NEURAL NETWORKS AND VECTOR AUTOREGRESSIVE MODEL IN FORECASTING YIELD CURVE

Zdravka Aljinović

University of Split, Faculty of Economics
Split, Croatia
zdravka.aljinovic@efst.hr

Tea Poklepović

University of Split, Faculty of Economics
Split, Croatia
tea.poklepovic@efst.hr

Abstract

Yield curve represents a relationship between the rate of return and maturity of certain securities. A range of activities on the market is determined by the abovementioned relationship; therefore its significance is unquestionable. Besides that, its shape reflects the shape of the economy, i.e. it can predict recession. These are the reasons why it is very important to properly and accurately estimate the yield curve. There are various models evolved for its estimation; however the most used is a parametric Nelson-Siegel model. What is also important is the ability of forecasting yield curve. Therefore in this paper after the estimation of weekly yield curves on Croatian financial market in years 2011 and 2012 with Nelson-Siegel model, yield curves are predicted using Neural networks and Vector autoregressive model. The obtained results are compared and conclusions regarding forecasting yield curves are given.

Keywords - yield curve, Nelson-Siegel model, Neural network, Vector autoregressive model

1 INTRODUCTION

The yield curve, as a picture of relationships between the yields on bonds of different maturities, provides a way of understanding the common markets' evaluation in the future, and whether the economy will be strong or weak [4].

Interest rates movements depend on the maturity period and the form of the yield curve has a great effect on the financial markets and the behaviour of financial intermediaries. Intermediaries will, to maximize their own profits, take into account the difference between short-term and long-term interest rates. A full range of activities in the financial markets is actually determined by the relationship between the interest rate and maturity.

Nelson Siegel model is extremely popular in the practice; both individual investors and the central banks use this model. This model is simple and stable for the evaluation, it is quite flexible and very well suited for assessing yields for more bonds or one bond and for the time series of returns, for a large number of countries and time periods and for different classes of bonds. It also has good prediction ability [8].

In this paper yield curves on Croatian financial market are calculated on weekly basis from 7th October 2011 to 24th August 2012 using Nelson-Siegel model and forecasted using Neural networks and Vector autoregressive model as it is given in Dedi et al [5].

In the first part of the paper theoretical overview of the yield curve is provided, followed by the explanation of the Nelson-Siegel model, most commonly used model for yield curve evaluation. Furthermore, Vector autoregressive and Neural network models are introduced. Finally, yield curves are calculated using Nelson-Siegel model and forecasted using both, Neural network and Vector autoregressive models on Croatian financial market.

2 NELSON-SIEGEL MODEL

Often used model for developing yield curve in the practice is the Nelson-Siegel model [9]. Nelson and Siegel introduced a simple, parsimonious model, which can adapt to the range of shapes of yield curves: monotonic, humped and S shape.

A class of functions that readily generates the typical yield curve shapes is that associated with solutions to differential or difference equations [1]. If the instantaneous forward rate at maturity *T*, *f(t,T),* is given by the solution to a second-order differential equation with real and unequal roots, it is of the form:

$$f(t,T) = \beta_0 + \beta_1 e^{-\frac{T-t}{\tau_1}} + \beta_2 e^{-\frac{T-t}{\tau_2}} \tag{1}$$

where τ_1 and τ_2 are time constants associated with the equation, and β_0, β_1 and β_2 are determined by initial conditions.

Now, zero-coupon rates $R(t)$ can be calculated by averaging the corresponding instantaneous forward rates:

$$R(t,T) = \frac{1}{T-t}\int_t^T f(x,T)\,dx \tag{2}$$

A more parsimonious model that can generate the same range of shapes is given by the equation solution for the case of equal roots:

$$f(t,T) = \beta_0 + \beta_1 e^{-\frac{T-t}{\tau}} + \beta_2 \frac{T-t}{\tau} e^{-\frac{T-t}{\tau}} \tag{3}$$

By substituting (3) into (2) and integrating, it is obtained:

$$\begin{aligned} R(t,T) &= \frac{1}{T-t}\int_t^T f(x,T)\,dx = \frac{1}{T-t}\int_t^T \left(\beta_0 + \beta_1 e^{-\frac{T-t}{\tau}} + \beta_2 \frac{T-t}{\tau} e^{-\frac{T-t}{\tau}}\right) dx = \\ &= \frac{1}{T-t}\left(\beta_0 (T-t) - \beta_1 \tau e^{-\frac{T-t}{\tau}} + \beta_1 \tau + \beta_2 \tau \left(-\frac{T-t}{\tau} e^{-\frac{T-t}{\tau}} - e^{-\frac{T-t}{\tau}} + 1\right)\right) \end{aligned} \tag{4}$$

After a simple rearrangement of this expression, the yield to maturity is given by:

$$R(t,T) = \beta_0 + (\beta_1 + \beta_2)\frac{1-e^{-\frac{T-t}{\tau}}}{\frac{T-t}{\tau}} - \beta_2 e^{-\frac{T-t}{\tau}} \tag{5}$$

So, the forward and zero-coupon yield curves are functions of four parameters: β_0, β_1, β_2 and τ.

It can be seen that

$$\lim_{T\to\infty} R(t,T) = \beta_0 \tag{6}$$

and β_0 corresponds to zero-coupon rates for very long maturities.

At the short end of the curve it is:

$$\lim_{T\to t} R(t,T) = \beta_0 + \beta_1 \tag{7}$$

which implies that the sum of parameter values β_0 and β_1 should be equal to the level of the shortest interest rates.

It can be seen that if β_1 is negative, the forward curve will have a positive slope and other way round. The parameter β_2, indicates the magnitude and the direction of the hump and if it is positive, a hump will occur at τ whereas,. In case it is negative, a U-shaped value will occur at τ. So it can be concluded that parameter τ which is positive, specifies the position of the hump or U-shape on the entire curve. Consequently, Nelson and Siegel propose that with appropriate choices of weights for these three components, it is possible to generate a variety of yield curves based on forward rate curves with monotonic and humped shapes [1].

The Nelson-Siegel model, which has only four parameters, enables us to estimate the yield curve, without being over-parameterized, when the number of observed bond prices is limited [7]. In the practice Nelson-Siegel model is preferred for the use especially where there are few input data [10]. Nelson and Siegel [9] demonstrated that their proposed model is capable of capturing many of the typically observed shapes that the spot rate curve assumes over time [3]. A significant weakness of the Nelson-Siegel model, resulting from its low elasticity, is goodness of fit that is lower than in the case of polynomial models. When the curve is fitted to an irregular set of data points this can result in relatively large deviations of model values from actually observed rates [8].

3 THE BASICS OF THE FORECASTING METHODS

3.1 VECTOR AUTOREGRESSIVE MODEL

Vector Autoregressive (VAR) model is a multivariate time series model that consist of multiple equations [2]. VAR model defined with *n* endogenous variables and *k* lags can be written as:

$$Z_t = a_0 + A_1 Z_{t-1} + ... + A_k Z_{t-k} + BD_t + \varepsilon_t \tag{8}$$

where Z_t is *n*-dimensional vector of potentially endogenous variables, $A_1,..., A_k$ are $n \times n$ coefficient matrices, D_t is a vector of other exogenous variables with coefficient matrix *B*. Vector a_0 is a vector of constants (intercept) and ε_t is vector of error terms, i.e. *n*-dimensional white noise process. The parameters of VAR model can be estimated using ordinary least squares method, where the optimal order, i.e. number of lags *k* can be found using information criteria: final prediction error (FPE), Akaike's information criterion (AIC), Schwarz's Bayesian information criterion (SBIC), and the Hannan and Quinn information criterion (HQIC). The advantages of VAR models are: simplicity of the model (it is not necessary to classify endogenous and exogenous variables in the model), ease of estimation (each equation can be estimated with ordinary least square method), quality of forecasted estimates.

3.2 NEURAL NETWORKS

Neural network (NN) is an artificial intelligence method, which has recently received a great deal of attention in many fields of study. It attempts to model the capabilities of the human brain [15]. Neural networks have been used for a wide variety of applications (engineering, law, computer science, medicine, manufacturing, transportation, finance etc.) where statistical methods are traditionally employed. Neural networks can be seen as a non-parametric statistical procedure that uses the observed data to estimate the unknown function [12]. Neural networks depend on data; they can learn from it and adjust to it, which implies that there is no need for *a priori* knowledge of the functional form of the relationship between variables. When appropriately specified, they are universal approximators, i.e. they can approximate any functional form between variables with high level of accuracy. Neural network architecture is very flexible. A wide range of statistical and econometric models can be specified modifying activation functions or the structure of the network (number of hidden layers, number of neurons etc.): multiple regression, vector autoregression, logistic regression, time series models, etc. Neural networks often give better results than statistical methods because of their possibility of analyzing the missing data, data with noise and learning from the previous data. Empirical researches show that neural networks are successful in forecasting extremely volatile financial variables that are hard to predict with standard statistical methods such as: exchange rates [6], interest rates [13] and stocks.

The same as human brain, neural network is an ensemble of interconnected neurons grouped in layers that send information to one another. Neural networks usually have two or more layers: input, hidden and output layer. Input neurons receive data from the external world and send it to one or more hidden neurons. In the hidden layer information from neurons are processed and sent to output neurons. Information than backpropagate through network and the values of weights between neurons are adjusted to the target output. The process in the network is repeated as much iterations (epochs) as needed to reach the output that is the closest to the targeted output. Learning in neural network is a process in which the system comes to the values of the weights between neurons. The weight is actually the power of the relationship between the two neurons. If the neuron *j* is connected to neuron *i*, w_{ij} is interpreted as the strength of the connection from neuron *j* to neuron *i*. The Figure 1 (a) presents a classic schematic representation of the neural network with one hidden layer and one neuron in a hidden layer, where output *y* and inputs $x_1, x_2, \ldots, x_j$ are $n \times 1$ vectors, and *n* is a number of observations.

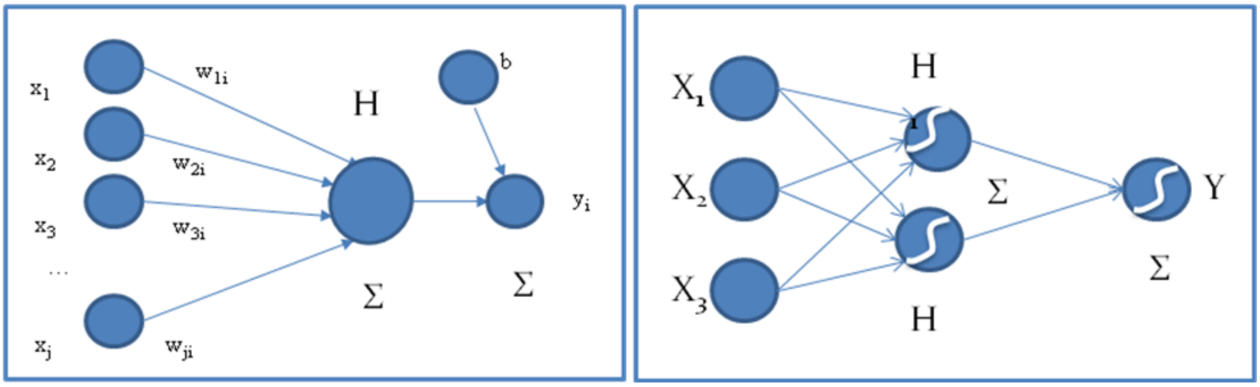

Figure 1 (a) Classic schematic representation of neural network; (b) Neural network MLP (3-2-1)

When a neuron receives inputs from connected units, the value of its input is calculated based on activation function. In most cases one of the inputs is called a *bias*, which is equal to one for all the observations. The value of the hidden neuron h_i can be calculated as:

$$h_i = \sum_{j=1}^{n} \left(w_{ji} \cdot x_j \right) \tag{9}$$

The output from the neuron is calculated based on the activation function:

$$y_i = f\left(h_i\right) \tag{10}$$

In the simplest neural network the activation function is linear, and the most commonly used are exponential, logistic and hyperbolic tangent function. Using nonlinear activation function allows a neural network to capture nonlinearity in data. Designing neural network is the process of trial and errors, where a vast number of neural networks are trained and the best neural network gives the minimum mean squared error (MSE).

The Figure 1 (b) is a representation of three-layer feedforward neural network, also called multilayer perceptron (MLP), with three inputs, two neurons in the hidden layer and one output, i.e. MLP (3-2-1).

4 EMPIRICAL RESULTS

In Croatia still does not exist an official yield curve due to a scarce issue of Croatian bonds denominated in Kuna and weak trade on a secondary market. In order to calculate yield curve on a Croatian financial market data from Zagreb money market, where data for treasury bills can be found, and Reuters data base, where data for government bonds can be found, is collected. Yield curves are calculated on a weekly basis from 7th October 2011 to 24th August 2012 using Nelson-Siegel model. Even though on these dates there was pour trade on treasury bills and bonds (on observed dates the number of securities traded was mostly 10), yield curves are successfully estimated using above mentioned formulas.

Parameters β_0, β_1, β_2 and τ are estimated for Nelson-Siegel model in MS Office Excel using least square method with quasi-Newton. In the case where it was particularly difficult to estimate parameters, using Simplex method in Statistica10 starting points for an estimation of parameters are generated. These appropriate start values are then used in subsequent quasi-Newton iterations[1]. Resulting yield curves are given in Figure 2.

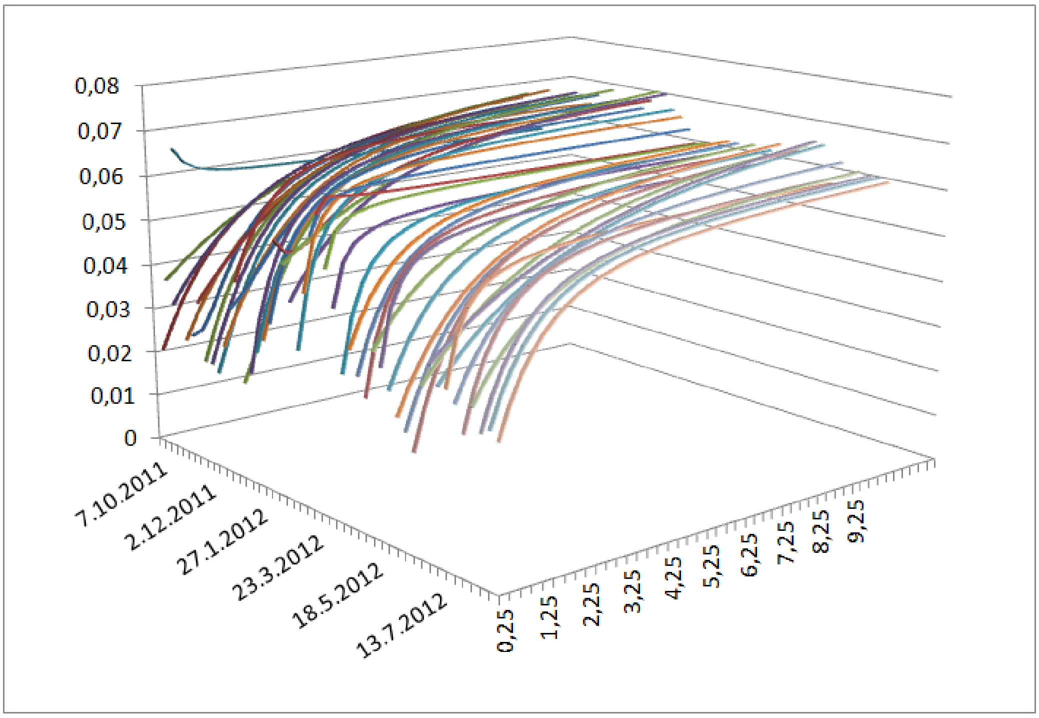

Figure 2 Yield curves on Croatian financial market using Nelson-Siegel model

After the estimation of the parameters using Nelson-Siegel model, yield curves are forecasted using Vector autoregressive and Neural network models, by predicting parameters β_0, β_1, β_2 and τ.

The parameters are predicted using Vector autoregressive (VAR) model in Stata11 by dividing the sample on two sets: first, the test set from 7th October 2011 until 15th June 2012 and second, the validation set from 22nd June 2012 until 24th August 2012. Based on final prediction error (FPE) and Akaike's information criterion (AIC) VAR (1) model is chosen, which is defined as:

$$Z_t = a_0 + A_1 Z_{t-1} + \varepsilon_t \tag{9}$$

VAR (1) model is tested on test set and predictions of four parameters β_0, β_1, β_2 and τ in out-of-sample forecast with mean square error (MSE) are given in the Table 1.

Table 1 Estimated and predicted parameters using VAR(1) with MSE

Parameters	β_0	$\hat{\beta}_0$	β_1	$\hat{\beta}_1$	τ	$\hat{\tau}$	β_2	$\hat{\beta}_2$
22.6.2012	0.0792	0.0684	-0.0422	-0.0507	2.3586	0.7737	0.0000	0.0032
29.6.2012	0.0769	0.0685	-0.0444	-0.0493	1.7828	0.8435	0.0000	0.0082
6.7.2012	0.0619	0.0693	-0.0490	-0.0495	0.1801	0.7983	0.0000	0.0054
13.7.2012	0.0694	0.0695	-0.0419	-0.0493	0.8197	0.8057	0.0000	0.0042
20.7.2012	0.0659	0.0695	-0.0497	-0.0491	0.4400	0.8234	0.0000	0.0043
27.7.2012	0.0688	0.0695	-0.0394	-0.0490	1.0312	0.8286	0.0000	0.0044

[1] *Simplex* method is generally less sensitive to local minima and is usually used in combination with the *quasi-Newton* method [11]

3.8.2012	0.0667	0.0695	-0.0462	-0.0490	0.5670	0.8290	0.0000	0.0044
10.8.2012	0.0669	0.0695	-0.0443	-0.0490	0.6059	0.8294	0.0000	0.0044
17.8.2012	0.0663	0.0695	-0.0458	-0.0490	0.5763	0.8297	0.0000	0.0044
MSE	0.00028		0.00029		4.14751		0.00022	

For prediction of Nelson-Siegel parameters using Neural networks (NN), training sample from 7th October 2011 until 11th May 2012, testing sample from 18th May 2012 until 15th June 2012 and validation sample from 22nd June 2012 until 24th August 2012 is used. MLP model with one hidden layer is used with the process of trial and errors for defining the right number of units in a hidden layer and the activation function. The best neutral network is obtained with seven hidden neurons (MLP 4-7-4), exponential activation function in hidden layer and logistic activation function in output layer. The predictions of four parameters β_0 , β_1, β_2 and τ in out-of-sample forecast and mean square error (MSE) are given in the Table 2.

Table 2 Estimated and predicted parameters using NN(4-7-4) with MSE

Parameters	β_0	$\hat{\beta}_0$	β_1	$\hat{\beta}_1$	τ	$\hat{\tau}$	β_2	$\hat{\beta}_2$
22.6.2012	0.0792	0.0697	-0.0422	-0.0490	2.3586	0.8707	0.0000	0.0053
29.6.2012	0.0769	0.0697	-0.0444	-0.0489	1.7828	0.8721	0.0000	0.0054
6.7.2012	0.0619	0.0697	-0.0490	-0.0489	0.1801	0.8754	0.0000	0.0054
13.7.2012	0.0694	0.0696	-0.0419	-0.0488	0.8197	0.8824	0.0000	0.0054
20.7.2012	0.0659	0.0697	-0.0497	-0.0488	0.4400	0.8780	0.0000	0.0054
27.7.2012	0.0688	0.0696	-0.0394	-0.0488	1.0312	0.8822	0.0000	0.0054
3.8.2012	0.0667	0.0697	-0.0462	-0.0488	0.5670	0.8759	0.0000	0.0054
10.8.2012	0.0669	0.0697	-0.0443	-0.0488	0.6059	0.8804	0.0000	0.0054
17.8.2012	0.0663	0.0697	-0.0458	-0.0488	0.5763	0.8795	0.0000	0.0054
MSE	0.00024		0.00024		4.00756		0.00026	

Neural network model gives marginally smaller MSE than VAR based method, which means that neural networks give better results than Vector autoregressive model in forecasting yield curve. Both models predict the values of parameters β_0 , β_1 and β_2 extremely well, ending with small mean square errors. However, due to the fact that estimated parameter τ is varying extremely through forecasting period, both models' forecasting abilities are weak, ending with much larger mean square errors. Figure 3 shows the yield curve estimated with Nelson-Siegel model on 22nd June 2012 and yield curves predicted with neural network and vector autoregressive model. It shows good short term forecast abilities of both models and marginally better results of NN model.

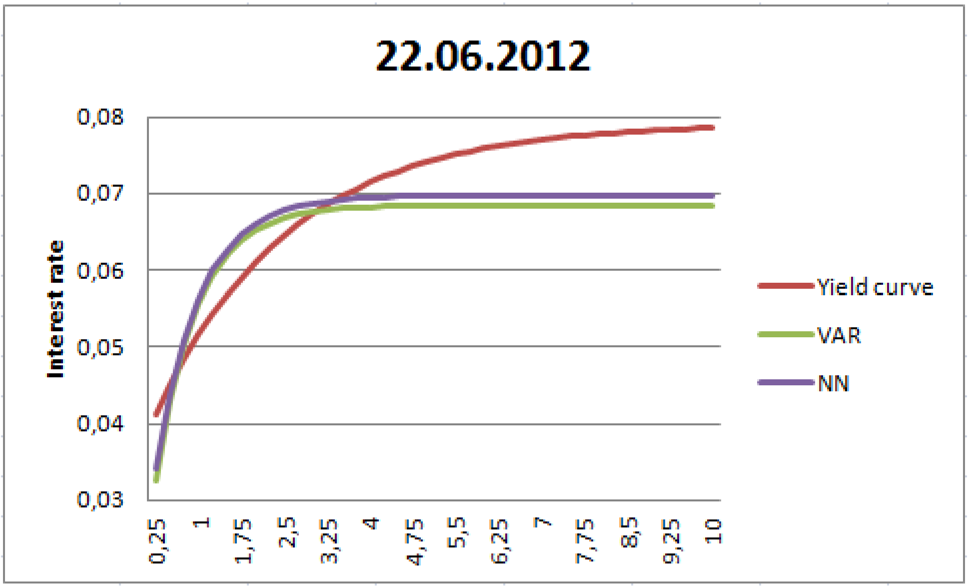

Figure 3 Nelson-Siegel yield curve and predicted yield curves using VAR and NN models on 22.06.2012.

Figure 4 shows the yield curve estimated with Nelson-Siegel model on 24th August 2012 and yield curves predicted with neural network and vector autoregressive model. It shows long term forecast abilities of both models and it can be concluded that both models perform poorly in longer term forecast horizons.

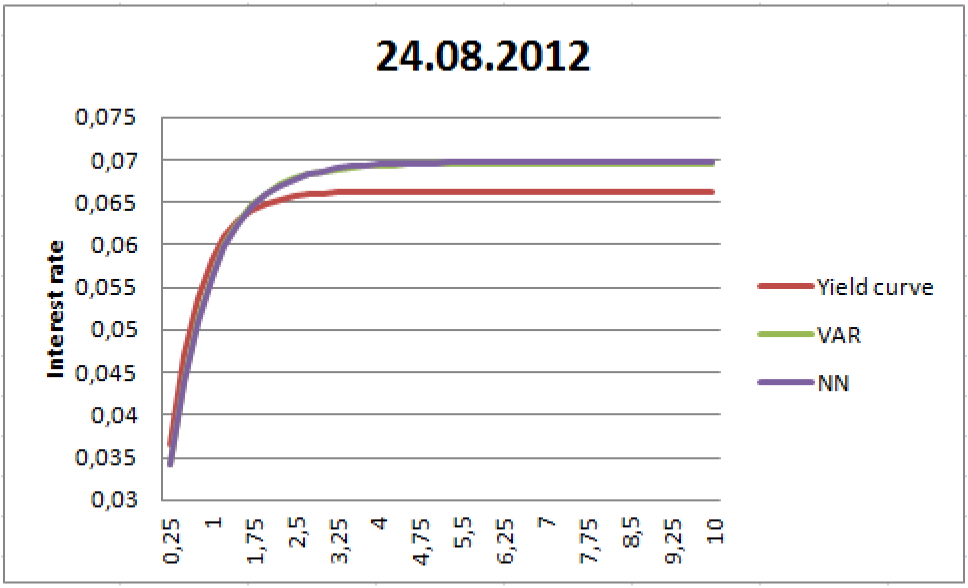

Figure 4 Nelson-Siegel yield curve and predicted yield curves using VAR and NN models on 24.08.2012.

5 CONCLUSION

It is well known that yield curve estimation is of crucial importance for all the participants on financial market and beyond. What is also very important is the ability of forecasting yield curves especially on emerging markets, like Croatian financial market, where a marginal bond trade exists. The main issue in the paper was to investigate the forecasting features of Neural networks and Vector autoregressive models. Empirical research on Croatian financial market shows rather good forecasting capacity of both methods, with slightly better results given by neural networks. The lack of reliability is evident primarily in long term forecasts.

6 REFERENCES

[1] Aljinović, Z., Marasović, B. and Škrabić, B. (2009), “Comparative Analysis of the Stochastic and Parsimonious Interest Rates Models on Croatian Government Market”, *International Journal of Human and Social Sciences*, 4:13, pp. 924-928

[2] Bahovec, V., Erjavec, N. (2009) „Uvod u ekonometrijsku analizu“, Element, Zagreb

[3] Berec R. (2010), “Extraction of Nelson-Siegel Factors from Bond Prices”, Comenius University in Bratislava - Faculty of Mathematics, Physics and Informatics, Master thesis

[4] Cano, C., Correa, R. and Ruiz, L. (2010), “La curva de rendimientos y la toma de decisiones financieras”, *Revista Moneda,* N° 145

[5] Dedi, R., Yoga, A.N., Rahmawati, K.D. (2011) „Forecasting the Indonesian Government Securities Yield Curve Using Neural Networks and Vector Autoregressive Model“, ISI 58th Session , Dublin, August 21-26

[6] Dunis, C.L., Williams, M. (2002) „Modelling and trading the euro/US dollar exchange rate: Do neural networks perform better?“, Journal of Derivatives & Hedge Funds, 8, 3, 211-239

[7] Kladivko K. (2010), “The Czech Treasury Yield Curve from 1999 to the Present”, *The Czech Journal of Economics and Finance (Finance a úver)*, 60(4), pp. 307-335.

[8] Marciniak M. (2006), “Yield Curve Estimation at the National Bank of Poland”, National Bank of Poland, Working paper N° 47

[9] Nelson, C. R. and Siegel, A. F. (1987), “Parsimonious modelling of yield curves”, *Journal of Business*, 60, pp. 474-489

[10] Rohde N. (2007), “Qualitätsvalidierung von Zinsstrukturkurven mit statistischen Modellen”, Humboldt - Universität zu Berlin, Institut für Informatik, Diplomarbeit

[11] StatSoft,http://documentation.statsoft.com/STATISTICAHelp.aspx?path=NonlinearEstimation/NonlinearEstimation/Dialogs/ModelEstimation/ModelEstimationQuickTab [Accessed 23.9.2012.]

[12] Tal, B. (2003) „Background Information on our Neural Network-Based System of Leading Indicators“, CBIC World Markets, Economics & Strategy, September

[13] Täppinen, J. (1998) „Interest rate forecasting with neural networks“, Government Institute for Economic Research, Vatt-Discussion Papers, 170

[14] Teräsvirta, T., van Dijk, D., Medeiros, M.C. (2005) „Linear models, smooth transition autoregressions, and neural networks for forecasting macroeconomic time series: A re-examination“, International Journal of Forecasting, 21, 755-774

[15] Warner, B., Misra, M. (1996) „Understanding Neural Networks as Statistical Tool“, The American Statistician, Vol. 50, No. 4, November

[16] Yao, J.T., Tan, C.L. (2001) „Guidelines for Financial Forecasting with Neural Networks“, Neural Information Processing, Shanghai