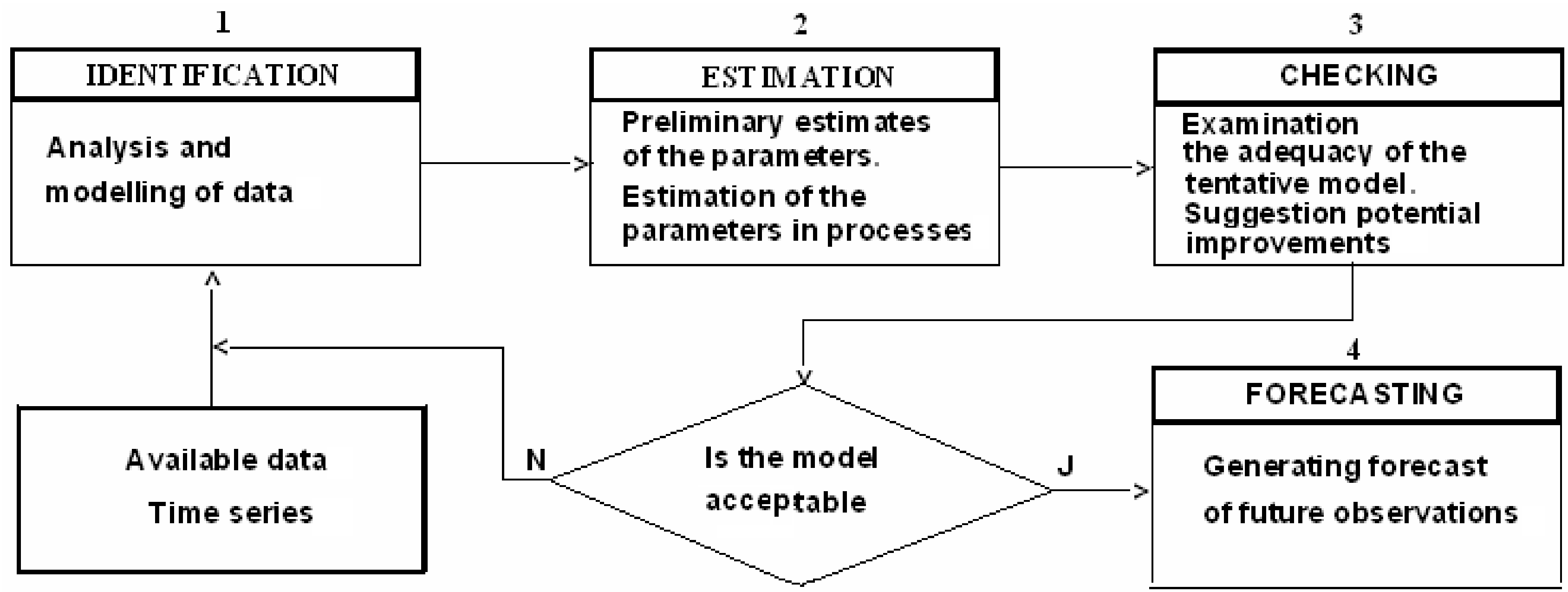
- The choice the independent variables
- specification of a functional form of the model
- collection, and analysis of a data set
- model estimation and statistical testing
- evaluation of the model´s forecasting over the ex post period

Time series approach

Box and Jenkins developed a new modeling approach based on time series analysis and derived from the linear filter known as AR or ARIMA (AutoRegressive Integrated Moving Average) models. The basic ARMA model of orders p, q (ARMA(p,q)) has the form:

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ \dots \ + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ \dots$$

The basic steps in Box-Jenkins Procedure are: (1) Identification, (2) Estimation, (3) Diagnostic checking, (4) Forecasting. See next figure



Flow chart of building an appropriate time series forecast model

ARCH-GARCH: Models

ARCH (GARCH) model for time sequence $\{y_t\}$ in the following form:

$$y_t = \nu_t \sqrt{h_t}, \qquad h_t = \alpha_0 + \sum_{i=1}^{m} \alpha_i y_{t-i}^2 + \sum_{j=1}^{s} \beta_j h_{t-j}, \tag{4}$$

Nelson (1991) proposed the following **exponential GARCH model** abbreviated as **EGARCH** to allow for leverage effects in the form:

$$\log h_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i \frac{|\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i}}{\sigma_{t-i}} + \sum_{j=1}^{q} \beta_j h_{t-j} \tag{5}$$

γ_i denotes the coefficient of leverage effects (see Zivot and Wang (2005, p. 243))

The basic GARCH model can be extended to allow for leverage effects. This is performed by treating the basic GARCH model as a special case of the **power GARCH (PGARCH)** model proposed by Ding, Granger and Engle (1993):

$$\sigma_t^d = \alpha_0 + \sum_{i=1}^{p} \alpha_i (|\varepsilon_{t-i}| + \gamma_i \varepsilon_{t-i})^d + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^d \tag{6}$$

where *d* is a positive exponent, and γ_i denotes the coefficient of leverage effects

(see Zivot and Wang (2005, p. 243)).

A classification table of statistical and NN methods for time series modelling and forecasting

E/Statistical-Probab. M.		SC	
Standard regression/ econometric models	Without or with Seasonal comp.	Integration of 3 IT:	ANN
			Fuzzy logic systems
Latest models: State-Space Models (K. filtration)	Structural models. EC, VEC models		Machine learning
Transfer Function M.		ANN:	
Models: ARIMA,		Perceptron-type NN	
ARCH-GARCH		**RBF NN**	classic
			soft (fuzzy logic)
Special models:	Bayes M.		**granular**
- ARIMA/ARCH-GARCH models require more costs of development, - there is not a convenient way to modify or update the estimates of the model parameters as each new observation becomes available and one has to periodically completely develop and refit the model.		Believe	
		Recurrent	
		Computational NN offer exciting adwantages such as learning, adaptation, fault-tolerace, parallelizm, generalization.	

ARMA-GARCH Models: Data

Our database is composed of of 610 monthly observations of the closing platinum prices (denoted yt (see Figure 1) (from 1960M01 to 2010M10). The time series is not stationary. After trend removing and differentiation it became stationary. The sample period for analysis from 1960M01 to 2008M12 was defined and the ex post forecast period from 2009M01 to 2010M10 (denoted as validation or ex-post data set)

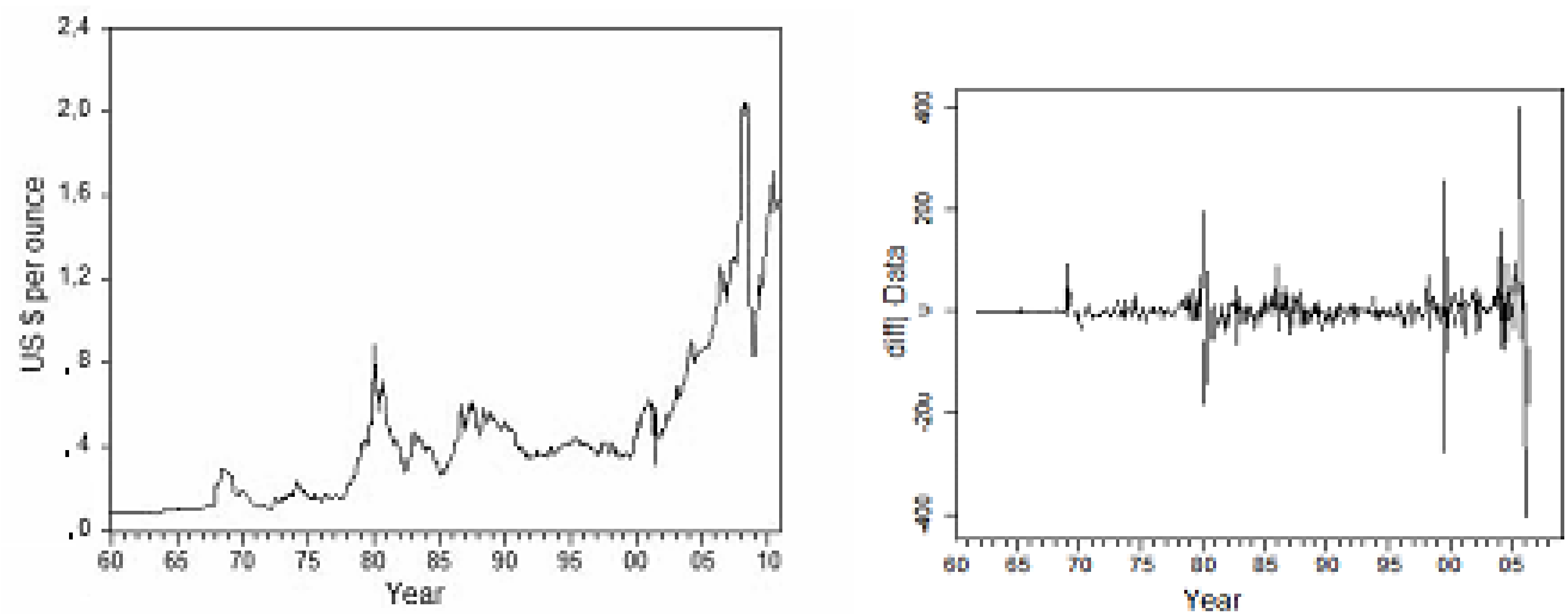


Fig. 1. Time series of monthly platinum prices 1960 – 2010 (left), the time plot of the data after the differencing (right).

Statistical modelling - ARMA-GARCH models

The appropriate ARMA model was identified by investigating the behavior of the autocorrelation (AC), partial autocorrelation (PAC) functions and Akaike Information Criterion (AIC) by using a computer program. We looked to determine the maximum lag for which the PAC coefficient was statistically significant and the lag given the minimum AIC. According to these criterions the ARMA(1,1) model was specified.

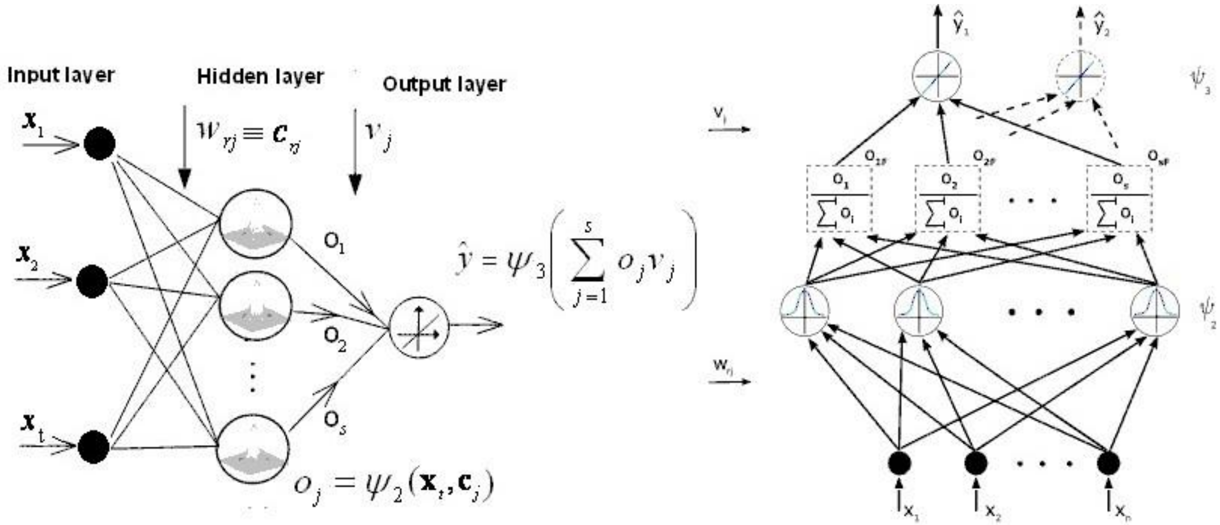
The LM (Lagrange-Multiplier) test performed on the time series of monthly platinum prices 1960 – 2010 clearly indicates presence of autoregressive conditional heteroscedasticity. The final models have the forms ARMA(1,1)/GARCH(1,1) GED.

$$y_t = -2{,}08398 + 0{,}8429.y_{t-1} + \varepsilon_t - 0{,}5328.\varepsilon_{t-1} \qquad (1)$$

$$h_t = 0.104759 + 2.7901\hat{\varepsilon}^2_{t-1} + 0.3912h_{t-1} \qquad (2)$$

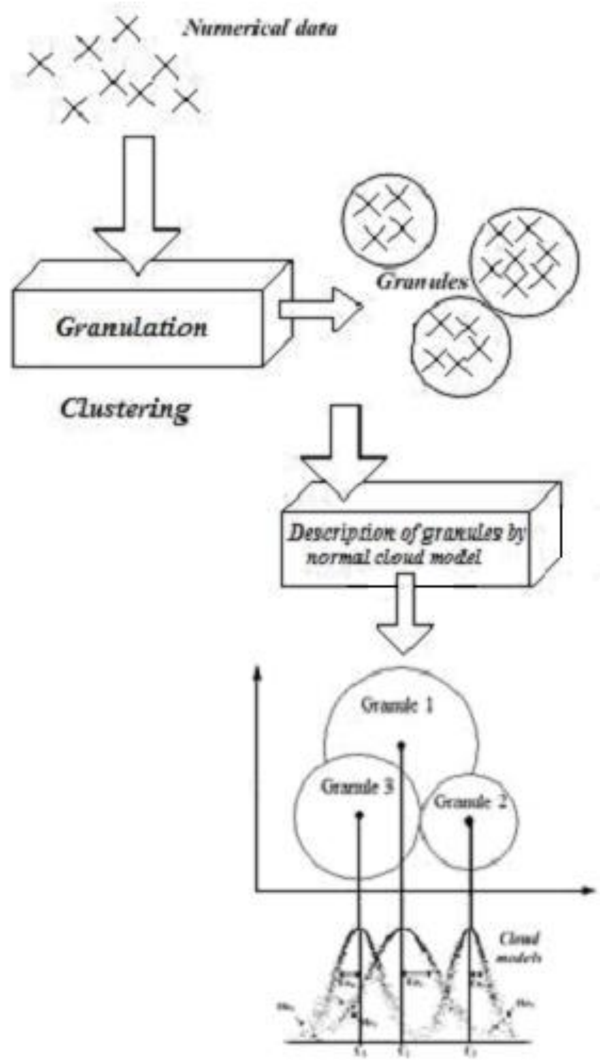
Neuronal Approach

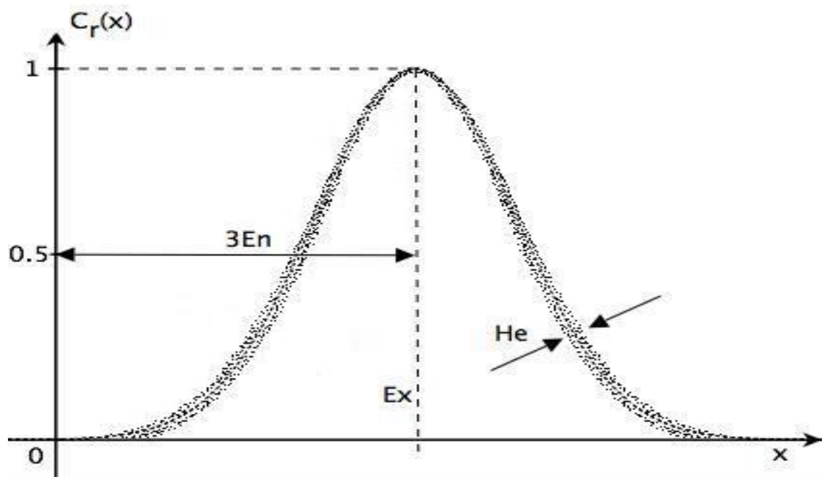
For the investigation with neural networks an RBF (classic, fuzzy logic) three layer feed-forward network is employed



The layer between the input and output layers is normally referred to as the hidden,
the hidden layer weights $\mathbf{w}_j$ represent the centres $\mathbf{c}_j$ of activation functions in the hidden layer,
to find the weights or centres of activation functions we use the adaptive version of *K*-means clustering algorithm for *s* clusters

Gaussian CAF





$\equiv \mathbf{c}_j$

$$\psi_2(\mathbf{x}_t, \mathbf{c}_j) = \exp\left[-(\mathbf{x}_t - E(\mathbf{x}_j)) / 2(En')^2\right] = \exp\left[-(\mathbf{x}_t - \mathbf{c}_j) / 2(En')^2\right]$$

Neural Approach

We also combined the soft RBF neural network with the statistical ARMA(1,1) model expressed by Eq. 1, in one unified framework. The scheme of such proposed hybrid model is depicted in Figure 2.

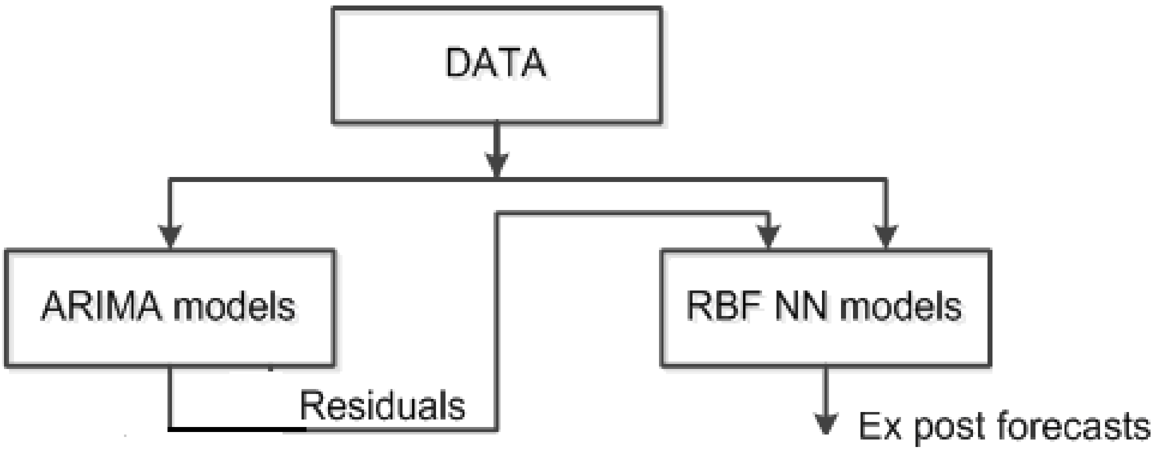


Fig. 2. The scheme of proposed hybrid forecasting model (see text for details).

ANN modelling (forecast comparison)

Table 1: Statistical summary measures of model´s ex post forecast accuracy for ARMA(1,1,1)+PGARCH(1,1) model and RBF nets.

Model	RMSE
ARMA(1,1)+GARH(1,1) GED	72.155
RBF NN (classic)	76.206
RBF NN (soft)	67.974
RBF NN (hybrid)	33.601

Results –empirical comparison

We showed that both the statistical and neural approach are very accurate and suitable for their deployment in managerial prediction systems. Thus, neural networks are usually used in the complicated problems of prediction because they minimize the analysis and modeling stages and the resolution time. Although we cannot generally to say that neural network models outperforms statistical models, we may say that neural network models have equivalent prediction performance comparing to statistical models.

We could also see, the soft RBF NN has such attributes as computational efficiency, simplicity, and ease adjusting to changes in the process being forecast

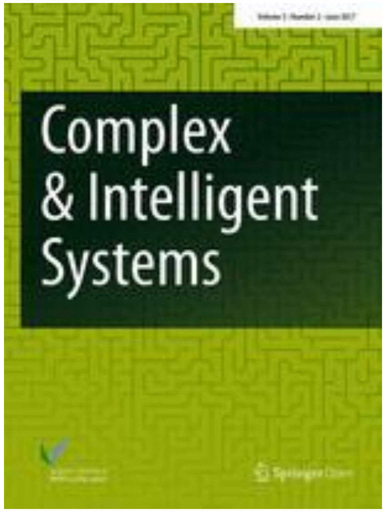
Other economic theories or assumption in particular as well as different time horizons will be considered for further investigations. In addition, the structure of neural networks is still to be investigated in the area of artificial intelligence theory.

Statistical Models and Granular Soft RBF Neural Network for Malaysia KLCI Price Index Prediction

•Authors: Dusan Marcek

•**Abstract**

Two novel forecasting models are introduced to predict the data of Malaysia KLCI price index. One of them is based on Box-Jenkins methodology where the asymmetric models, i.e. EGARCH and PGARCH models were used to form the random component for ARIMA model. The other forecasting model is a soft RBF neural network with cloud Gaussian activation function in hidden layer neurons. The forecast accuracy of both models is compared by using statistical summary measures of model's accuracy. The accuracy levels of the proposed soft neural network are better than the ARIMA/PGARCH model developed by most available statistical techniques. We found that asymmetric model with GED errors provide better predictions than with Student's *t* or normal errors one. We also discuss a certain management aspect of proposed forecasting models by their use in management information systems.



Complex & Intelligent Systems
ISSN: 2199-4536 (Print) 2198-6053 (Online)

Forecasting of Financial Data: A Novel Fuzzy Logic Neural Network Based on Error Correction Concept and Statistics

Authors:
Dusan Marcek

doi:https://doi.org/10.1007/s40747-017-0056-6

Abstract

First, this paper investigates the effect of good and bad news on volatility in the BUX return time series using asymmetric ARCH models. Then, the accuracy of forecasting models based on statistical (stochastic), machine learning methods, and soft/granular RBF network is investigated. To forecast the high-frequency financial data, we apply statistical ARMA and asymmetric GARCH-class models. A novel RBF network architecture is proposed based on incorporation of an error-correction mechanism, which improves forecasting ability of feed-forward neural networks. These proposed modelling approaches and SVM models are applied to predict the high-frequency time series of the BUX stock index. We found that it is possible to enhance forecast accuracy and achieve significant risk reduction in managerial decision making by applying intelligent forecasting models based on latest information technologies. On the other hand, we showed that statistical GARCH-class models can identify the presence of leverage effects, and react to the good and bad news.

INTERNATIONAL JOURNAL OF COMPUTATIONAL INTELLIGENCE SYSTEMS

Publisher ATLANTIS PRESS, 29 AVENUE LAUMIERE, PARIS, 75019, FRANCE
ISSN: 1875-6883

Granular RBF NN Approach and Statistical Methods Applied to Modelling and Forecasting High Frequency Data

Dusan Marcek[1,2], Milan Marcek[3,4], Jan Babel[5]

Abstract
We examine the ARCH-GARCH models for the forecasting of the bond price time series provided by VUB bank and make comparisons the forecast accuracy with the class of RBF neural network models. A limited statistical or computer science theory exists on how to design the architecture of RBF networks for some specific nonlinear time series, which allows for exhaustive study of the underlying dynamics, and determination of their parameters. To illustrate the forecasting performance of these approaches the learning aspects of RBF networks are presented and an application is included. We show a new approach of function estimation for nonlinear time series model by means of a granular neural network based on Gaussian activation function modelled by cloud concept. In a comparative study is shown, that the presented approach is able to model and predict high frequency data with reasonable accuracy and more efficient than statistical methods.

Keywords: Time series, classes of ARCH-GARCH models, volatility, forecasting, neural networks, cloud concept, forecast accuracy.

Thank you