

CORVINUS UNIVERSITY OF BUDAPEST
FACULTY OF BUSINESS ADMINISTRATION
DEPARTMENT OF FINANCE

**Forecasting the Hungarian term
structure of interest rates with
econometrics and neural network
methods**

Author: SZENCZI, ROLAND
MSC IN FINANCE
INVESTMENTS SPECIALISATION
2016

Supervisor:
BADICS, MILÁN

Declaration of Authorship

I, SZENCZI, ROLAND, declare that this thesis titled, "Forecasting the Hungarian term structure of interest rates with econometrics and neural network methods" and the work presented in it are my own. I confirm that:

- This work was done wholly or mainly while in candidature for a research degree at this University.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed:

Date:

CORVINUS UNIVERSITY OF BUDAPEST

Abstract

Investments Specialisation
Investments And Corporate Finances Department

Msc in Finance

**Forecasting the Hungarian term structure of interest rates with
econometrics and neural network methods**

by SZENCZI, ROLAND

Using the high modelling performance of the dynamic Nelson-Siegel model it is able to generate a well fitting curve for the Hungarian interest rate points by different maturities. The collection of interest rate curves' factors are the input for the econometrics models and the multilayer artificial neural network based on feed-forward architecture. The non-parametric neural network is able to recognise non-linear connections and hidden patterns, hence it can forecast the whole term structure of interest rates for 5-days and 10-days ahead. The estimated term structure of interest rates can be a good basis for trading.

Contents

Declaration of Authorship	iii
Abstract	v
1 Introduction	1
2 Term structure and its modelling	3
2.1 Importance of interest rate modelling	3
2.1.1 Interest rates in trading	3
2.2 Interest rate models	5
2.2.1 Stochastic models	5
2.2.2 Statistical approaches	7
2.2.3 PCA and Nelson-Siegel model	7
2.2.4 Nelson-Siegel approximation	8
2.2.5 Extended Nelson-Siegel model	10
3 Models for predictions	11
3.1 Random walk model	11
3.2 Autoregressive model	12
3.3 Vector autoregression	13
3.4 Neural Networks	13
3.4.1 Random weights	16
3.4.2 Over fitting and data splitting	16
3.4.3 Select the right neural network	17
3.4.4 Delayed inputs for neural networks	17
3.4.5 Importance of using neural network	18
4 Research methodology	19
4.1 Data source	19
4.1.1 Cubic spline interpolation	19
4.1.2 The analysed data	20
4.2 Exponential decay parameter	26
4.3 Standard Nelson Siegel versus extended model	27
4.3.1 Grid search for Svensson model	27
4.3.2 Results of comparing the standard and extended Nelson-Siegel models	28
4.4 β coefficients of Nelson-Siegel	29
4.5 Analysing of β parameters	30
4.6 Errors in forecast methodologies	33
5 Forecasting of interest rates	35
5.1 Forecasting plan	35
5.2 Stationary and outliers	36
5.3 Autoregressive model	37

5.4	Vector autoregression of beta parameters	38
5.5	Forecasting with neural network	39
5.5.1	Selecting the right architecture	39
5.5.2	Training method of the neural network	39
5.5.3	Data preprocessing for neural network	40
5.5.4	Building the neural network	41
6	Comparing of results	43
6.1	Evaluation of the predictions	43
6.2	Expectations	44
6.3	Results	45
6.4	Comparing the β parameters' forecasts	45
6.4.1	RMSE of β parameters	46
6.4.2	Diebold-Mariano test	47
6.5	Model error	48
6.5.1	Comparing by RMSE after substitution into Nelson-Siegel model	48
6.5.2	Residuals from predicted β variables	50
6.5.3	RMSE by maturities	52
7	Conclusion	53
A	Appendix	55
	Bibliography	59

List of Figures

2.1	Loadings of the three β parameters with different maturities and $\lambda = 0.05$	9
3.1	Simplified architecture of the feedforward neural network. There are connections for sending data forward without feed-backs.	14
3.2	Map about the activation function in neural network	16
4.1	Hungarian interest rates	20
4.2	2001-2003 period of term structures of Hungarian interest rates	21
4.3	2004-2007 period of term structures of Hungarian interest rates	22
4.4	2008-2010 period of term structures of Hungarian interest rates	22
4.5	2010-2016 period of term structures of Hungarian interest rates	23
4.6	Zero coupon points interpolated by spline regression and the 25th and 75th percentiles.	24
4.7	Zero coupon points for the period 2001-2003 interpolated by spline regression and the 25th and 75th percentiles.	24
4.8	Zero coupon points for the period 2004-2007 interpolated by spline regression and the 25th and 75th percentiles.	25
4.9	Zero coupon points for the period 2008-2010 interpolated by spline regression and the 25th and 75th percentiles.	25
4.10	Zero coupon points for the period 2010-2016 interpolated by spline regression and the 25th and 75th percentiles.	26
4.11	Dynamically evolving λ_t parameters	26
4.12	Model comparison by time periods (I: 2001-2003, II: 2004-2007, III: 2008-2009, IV: 2010-2012, V: 2013-2016) and short term maturities	28
4.13	4 different types of the term structure of Hungarian interest rates	30
4.14	Autocorrelation of β_0 variable for 14 different periods	31
4.15	Autocorrelation of β_1 variable for 14 different periods	31
4.16	Autocorrelation of β_2 variable for 14 different periods	32
6.1	RMSE by maturities and models for different horizons	52
A.1	Summarising the basic properties of Nelson-Siegel model's parameters	55
A.2	Standard deviation of interest rate points for the 17 standard maturities. M = Months	56

List of Tables

2.1	Notional outstanding on OTC markets, Source: BIS (2016)	4
4.1	Cross correlations for lag=1	32
4.2	Statistical description of β variables	32
6.1	The RMSE numbers above are multiplied by 1.0^4 . 1-day-ahead forecasts' RMSE by predicted β variables and models	46
6.2	1-day-ahead forecasts' RMSE by predicted β variables and models	46
6.3	5-days-ahead forecasts' RMSE by predicted β variables and models	47
6.4	10-days-ahead forecasts' RMSE by predicted β variables and models	47
6.5	1-day-ahead forecasts' RMSE by periods and models	48
6.6	5-days-ahead forecasts' RMSE by periods and models	49
6.7	10-days-ahead forecasts' RMSE by periods and models	50
6.8	Lowest mean autocorrelations of residuals for short (3 months - 24 months) and long maturities (36 months - 120 months) for lag=1 by periods and models	51
6.9	Worst mean autocorrelations of residuals for short (3 months - 24 months) and long maturities (36 months - 120 months) for lag=1 by periods and models	51
A.1	Mean autocorrelation of 1-day-ahead predictions' residuals for short end maturities (30 months - 120 months) for lag=1 by periods and models	57
A.2	Mean autocorrelation of 1-day-ahead predictions' residuals for long end maturities (3 months - 24 months) for lag=1 by periods and models	57
A.3	Mean autocorrelation of 5-days-ahead predictions' residuals for short end maturities (3 months - 24 months) for lag=1 by periods and models	57
A.4	Mean autocorrelation of 5-days-ahead predictions' residuals for long end maturities (30 months - 120 months) for lag=1 by periods and models	57
A.5	Mean autocorrelation of 10-days-ahead predictions' residuals for short end maturities (3 months - 24 months) for lag=1 by periods and models	57
A.6	Mean autocorrelation of 10-days-ahead predictions' residuals for long end maturities (30 months - 120 months) for lag=1 by periods and models	58

Chapter 1

Introduction

This thesis is focusing on the forecasting of Hungarian term structure of interest rates using dynamic Nelson-Siegel approximation with neural network. This is the first research which is trying to capture the daily Hungarian zero-coupon rates and using neural network to forecast the term structure of interest rate for the coming days.

Forecasting of interest rate points is important for the risk management, derivative instruments pricing, monetary decision makers and investment managers. If we want to discount an expected cash flow or find the fair price for an option we always need the right interest rates thus several economic researches are dedicated to the interest rate modelling techniques. The forecasting methods have been paid particular attention from researchers as well like Reppa's (Reppa, 2009) and Kopányi's researches (Kopányi, 2009). Now I am focusing on the forecasting and use new technologies like the neural network to explore this new field of interest rates estimation.

In the economic studies there are two main approaches to build interest rate models: the no-arbitrage models like model of Hull and White and the affine equilibrium models like Vasicek and Cox Ingersoll Ross. In this research the most popular empirical method will be used: the dynamic Nelson Siegel model which is using a dynamic three dimensional exponential approximation (Diebold and Li, 2006). The level, slope and curvature parameters of the interest rate term structure are able to determine the entire curve according to the empirical results of Diebold and Li (Diebold and Li, 2006) and I am using them to forecast the next day interest rate points. My thesis is based on the assumption that the dynamic Nelson Siegel models parameters are highly autocorrelated and a neural network is able to predict the coming day's interest rate term structure understanding the complex patterns and non-linearity of the change in parameters.

The three factors of Nelson-Siegel approximation will be forecasted by models random walk, autoregressive, vector autoregressive, neural networks based on single-input and multi-input. The previous three methods are popular in both literature and practice (Diebold and Li, 2006), and in addition, I will use a neural network too. I expect that the residuals of the autoregressive and vector autoregressive models will show higher autocorrelation on the different ranges of the time period, than the neural network's residuals and these residuals can provide more information for the proper understanding of the models.

In the consequence of existing cross-correlation it is useful to test a neural network with single input vector - similarly to the autoregressive model - and a neural network with multiple input vectors like in the case of vector-autoregressive model. The cross-correlation can provide relevant additional

information relating to the change of variables. In the multi-inputs comparing case, the neural network should understand better the non-linear connections, but if the relationships between the parameters are linearly, the vector autoregressive model can describe more accurate the movements of Nelson-Siegel model's variables.

In addition, if the cross correlation of the Nelson-Siegel model's three parameters were changing over time, then the autoregressive, vector autoregressive and single-input vector neural network forecasting models must be dominated by the multi parameters-neural network solution. The random walk can be a strong competitor of the neural network, but there could be different forecasting powers on the short and long end of the term structure.

This is the first time that the Hungarian daily interest rate points are estimating with neural networks including dynamic Nelson-Siegel model outputs.

Chapter 2

Term structure and its modelling

2.1 Importance of interest rate modelling

Interest rate is used for discounting future cash flows in bond and swap pricing, also used for the evaluation of companies. The changes in interest rates affect the interest sensitive exposures, inflation, expectation of investors and stock prices (Mankiw, 2014). Consequently, the term structure of interest rates has its own role in every field of economics like in investments, macroeconomics, corporate finances.

The synthetic financial products transform the fluctuation of interest rates into credit risk (Coons, 2015), for this reason there are researches, which suggest the implementing of forecasting methods for term structure of interest rates to minimise risk in banks (Coons, 2015).

Companies pay higher costs on borrowing if the interest rates are rising. Firms with lower yields are not able to lend money with higher costs. The private investors prefer to invest in treasury bonds and sell their stocks, if the interest rate points are high. The increasing stock supply on financial markets are generating a drop in stock prices. The value of projects and the project owner companies decrease because of higher discount factors calculated from rising interest rates. This is a potential macroeconomic scenario, where the interest rates are increasing (Mankiw, 2014). The consumers reduce their demand and try to invest in bonds, but this process has a negative effect on the aggregated demand. If the aggregated demand is falling, the economic growth slows down or turn to negative and results recession (Mankiw, 2014).

The falling interest rates makes the loans more cheaper and this can be favourable for the investors. The cheap credit can generate leverage in investments, which is risky, but theoretically the consumers demand increases, because it worth to consume rather than invest into the low interest rates (Mankiw, 2014).

The above listed potential events describe the importance of analysing the term structure of interest rates. We cannot overlook the fact that the change in yield curve has direct impact on our wealth.

2.1.1 Interest rates in trading

According to statistics of Bank of International Settlements the interest rate derivatives (IRD) market is the largest OTC derivatives market in the world (BIS, 2016).

Products	Notional outstanding (BN\$)
Total interest rate contracts	434,740
Total foreign exchange contracts	74,519
OTC, credit default swaps	14,596
Total equity-linked contracts	7,545

TABLE 2.1: Notional outstanding on OTC markets, Source: BIS (2016)

The total notional outstanding of the interest rate contracts with 434.740 trillion dollars represent a huge number as compared to the smaller foreign exchange and CDS markets as the table 2.1 shows it.

There are variable reasons why a company or an institutional investor uses an IRD. The institutional investors trade IRD to hedge their interest rate related risks generated by other trades or speculate for unexpected movements, while the entrepreneur's world tries to reduce their risk related to change of interest rate to protect their cash flow (*Credit Suisse - Debt capital and interim financing to hedge interest rate risks*. 2016).

Forecasting the right IR term structure represents a big added value for different companies and its different departments, too. The hedge funds, pension funds, insurance companies, banks and its Asset-Liabilities management and risk management are able to directly or indirectly use the predictions for their daily work (SHOOK, 2013).

If the fixed income investors expect a fall of interest rates they try enter a floating-for-fixed interest rate swap and waiting for the decrease of interest rates level to pay the lower amount and receive the bigger, fixed cash (Singh, 2011). This shows the importance of the forecasting of interest rate term structures. The right prediction of the change in structure's level can be a proper basis for trading. The trading strategy can work by the right forecasting of the change of IR term structure's shapes (Group, 2013).

The market of interest rates derivatives reduced in the previous months because of the clearing and portfolio compression activities at the end of 2015 (ISDA, 2016). The BIS reported a drop of 14.0 % in the notional outstanding of interest rate derivatives products in the first half of 2015. The notional outstanding decreased from \$ 505.4 trillion to \$ 434.7 trillion in this time (ISDA, 2016). In this respect, it should be recalled that while there is a strong decreasing in the outstanding of IRD products, the IRD trading activity (on a gross basis without any netting out of clearing and compression) increased by 4.7 % in the same time (ISDA, 2016).

The above described effect can be caused by the new regulation rules like the requirements related to leverage ratio in Basel III (ISDA, 2016). The new compression technologies could likely have significant effects on the decreasing of notional outstanding of IRD products.

The IRD market has its importance despite the fact that the notional numbers are falling. The IRD trading activity is not decreased which is a good reason to research the working of interest rate term structures.

Fabozzi wrote a study about implementing Nelson-Siegel for trading with IRD product. He forecasted the parameters of Nelson-Siegel model and traded by butterfly swaps and calculated that betting on β_0 produces a Sharpe ratio approximately equal to 2 (Fabozzi, Martellini, and Priaulet, 2005), hence Fabozzi concludes that the forecasting the right shape or the

level of term structure is a good tool to build a working trading strategy. It is not the purpose of my research to find trading strategies, I just want to build an accurate forecasting methodology for Hungarian interest rates term structure, which has a relevance in the trading.

2.2 Interest rate models

The interest rate models can be categorised into three main groups: the stochastic models, the statistical models and the economical one. All of them try to capture the relation between the interest rates and the maturities, but the methodologies are based on different approaches. The models have different objectives which is very important for the model selection step. There are models which are implementing clear rules to ensure the arbitrage-free fitting like HJM model, but another models do not pay particular attention to the arbitrage-free curves like Nelson-Siegel.

2.2.1 Stochastic models

The family of stochastic models has two main categories: the affine equilibrium and the no-arbitrage models. The affine equilibrium models focusing on the change of short rate. If the calculation is implemented in risk neutral measure:

$$dX(t) = [\mu_0(t) + \mu_1(t)]dt + \sqrt{\sigma_0^2(t) + \sigma_1^2(t)}X(t)dW(t) \quad (2.1)$$

The affine models satisfy a stochastic differential equation (2.1), where an affine drift and an affine square of the diffusion coefficient are in the function to generate short rates. The $dW(t)$ is the Wiener process in the equation 2.1, which is a function of time (t) and continuous with probability equal to 1. The independent random variables of the Wiener process are following a normal distribution $N(0, \Delta t)$. If the process is the Geometrical Brownian Motion the $\mu_0 = \sigma_0^2 = 0$ equation is true (Pacati, 2012).

The one-factor short rate models can be described by a nested stochastic differential equation:

$$dr = (\alpha + \beta r)dt + \rho r^\gamma dZ \quad (2.2)$$

where α, β, ρ and γ are representing constant values and dZ is the standard Wiener process. The Vasicek's model is correspond to $\gamma = 0$, and Cox-Igeroll-Rox's model is a modification of the 2.2 with $\gamma = 1/2$ parameter. Merton's stochastic model can be implemented by $\gamma = 0$ and $\beta = 0$ (2.3).

$$dr_t = \mu dt + \sigma dZ \quad (2.3)$$

Vasicek's model describes the short rate based on the Ornstein–Uhlenbeck stochastic process (Vasicek, 1977) and it is the most popular model among interest rate's models (Orduna, Lin, and Larochelle, 2015).

$$dr_t = \alpha(\gamma - r_t)dt + \sigma dW_t \quad (2.4)$$

where W_t is a Wiener process in Q-dynamics, σ is the volatility of the interest rate, α is the speed of reversion, β is the mean. σ is very important because it defines the amplitude of the randomness in the process. The important part of the Vasicek's model is the implementing of Ornstein–Uhlenbeck stochastic process because this gives the theoretical background to the modelling. Ornstein–Uhlenbeck's models based on the mean-reverting function which means that the stochastic process generates numbers tending to the average over time (Finch, 2005).

$$dX_t = -\beta(X_t - \alpha)dt + \sigma dW_t \quad (2.5)$$

In equation 2.5 the β parameter is the measure of the function's reaction to the deviations, α is the mean and the σ is the volatility. The importance of the mean-reverting or Ornstein–Uhlenbeck stochastic process lies in the temporal dependency of interest rates. The too high interest rates should have a negative trend and the too low interest rates will turn to the positive direction and tend to the average. This reversion process has been observed in practice (Martellini, Priaulet, and Priaulet, 2003).

There are no-arbitrage models as well which are focusing on the bond pricing. The Vasicek and Cox Ingersoll Ross models are not the best tools to price bonds because the estimated price can differ from the observable market price. This is the reason why no-arbitrage models' development has been started.

The first model was developed by Thomas Ho and Sang Bin Lee in 1986.

$$dr_t = \Theta_t dt + \sigma dW_t \quad (2.6)$$

The Ho Lee model can describe the change of interest rates by a binomial tree and using one single factor of uncertainty. However, it does not contain mean reversion and furthermore it can produce negative interest rates too. The nominal negative interest rate was not considered possible previously, hence this model was not preferred in the practice. The constant σ is not appropriate for the empirical experiences as well (Martellini, Priaulet, and Priaulet, 2003).

Heath, Jarrow and Morton modified the Ho Lee model to implement a n-dimensional uncertainty into the model. They are starting the calculation with the currently observable yield curve which is the basis for the bond pricing. The Heath-Jarrow-Morton model describe the interest rates by the change in bond prices dependent from time- and maturity- variant $\mu(t, T)$ and $\sigma(t, T)$:

$$\frac{dB(t, T)}{B(t, T)} = \mu(t, T)dt + \sigma(t, T)dW_t \quad (2.7)$$

There is a big difficulty in the Heath-Jarrow-Morton model: the compute-intensive methodology, because of the path-dependency (Martellini, Priaulet, and Priaulet, 2003). This problem derives from the non-markovian process property. If the sigma can change in every n-th point and differs by the given maturities it is very computable to get the interest rate curve.

The models above are representing the most popular stochastic solutions for modelling the term structure of interest rates, but there are critics on them. Vasicek model is not able to fit well the short end of the term structure (Balter, Pelsser, and Schotman, 2014), thus it is not the best model for

trading activity. The Heath-Jarrow-Morton model can produce the same term structure from different parameters (Schumacher, 2009), which is not consistent for modelling the dynamics of the model's input variables. For this reason I think, that new approaches must be analysed.

2.2.2 Statistical approaches

To capture the dynamic of the interest rate term structure it is required to decrease the number of parameters in the model and explain the changes to the greatest possible extent. This is the basics of every model because the calculation is costly and time-consuming. The principal component analysis (PCA) is implemented in many researches (Martellini, Priaulet, and Priaulet, 2003).

PCA is a method which helps to find the dominant axes to describe a sample. The axes or dimensions can not be correlated with each other, like the 2 dimensional coordinate system where x and y are independent axes.

PCA is a simple, non-parametric method to reduce dimensionality and find perfectly uncorrelated dimensions to build a model. PCA has been implemented for different economic researches in the previous years, like analysing of stock market data, exchange rates or commodity markets (Suryanarayana and Mistry, 2016).

One of the biggest advantage of the PCA is the ability to find important dimensions by its role in describing the variables of the analysed sample. The number of components can be reduced by PCA because in many examples there is a good chance that a large percentage of the variance can be described by a few variables (Suryanarayana and Mistry, 2016). The reduction of dimensionality can help to compact the relevant information into the minimum of numbers and later, when I want to predict the variables, the computation time can be shorter. However, it is difficult to define the number of sufficient variables, if there were three factors which could explain 98 % of the variance, but adding one more factor does not increase significantly the explanatory power, while the environment of the developing - like the trading activity - requires a high accuracy. This consideration is important for the trading activity.

2.2.3 PCA and Nelson-Siegel model

Nelson-Siegel is not arbitrage-free (Bjork and Christensen, 1999), but it is possible to fit it like an arbitrage-free model, because Ken Nyholm and his colleagues demonstrated that the NS can produce interest rate term structure with arbitrage-free points on 95 % of confidential level (Coroneo, Ken Nyholm, and Vidova-Koleva, 2008).

Nelson-Siegel's three parameters are able to describe the 99 % of the total variance (Afonso and Martins, 2010), which is proper for my research. If the PCA was able to produce equal to or bigger than 99 % result, it would be not enough for a proper decision. Modelling the term structure of interest rates and forecasting it, are two different methods, hence an economic point of view is required for better decision.

Mönch proved that the models based on principal component analysis outperforms the yield-based models, but he confirms the high performance

of dynamic Nelson-Siegel model in his research especially for longer horizons (Mönch, 2005). The Nelson-Siegel model, despite the fact that it is not a proper model in statistical term, can forecast better on out-of-sample and has a good ability to forecast long term (Mönch, 2005).

Theoretically the cross correlation persists time to time in dynamic Nelson-Siegel model (Diebold and Li, 2006), but I prefer this property of the model in contrast of the total uncorrelated dimensions of PCA. The cross-correlation can be useful for VAR model and those version of neural network which is developed for receiving more input parameters.

2.2.4 Nelson-Siegel approximation

A wide variety of models are available to capture interest rates. The Nelson-Siegel model is neither arbitrage-free and nor affine model, but very popular in the practice of monetary decision makers and governments (Aljinovic, Poklepovic, and Katalinic, 2012). The model is fitting three factors to describe the entire term structure of the interest rate curve:

$$p_t(\tau) = \beta_{0t} + \beta_{1t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) + \beta_{2t} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) \quad (2.8)$$

where $\beta_0 > 0$ and $\beta_1 + \beta_2 > 0$ and $\tau, t > 0$ (Nelson and Siegel, 1987).

β_0 is the level of the term structure. If this parameter would be increasing then the whole curve – with every point – rise by a quantity that is equivalent to the quantity of change in β_0 . Factor β_0 is independent from other β parameters in the model, but in the practice we have to measure the cross correlation of factors. The power of this cross correlation can change over time; there can be stronger ranges on the time series and weaker periods too. The financial crisis in 2007/2008 was supposed to be able to modify the cross correlation, but there could have been other shocks on the market as well.

The equation shows that τ and λ are playing an important role in the model because they affect the β_1 and β_2 factors. These two variables combined constitute the loadings for the β factors, thus the factor loadings are weights in the equation. The β_0 has a special loading which is a constant 1. The equation 2.9 shows the loadings for the different β factors.

$$\underbrace{\begin{bmatrix} 1 & \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} \right) & \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) \end{bmatrix}}_{\text{Loadings}} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix} = p_t(\tau) \quad (2.9)$$

β_1 parameter is measuring the slope of the curve. We can accept the existence of this factor if we imagine an interest rate curve which must not be a horizontal line like in the basic financial examples. If we lend our money to a counterparty who generates risk to our portfolio, we will increase the expected premium on longer maturities. This simple expectation can prove the existence of a curve with positive slope. Of course there is negative slope when the expected inflation decreasing is huge. The β_1 factor does not independently affect the points as the previous factor β_0 , because it is linked to the λ_t and the τ factor in the loading.

Parameter β_2 is the curvature of the interest rate curve. The curvature can symbolise the changing of the slope. If the slope is the velocity in the world of physics, the curvature is the acceleration. This factor affects the interest rate curve by the λ_t and the τ factor similar to the β_1 , but the $e^{-\lambda_t \tau}$ part decreases a bit the effect of the β_2 factor.

τ is the maturity in the equation and λ is the decay parameter. If the decay parameter is big the curve will fit better on the short end and if it is small the model will have smaller residuals on the long end. The λ parameter maximises the β_2 value therefore $1/\lambda^*$ is the mid-term point of the curve. This methodology is presented on the figure 2.1, where I chose $\lambda_t = 0.05 = \lambda^*$. The figure is showing that $1/\lambda^* = 20$, thus the β_2 loading is maximised at the 20th month.

The connection between the β loadings with τ is shown in figure 2.1. This picture can help to understand the different sensitivities of the variant maturities for the β variables. The level of loading represents the influence of the given β variable in the Nelson-Siegel model for a given maturity. The level has its importance in the model on every maturity, but the β_1 loading is falling dramatically in the direction of bigger maturities. Consequently, the sensitivity of the model to β_1 parameters' error decreases by longer maturities. The β_2 loading reaches its maximum at middle-term point and it maximise its weight for β_2 variable in the model.

λ is a trade-off between fitting the long and short end, therefore, if I want to fit my model for both ends I have to find an optimal λ . This optimisation is discussed in more detail in the next section.

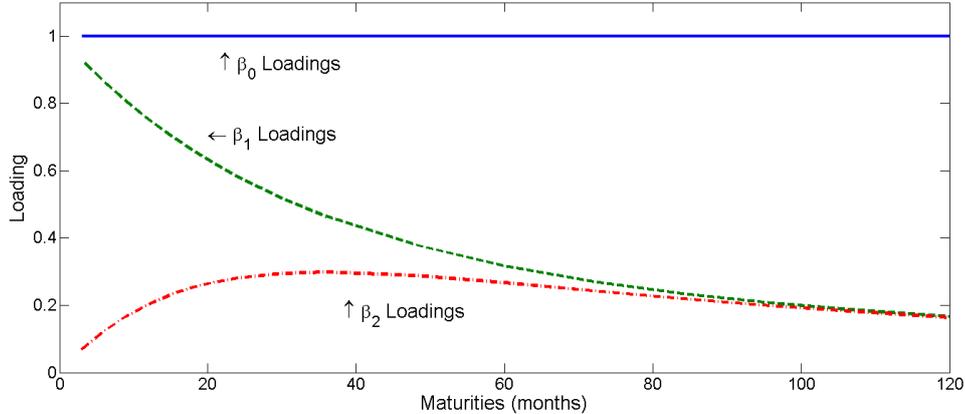


FIGURE 2.1: Loadings of the three β parameters with different maturities and $\lambda = 0.05$

The represented level of the variables on the figure 2.1 shows not only the maximisation or change of different parameters, it reflects the weight for given β . The prediction's error of β_1 has double effect on the first maturity than the 50th, while β_0 has the same weight for every maturity. β_2 can highly influence the results at the middle-term point and on the long end of the term structure it has the same effect as β_1 .

2.2.5 Extended Nelson-Siegel model

Diebold and Li implemented a Nelson Siegel approximation to model the term structure of zero coupon yield (Diebold and Li, 2006), but in the economic literature there is another method which is frequently referred to as the Svensson extension of Nelson Siegel model (Gilli, Grosse, and Schumann, 2010). If the points of zero coupon yield are creating a two-humped curve, it is recommended to fit Nelson Siegel Svensson model and test its RMSE comparing to the standard model. Many central banks in the European Union use the Svensson extension for modelling term structure, such as in France, Finland, Spain and United Kingdom (Aljinovic, Poklepovic, and Katalinic, 2012). Even the European Central Bank publishes the daily interest rate term structures approximated by Nelson Siegel Svensson model (Coroneo, Nyholm, and Vidova-Koleva, 2008).

β_2 is the parameter where the curve has a hump which is a global maximum of the curve. In the Svensson model there is another parameter to maximise the curvature, hence there will be two humps and the β_2 is not necessary the global maximum in this case. The new parameter, β_3 is able to maximise the curvature and fit better the long end. If the long end is well fit the short end can be modelled more accurately too, because the short and long end of the curve are dependent from each other during the term structure fitting method. If the long end curve has bigger RMSE, and the optimisation method tries to modify the level or slope variable to rebalance the entire curve and minimise the RMSE to detriment of short fit's goodness.

I implement a dynamic version of the Svensson model similar to Diebold and Li to model the term structure of interest rates by an extended four factor model. I think it is required to compare the fit ability of the two models because the Hungarian interest rate term structure had significant changes in the last years hence there can exist different periods with different corresponding models. The model with lowest RMSE can be the selected one for the research.

$$p_t(\tau) = \beta_{0t} + \beta_{1t} \left(\frac{1 - e^{-\lambda_{1t}\tau}}{\lambda_{1t}\tau} \right) + \beta_{2t} \left(\frac{1 - e^{-\lambda_{1t}\tau}}{\lambda_{1t}\tau} - e^{-\lambda_{1t}\tau} \right) + \beta_{3t} \left(\frac{1 - e^{-\lambda_{2t}\tau}}{\lambda_{2t}\tau} - e^{-\lambda_{2t}\tau} \right) \quad (2.10)$$

Chapter 3

Models for predictions

3.1 Random walk model

The random walk model allows an object to move with equal probability in any direction (Grinstead and Snell, 1997). The term of “random walk” is originated from Karl Pearson’s research (Rycrof, 2005) where he found that the basic forecasting models require an idealistic system and this idealistic system should be based on the random walk (Pearson, 1906).

M. G. Kendall wrote his research in 1953 about the analysis of economic time series where he focused on the change of stock prices on financial markets and tried to find the best fitting model for prediction. He assumed that in each period the variable takes a random step forward from its previous value and the steps are independently and identically distributed in size. Hence he implemented random walk model for stock price prediction (Kendall and Hill, 1953).

The random walk model is very important in finance because this is the basic stochastic model for predictions (Fama, 1995). The random walk assumes that the best estimate of tomorrow’s price is today’s price on stock markets, which is the fundament of efficient market hypothesis, too (Fama, 1969).

$$\tilde{y}_{n+k} = y_n + \epsilon_{n+k} \quad (3.1)$$

The equation 3.1 shows the predicted \tilde{y} for k -steps ahead from point n . The ϵ_{n+k} represents white noise with mean zero and variance σ^2 . This theory about random migration implies the hypothesis of martingale but martingale follows a geometric random walk (Hong, 2009). This connection is very important in the world of investment analysis, because Samuelson proved that the stock prices follow martingale-process (Samuelson, 1965).

The stock prices generate an unpredictable path and the probabilities of moving up and down are the same, for this reason the expected value of tomorrow’s prices and the today’s prices are equal. The multi-day forecast of random walk shows a horizontal line which is originated from the most recent available value and is not calculated from the historical data. The mean reversion model is similar, but it applies the historical time series for prediction (Fama, 1995).

Random walk is important not only for stock prices, but also for interest rates. The predictions of interest rates, based on the expectations hypothesis and the random walk produces the same results which is empirically proven by a research published in the journal of the European Central Bank (Guidolin and Thornton, 2008). If the short rates are moving by random

walk, the expectations hypothesis implies that the long term interest rates change by random walk as well (Mishkin, 2007).

Pooter proved empirically that the best model for estimating the three-factor Nelson-Siegel's parameters is the random walk (Pooter, 2007). Pooter built different models like AR, VAR and random walk to predict the zero-coupon yield curve and forecasted it for 3, 6 and 12 months ahead.

In the research of North Carolina State University, it is pointed out that the random walk model has a better forecasting power to predict interest rates 6-month-ahead as compared to models developed by independent economists (Pearce, 2005). The only reason why these economists use wrong models and ignore the problem of significant errors in forecasts is that they are motivated to develop their own model and create a story besides the predictions. The random walk model's forecasts do not have any story and it must be very hard to sell it for consumers who want to feel the added value of professional forecasting (Pearce, 2005).

Accepting the predictive performance of the random walk I use it to forecast the β parameters of the Nelson-Siegel model. The random walk was a very strong competitor of the neural network in the research of Malinska based on Diebold and Li's model.

3.2 Autoregressive model

In the autoregressive model, the linear combination of past values predicts the next element of the time series. This simple solution is created for linear modelling of time series, hence it can be relevant to forecast β parameters. Previously I mentioned that the random walk is a very strong competitor in the prediction power comparisons, thus I highlight the advantages of the autoregressive models and the reason why is it can be a relevant model in this research.

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t \quad (3.2)$$

The equation 3.2 shows a discrete-time autoregressive model which has an order (p) and a ϕ parameter which is constant and it is optimised to fit well the model. There is no standard method to define the right p -order for an autoregressive model but the Akaike Information Criterion can help to find it by some iteration. The constant (c) value is a bias, similar to the linear regression's constant parameter. e_t symbolises the error which should be a white noise for the proper modelling (Kunst, 2004). The analysis of e_t can help a lot to understand the modelling sample. If I implement an autoregressive model with $\phi = 0$ parameter, the result (y_t) is a white noise. If $\phi = 1$ and $c = 0$ in the AR model, then y_t is equivalent to a random walk because the direction of change is not predictable from the current level of y_t .

The autoregressive model is a very popular tool for time series analysis and can be very fast by selecting the least-squares regression for fitting method (Kunst, 2004). The random walk model assumes that the markets are driven by unpredictable, random events, but the autoregressive model accepts that there are factors which can be detected behind the observed movements. The monetary decision makers or the government can influence the interest rates - intentionally or unintentionally - hence there should

be factors which can help to predict the next value. The AR can help to catch basic patterns and trend on achieving a higher performance in prediction than random walk, and it has relevance to use it for prediction of term structure of interest rates (Aas and Dimakos, 2004).

3.3 Vector autoregression

The vector autoregression model (VAR) is a generalised univariate autoregressive model created to predict a set of variables. The VAR model is one of the best solutions for financial and economical time series (Jiahui Wang, 2006). The multivariate time series need an autoregressive model to forecast values and the VAR is developed for this task. Each variable is modelled by the linear combination of all previous variables including its own lagged values of the dependent variable. The variables in the model influence each other equally and symmetrically (Jiahui Wang, 2006).

$$y_{t,1} = \alpha_1 + \phi_{11}y_{t-1,1} + \phi_{12}y_{t-1,2} + \phi_{13}y_{t-1,3} + e_{t,1} \quad (3.3)$$

$$y_{t,2} = \alpha_2 + \phi_{21}y_{t-1,1} + \phi_{22}y_{t-1,2} + \phi_{23}y_{t-1,3} + e_{t,2} \quad (3.4)$$

$$y_{t,3} = \alpha_3 + \phi_{31}y_{t-1,1} + \phi_{32}y_{t-1,2} + \phi_{33}y_{t-1,3} + e_{t,3} \quad (3.5)$$

The equations 3.3, 3.4 and 3.5 show the VAR(p) model where $p = 1$. The constant value for the regression in this equation is the α , which is optimised only for one variable, and independently each predicted $y_{t,x}$ has its own α_x .

The Nelson-Siegel model has three different β parameters, where the cross correlation can help the prediction. In economics the cross correlation of factors in a model are not absolutely wrong, because the forecasting performance is being in the focus, not the principal component analysis. From this point of view a VAR model can be a relevant tool for prediction on time series in this research. Another advantage of the VAR model is the ability to analyse the impacts of innovations and shocks (Jiahui Wang, 2006), thus the VAR is used for structural interference and policy analysis in macroeconomics (Stock and Watson, 2001).

Vector autoregressive models are able to capture the stochastic trends as well. Theoretically the vector autoregressive model was developed for stationary time series but later in the 1980s there was a big surprise that the stochastic trends started to play an important role in economic time series, and the vector vector autoregressive model's reliability was questioned (Luetkepohl, 2011). Engle, Granger and Johansen proved that the stochastic trends can be modelled by vector autoregressive model, too (Luetkepohl, 2011), hence vector autoregression is a reliable and important model in economics.

3.4 Neural Networks

Artificial neural network is a machine learning solution inspired by the model of human central nervous system (Hai-Jew, 2014). In statistics there

two requirement that a system could be called as neural network:

1. Capability to generate weights like numerical parameters that are adjusted by a learning algorithm
2. Implementing of non-linear approximating methods for the inputs.

(Hai-Jew, 2014)

The multilayer neural network contains three different parts: the input-, hidden- and the output-layer. The hidden layer can be multiplied but there are very rare cases when it is really needed in practice (Krasnopolsky, 2006). The input layer contains the input nodes which are able to receive the numbers from the user and send forward to the hidden layer. The hidden layer contains nodes to connect them in the functioning of the network. The nodes are able to create connections with different nodes and are able to exchange information, and push weighted number forward to the next node (figure 3.1). The output layer contains nodes set up by the user to receive the result from the neural network (Patricia Melin, 2015).

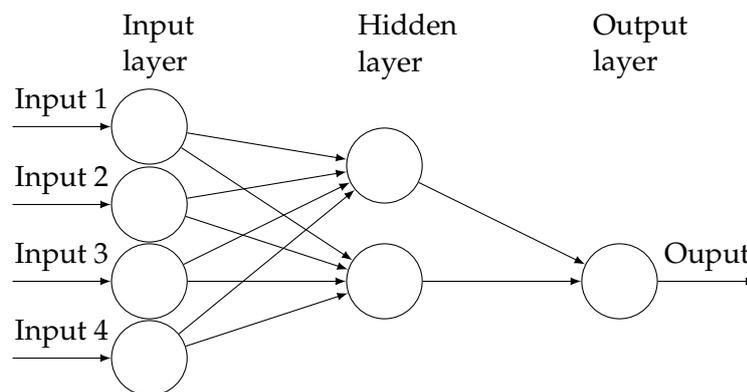


FIGURE 3.1: Simplified architecture of the feedforward neural network. There are connections for sending data forward without feedbacks.

Neural networks are working without previous knowledge about the output or the connection of the input parameters. This is a learning system, where every running of the neural network represents a totally new discovery process without memory. There are no in-built functions for special cases like measure the correlation between different variables, or help to remove unnecessary parameters by principal component analysis. The machine learning based on the recognition of patterns which can be discovered on time series or cross sectional data too. There are events which can recur on the markets and they can draw similar or the same patterns on the time series of interest rate term structure factors. These patterns are able to be recognised by a learning system like neural network and make forecasts for the coming day.

The neural networks outperform the econometric models in forecasting, because the artificial neural networks are able to work with noisy and non linear data (Bajracharya, 2011). The only advantage of econometric models in forecasting is the better understanding of the predictions (Bajracharya,

2011). The econometric models can analyse the temporary relationships between variables, but this is not equal to the prediction ability (Bajracharya, 2011).

Each neuron in a neural network has the same function, usually a sigmoid function. Previously I mentioned that there are no prepared statistical tools or economical knowledge built in the network, hence there must be universal sensors which are able to recognise the patterns. These sensors are the nodes – or neurons –, where every node represents a function with a basic calculation.

The sigma function – sigmoid – is the activation function which is included in the neuron. The output of neuron is referred to as activation and every activation can be an input for another neuron too. The equation 3.6 shows the sigmoid activation function, where the a_{jm} neuron is the m th node on the layer j receive a message from the layer i . The arriving input information (S_{jm}) is the aggregation of the previous activation functions ($x = 0, 1...n$) and the weights between i and j layers (w_{ij_x}).

$$a_{jm} = \frac{1}{1 + e^{-S_{jm}}} \quad S_{jm} = \sum_{x=0}^n w_{ij_x} a_{i_x} \quad (3.6)$$

The sigmoid function can activate the data and send forward it to the next layer, where the data multiplied by weights (figure 3.2). This process is not satisfactory for a complete calculation, because there are cases, when the result of the sigmoid should be shifted horizontally to positive or negative direction. This is precisely why there are biases in the neural networks. The bias is representing a constant value and has a weight similar to the neurons. The value of its weight is generated by the learning algorithm, because the bias functions in the same way like a neuron from the point of view of training method. There is constant value in the linear regression too, and the bias has the same role in neural networks too. The constant value of bias can help to get better fit from the neural network by shifting results.

There are critics on the neural network that it is a black box which is not understandable for the users (Rojas, 2013). I can accept the truth that it would be very hard to present the error minimising algorithm steps and simultaneously analyse the partial results of the neural network for every epoch - calculation of new weights - to be sure that the neural network is not cheating or wrong. Nonetheless, it is better to understand the working of the neural network and analyse its architecture, because it can help to find the right implementation environment where this system performs reliable results. Neural network is an alternative solution to test it for non-linear problems, but it is not a universal tool in spite of universal neurons.

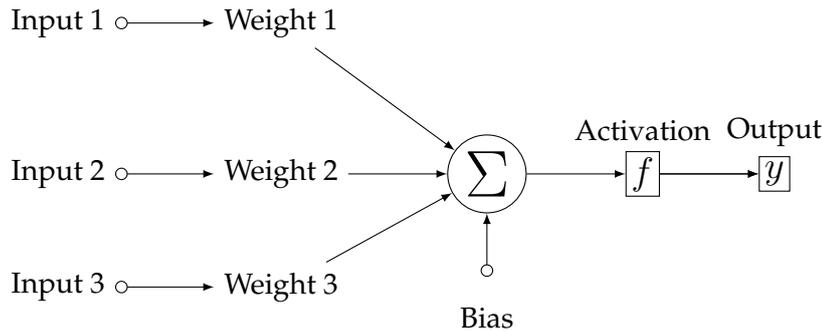


FIGURE 3.2: Map about the activation function in neural network

3.4.1 Random weights

When we start a neural network, the learning algorithm generates random values for the weights. This strategy is based on the error optimisation problem, where the random initial numbers can guarantee that the error surface will change after every running of neural network and the optimisation process will not get in stuck. The learning algorithm tries to modify the weights to minimise the error, but the initial random numbers determine the result. Every running of a neural network gives different results for the same input because of the random initial weights. To manage this problem, the neural network should be run more times and calculate the mean of the results. The mean should converge to the real expected output, but there is no real guarantee that the global minimum is available by the given learning algorithm.

3.4.2 Over fitting and data splitting

Neural network solutions has a big disadvantage: the over fitting. The learning algorithm finds the best weights for the neurons and the network will be not able to focus on the trend of time series, but uses its memory to predict the coming day's number. This is a very relevant question in this thesis, because the patterns an trend are of equal importance. If I have a time series of $train = [0, 1, 0, 1, 0, 1]$ and I try to predict the seventh number by linear regression it must be 0.8, but if I recognise the pattern, I will forecast 0. The difference is huge between the two logic and if there are no other information about the rule, we, human, are not able to decide which method would be better to forecast this special series.

To avoid the effect of over training, the input data shall be allocated into three distinct subsets (Tetko, Livingstone, and Luik, 1995). If we want to find the right forecast for the previous example, it is required to check another series to recognise the rule. The example showed only the training subset, the first subset of the required data for neural network's building process. The second series is called to test subset, which helps to calculate the error and understand the logic of the trend or patterns. If I would show $test = [0, 1, 0]$ for the previous example, it will be clear that the right pattern shows a repeating 0,1 values after each other. The right prediction for the seventh number must be 0 in this case, and the neural network can

recognise the right prediction logic. The third split of data is the validation subset, which helps to select the right model by calculating the predictive power on an out-of-sample (Kaastra and Boyd, 1995). The model with bigger prediction performance is the result which will be presented for the user. If the neural network receive a validation subset with data $validation = [0, 1]$, the algorithm can try its forecasting models and calculate the performance. By the three distinct subsets the network can recognise the logic in the sample, hence the over fitting issue is avoided.

The ratio between the three subsets is usually 80% for the training set, 10% for the test and 10% is allocated for the validation sample. There is no best practice in the literature which would be able to provide a guarantee to find the optimal ratio, hence I selected the 8:1:1 version. The elements of subsets could be selected randomly too, but this would be bad for time series analysis. The autocorrelation which could help to find the trend and the patterns described by the moving of Nelson-Siegel's parameters are very important for the training, hence the subset is allocated by maintaining the continuity of the time series' elements.

3.4.3 Select the right neural network

There are two main types of neural networks: the feed-forward and feedback artificial neural networks. The neurons of a given layer in a feed-forward neural networks are always connected with the previous and the next layer's nodes, but the nodes on the same layer do not have any connection. The activation functions send information only forward and there is no opportunity to get feedback from another neuron.

The recurrent neural network's nodes can receive message from the layer ahead, hence there are feedbacks in the system. This feedback solution provides an inner memory for the system. The disadvantage of this network is related to the scalable, because this solution is not proper for big data.

3.4.4 Delayed inputs for neural networks

The focused time-delay neural network falling within the category feedforward neural networks is created especially for analysis of time series. The first application of the focused time-delay neural networks was the speech recognition, where the researchers created a three layer neural network to find the same phonemes in different speeches (Waibel et al., 1989). This pattern finding solutions can be useful for the financial world too, if the market produces similar movements. If an algorithm can recognise the same phoneme in different dialects and accents, this solution can be proper for financial problems too.

Technically the focused Time-delay neural network is using the same architect like the simple neural networks, the only difference is the method of receiving the inputs on the hidden layer. The hidden layer receives the information at different points in time, hence it can learn the connection between the delayed elements. The effect of delay performs a memory inside the neural network, which can help to recognise the previous numbers and

patterns (Vries and Principe, 1990). This methodology is similar to the autoregressive model, thus the focused Time-delay neural network can be a relevant competitor for the performance comparing.

The non-linear autoregressive neural network model based on the same concept like the focused Time-delay neural network, it uses a time delay for the input vector, but it is better for multi-variable inputs (Junior and Barreto, 2009) and able to model dynamic state spaces (Sherwood, Derakhshani, and Guess, 2008). For this reason, I implement non-linear autoregressive neural network for this research representing the forecasting power of neural networks.

3.4.5 Importance of using neural network

The application of artificial neural networks is extremely diverse, because there are lots of industries where the researchers tried to implement machine learning techniques for complex problems. The military, physicists, medical researchers and the financial world use neural networks for their daily work. The construction of a neural network can be motivated by an optimisation problem, complexity of data mining, robust pattern detection, data compression or signal filtering (Maren, Harston, and Pap, 2014).

The increasing amount of data points out that the neural networks are relevant to read and handle effectively big data. The faster processors and bigger memories in computers can help for the students at home or at university to build their own neural networks for special problems. The software packages like MATLAB, SPSS gives standardised neural network solutions for basic and advanced researches too.

The first relevant application of neural network was presented in 1960 by B. Widrow, who built an adaptive signal filtering solutions. The biggest advantage of this network was based on fast and easy implementation (Maren, Harston, and Pap, 2014). This was the event, when the first step has been made that the neural network will be an everyday tool for researches. Werbos, Parker and Rumelhart presented the first version of their backpropagation neural network in 1974. This net was able to recognise patterns, remove noise from signals, segment images and signals. The first successful speech recognition happened in 1987, when Tank and Hopfield developed its time-delay neural network (Maren, Harston, and Pap, 2014).

The neural network can be used in many segment of the financial world like in retail and investment banks or at hedge funds.

The retail banking sector implement neural network solutions to automate their credit scoring process and minimise the credit risk by learning algorithms. Basel III regulation accepts the sophisticated models like neural networks too, hence the biggest banks try to develop their models for better risk management (Pacelli and Azzollini, 2011). The operation risk can also benefit from machine learning, because the pattern of frauds, both inside and outside, can be detected by neural networks. The credit card frauds are detectable by the different consumption behaviour of the fraudster and the real owner of the card, because the neural network can analyse and identify the significant change in the patterns (Patidar and Sharma, 2011).

Chapter 4

Research methodology

4.1 Data source

The thesis is based on the Hungarian zero coupon interest rates from January 2001 to 22nd February 2016. This time period is long enough to represent the changes of bonds with a 10-year old maturity too.

The Government Debt Management Agency collects and publishes the zero coupon rates every day. I built a C# script to download the zero coupon time series from January 2012 because they are not zipped on the website of the Government Debt Management Agency and the downloading manually the daily data would have been too slow. The downloaded zipped archives and the new data collection generated by my C# script are merged in MATLAB in one dataset.

There are zero coupon points for 3783 days in the analysis. I selected 17 maturities for the research. Selecting maturities is important for the research, first, because eases the calculation processes, and second, because helps focusing on the trading activity. Fixed maturities are created by months and the nearby interest rate points were interpolated to the right pool, where an expected maturity point was not traded.

The selected maturities¹ are corresponding to Diebold and Li's researches, hence my results can be compared with the original and the cited articles as well. The maturities defined by Diebold and Li are available in trading systems too, therefore I can develop further the research into the trading strategies' direction.

4.1.1 Cubic spline interpolation

I used cubic spline function $f : \mathbb{R} \rightarrow \mathbb{R}$ to interpolate the missing maturities in the time series because this method is popular in the economic literature and monetary researches. Spline is a non-parametric polynomial interpolation method, which generated from distinct polynomial segments that are connected at so-called points (Choudhry, Pienaar, and Lizzio, 2004).

$$f(x) = \begin{cases} p_1(x) & x_{(1)} \leq x < x_{(2)} \\ p_2(x) & x_{(2)} \leq x < x_{(3)} \\ \vdots & \\ p_{n-1}(x) & x_{(n-1)} \leq x < x_{(n)} \end{cases} \quad (4.1)$$

¹ $\tau = [3,6,9,12,15,18,21,24,30,36,48,60,72,84,96,108,120]$

where $p_i(x)$ is the polynomial fit to the subinterval $[x_{(i)}, x_{(i+1)}]$, and x is the range where the spline regression will be generated. The first and second differential of the joined spline polynomials are equal the knot points, which guarantees the smooth connections between the spline curves. Spline regression is able to model complex connections by joining different polynomials which represent different subintervals (Katz, 2011). This potential of the spline regression can be used for the interpolation of interest rate points.

4.1.2 The analysed data

I generated a 3D surface in MATLAB to visualise the variety of zero coupon curves from January 2001 to February 2016.

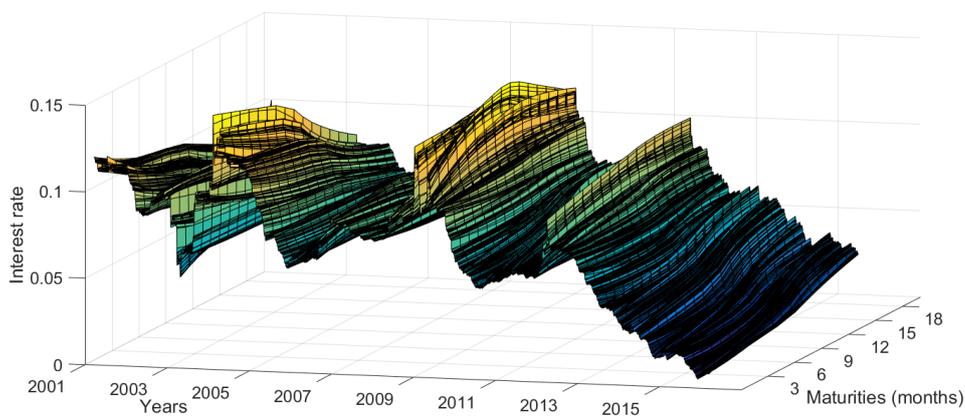


FIGURE 4.1: Hungarian interest rates

The surface of figure 4.1 shows the changing level, slope and curvature of the term structure of zero coupons. The interest rate points are shifting down which means that the level factor is decreasing over time. The level factor must have bigger autocorrelation than the slope and curvature factors, but innovations are observable on the time series, where the constant growing or falling level changes into a new direction and preserve it in the long run. The slope and curvature factors must be far more dynamic and fluid than the level.

For further analysis I separated the different periods of the interest rates development. There are four well separable ranges on the time series because of the peaks generated by economic shocks.

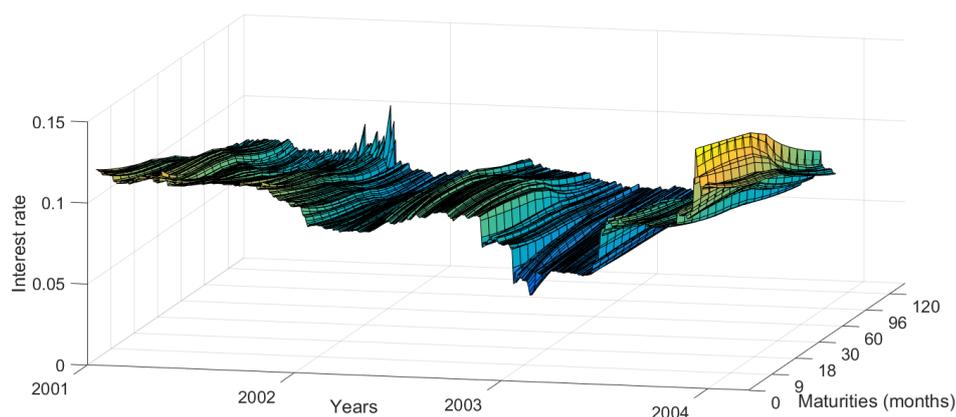


FIGURE 4.2: 2001-2003 period of term structures of Hungarian interest rates

In 2001 Hungary was one of the fastest growing economy among the OECD countries (OECD, 2002) with a reduced inflation from 28 % in 1995 to 10 % in 1999. In 2000 the inflation could not fall because of the international food and energy price increases. The year 2001 was an important milestone in the history of Hungarian monetary economy, because a Hungarian Parliament modified the policy of the central bank. From the year 2001 the official primary objective of the National Bank of Hungary was the price stability (OECD, 2002).

Figure 4.2 shows the effect of the Hungarian austerity measures commonly known as Bokros package. The package was announced in the first quarter of 1995 and avoiding the national bankruptcy was its first objective. The inflation expectations started to decrease resulting from the success austerity measures. On the observed period the inflation decreased from 9.1% to 4.7% (KSH, 2011). The falling inflation expectations of the investors can modify the slope and curvature of the interest rate term structure. The expectation about the decreasing inflation rates is able to turn the value of slope to negative and the value of curvature as well.

The first quarter of 2003 is showing horizontal interest rate term structures, where the spread between the maturities 3 month and 120 month is only 35 bps comparing to the average of the first range which is 245 bps. This effect symbolises well the effect of the markets liquidity changes. If the demand on long end is decreasing the lending is getting cheaper hence the interest rate is decreasing on long end. This market movement can cause horizontal interest rate term structure.

The value of the level is changing very slowly over time, only the shocks can turn them into another direction. The big jump at the middle-term and end of 2003 is imputable to the Hungarian Nation Bank's decision about the 300 bps increasing of the interest rate.

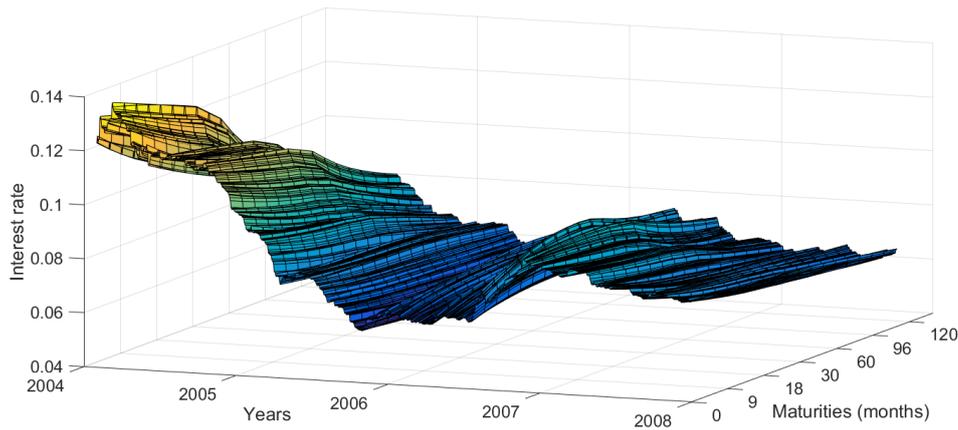


FIGURE 4.3: 2004-2007 period of term structures of Hungarian interest rates

Figure 4.3 shows interest rate term structure from 2004 to end of 2007. The term structures with negative slope is changing gradually and turns into positive from end of 2005. The inflation rate increased in period from 2005 to 2007 from 3.5% to 7.9% (KSH, 2011), hence the the slope of the interest rates term structure turned to positive. The change of inflation expectations of investors can cause modification mostly related to the slope and curvature as seen on the figure 4.3.

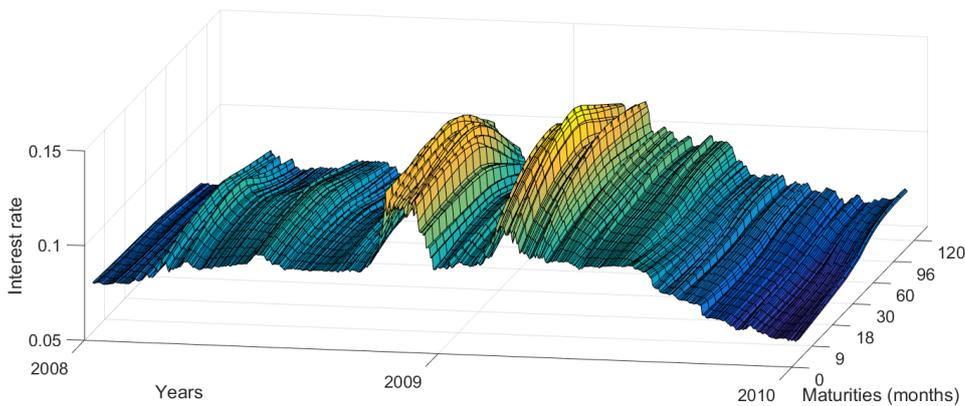


FIGURE 4.4: 2008-2010 period of term structures of Hungarian interest rates

Figure 4.4 shows the period of the financial crisis. In October of 2008 there was a radical decision made by the Hungarian National Bank to increase the rate by 300 bps. This extreme change is visible on the figure 4.4. The risk increased the investors' inflation expectations hence the slope is positive and the curvature is higher at the end of 2009.

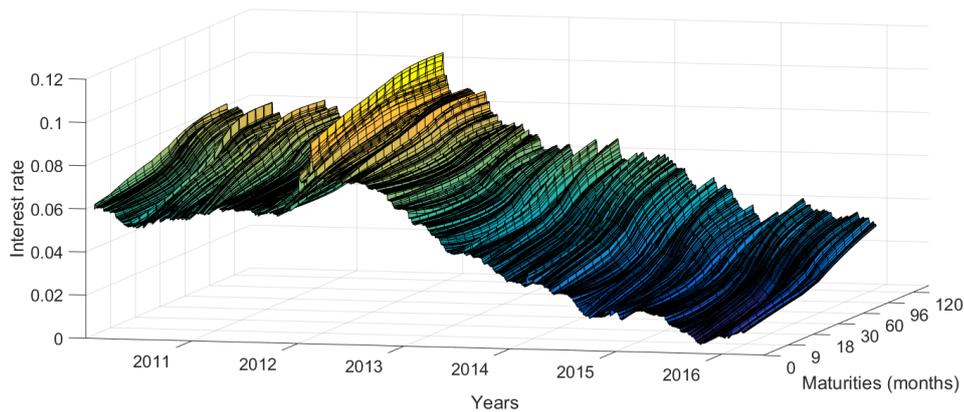


FIGURE 4.5: 2010-2016 period of term structures of Hungarian interest rates

The most recent interest rate term structures are showing concave curves with decreasing level in figure 4.5, which is imputable to the lowering of rates.

The less variant part of the term structures is the long end. This effect is visually well apparent on figure 4.1. Focusing on the 2012-2015 range of the figure it is visible that the points with more than 30 months maturity are more slowly decreasing than the short end.

The short end of the Hungarian term structure is sensitive for the activities of the central bank, but the long end is dependent on change in yield of markets in Western Europe (Nagy, 2015). If this rule was true, the volatility of term structure's short end would reflect the effect of the National Bank of Hungary's interventions.

Due to the transformation of the central bank's two-week deposit into three months in 2015, the banks are motivated to invest in bonds, because the bonds are more favourable on both legislation and tax grounds (Nagy, 2015).

The demand is higher on the short end, and the maturities like 3 and 5 years are not really preferred by the banks (Nagy, 2015). This trend shows that the banks try to avoid the risk holding Hungarian bonds on long term. The increased demand pushed down the short end of the yield curve, hence the form of the whole term structure is similar to character 'S' (Nagy, 2015).

This observations can be helpful for modelling and interpreting because the auto- and cross correlation are playing an important role in the research. Moreover, the non-linear connections like the changing correlation between factors over time can be a good reason to use neural network to forecast interest rates term structure and compare it with the results of vector autoregressive model, because the latter should be able to model the non-linear connections, as well.

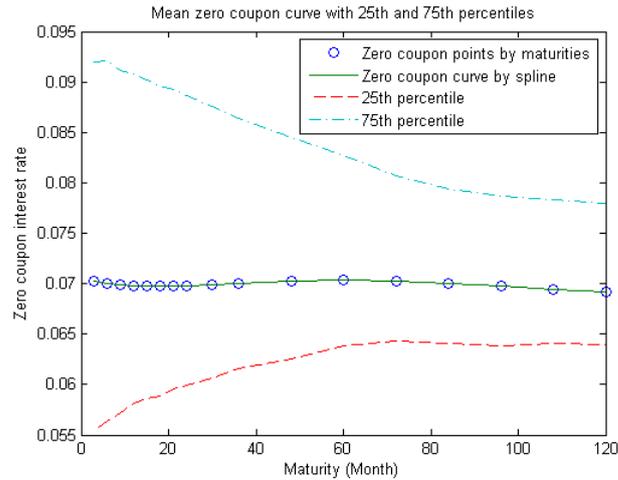


FIGURE 4.6: Zero coupon points interpolated by spline regression and the 25th and 75th percentiles.

Diebold and Li are expecting a concave yield curve from the mean interest rate points (Diebold and Li, 2006), but this is not true for the Hungarian mean curve as the figure 4.6 shows. The difference between the American and the Hungarian yield curve can be based on the dissimilar investment climate like inflation and its expected future value.

Figure 4.6 is perfectly showing that the the long end of the yield curve is less volatile then the short end. The difference between the 25th and 75th percentile lines are decreasing by converging to the mean, therefore the expectation of Diebold and Li is true on the less volatile long end.

The mean interest rate points interpolated by spline between 2001 and 2003 contradicts the expected concave curve by Diebold and Li. This negative slope and curvature is generated by the decreasing Hungarian inflations expectations mentioned above. This curve is the reason why the average of the whole period (2001-2016) is horizontal: the different curves of the periods are aggregating into a smooth horizontal interest rate curve which is not consistent with Diebold and Li's research. What is true that the shapes can be inverted on certain periods of time as can be seen on figure 4.7.

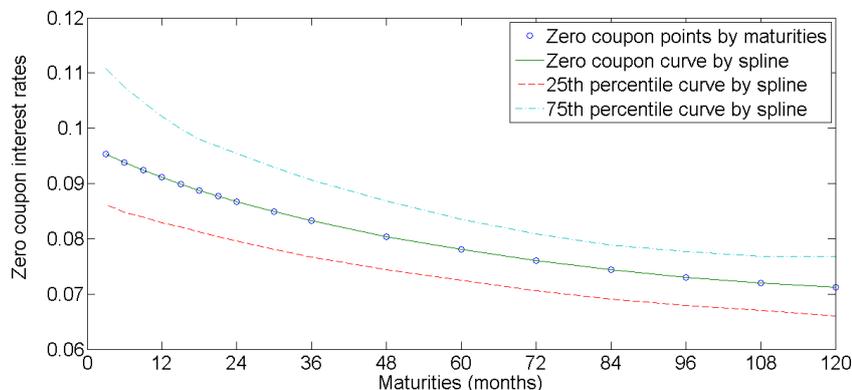


FIGURE 4.7: Zero coupon points for the period 2001-2003 interpolated by spline regression and the 25th and 75th percentiles.

The 2004-2007 period (figure 4.3) is more volatile which can be imputable to the longer selected range of time series, but the slope is less negative which means the inflation expectations were deteriorated since the previous time period, hence the stable period after Bokros package is ended and there are negative outlooks on the Hungarian markets. The curve is less convex like before although this period is not satisfy the expectations of Diebold and Li.

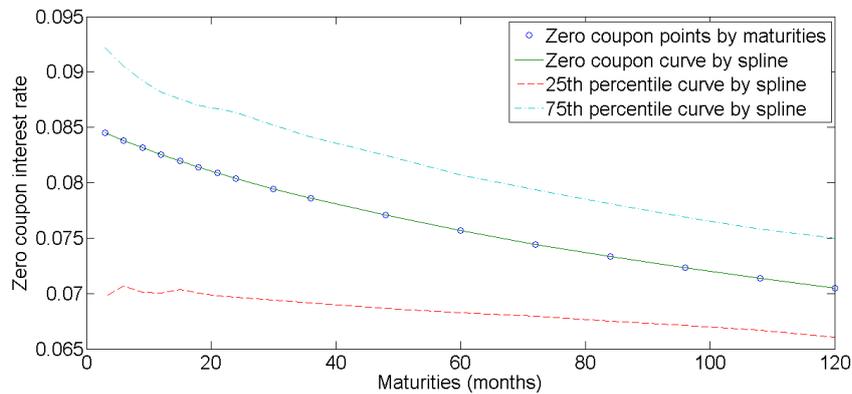


FIGURE 4.8: Zero coupon points for the period 2004-2007 interpolated by spline regression and the 25th and 75th percentiles.

The 2008-2009 period is more horizontal and concave as it expected by Diebold and Li as can be seen on figure 4.4. The level of the interest rate increased during the crisis as mentioned above. The 75th percentile curve has a bigger curvature than the 25th percentile which is almost horizontal. This effect is derived from the changes in liquidity on Hungarian markets. If the investors prefer less the long end it can happen that the interest rate points decreases on those part and the curve will be horizontal for a while.

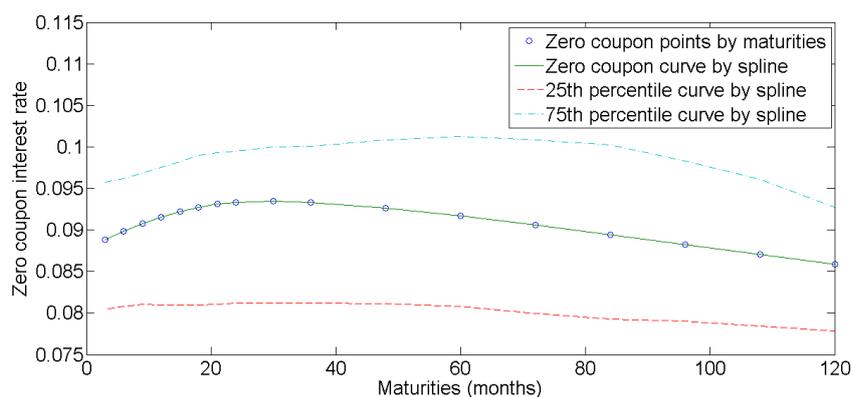


FIGURE 4.9: Zero coupon points for the period 2008-2010 interpolated by spline regression and the 25th and 75th percentiles.

The 2010-2016 period is concave and has positive slope which is satisfy the facts presented by Diebold and Li. The aggregation of periods with very different properties generates a horizontal line as seen on figure 4.6.

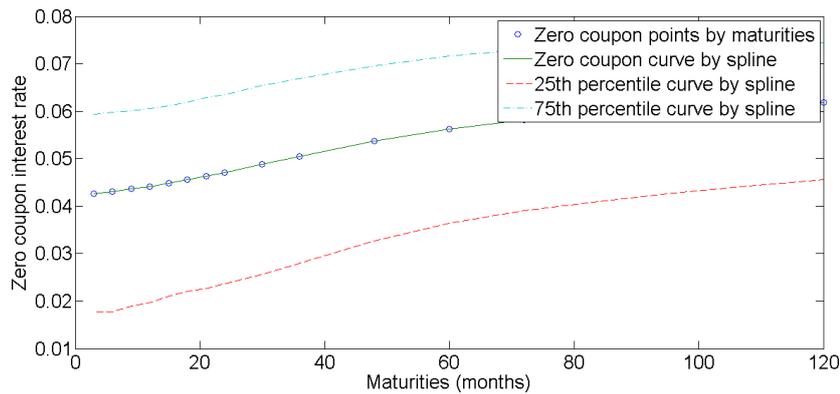


FIGURE 4.10: Zero coupon points for the period 2010-2016 interpolated by spline regression and the 25th and 75th percentiles.

4.2 Exponential decay parameter

In the formula there is the λ_t parameter which is evolving over time, but this is weakening the forecasting power of the model (Malinska and Barunik, 2015). The λ_t variable is the middle-term of term structure, thus it should be a constant number. If I let it change dynamically over time the fitting will be better but the unexpected jumps can cause false forecasting (Vela, 2013).

Diebold and Li fixed the decay parameter in their research (Diebold and Li, 2006), and their methodology is cited often in the literature related to modelling any kind of interest rates by Nelson-Siegel model like zero-coupon yield curve (Pooter, 2007) or interest rates of crude oil (Malinska and Barunik, 2015).

Figure 4.11 shows the dynamically evolving λ_t parameter, which is very volatile. For every λ_t could be drawn a new figure about the change of loadings like figure 2.1. This makes worse the forecasting ability of the model (Malinska and Barunik, 2015), thus I do not accept the changing λ_t . For every analysed and predicted day I would like to use the same loadings process to keep the β parameters in the same λ dimension.

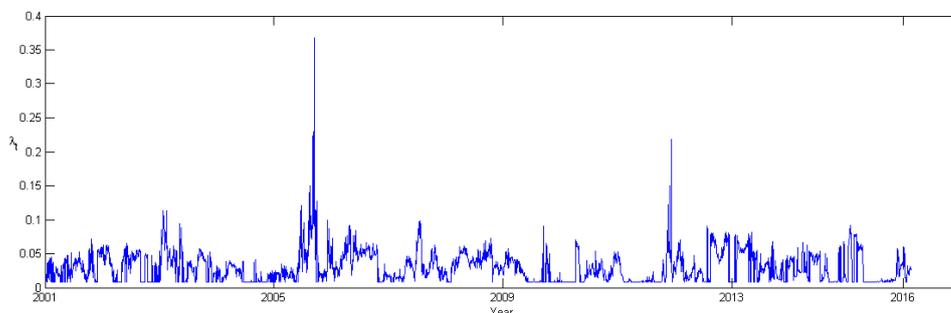


FIGURE 4.11: Dynamically evolving λ_t parameters

I decided to find a constant lambda for the modelling, because I am focusing on the predictive power of the solution. Moreover, finding the optimal lambda is a non-linear problem, but there is no best practice in the

economic literature how to optimise it well. This gives me the opportunity to test different methods.

$$\lambda_t = \lambda^* \quad (4.2)$$

If I optimise the λ_t and calculate the right β parameters, I have to search in a three dimensional world for the fitting coefficients. The surface of the least squares estimation can be too complex and very hard to build a good algorithm. In the literature the differential evolution was referred and implemented with success (Gilli, Grosse, and Schumann, 2010).

Diebold and Li assert that a well fitted model has a loading factor which is maximising the β_2 on the middle-term of term structure, thus it is required to nominate a point on the time line as mid-term. The λ_t is reciprocal of the selected mid-term, as mentioned above (Diebold and Li, 2006). The investors are separating the investments into short-, middle- and long-term, but this depends on the frequency of trading too. Diebold and Li is calculating the average of 24 and 36 months, because they nominated the 2-3 years range to middle-term period. They chose in their research the λ to 0.0609, thus maximised the β_2 on 30 months. To get the right λ value the equation of 4.3 can be useful. Maximising the loading of β_2 will give the optimal λ for the modelling.

$$\lambda^* = \arg \max_{\lambda} \left(\frac{1 - e^{-\lambda_t \tau}}{\lambda_t \tau} - e^{-\lambda_t \tau} \right) \quad (4.3)$$

In equation 4.3 the τ can be chosen as a constant maturity like 24,36,48 months. I tested it on the scale of 24 and 48 months and it gives the range [0.0373 – 0.0747]. Diebold and Li's $\lambda = 0.06$ solution seems correct, because the arg max function results $\lambda = 0.0597$ when the maturity is 30.

A well optimised model should have a constant λ to increase the forecasting ability, and if I want to accept the Nelson-Siegel model, I have to be sure that the loading factors are not correlated linearly. They are weights which can decrease the quality of principal component analysis which is the logical background of the Nelson-Siegel model.

It is empirically proven that the loadings of the β_1 and β_2 factors can be correlated around zero (Gilli, Grosse, and Schumann, 2010). The correlation between two loadings could be tested but I found that the λ should be not derived from a statistical approach losing the economical view of the modelling. If the loadings would be totally uncorrelated the economical prediction power of Nelson-Siegel could be worst.

Finally, I chose the arg max method to find the right λ parameter. I accept that the middle-term point should be the 30th month and built a script in MATLAB to find the λ which maximises the β_2 above the 30th month. I chose $\lambda = 0.0598$ because it satisfies the requirements of the Nelson-Siegel model's economical requirements.

4.3 Standard Nelson Siegel versus extended model

4.3.1 Grid search for Svensson model

For the Svensson model it is necessary to find fixed lambda parameters similar way as in the case of standard Nelson-Siegel method. The fixed

<i>T</i>	<i>3Months</i>	<i>6Months</i>	<i>9Months</i>	<i>12Months</i>	<i>15Months</i>
<i>I.</i>	<i>NSS, 79%</i>	<i>NSS, 76%</i>	<i>NSS, 87%</i>	<i>NSS, 82%</i>	<i>NSS, 82%</i>
<i>II.</i>	<i>NSS, 99%</i>	<i>NSS, 80%</i>	<i>NS, 61%</i>	<i>NS, 89%</i>	<i>NSS, 95%</i>
<i>III.</i>	<i>NS, 93%</i>	<i>NSS, 78%</i>	<i>NS, 84%</i>	<i>NS, 85%</i>	<i>NS, 88%</i>
<i>IV.</i>	<i>NS, 90%</i>	<i>NSS, 74%</i>	<i>NSS, 93%</i>	<i>NS, 88%</i>	<i>NS, 86%</i>
<i>V.</i>	<i>NS, 3%</i>	<i>NS, 5%</i>	<i>NSS, 54%</i>	<i>NS, 8%</i>	<i>NS, 8%</i>

FIGURE 4.12: Model comparison by time periods (I: 2001-2003, II: 2004-2007, III: 2008-2009, IV: 2010-2012, V: 2013-2016) and short term maturities

lambda makes the problem to linear hence the computation difficultness of the approximation decreases. Find a fixed decay parameter for standard model was relatively easy comparing to the two lambda parameters of Svensson. The best practice is the so called Grid search method which is able to implement easily in MATLAB to find the optimal lambda pair. I collected RMSE results into a matrix where the rows are symbolising the maximised month of the parameter λ_1 from 10 to 44 and the columns were the maximised maturity months of λ_2 from 55 to 120. The logic of this selection is based on the experience from standard model's lambda which is the middle-term factor, hence I gave a range for the first lambda in the middle-term period. The second lambda has a range in the long end of the curve.

I found that there are more local minimum which is corresponding to the theory that finding lambda parameter is a non-linear problem. The $\lambda_2 = [75, 85]$ range and $\lambda_1 = [20, 24]$ seems to be a local minimum, but $\lambda_2 = [110, 115]$ range and $\lambda_1 = [16, 18]$ is another local minimum area. The previous local minimum is deeper, hence it is the global minimum. Finally I selected $\lambda_1 = 23$ and $\lambda_2 = 85$, because this combination gave the smallest aggregated error, and empirically the 23th month can be a middle term point on the term structure. The 85th month of maturity maximise a long term point on the β_3 curve, which suggest that there must be generally a hump in seventh year of maturities.

4.3.2 Results of comparing the standard and extended Nelson-Siegel models

There are different periods where the standard Nelson-Siegel or the extended model can be better, hence I created a matrix by MATLAB to compare the sum of relative errors generated by the two models. This thesis has a preference to provide knowledge for trading activity thus first of all I compare the results about short end of the curve, but the long end has similar errors too. In the table the method's name symbolise the winner model with lower MSE, and the number next to the model's name is the ratio between the MSE of the winner and the dominated one. This can help to understand the power of fit for different periods and maturities.

The first period – the after-Bokros-package period – shows a good fit of the Nelson-Siegel-Svensson model for every maturity, but this fact changes if we check the second, pre-crisis period too. There are maturities, where the Svensson model is dominated by the standard Nelson-Siegel model. The MSE ratio numbers show that the Svensson model is highly better for

the first period, but in the second period the 3- and 15-months maturities have similar MSE for both models.

The standard Nelson-Siegel model has a weak dominance in the third period for 3-months maturity, but the 6-, 9- and 12-month maturities are clearly better fit by it.

The fourth and fifth period were analysed in a block previously, but now I cut this period to collect more information about the time series of 2010-2016. The compared model's MSE values are not highly differ from each other in the 2010-2012 period, but the standard model dominates the Svensson. The fifth period's 3- and 5-months maturities can be fit highly better by standard model, because the ratio between the good models MSE and the dominated method is 3% for the 3-months maturity and 5% for the 5-months, which symbolise an extreme difference in fitting.

These results suggests that I have to use the standard Nelson-Siegel model for the Hungarian interest rate term structure's modelling. Diebold and Li mentioned in their research that the Nelson-Siegel extensions are not necessary better methods, because the goodness is dependent on the type of term structure too (Diebold and Li, 2006). The Hungarian term structure changed after 2004, and the standard model dominates the Svensson extension's result both on short and long end of the term structure. From the view of forecasting it is better to use those model which describe better the recent term structures.

The curve of Nelson-Siegel model are generally fit more properly than the curve of Svensson model, but there are special cases, when the extended model is highly dominates the standard one. In my thesis I use the standard Nelson-Siegel model to fit and forecast the term structure of Hungarian interest rate, but it should be noted that Svensson's model was better for Hungarian zero coupon yield's modelling over a decade ago than the standard method.

4.4 β coefficients of Nelson-Siegel

After receiving the optimal λ parameter I could start the Nelson-Siegel approximation. It is recommended to pay attention to the statistical analysis of residuals because the goodness of fit is clearly shown by residuals. The mean, standard deviation and autocorrelation test can be very useful to find weakness of a model.

The fitting can be diverse depending on the shapes of the curve. The smooth curves can be well fitted by Nelson-Siegel, but there are curves where the slope and the curvature is higher, hence the approximation method has weaker performance. In average the dynamic Nelson-Siegel method produce a high quality fitting because the average of residuals percentage is very slow: it is between -0.8824% an 0.7384% which is acceptable for the modelling (A).

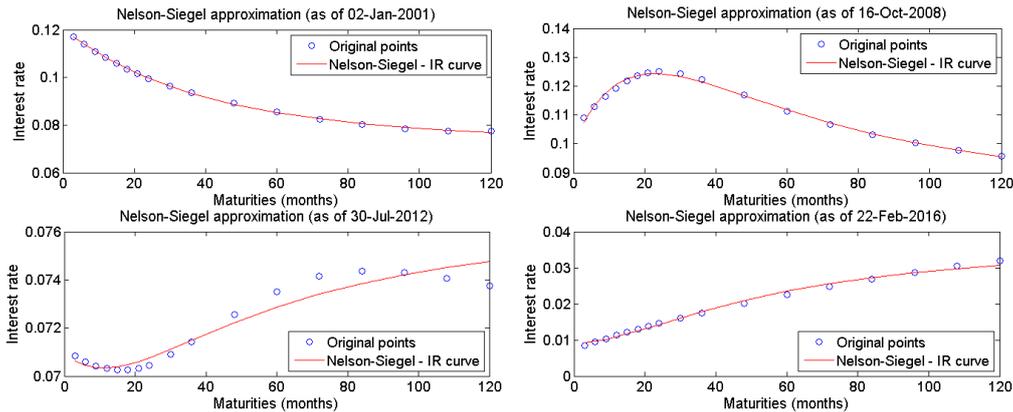


FIGURE 4.13: 4 different types of the term structure of Hungarian interest rates

Day 30-Jul-2012 from figure 4.13 has a complex form because the original interest rate points are similar to a rotated 'S' form. It has two curvature, but one of them is convex, the other one is concave. The dynamic Nelson Siegel approximation fit a curve, which has a minimal average error taking into account every interest points, but if we want to select a given maturity, the percentage error can be high.

Another important example is the curve from 22-Feb-2016 from figure 4.13. The interest rate curve is fit well by Nelson-Siegel, if we take a closer look at the figure, but after a percentage error calculation the result will be shocking: the first maturity ($\tau = 3$) is more than 20%.

In this research I did not want to find the best interpolation method, because the first objective is the prediction power of the model what I am looking for. Dynamic Nelson-Siegel produces small errors in average and this result is enough to develop further the model.

4.5 Analysing of β parameters

The analysis of β variables constitutes an important element of the model building process. The decisions about the order of autoregressive models and delays in non-linear autoregressive neural networks or the application of a given model is based on the previous analysis of input data. I collected information about the autocorrelation and cross-correlation of β inputs.

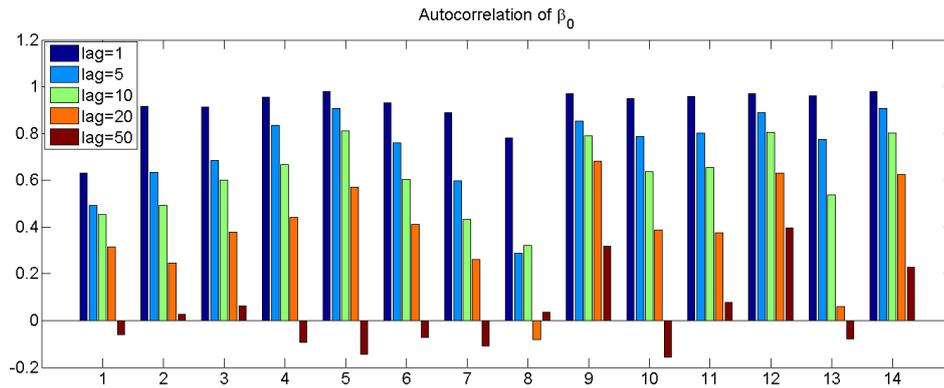


FIGURE 4.14: Autocorrelation of β_0 variable for 14 different periods

The autocorrelation of β_0 variable is very high (figure 4.15), because it is 0.9864 for the $lag = 1$. The decreasing of autocorrelation is very slow, the value for $lag = 20$ is 0.8897. This can be an important indication that the autoregressive model and vector autoregressive models should have $p = 1$ order.

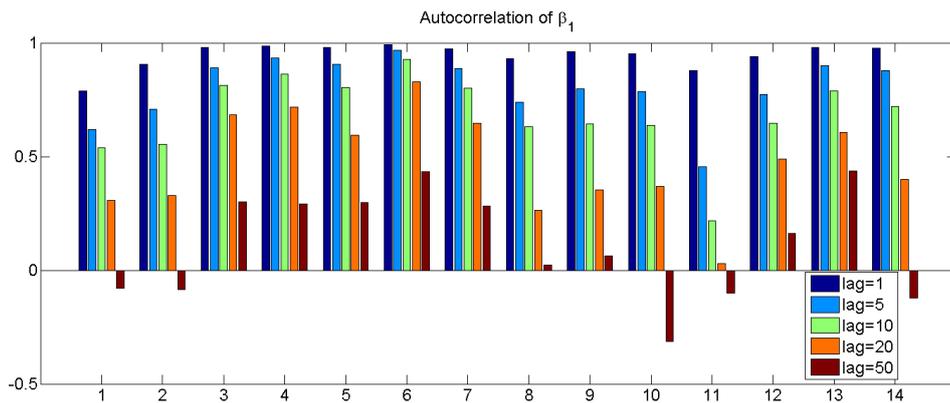


FIGURE 4.15: Autocorrelation of β_1 variable for 14 different periods

The β_1 parameter has bigger autocorrelation than previously analysed β_0 . Its value is 0.9971 for $lag = 1$ and it decreases very slowly: the 20th lag shows 0.9577. The β_2 variable is highly auto-correlated because of its 0.9871 value for $lag = 1$ which falls slowly as in case of previously presented β parameters' autocorrelation by lags.

The autocorrelation plots in appendix shows (Appendix 4.16) the very different distribution of autocorrelations for $lag = [1, 5, 10, 20, 50]$ on time series divided into equal periods, and it seems that the autocorrelation of $lag = 1$ is the most stable value. This observation can be useful for the autoregressive modelling later.

Variables	β_0	β_1	β_2
β_0	0.9864	-0.0541	0.1907
β_1	-0.0541	0.9971	0.6814
β_2	0.1907	0.6814	0.9871

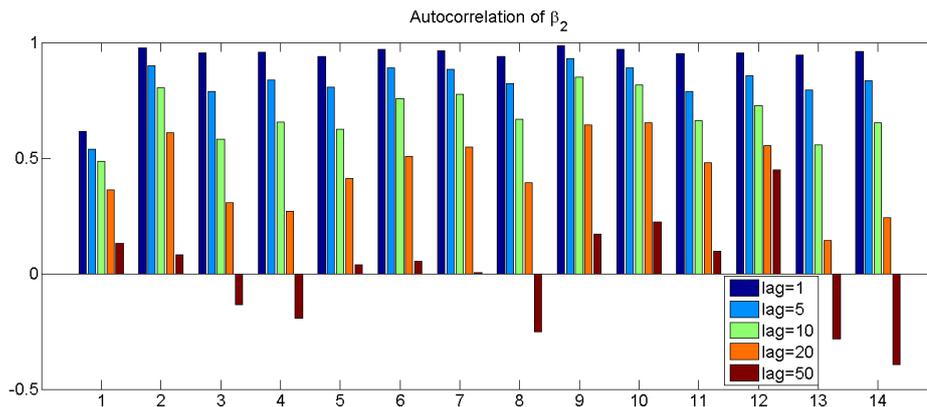
TABLE 4.1: Cross correlations for lag=1

The table 4.1 shows the changes of cross correlation of β variables, when $lag = 1$. The cross correlation of β_2 and β_3 variables was relative small for $lag = 1$ in periods 2005-2007 and 2010-2016, but in the financial crisis it was changed and its absolute value increased to 0.65. Parameters β_0 and β_2 have their weakest connection during the crisis, while in other periods they are connected more stronger. To forecast the right factors of the interest rates term structure in an always changing environment, a complex solution must be needed, which gives meaning to the implementation of machine learning techniques.

Variables	Mean	Minimum	Maximum	Std.Dev.	KPSS Test (p)
β_0	0.0696	0.0362	0.1179	0.0122	0.001
β_1	0.0002	-0.0511	0.0833	0.0285	0.001
β_2	0.0009	-0.0672	0.1540	0.0337	0.001

TABLE 4.2: Statistical description of β variables

The average of β_0 parameters is above zero because there is no negative interest rates on the analysed time series. The standard deviation of the first parameter in Nelson-Siegel is the lowest compared to other β variables, as it can be seen from the table 4.2. This suggests that the modelling of the β_0 variable should be easier than the others. The most volatile variable is the β_2 , which describes the curvature.

FIGURE 4.16: Autocorrelation of β_2 variable for 14 different periods

β_1 and β_2 (figure 4.16) have means close to zero which corresponds to the draw about the average term structure of Hungarian interest rates from 2011 to February 2016. The average term structure is nearly horizontal, hence it has no curvature and zero slope.

I tested the β variables' stationary by Kwiatkowski-Phillips-Schmidt-Shin test with null hypothesis in which the time series is stationary. The test rejected the null hypothesis for every β , because the p-value is 0.001. This information is important for the modelling because the time series in finance or in economics are very often non-stationary and this problem must be taken into account for pre-processing of data for models (Virili, 2000).

Substituting the β parameters into Nelson-Siegel model gives the term structure of interest rates. The statistical description of this results can help to check the reliability of the fitting. The mean of residual percentages are close to zero on every maturity, and the standard deviation is relatively small (Appendix table A.1). The extreme errors are calculated by the minimum and maximum value of residuals to make visible the accuracy of the fitting by maturities. The end of long end and the first two maturities show bigger standard deviations than the middle of the term structure (Appendix table A.1). This must be caused by the selected λ , which tries to balance the fitting's error on the term structure and its value is optimised to the middle-term. For this reason the first and the last maturities could have bigger errors.

4.6 Errors in forecast methodologies

There are two types of prediction errors. The forecasting error of β_i parameters on the one hand and the model error calculated from the difference of interest rate points of Nelson-Siegel's term structure and the original interest rate points.

The forecasting error (4.4) of β_i parameters represents the prediction ability of the selected model like random walk, AR(p), VAR(p) or neural networks.

$$e_{forecasting_{i,t}} = (\beta_{i,t} - \tilde{\beta}_{i,t})^2 \quad (4.4)$$

The model error (4.5) is based on the predicted term structure which is the result of the substitution of the forecasted β_i parameters into Nelson-Siegel model.

$$e_{model_t} = \sum_{m=1}^{17} (IR_{m,t} - \tilde{IR}_{m,t})^2 \quad (4.5)$$

The index m is the maturity in the equation and there are 17 different standard maturities which involve the interest rate points. $IR_{m,t}$ symbolises the interest rate point for t th day and maturity m . e_{model_t} is the square of the difference between the original interest rate points at time t and the forecasted points at time t .

The two errors are not necessarily correlated. The Nelson-Siegel model has an error because of the fit of term structure. The λ optimisation process can be accurate but there are very different term structures with variant optimal middle-term points. Selecting the right λ is difficult because of its non-linearity connection with the fitting's mean square error. There are always discrepancies on different ranges of the analysed time series. If the term structure curve is not fit well, the errors generated by the original and

predicted β can produce a better fit term structure closer to the real one than the expected NS-curve.

The primary goal of the research is finding a model which can predict accurate interest rate points. If a model has a consistent forecast error which produces lower model error, it means that this model is proper for the accurate prediction. The relative lower forecasting error with higher model error compared to other models is not acceptable because the dominance of models is dependent from the model error. This is a trade-off between the well fitting and the good forecasting performance, where this research prefers the better forecasting. The dimensions which are narrowed can cause worst fitting, but the modelling the dynamics of the variables in a reduced range is more easy.

Chapter 5

Forecasting of interest rates

5.1 Forecasting plan

The time series dividend by four parts - which was mentioned and analysed above -, can provide an objective picture about the performance of forecasting models. The different periods like stabilised economy with lower inflation expectation and crisis, post-crisis and most recent month are needed different modelling tools. The changing of auto- and cross-correlation of the β factors can result that there are periods where only one model is working, and there are other ranges on the time series where both or nothing is able to forecast the next coming day.

There are expectations from the different methods. Diebold and Li showed in theirs research, that the Nelson-Siegel AR(1) can be a very good forecasting method with lower RMSE comparing to VAR(1), Slope Regression, Fama-Bliss forward rate regression, Cochrane-Piazessi forward curve regression (Diebold and Li, 2006).

The predictions of the vector autoregressive forecasting method can be better if the cross correlation of the β parameters are significant (Malinska and Barunik, 2015). The VAR models are usually over-performed by other models according to the economic literature (Diebold and Li, 2006), but it should be noted that if there was no cross-correlation among β variables, or it does not contain any relevant information, a principal component analysis would have relevance.

There are assumptions related to the prediction power of different models. It is expected that the best performing model to predict the coming day's term structure will be the random walk. The 5th, 10th and 15th days can be estimated better by neural network than random walk, AR or VAR.

The random walk cannot predict changes - it draws only a straight horizontal line for every prediction's point independently from the number of days - but the other compared models are able to find patterns and they are able to forecast them for future periods. It was previously mentioned that the random walk is very accurate for financial forecasting, both for short and long term estimations, but the Hungarian interest rates term structure has variant shapes. That is there are diverse periods with different types of structures, hence a complex model must be needed. The neural network can work in many ways: the parameters can be estimated individually – by one input node – and can be added in the same time by three input nodes to the network.

There are scenarios where the different model's results should be combined with each other to find the best solution. The β parameters have their

own patterns and are volatile, hence it is also possible that not the same model can predict all of the β variables.

5.2 Stationary and outliers

The first object of the analysis is to make the comparing of models as fair as possible. The inputs data should be the same for every models, hence I have to collect the requirements of variant models about the input data. The aggregated needs can provide information about the fair data preprocessing.

First, I wanted to remove the linear trends from the data by calculating the change of variables. Computing changes of variables for neural network inputs is the most effective data preprocessing solution and it can help to eliminate the liner trends (Kaastra and Boyd, 1995), hence I implemented it. Calculating the logarithmic transformation of the changes can help to reduce the right-tail effect (Kaastra and Boyd, 1995), but the histogram of beta changes did not show any significant positive skewness, thus I rejected to use this adjustment.

$$\Delta\beta_{i,t} = \frac{\beta_{i,t}}{\beta_{i,t-n}} - 1 \quad (5.1)$$

The variable n in the equation 5.1 is the number of days for the period of change. In this research I used $n=1$, $n=5$ and $n=10$, to forecast 1 day, 5 days and 10 days ahead. Parameter i represents the index of the given β variable from Nelson-Siegel model.

I analysed the stationary of the possible inputs, because the autoregressive and vector-autoregressive models are not able to perfectly model the non-stationary time series (Tadjuidje Kamgaing, 2005). The Kwiatkowski, Phillips, Schmidt, and Shin (KPSS) test for stationary showed previously that the time series of β_i parameters - where i represents the three variant Nelson-Siegel factor - did not satisfy the requirement of stationary. Now I used the KPSS-test for the first differences and the changes of β variables, but even these transformations are not stationer. Consequently, the changes of variables with the property of non-stationary is a reason to use neural networks, because the universal learning ability of networks can help to model complex non-linear connections which is not suitable for AR or VAR (Tadjuidje Kamgaing, 2005).

Second, I checked the outliers in the time series of β parameters.

Morris Hansen' definition about outliers is based on the model estimation's performance (Hansen, W., and Tepping, 1983). Hansen thought that an element in a dataset must be an outlier if the estimation of the model could increase more than 10 percent by removing the given element (Ghosh and Vogt, 2012). I tested this statement on my time series and I found that removing the effect of elements outside the 2nd and 98th percentiles can result 3 or 4 times lower RMSE on certain ranges of time series for the neural network. For this reason, it may seem like a good idea to select the range of input variables enclosed by the boundaries of 2nd and 98th percentiles of the dataset as the normal set and every element outside of it is categorised into the outlier set, but I did not implement it.

Various means are available such as cutting by percentiles, standard deviations multiplied by two or three, using signal processing solutions or Sprent-test to find outliers. If I select one method, I have to use it for every model's input, hence a universal solution is needed.

I found Sprent-test proper for finding extreme values in time series instead of cutting by percentiles, because it is empirically tested for it and has proven to be successful in filtering outliers in high frequency data as well (Venturini, 2011). Finding a method which is fast and reliable for time series (Venturini, 2011) is preferred by me for this research and for further development of the analysed models, especially for the neural networks.

$\Delta\beta_{i,t}$ represents the analysed variable in the window, where i refers to the elected factor (level, slope, curvature), and t is the index of the variable in the equation 5.2. The inputs are split into windows, because for every estimation the previous n days is selected to forecast $n + 1$ from the out of sample, thus the outliers are calculated on the moving windows.

$$Score_{e_{i,t}} = \frac{\Delta\beta_{i,t} - med(\Delta\beta_i)}{med|\Delta\beta_i - med(\Delta\beta_i)|} \quad (5.2)$$

The neural networks' outputs can be easily distorted by outliers, because of wrong pattern recognition (Nguyen and Chan, 2004). The network reproduces the learnt patterns independently from its relevance, thus the inputs must be adjusted to avoid extreme errors ¹ The performance of the neural network is highly dependent on the inputs as well, hence a well prepared data can reduce the running time and increase the propagation's quality (Yu, Wang, and Lai, 2007).

A modified version of winsorization methodology is applied for the limitation of the extreme values in dataset. Winsorization replaces any value above the selected p percentile by the value of p percentile and any value below the $1 - p$ by the $1 - p$ percentile (Ghosh and Vogt, 2012). This method preserves the size of the original dataset by substituting the given percentile of the dataset (Ghosh and Vogt, 2012). I preferred this solution instead of trimming because I did not want to remove numbers without replacing them, violating the coherence and sequence of the time series. There is only one modification in the winsorization, what I implemented: I use the Sprent-tests' results (5.2) to find the upper (lower) outliers and I replace them by the given value which corresponds to the maximum (minimum) accepted Sprent-score. I selected the $|\text{Sprent Score}| = 4$ which corresponds more than 2 standard deviations distance from the mean, but smaller than three standard deviations.

5.3 Autoregressive model

Diebold and Li implemented a Nelson-Siegel function with $AR(p)$, where $p = 1$ to predict the future interest rate term structure (Diebold and Li, 2006). This forecasting can be possible by the high autocorrelation of the β parameters.

¹I tested the neural networks with unadjusted data inputs to forecasts 10 days ahead. Except the end of 2014 and 2016 period, - which is very stable in every parameter - the RMSE of neural networks in the forecasting of interest rate points were in average 4-5 times bigger than the VAR(1) forecasts' RMSE, where the input data for VAR model was unadjusted too.

I selected a moving window sized to 300 days to build the model. I defined the minimum required size of the moving window to 250 days because I wanted to derivate information from a full trading year. Choosing a too long period for autoregressive model is not preferred and the empirical analysis shows that 300 days is a proper window size for modelling the Hungarian zero-coupon interest rates as seen previously on its 3D plots.

Considering the possible over training of neural network I did not want to select a too long window's size, because I applied the same window's size for every model – including the networks – in order to ensure a fair comparison. The 300 days long window covers a full trading year expanded by 50 more days representing 10 more weeks from the historical data to attach a short period before the point one year ago.

Finding the right p-order for the AR model is feasible by the Akaike Information Criterion. Akaike Information Criterion is based on the calculation of Kullback-Leibler distance between the original and estimated value (Burnham and Anderson, 2003).

$$AIC = 2k - 2\ln(L) \quad (5.3)$$

L is the likelihood function in the equation 5.3 and k is the number of estimated parameters. The model which has the lowest AIC is the suitable one for capturing the real variables (Burnham and Anderson, 2003). I calculated Akaike Information Criterion for every β parameters from Nelson-Siegel model to find the best fit. The Akaike Information Criterion's value decreases strictly for every parameter for every analysed forecast time horizon – except the one-day ahead forecast of β_0 , where the $p=2$ is a bit higher than the Akaike Information Criterion's value of $p=1$ –, hence I decided to use $p=1$ from the consideration that the autocorrelation is extremely high for $\text{lag}=1$. The decreasing of value AIC is slowly, thus it would be too hard to prove where is the right p-order, or it would be too arbitrary to select an order between very similar Akaike Information Criterion's values.

Previously in the chapter of β parameters I highlighted that the three variable of Nelson-Siegel model is relatively stable for $\text{lag} = 1$, but for higher lags there are huge differences between the equally distributed periods. For this reason I implement the AR model with order $p=1$ for the analysed time series and compare its prediction's results with other models.

I used the same windows size for the neural network which was applied for the autoregressive and the vector autoregressive methods. The lag or delay parameter was the same to prepare the same parameters for the forecasting. The fair comparing achieved by the same testing environment gives the opportunity for the research to analyse the performance of the different models more accurate.

5.4 Vector autoregression of beta parameters

I chose $p = 1$ for the VAR model similarly to the p-parameter of AR. The biggest advantage of the VAR model is based on the analysis of the cross variance of beta parameters. If there was no cross-correlation between the

β parameters, the VAR would not be able to forecast the variables significantly better than the AR. This is the point where it seems that avoiding the principal component analysis was a good decision.

The non-stationary time series of Nelson-Siegel parameters theoretically cannot degrade the modelling performance of VAR (Sims, Stock, and Watson, 1990), but I tested the co-integration of the level and the first difference of β variables by Johansen-test. I tested the change of β parameters for the co-integration as well, because the $\delta\beta_i$ is a standard pre-processing methodology for neural networks, too. The co-integration test provides information about the ability of a new time series to use it or not with another time series (Johansen, 1991), hence I can decide the necessity of implementing a VAR model which is not based only on the level of β .

The tests show that I can reject the integration of first differences and the changes too. Modelling the level of β variables should produce reliable results according to the study of Sims, Stock and Watson (Sims, Stock, and Watson, 1990).

5.5 Forecasting with neural network

The day-to-day change of beta parameters can be very auto-correlated because of the strong autocorrelation of interest rate points (Piazzesi, 2002). This autocorrelation can be stronger on the long end of the term structure (Piazzesi, 2002), thus there is a big risk in using machine learning technologies for forecasting interest rate term structures. If the relationship between two consecutive days is very strong, an over fitted neural network must predict wrong value ahead. The pattern recognise ability can be seriously detrimental to the forecasting performance of neural network if the patterns generated by random noise and the main factor behind the changes is a simple linear trend on the selected window. This is my hypothesis before the neural network's running.

5.5.1 Selecting the right architecture

There are no standards how many layers and nodes are required to construct the best architecture, hence the practice is the best methodology to find it (Kumar et al., 2011). The only rule which is useful to bear in mind that the number of layers and nodes should be minimised for faster running and avoid unnecessary calculations. This logic is similar to the principal component analysis, where the methodology focuses on the decreasing of variables number to find the simplest set of dimensions to describe the sample. The neural network needs nodes to activate the input data, but the increasing the number of neurons do not necessarily lead to a better propagations result.

5.5.2 Training method of the neural network

I chose the Levenberg-Marquardt training method for the forecasting neural network, because this algorithm that takes on board the best practices from the different learning solutions like steepest descent method and the Gauss-Newton algorithm (Yu and Wilamowski, 2010). The method of steepest descent is a minimum finding solution, where the new search direction

is orthogonal to the previous one. The steepest descent algorithm tries to minimise the length of the steps to find the shortest way to the global minimum point of a non-linear problem. This algorithm is very stable, which is the main advantage of this method, moreover, its iteration can be very fast, if the system is well scaled. If the problem is not scaled well, the algorithm can perform infinite iterations for finding the global minimum point. The main disadvantage of this method is the slow convergence particularly in complex, un- or miss- scaled problems (Hjorteland, 1999).

Gauss–Newton algorithm has a huge advantage in the world of big data: it is very fast (Yu and Wilamowski, 2010). Newton’s algorithm uses the second order derivative of the total error function to find the global minimum, while the steepest descent searches the minimum only by the first order derivative (Yu and Wilamowski, 2010).

The Levenberg-Marquardt training method combines the advantages of the steepest descent and the Gauss–Newton method, and these advantages are preferred for this research. The daily estimations, the length of the analysed time horizon and complexity of the problem required an error minimising algorithm which is widespread solution.

5.5.3 Data preprocessing for neural network

The change of β variables are the input for the neural network like in the case of other models in the comparison.

The input variables must be filtered by outliers because the proper training of neural network requires clean data. Learning patterns containing outliers is very disadvantageous for the neural network because it could forecast false movements of the β parameters (Nguyen and Chan, 2004). The unexpected events on market and the shocks from the monetary decision makers can generate extreme changes in the factor of level. This step was completed previously, because I use the same inputs for every model in my research.

Input variables are min-max normalised to the $[0, 1]$ range for faster and better training process. This step is the part of the data pre-processing which is needed for the proper training of neural network. If there are large differences between the numbers or there are outliers, it requires much bigger axes to describe the values than if the inputs were normalised. If the axes are too big to locate the numbers, the connections will be more complex in the neural network (Li, Chen, and Huang, 2001). This would cause a slower and less efficient training process. The min-max normalisation preserves the relationships between the values, but reduces the needed axes’ size (Li, Chen, and Huang, 2001).

The neural network starts its training from random weights. The random starting points can influence the training algorithm to find local or global minimum of error. I thought that the best method to avoid falling in the trap of suboptimal solutions is the implementation of a neural network with repeated learning. If the learning process is reiterated, the random weights are generated once again and the training algorithm run from a randomly selected new point to find the global minimum of error. The variant results from the iterations create a vector in MATLAB and I calculate a mean from them. The mean is the real result, which is denormalised in order to get the predicted change of the given β .

I tested the number of required reiteration, and the 30 reruns were stable for every neural network.

$$\tilde{\beta}_{i,t} = \beta_{i,t-n}(1 + \Delta\tilde{\beta}_{i,t}) \quad (5.4)$$

The estimated β is equal to the value as defined in the equation 5.4. The change of the given β ($\delta\tilde{\beta}_{i,t}$) is predicted by the neural network and multiplied by the previous β , which is the basis. The n variable defines the step of the forecasting.

5.5.4 Building the neural network

Finding the best parametrisation for a neural network is always a very hard and time consuming process. There are basic rules which can provide some useful advice for starting, but the critical thinking is always required for implementing them.

In the practice of the building of the neural networks there is a standard rule to set the number of nodes. This is important to save time by avoiding the unnecessary tests and reduce the risk of over-training.

$$N_{hidden} = \frac{N_{training}}{(\alpha(N_{input} + N_{output}))} \quad (5.5)$$

The N_{hidden} equation (5.5) can give a hint to find the optimal number of nodes for the hidden layer. The $N_{training}$ symbolises the size of the training set, which is 210 because of the 70 % training ratio. The parameter α is the arbitrary scaling factor which is located usually between 2 and 10. I tested the different α variables on the time series and I found that the 16 nodes solution for β_0 and β_1 is the best, and the β_2 required 40 nodes for the highest performance in forecasting.

Choosing more nodes increases the required time for running and leads to an increased risk of over-training. These costs represent barriers for the developing of the neural network. I tested the 80:20:20 training ratio next to the default proposed 70:15:15 and I found that the last setting is better for this time series. The reason is based on the risk of over-training. If I selected too many elements into the training set, the neural network could be probably over-trained.

The properties of the neural network:

- Architecture: Feedforward Nonlinear Autoregressive Network
- Single β prediction:
 1. 1 input node on the input layer
 2. 16 nodes for β_0 and β_1 on the hidden layer
 3. 40 nodes for β_2 on the hidden layer
 4. 1 output node on the output layer
- Multi β prediction:
 1. 3 input node on the input layer
 2. 25 nodes on the hidden layer

3. 3 output node on the output layer

- Delay = 1 day
- Training ratio: 70 %, Validation: 15 %, Test: 15 %
- Training algorithm: Levenberg-Marquardt
- Performance function: mean squared normalized error
- Activation in hidden neurons: logarithmic-sigmoid function
- Activation in output neuron: linear transformation function
- Maximum number of epochs: 1000
- Number of reiterations: 30
- Software: MATLAB R2016a
- Applied function for neural network architecture: narnet()

I selected the same delay for every β variables because the tests show that 1 day delay is proper for this forecasting. There are certain ranges where a 12 days delay would be better for β_2 , but the main object of this research is to find a model which is consistently proper for predicting the IR term structure. When a model is applied, the users cannot decide if the solution is wrong or there is an innovation in the input time series or it is too volatile to predict the future. They simply accept that there are days when the predictions do not work – for any reason -, and wait for the next forecasting. On long term they should update the neural network by modification of node's number or in an extremely unlikely case putting a new layer into the network, but this is the model development process. The model should be consistent in time and the certain – short – ranges with other neural network settings are ignored, because my goal is to find a solution with relevance for using today with reliable results from back-tests.

Chapter 6

Comparing of results

6.1 Evaluation of the predictions

The observations can differ from the expected values, and an appropriate error measurement is needed to evaluate the models. I impose a set of stringent requirements: firstly, to ensure that the error measurement system is able to show the asymmetry of errors, secondly, the errors should be comparable with each other. The model validation should be based on reliable error indicators, otherwise the application of the model in real life would generate loss or not the largest available profit compared with alternative models.

There is no consensus about which error measure is the best because there are researches with contradictory results about this topic (T.Chai and Draxler, 2014). Theoretically the RMSE (6.2) cannot dominate the performance of MAE indicator (6.1), and it would be better to avoid RMSE because of its misleading measurement (Willmott, Matsuura, and Robeson, 2009). Chai and Draxler proved that the RMSE is an accurate indicator to measure statistical error if the error follows a Gaussian distribution.

$$MAE = \frac{1}{N} \sum_{i=1}^N |x_{pi} - x_{di}| \quad (6.1)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (c_i - \bar{c}_i)^2} \quad (6.2)$$

One of the biggest advantages of the RMSE is based on the calculation of quadratic loss function, and the biggest disadvantage of RMSE is the absolute measure of errors (T.Chai and Draxler, 2014). The absolute values cannot help to understand the direction of the differences and its changes on the time series. I chose the RMSE to quantify the uncertainty in the predictions of models because it is scientifically proven that the RMSE is able to present the uncertainty of models. It is also implemented and applied in many studies like that of Diebold and Li (2006), and Barunik-Malinska's (2015), which are highly cited researches related to forecasting by Nelson-Siegel model.

I prefer the property of RMSE that is very sensitive for the outliers. The outliers can produce very high differences, especially if the level's forecasts are wrong, hence I apply RMSE to measure the error of single β parameter's forecasts and the model's error. The RMSE would be not enough for the right evaluating process because of its deficiency to recognise the direction of the error. It is possible that when I predict the β parameters, the

consistent errors which are concentrated in a given direction produce lower model error – substituting them into the Nelson-Siegel model – than those models which have a white-noise error. For this reason I measure the residuals of β prediction and draw them on plots to analyse its trend on the time series.

Description of idealistic β parameters:

1. Uncorrelated residuals of β estimations 2. Errors follow a Gaussian distribution for every β

The first point is important for the best explanatory power of the model. If there were high autocorrelations between the residuals, there must be a factor not included in the model. This case is not acceptable for a good model. The second point is important for the perfect functioning of RMSE, because the RMSE can be a reliable error indicator, if the error follows normal distribution.

I defined the proper model description as well:

1. Better evaluated forecasting performance by Diebold-Mariano test statistics on β variables 2. Lowest RMSE calculated from estimated interest rates by maturities (preferably for the short term maturities)

RMSE can show which model can predict the future value more accurately and its sensitivity for outliers can be a useful feature.

Diebold-Mariano test statistics provide information about the accuracy of models (Diebold, 2013). The DM test is based on the calculation of a loss function from the errors, then it calculates the loss differential which should be covariance stationary if the two predictions are equally accurate (Diebold, 2013). I implemented a method with null hypothesis for my research: the neural networks prediction and the alternative model are equally accurate. The alternative hypothesis submits that the alternative model is less accurate than the neural network. This test was applied for every standard maturity because the more volatile and more important short end of the term structure is highly critical to be checked precisely.

6.2 Expectations

For one-day prediction the random walk must be the best model, because the high uncertainty is a good basis to implement this model. The second best model for one-day prediction should be the vector autoregressive or autoregressive, because I do not consider, that the performance of the neural network would be satisfying in an environment, where catching patterns can be dominated by random walk.

The five-days forecasting is the environment, where the autoregressive, vector autoregressive model and neural network can show their performance and dominate the random walk on certain periods of the analysed time series. There could be differences in the performance by maturities, but this depends also on the error generated by the fitting of Nelson-Siegel model (Diebold and Li, 2006). The best model for the five-days prediction should be the neural network, and the vector autoregressive model could be the second best.

The ten-days prediction could produce highly better by neural network than vector autoregressive or autoregressive model. This comparing should

show that the random walk is dominated by other models, because its horizontal forecasts – without any trend – must generate bigger error than a well parametrised neural network or an AR/VAR model which is able to catch trends.

I expect that the biggest differences in the performance of models will be visible on the short end of the interest rate's term structure. This assumption is based on the previous analysis of the Hungarian term structure of interest rates' changes, where I conclude that the short end is more volatile than the relative stable long end. The less volatile end of the term structure could be estimated by linear models very good, thus the autoregressive model and random walk should have closely the same results.

The prediction power of dynamic Nelson-Siegel model is better for longer horizons (Malinska and Barunik, 2015), (Mönch, 2005), for this reason the forecast's RMSE should decrease by expanding the prediction's horizon.

6.3 Results

The analysis of results is based on the maturities and different periods of the time series. Both analyses aim at making sure that, the models are consistently accurate in time and reliable for every standard maturity – especially for the sort end of the term structure.

There were expectations that VAR can catch the information integrated in the cross correlation, and the random walk is accurate both for short term forecasts, while neural network can find patterns, which useful if the thesis is true about the history, that it repeats itself.

The analysed time series of Hungarian interest rates contains different periods, thus the performance of models could be tested in variant environments. This is useful if the model is built with a view to apply it for real trading. The more volatile periods with their non-linear connections and the periods of slowly changing or largely unchanged term structures without shocks provides the opportunity to highlight the advantages and disadvantages of variant models.

I divided the time series into five parts, where the RMSE values can be connected to known periods like the range where the negative inflation expectations were dominant, and the range of the financial crisis. This clusters are created by the analysing of the Hungarian interest rate term structure's changes. I wanted to separate them by their attributes.

6.4 Comparing the β parameters' forecasts

The right β prediction is the basis of the proper forecasting, because the errors are multiplied by the loadings of β variables after the substitution into the Nelson-Siegel model. Minimising the prediction error of β parameters is the first point, where the forecasting issues can be managed, but selecting a model for applying it in the real life should be based on other considerations as well. For this reason there is a model error calculation, too, which can show the error generated by the combination of parameter prediction error and any fitting related incompatibility of Nelson-Siegel model.

6.4.1 RMSE of β parameters

The RMSE calculated by models and β variables can show the forecasting error of the different models. It is important to note, that the β loadings are playing an important role in this analysis, because the errors are multiplied by them in Nelson-Siegel model, hence a wrong β_0 with accurate β_1 and β_2 must be not enough for an accurate and reliable forecast.

Table 6.1 shows the one-day-ahead forecasts' RMSE values for the different β parameters calculated on the full time series.

Variable	NN (All β)	NN (Single β)	VAR(1)	AR(1)	RW
β_0	0.0017	0.0016	0.0016	0.0016	0.0016
β_1	0.0018	0.0018	0.0018	0.0018	0.0018
β_2	0.0052	0.0048	0.0048	0.0046	0.0047

TABLE 6.1: The RMSE numbers above are multiplied by 1.0^4 . 1-day-ahead forecasts' RMSE by predicted β variables and models

I checked the mean of errors to collect information from the asymmetry of errors. Theoretically the mean of errors must be zero, because the direction of wrong predictions can be negative or positive with equal probabilities, but in the practice there may be differences.

The random walk and AR(1) models have the same RMSE of β_0 and β_1 , and the only difference on β_2 parameters RMSE is not so significant because of the β_2 loadings effect. The random walk is not the best model in comparing of RMSE values on β variables, but the two times bigger mean of errors on β_0 comparing to AR(1) models RMSE can effect significant difference in the Nelson-Siegel model. This information can be useful for the next analysis.

Variable	NN (All β)	NN (Single β)	VAR(1)	AR(1)	RW
β_0	0.3138	0.0486	0.2589	0.0344	0.0689
β_1	0.0171	0.2673	-0.2937	0.1249	0.1484
β_2	0.5074	-0.5434	-0.2244	-0.0938	0.1455

TABLE 6.2: 1-day-ahead forecasts' RMSE by predicted β variables and models

The 5-days-ahead forecasts have different order for the best models by RMSE. The coming day can be similar to the previous one, but increasing the forecasting horizon the prediction must be more difficult and the RMSE values should show bigger differences among models RMSE values.

The model with the lowest RMSE in table 6.3 is the neural network with single β input. In every row the single-input neural network dominates and its nearest competitor is the multi input neural network. This example proves the long term prediction performance of the neural network, because the differences between the models-errors is bigger than in the case of one-day-ahead forecasts, and the single-input neural network can be an absolute winner.

It is visible, that the random walk produces the biggest RMSE among forecasting models (table 6.3). This caused by the horizontal line, which

represents the predicted points of random walk for β variables. The error between the horizontal line and the observable points can be get ever bigger over time, if the time series has a trend on the analysed period of time.

Variable	NN (All β)	NN (Single β)	VAR(1)	AR(1)	RW
β_0	0.0029	0.0024	0.0029	0.0029	0.0033
β_1	0.0033	0.0031	0.0037	0.0037	0.0042
β_2	0.0084	0.0076	0.0092	0.0086	0.0090

TABLE 6.3: 5-days-ahead forecasts' RMSE by predicted β variables and models

In the table below (table 6.4) there are 10-days-ahead forecasts' RMSE values for the different analysed models by β variables. The best model is the single β -input neural network, and the second best is the multi-input neural network. The differences among the RMSE values of models for the estimation of level can be a basis for higher benefit for neural networks' model-error. Because of the highest sensitivity of β_0 loading this differences can cause a big advantage for the neural networks.

The RMSE of multi-inputs neural network is bigger than the single-input neural network's one and the RMSE of VAR(1) model is higher than the AR(1) model's one for every β variable. This effect must be caused by the low value of the information contained by the cross correlation and its change over time. The single-parameter predictions outperformed the multi-inputs models, hence it is possible to build accurate model without any knowledge about the cross correlation of variables.

Variable	NN (All β)	NN (Single β)	VAR(1)	AR(1)	RW
β_0	0.0031	0.0027	0.0035	0.0035	0.0041
β_1	0.0037	0.0037	0.0048	0.0048	0.0055
β_2	0.0093	0.0090	0.0126	0.0118	0.0124

TABLE 6.4: 10-days-ahead forecasts' RMSE by predicted β variables and models

6.4.2 Diebold-Mariano test

For the one-day-ahead forecasting I tested Diebold-Mariano for the predicted β parameters with null hypothesis that the random walk and the compared model are equally accurate. The alternative hypothesis assumes that the compared model is less accurate than the random walk. This test corresponds to the theory of my research, that the random walk is a strong forecasting model because of the uncertainty of the interest rates' movements.

The Diebold-Mariano test shows, that the one-day-ahead forecasts of random walk dominates every model in accuracy of β prediction. The fail of the neural network based on multi β prediction shows that neither the linear, nor the non-linear models can forecast more accurately the short-term movements of interest rates.

The 5-days-ahead forecasts' Diebold-Mariano-test shows an important change comparing to the one-day-ahead prediction. The neural network

based on multi- β prediction and the single- β neural network are not similarly accurate for 5-days-ahead forecasts horizon. The single- β neural network dominates all of the models in every analysed period.

The multi β neural network and the single β neural network are not equally accurate in the prediction of β_0 and β_1 for 10-days horizon. The single- β neural network dominates every model in this case, as well, hence the its accuracy for longer forecasting horizons is proved by RMSE calculated by β variables and by Diebold-Mariano test, too. In the next section the model error is playing an important role, because the substitution into the Nelson-Siege model to get the interest rates term structure can change the performances of the models.

6.5 Model error

6.5.1 Comparing by RMSE after substitution into Nelson-Siegel model

I calculated RMSE of predicted interest rate points for every model by five distinct periods to show the difference in prediction performance on variant ranges of the time series.

The best model for one-day-ahead predictions is the random walk. In every period the RMSE of the random walk was smaller than or equal to the other model's one. The nearest competitor of random walk is the autoregressive model, because the models have the same RMSE in two distinct period from five. However, the single- β neural network's and the autoregressive, vector autoregressive models have the same RMSE rounded to the fourth decimal place for the full time series, It must be noted, that the AR(1) has a bit smaller number. These results confirmed that the random walk can be an accurate prediction model.

Start	End	NN (All β)	NN (single β)	VAR(1)	AR(1)	RW
05-Apr-2002	19-Jan-2005	0.0017	0.0016	0.0016	0.0016	0.0014
19-Jan-2005	31-Oct-2007	0.0008	0.0007	0.0008	0.0007	0.0007
31-Oct-2007	30-Mar-2010	0.0022	0.0021	0.0021	0.0021	0.0020
30-Mar-2010	14-Jan-2013	0.0014	0.0013	0.0013	0.0012	0.0012
14-Jan-2013	22-Feb-2016	0.0009	0.0009	0.0008	0.0008	0.0008
05-Apr-2002	22-Feb-2016	0.0015	0.0014	0.0014	0.0014	0.0012

TABLE 6.5: 1-day-ahead forecasts' RMSE by periods and models

In the last period every model has similar RMSE to the random walk, but in the other years the neural network based on multi- β predictions are dominated by every competitor. The biggest RMSE in the table is generated by multi- β prediction in the period of financial crisis, which shows that the forecasts in the range of highest volatility is so unpredictable, that the learnt pattern for single factors cannot help. A simple random walk can be the best tool to predict variables on financial time series in crisis, because it avoids the consequent fails in predictions and this can guarantee a good result in average.

Previously I showed table 6.2 about the mean of errors calculated by β variables and models. The autoregressive model and random walk had

very similar RMSE values for every β , hence I wanted to check the asymmetric in errors. It is considered from the table 6.2, that the directions of random walk's errors concentrated more on the positive range compared to the autoregressive model. The same β -RMSE with more overestimation produced a better model-RMSE for random walk. It can be inferred from this fact, that the average increasing of the level of interest rates provides a good basis for overestimations of β_1 parameters. The overestimation of other variables are not so significant because of the β loadings and its relative small differences between the autoregressive and random walk models.

The 5-days-predictions show highly different results from the one-day's one. The models has more variant RMSE results and the order of best models changed. Predictions for longer horizons can favour for the Nelson-Siegel model, as previously mentioned, hence there are good expectations from the pattern recognition ability of neural networks too.

The RMSE is higher for every model in the period of financial crisis as set out in the table below. The reason behind this effect must be based on the higher standard deviation of interest rate points (Appendix A.2) measured in this range. The standard deviation of interest rates before and after the financial crisis is half as much as in it for most of the maturities. The RMSE of 5-day-forecasts is more than two times bigger in the crisis than before and after, thus the connection between the standard deviation of interest rates and the RMSE of β forecasting is not linear in this case. For the one-day-prediction the increasing of RMSE in the financial crisis is similar to the increasing of standard deviation of interest rates.

Start	End	NN (All β)	NN (Single β)	VAR(1)	AR(1)	RW
05-Apr-2002	19-Jan-2005	0.0033	0.0028	0.0035	0.0033	0.0034
19-Jan-2005	31-Oct-2007	0.0016	0.0011	0.0014	0.0014	0.0016
31-Oct-2007	30-Mar-2010	0.0041	0.0036	0.0043	0.0042	0.0046
30-Mar-2010	14-Jan-2013	0.0029	0.0020	0.0024	0.0024	0.0027
14-Jan-2013	22-Feb-2016	0.0016	0.0013	0.0015	0.0014	0.0018
05-Apr-2002	22-Feb-2016	0.0029	0.0024	0.0029	0.0028	0.0030

TABLE 6.6: 5-days-ahead forecasts' RMSE by periods and models

The table 6.6 shows the results for the 5-day forecasts by different models implemented in this research. The 5-day prediction compared to the 1-day forecasts shows change in order of model RMSE values. The best model for every period is the single- β inputs neural network, and the random walk has the lowest performance. It is important to note that the multi- β neural network and the vector autoregressive model has a relative low performance, which can be caused by the wrong application of cross-correlation. I expected that the cross-correlation could create added value for those models, which are able to incorporate it. Analysing the five-days-ahead forecasts, it seems, that this hypothesis must be rejected.

In average the RMSE is relative decreasing after the crisis and its value in period 2005-2007 is approximately equal to the period 2013-2016 for every model. This corresponds to theory that the less volatile periods are better predictable or there were fewer unexpected movements in the term structure in these ranges comparing to the crisis.

Start	End	NN (All β)	NN (Single β)	VAR(1)	AR(1)	RW
05-Apr-2002	19-Jan-2005	0.0040	0.0035	0.0046	0.0044	0.0047
19-Jan-2005	31-Oct-2007	0.0019	0.0015	0.0022	0.0021	0.0024
31-Oct-2007	30-Mar-2010	0.0050	0.0046	0.0066	0.0062	0.0067
30-Mar-2010	14-Jan-2013	0.0035	0.0024	0.0034	0.0031	0.0035
14-Jan-2013	22-Feb-2016	0.0019	0.0015	0.0020	0.0018	0.0024
05-Apr-2002	22-Feb-2016	0.0035	0.0029	0.0041	0.0038	0.0042

TABLE 6.7: 10-days-ahead forecasts' RMSE by periods and models

The 10-days-ahead forecasts' RMSE shows (table 6.7) a bigger differences – as expected – among the RMSE values. The single- β neural network holds its position on the top of the performance ranking. There is an interesting effect: the VAR(1) has similar result to the random walk, thus they are the worst models to forecast the 10-days-ahead interest rates. The multi-input neural network comparing to the single-input's one has a lower performance, hence it seems, that the models based on multi- β inputs forfeit the advantage of getting patterns in cross-correlation for longer horizons.

The multi- β neural networks is the best model in the comparison, but this benefit is not significant because of the single- β neural network's almost same performance.

The random walk has better RMSE for every period comparing with AR(1) and VAR(1), hence it is the nearest competitor of the neural networks in prediction's error.

6.5.2 Residuals from predicted β variables

Previously I defined the properties of a perfect model, hence I test the change of autocorrelation and the normality of residuals.

First, I compared the autocorrelation of residuals calculated by models, periods and different ends of the term structure.

The table blow shows the best models which corresponds to the previously results by analysing of RMSE. For every maturity in the 5-days-ahead forecastings results, the multi β prediction has the lowest autocorrelation in residuals calculated by the difference of original interest rate and the forecasted one. The 10-days-ahead and the one-day-ahead predictions' results confirm the performance of selected models.

As the table 6.8 shows the autocorrelation of residuals in every model - except the random walk - both for short and long end of the term structure is high. Relatively the long end has lower autocorrelation in residuals in the 2002-2005 period, but later this trend was changed. I expected, that the quality of forecasting must be better for the long end among every model, because of the low standard deviation of the interest rates on long end. It is visible in the table 6.8, that there are bigger autocorrelations for the long end in period 2013-2016, which do not correspond to my expectation, but the forecasting performance can be proper independently from this result.

Start	End	1 day - Short	1 day - Long	5 days - Short	5 days - Long	10 days - Short	10 days - Long
05-Apr-2002	19-Jan-2005	RW (0.1527)	RW (-0.0631)	NN-S (0.566)	NN-S (0.4276)	NN-All (0.7588)	NN-S (0.6739)
19-Jan-2005	31-Oct-2007	RW (0.1108)	RW (0.098)	NN-S (0.461)	NN-S (0.4049)	NN-All (0.5324)	NN-All (0.544)
31-Oct-2007	30-Mar-2010	RW (0.0763)	RW (0.1652)	NN-S (0.487)	NN-S (0.5108)	NN-All (0.6804)	NN-All (0.6617)
30-Mar-2010	14-Jan-2013	RW (0.0408)	RW (0.0378)	NN-S (0.4214)	NN-S (0.3367)	NN-All (0.6311)	NN-S (0.6367)
14-Jan-2013	22-Feb-2016	RW (0.037)	RW (0.0268)	NN-S (0.3978)	NN-S (0.3994)	NN-S (0.6132)	NN-S (0.6444)

TABLE 6.8: Lowest mean autocorrelations of residuals for short (3 months - 24 months) and long maturities (36 months - 120 months) for lag=1 by periods and models

Every model in the period of financial crisis has an increase in the autocorrelation of residuals. This shows that one or more new variables are required for better forecasting especially in this range.

The table 6.9 presents the worst models in term of residuals' quality, where I collected the greatest autocorrelations by its absolute values. The random walk model dominates in the worst table, but the vector autoregression has similarly bad results for one-day-ahead forecasting horizon in recent years. The change in cross-correlation of variables can reduce the performance of vector autoregressive model.

Random walk shows huge autocorrelation for longer forecasting horizons, which corresponds to the previous results related to RMSE and accuracy, as well. The huge autocorrelation of residuals highlights that the model consistently fails to predict the future. The residuals should be modelled to understand the information which is not involved into the basic model. For this reason there are hybrid neural networks, where the non-linear connection of residuals are analysed (Adhikari and Agrawal, 2013).

Start	End	1 day - Short	1 day - Long	5 days - Short	5 days - Long	10 days - Short	10 days - Long
05-Apr-2002	19-Jan-2005	VAR (0,8365)	VAR (0,7948)	RW (0,8449)	RW (0,7745)	RW (0,9196)	VAR (0,8675)
19-Jan-2005	31-Oct-2007	AR (0,766)	AR (0,7358)	RW (0,8567)	RW (0,8231)	RW (0,9423)	RW (0,9215)
31-Oct-2007	30-Mar-2010	AR (0,627)	AR (0,633)	RW (0,8439)	RW (0,8251)	RW (0,9417)	RW (0,9142)
30-Mar-2010	14-Jan-2013	AR (0,6442)	AR (0,7244)	RW (0,7966)	RW (0,7864)	RW (0,8994)	RW (0,8852)
14-Jan-2013	22-Feb-2016	VAR (0,7694)	VAR (0,8404)	RW (0,8271)	RW (0,8211)	RW (0,9245)	RW (0,9087)

TABLE 6.9: Worst mean autocorrelations of residuals for short (3 months - 24 months) and long maturities (36 months - 120 months) for lag=1 by periods and models

I checked the normality of the residuals of models by Kolmogorov-Smirnov test, because previously I defined that a good residual follows a Gaussian process. For all models I had to reject the null hypothesis, because p-value was extremely low. I could reject the normality on level 99.99 %.

I found that the extremely small p-value for null hypothesis is caused by the huge kurtosis of the distribution of residuals. The one-day-ahead forecasts smallest kurtosis is 27.7053 for β_0 , 21.0632 for β_1 by all-inputs neural network, 26.7131 for β_2 by VAR(1). The 5-days-ahead forecasts have smaller minimum for kurtosis, but wider range for kurtosis-values, because the range is between 11.5733 and 44.7668. The 10-days-ahead forecasts have a smaller range: 6.1188 and 16.3965, but this values are far from the optimal -2 and +2 range of kurtosis to achieve the normal distribution. The skewness is close to zero in every case, thus the kurtosis is the only reason why the errors are not following Gaussian process.

I think, that the extreme concentration of the residuals at zero, can be an advantage for the analysis, because the small errors of predicted β variables

can preserve the rankings by the β -prediction error on the level of model-error.

6.5.3 RMSE by maturities

Checking the RMSE by maturities is important for the understanding the models' sensitivity for the volatility connected to different maturities. From the trading perspective it is important to note, that the short end has higher focus. If a model had low RMSE on long end, and in average a relative good RMSE comparing to other models, but it had a bigger RMSE on short end of term structure, it can be, that I had to reject it because of its incompetence to forecast accurately the short end.

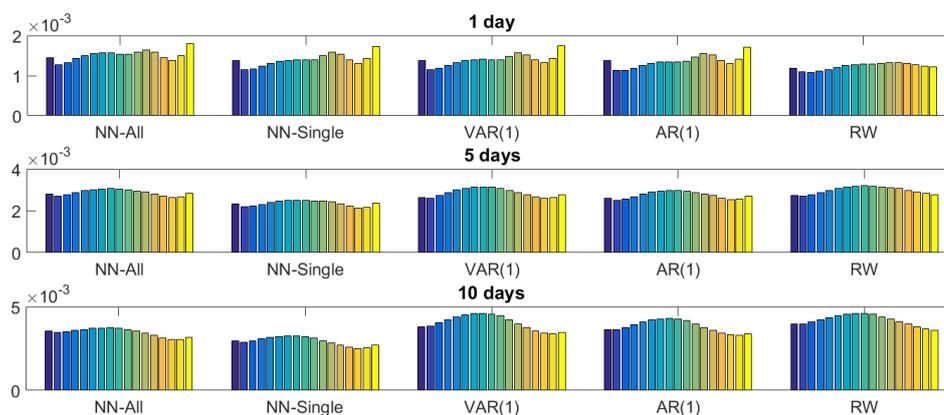


FIGURE 6.1: RMSE by maturities and models for different horizons

The best predicted maturity is the 6th month on every horizon, but the worst predicted one changes by increasing the horizon. It is an interesting effect that the highest point of RMSE shifts from the 72nd months left to the 30th month maturity by increasing the forecast's horizon. If a fixed income or an IRD trader would like to know the future term structure, he or she should take into consideration the changing distribution of RMSE by the length of forecast. This effect has a presence in all models.

The random walk has the most symmetric shape on every time horizons, because it is independent from consequence errors, but not independent from the standard deviation of interest rates by maturities. However, the standard deviation of interest rates is falling from the smallest maturity to the last one (Appendix table A.2), but the shape of the random walk's error by maturities is not similar to it. The random walk is strictly decreasing from the highest RMSE value in the direction of long maturities. The less volatile interest rates must be an ideal environment for random walk's forecasting.

The neural network based on multi β inputs dominates every other model on every maturity. Its RMSE distribution is similar to other models, thus it has the same sensitivity to the effect of hidden factors, unexpected events, but to a lesser extent.

Chapter 7

Conclusion

The Hungarian term structure of interest rates was a good basis for back-tests, because it represents all of the main types of the interest rate's curves on variant ranges of the time series.

The best model for one-day-ahead prediction is the random walk. The unpredictable coming day's interest rate can be forecasted by the simplest method. The neural networks can not achieve the level of the random walk model's performance on this short horizon. The pattern recognition ability requires a longer forecasting horizon to produce the best performance.

The random walk model is not proper for the predictions on longer horizons, because there are trends, which can be recognised even by autoregression, and the high autocorrelation of the β variables suggests, that this trends can be permanent on 5-days or 10-days horizons. The horizontal line which represents the random walk's forecast is not sufficient to predict the Hungarian term structure of interest rates for 5- or 10-days ahead.

For 5- and 10-days forecasts the single-input neural network is the best solution. I had a hypothesis about the strong forecasting performance of the multi- β neural network on longer horizons, but I found, that all of the multi-input models like vector autoregression and multi- β neural network are outperformed by their single-input competitor like autoregressive model and single-input neural network. It follows that, the cross-correlation of β variables can not help to build a better model. If this thesis is true, the principal component analysis would have a relevance in the study of term structure's forecasting.

The rank of models based on RMSE of β variables and the rank based on the estimated interest rates is the same. This could be caused by the well fitting Nelson-Siegel model, which was proved in the research, that it dominates the Nelson-Siegel-Svensson model's performance. The proper fitting of the term structure by Nelson-Siegel model can reduce the predicted β variables' error, but because of the narrowed dimensions the forecasting performance may decrease. The same orders of the performance by β and model-error show, that I could find a proper yield curve model and optimal λ combination. The trade-off between modelling and forecasting was managed successfully in the research.

The maturity 6th month has the best RMSE calculated by every model. This is important for the IRD-traders, who are interested in the short end of the term structure, which is relevant in Hungary.

I am satisfied with the results, because I could prove the accuracy and reliability of neural networks, and I could highlight the relevance of the simplest stochastic model on the Hungarian interest rates: the random walk.

This model is more reliable than the autoregressive model for one-day-ahead, and significantly better than the neural networks' forecasts for one-day horizon. Two very different approaches were presented, and both of them are playing an important role in the world of forecasts.

In the future there are studies what should be carried out related to the Hungarian term structure of the interest rates. First of all, I think, it would be important to develop a time-delay neural network for single- β predictions. This architecture can be a good competitor for the non-linear autoregressive neural network. The multi- β neural network requires another parametrisation, because the rerunning of the back-test is very long time, thus I could not test every setting. Another outlier handling could help to clean the input data for the models. Beyond the possible methodological changes, it would be useful to test the principal component analysis for this time series and produce forecasts.

The results from this research provide a good basis for further studies and developing basic trading strategies by the predicted term structures. The programmed MATLAB scripts and the C # code for downloading the zero-coupon interest rates could prosecute technically a new research related to Hungarian term structure of interest rates.

The econometric methods and neural networks are important in the modelling and forecasting, as well. This research proved, that the prediction performances of these models are dependent on the different forecasting horizons, too. The short term is proper for econometric methods, and the longer forecasting horizons can be estimated better with single-input neural network. I hope, that the neural networks will be promoted for further researches related to term structures in the future.

Appendix A

Appendix

<i>Maturity</i>	<i>Mean</i>	<i>Std.Dev.</i>	<i>Min.</i>	<i>Max.</i>	<i>MAE</i>	ρ_1	ρ_{15}	ρ_{30}
3.0	0.3927	1.604	-12.3	18.96	0.000518	0.9146	0.5148	0.2189
6.0	0.3139	0.8845	-3.249	11.52	0.0001552	0.9642	0.7432	0.5333
9.0	0.0988	0.5724	-1.738	4.928	0.0001917	0.9739	0.7696	0.6412
12.0	-0.1428	0.7378	-5.87	4.708	0.0002974	0.9272	0.5164	0.2663
15.0	-0.342	1.004	-9.958	5.535	0.0003447	0.9331	0.5666	0.3256
18.0	-0.456	1.113	-10.59	5.482	0.0003402	0.9404	0.6329	0.4275
21.0	-0.4673	1.023	-8.719	4.837	0.0003069	0.9416	0.6669	0.4946
24.0	-0.3996	0.8118	-6.864	3.658	0.0002616	0.9374	0.6725	0.5281
30.0	-0.162	0.4225	-3.076	1.77	0.0001712	0.9247	0.6455	0.5273
36.0	0.1037	0.5217	-2.245	4.359	0.0002433	0.9428	0.6046	0.4167
48.0	0.5415	1.054	-6.103	7.065	0.0005165	0.9285	0.5723	0.3871
60.0	0.7384	1.224	-7.109	6.723	0.0006561	0.9029	0.534	0.3404
72.0	0.6279	1.032	-5.949	6.064	0.0006029	0.8449	0.4732	0.2679
84.0	0.3242	0.7043	-7.785	8.152	0.0003928	0.6571	0.3822	0.2272
96.0	-0.06388	0.5376	-8.89	10.26	0.0001535	0.2674	0.1818	0.1744
108.0	-0.4625	0.876	-9.005	12.62	0.0004569	0.5778	0.3255	0.1974
120.0	-0.8824	2.412	-65.82	20.92	0.000951	0.3379	0.1046	0.07822

FIGURE A.1: Summarising the basic properties of Nelson-Siegel model's parameters

Start	End	NN (All Beta)	NN (Single Beta)	RW	VAR(1)	AR(1)
05-Apr-2002	19-Jan-2005	0.2219	0.2191	0.1527	0.8365	0.583
19-Jan-2005	31-Oct-2007	0.1477	0.2268	0.1108	0.4857	0.766
31-Oct-2007	30-Mar-2010	0.1577	0.2174	0.0763	0.3265	0.627
30-Mar-2010	14-Jan-2013	0.0927	0.1209	0.0408	0.41	0.6442
14-Jan-2013	22-Feb-2016	0.2487	0.2687	0.037	0.7694	0.7148

TABLE A.1: Mean autocorrelation of 1-day-ahead predictions' residuals for short end maturities (30 months - 120 months) for lag=1 by periods and models

Start	End	NN (All Beta)	NN (Single Beta)	RW	VAR(1)	AR(1)
05-Apr-2002	19-Jan-2005	0.1518	0.1463	-0.0631	0.7948	0.581
19-Jan-2005	31-Oct-2007	0.1561	0.2829	0.098	0.5438	0.7358
31-Oct-2007	30-Mar-2010	0.1899	0.3193	0.1652	0.3749	0.633
30-Mar-2010	14-Jan-2013	0.139	0.1358	0.0378	0.6339	0.7244
14-Jan-2013	22-Feb-2016	0.1754	0.2359	0.0268	0.8404	0.7684

TABLE A.2: Mean autocorrelation of 1-day-ahead predictions' residuals for long end maturities (3 months - 24 months) for lag=1 by periods and models

Start	End	NN (All Beta)	NN (Single Beta)	RW	VAR(1)	AR(1)
05-Apr-2002	19-Jan-2005	0.642	0.566	0.8449	0.8171	0.78
19-Jan-2005	31-Oct-2007	0.4989	0.461	0.8567	0.7565	0.7794
31-Oct-2007	30-Mar-2010	0.6347	0.487	0.8439	0.8067	0.8009
30-Mar-2010	14-Jan-2013	0.5042	0.4214	0.7966	0.7758	0.7311
14-Jan-2013	22-Feb-2016	0.6135	0.3978	0.8271	0.7897	0.6913

TABLE A.3: Mean autocorrelation of 5-days-ahead predictions' residuals for short end maturities (3 months - 24 months) for lag=1 by periods and models

Start	End	NN (All Beta)	NN (Single Beta)	RW	VAR(1)	AR(1)
05-Apr-2002	19-Jan-2005	0.5852	0.4276	0.7745	0.7607	0.696
19-Jan-2005	31-Oct-2007	0.5102	0.4049	0.8231	0.7482	0.7538
31-Oct-2007	30-Mar-2010	0.6435	0.5108	0.8251	0.7744	0.7771
30-Mar-2010	14-Jan-2013	0.5795	0.3367	0.7864	0.7116	0.7039
14-Jan-2013	22-Feb-2016	0.6042	0.3994	0.8211	0.729	0.6813

TABLE A.4: Mean autocorrelation of 5-days-ahead predictions' residuals for long end maturities (30 months - 120 months) for lag=1 by periods and models

Start	End	NN (All Beta)	NN (Single Beta)	RW	VAR(1)	AR(1)
05-Apr-2002	19-Jan-2005	0.7588	0.8221	0.9196	0.9096	
19-Jan-2005	31-Oct-2007	0.5324	0.7696	0.9423	0.9131	0.9019
31-Oct-2007	30-Mar-2010	0.6804	0.7823	0.9417	0.925	0.9244
30-Mar-2010	14-Jan-2013	0.6311	0.6758	0.8994	0.8927	0.9232
14-Jan-2013	22-Feb-2016	0.7231	0.6132	0.9245	0.9021	0.8741

TABLE A.5: Mean autocorrelation of 10-days-ahead predictions' residuals for short end maturities (3 months - 24 months) for lag=1 by periods and models

Start	End	NN (All Beta)	NN (Single Beta)	RW	VAR(1)	AR(1)
05-Apr-2002	19-Jan-2005	0.6928	0.6739	0.8639	0.8675	0.8502
19-Jan-2005	31-Oct-2007	0.544	0.6997	0.9215	0.89	0.896
31-Oct-2007	30-Mar-2010	0.6617	0.7231	0.9142	0.8887	0.8895
30-Mar-2010	14-Jan-2013	0.6379	0.6367	0.8852	0.859	0.8577
14-Jan-2013	22-Feb-2016	0.6991	0.6444	0.9087	0.8811	0.8672

TABLE A.6: Mean autocorrelation of 10-days-ahead predictions' residuals for long end maturities (30 months - 120 months) for lag=1 by periods and models

Bibliography

- Aas, Kjersti and Xenia K. Dimakos (2004). *Statistical modelling of financial time series: An introduction*. URL: www.nr.no/files/samba/bff/SAMBA0804.pdf (visited on 04/12/2016).
- Adhikari, Ratnadip and R. K. Agrawal (2013). *A combination of artificial neural network and random walk models for financial time series forecasting*. URL: link.springer.com/article/10.1007%2Fs00521-013-1386-y (visited on 04/10/2016).
- Afonso, Antonio and Manuel M.F. Martins (2010). "Level, slope, curvature of the sovereign yield curve, and fiscal behaviour". In: *WORKING PAPER SERIES* 1276. URL: www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1276.pdf.
- Aljinovic, Zdravka, Tea Poklepovic, and Kristina Katalinic (2012). "BEST FIT MODEL FOR YIELD CURVE ESTIMATION". In: *Croatian Operational Research Review* Vol. 3, pp. 28–40. URL: hrcak.srce.hr/file/142293.
- Bajracharya, Dinesh (2011). *ECONOMETRIC MODELING VS ARTIFICIAL NEURAL NETWORKS*. URL: bada.hb.se/bitstream/2320/7986/1/2010MI17.pdf (visited on 04/11/2016).
- Balter, Anne, Antoon Pelsser, and Peter C. Schotman (2014). "What Does a Term Structure Model Imply About Very Long-Term Discount Rates?" In: *SSRN Electronic Journal* 1.
- BIS (2016). *BIS Statistics*. URL: stats.bis.org/statx/srs/table/d5.1 (visited on 04/15/2016).
- Bjork, T. and B.J. Christensen (1999). "Interest Rate Dynamics and Consistent Forward Rate Curves". In: *Mathematical Finance* 9, 323–348.
- Burnham, Kenneth P. and David R. Anderson (2003). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer.
- Choudhry, Mooraad, Rod Pienaar, and Michele Lizzio (2004). *Advanced Fixed Income Analysis*. Butterworth-Heinemann, Oxford, p. 118.
- Coons, James W. (2015). *Predicting Turning Points in the Interest Rate Cycle*. Routledge, Abingdon-on-Thames.
- Coroneo, Laura, Rositsa Ken Nyholm, and Vidova-Koleva (2008). "How Arbitrage-Free is the Nelson-Siegel Model?" In: *ECB Working Paper* 874.
- Coroneo, Laura, Ken Nyholm, and Rositsa Vidova-Koleva (2008). "HOW ARBITRAGE-FREE IS THE NELSON-SIEGEL MODEL?" In: *European Central Bank - Working Paper Series* 874. URL: www.ecb.europa.eu/pub/pdf/scpwps/ecbwp874.pdf.
- Credit Suisse - *Debt capital and interim financing to hedge interest rate risks*. (2016). URL: www.credit-suisse.com/ch/en/unternehmen/oerk/derivate.html (visited on 04/15/2016).
- Diebold, Francis X. (2013). *Comparing Predictive Accuracy, Twenty Years Later: A Personal Perspective on the Use and Abuse of Diebold-Mariano Tests*. URL: www.ssc.upenn.edu/~fdiebold/papers/paper113/Diebold_DM%20Test.pdf (visited on 04/05/2016).

- Diebold, Francis X. and Canlin Li (2006). "Forecasting the term structure of government bond yields". In: *Journal of Econometrics* 130, pp. 337–364. URL: www.ssc.upenn.edu/~fdiebold/papers/paper49/Diebold-Li.pdf.
- Fabozzi, F. J., L. Martellini, and P. Priaulet (2005). "Predictability in the Shape of the Term Structure of Interest Rates". In: *Journal of Fixed Income* 59, 40–53.
- Fama, Eugene F. (1969). "Efficient Capital Markets: A Review of Theory and Empirical Work". In: *Journal of Finance* 25, pp. 383–417. URL: www.e-m-h.org/Fama70.pdf.
- (1995). "Random Walks in Stock Market Prices". In: *Financial Analysts Journal* 51, pp. 75–80. URL: web.williams.edu/Mathematics/sjmiller/public_html/341Fa09/handouts/Fama_RandomWalksStockPrices.pdf.
- Finch, Steven (2005). *Ornstein-Uhlenbeck Process*. URL: [people.fas.harvard.edu/~sfinch/ou.pdf](http://people.fas.harvard.edu/~sfinch/csolve/ou.pdf).
- Ghosh, Dhiren and Andrew Vogt (2012). *Outliers: An Evaluation of Methodologies*. URL: www.amstat.org/sections/srms/proceedings/y2012/files/304068_72402.pdf (visited on 04/05/2016).
- Gilli, Manfred, Stefan Grosse, and Enrico Schumann (2010). "Calibrating the on-Siegel-Svensson model". In: *COMISEF WORKING PAPERS SERIES* 31, p. 7. URL: comisef.eu/files/wps031.pdf.
- Grinstead, Charls M. and J. Laurie Snell (1997). *Introduction to Probability*. American Mathematical Society, p. 471.
- Group, CME (2013). *Yield Curve Shifts Create Trading Opportunities*. URL: www.cmegroup.com/trading/interest-rates/files/Yield_Curve_Strategy_Paper.pdf (visited on 04/13/2016).
- Guidolin, Massimo and Daniel L. Thornton (2008). "Predictions of Short-Term Rates and the Expectations Hypothesis of the Term Structure of Interest Rate". In: *European Central Bank - Working Paper Series* 977. URL: www.ecb.europa.eu/pub/pdf/scpwps/ecbwp977.pdf.
- Hai-Jew, Shalin (2014). *Enhancing Qualitative and Mixed Methods Research with Technology*. IGI Global, Kansas.
- Hansen, M., Madow W., and B. Tepping (1983). "An Evaluation of Model-Dependent and Probability-Sampling Inferences in Sample Surveys". In: *J. Amer. Stat. Assoc* 78, pp. 776–793.
- Hjorteland, Trond (1999). *Method of Steepest Descent*. URL: trond.hjorteland.com/thesis/node26.html (visited on 04/01/2016).
- Hong, Yongmiao (2009). *Martingale Theory in Economics and Finance*. URL: economics.ucr.edu/seminars_colloquia/2009/econometrics/UCR_Lecture1A_2009.pdf (visited on 04/11/2016).
- ISDA (2016). *ISDA Interest Rate Derivatives Market Analysis*. URL: [www2.isda.org/attachment/ODEwNg==/Market%20Analysis%20press%20release%2020-01-2016%20revised%20\(2\).pdf](http://www2.isda.org/attachment/ODEwNg==/Market%20Analysis%20press%20release%2020-01-2016%20revised%20(2).pdf) (visited on 04/15/2016).
- Jiahui Wang, Eric Zivot and (2006). *Modeling Financial Time Series with S-PLUS*. Springer.
- Johansen, Søren (1991). "Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models". In: *Econometrica* 59, 1551–1580.

- Junior, Jose Maria P. and Guilherme A. Barreto (2009). *Multistep-Ahead Prediction of Rainfall Precipitation Using the NARX Network*. URL: <http://www.mafy.lut.fi/timeseries/ESTSP/PDF/09.pdf> (visited on 03/10/2016).
- Kaastra, Iebling and Milton Boyd (1995). "Designing a neural network for forecasting financial and economic time series". In: *Neurocomputing* 10, pp. 215–236.
- Katz, Mitchell H. (2011). *Multivariable Analysis: A Practical Guide for Clinicians and Public Health Researchers*. Cambridge University Press, Cambridge, p. 82.
- Kendall, M. G. and Bradford Hill (1953). "The Analytics of Economic Time Series, Part 1: Prices". In: 116, pp. 11–34. URL: [ww.e-m-h.org/KeHi53.pdf](http://www.e-m-h.org/KeHi53.pdf).
- Kopányi, Szabolcs (2009). *A HOZAMGÖRBE DINAMIKUS BECSLÉSE*. URL: http://phd.lib.uni-corvinus.hu/509/1/kopanyi_szabolcs.pdf (visited on 04/01/2016).
- Krasnopolsky, Vladimir (2006). *Elements of Nonlinear Statistics and Neural Networks*. URL: polar.ncep.noaa.gov/mxab/people/vladimir/nntutor/Nonlinear_Statistics_and_NNs.pdf (visited on 03/09/2016).
- KSH (2011). *KSH - Inflation rates*. URL: www.ksh.hu/docs/hun/eurostat_tablak/tab1/tsieb060.html (visited on 03/04/2016).
- Kumar, P. N. et al. (2011). *Financial Market Prediction Using Feed Forward Neural Network*. Springer Berlin Heidelberg, Berlin, pp. 77–84.
- Kunst, Robert M. (2004). *Ökonometrische Prognose*. URL: homepage.univie.ac.at/robert.kunst/prognos3.pdf (visited on 04/12/2016).
- Li, Hongxing, C.L. Philip Chen, and Han-Pang Huang (2001). *Fuzzy Neural Intelligent Systems: Mathematical Foundation and the Applications in Engineering*. CRC Press, Boca Raton.
- Luetkepohl, Helmut (2011). *Vector Autoregressive Model*. European University Institute.
- Malinska, Barbora and Jozef Barunik (2015). "Forecasting Term Structure of Crude Oil Markets Using Neural Networks". In: URL: arxiv.org/pdf/1504.04819.pdf.
- Mankiw, N. Gregory (2014). *Principles of Macroeconomics*. Cengage Learning, Boston.
- Maren, Alianna J., Craig T. Harston, and Robert M. Pap (2014). *Handbook of Neural Computing Applications*. Academic Press, Cambridge, pp. 1–5.
- Martellini, Lionel, Philippe Priaulet, and Stephane Priaulet (2003). *Fixed-Income Securities: Valuation, Risk Management and Portfolio Strategie*. John Wiley and Sons, Chichester. URL: www.untag-smd.ac.id/files/Perpustakaan_Digital_1/FINANCE%20Fixed-Income%20Securities%200470852771.pdf.
- Mishkin, Frederic S. (2007). *A Rational Expectations Approach to Macroeconomics: Testing Policy Ineffectiveness and Efficient-Markets Models*. University of Chicago Press, Chicago, p. 71.
- Mönch, E. (2005). "Term Structure Surprises: The Predictive Content of Curvature, Level, and Slope". In: *ECB - Working paper* 544.
- Nagy, András (2015). *Hirtelen torzult el a hozamgörbe*. URL: http://www.erstebroker.hu/hu/blog_hirtelen_torzult_el_a_hozamgorbe_20150617.html (visited on 03/16/2016).

- Nelson, Charles R. and Andrew F. Siegel (1987). "Parsimonious Modelling of Yield Curves". In: *The Journal of Business* 60(4), pp. 473–489. URL: www.math.ku.dk/~rolf/teaching/NelsonSiegel.pdf.
- Nguyen, H. H. and C. W. Chan (2004). *A Comparison of Data Preprocessing Strategies for Neural Network Modeling of Oil Production Prediction*. ICCI '04. Washington, DC, USA: IEEE Computer Society, pp. 199–207. ISBN: 0-7695-2190-8. DOI: [10.1109/ICCI.2004.2](https://doi.org/10.1109/ICCI.2004.2). URL: <http://dx.doi.org/10.1109/ICCI.2004.2>.
- OECD (2002). *OECD Economic Surveys: Hungary 2002*. OECD Publishing.
- Orduna, Francisco, Marshall Lin, and Jean-Philippe Larochelle (2015). *REAL-WORLD INTEREST RATE MODELS AND CURRENT PRACTICES*. URL: soa.org/library/newsletters/the-modeling-platform/2015/july/modeling-platform-2015-vol-1-iss-1-larochelle-lin-orduna.aspx (visited on 04/13/2016).
- Pacati, Claudio (2012). *Short Rate Models*. URL: www.econ-pol.unisi.it/fm10/shortRateModels.pdf (visited on 03/04/2016).
- Pacelli, Vincenzo and Michele Azzollini (2011). "An Artificial Neural Network Approach for Credit Risk Management". In: *Journal of Intelligent Learning Systems and Applications* 3, pp. 103–112. URL: file.scirp.org/pdf/JILSA20110200003_68195552.pdf.
- Patidar, Raghavendra and Lokesh Sharma (2011). "Credit Card Fraud Detection Using Neural Network". In: *International Journal of Soft Computing and Engineering* 1, pp. 32–38. URL: www.ijscce.org/attachments/File/NCAI2011/IJSCE_NCAI2011_025.pdf.
- Patricia Melin Oscar Castillo, Janusz Kacprzyk (2015). *Design of Intelligent Systems Based on Fuzzy Logic, Neural Networks and Nature-Inspired Optimization*. Springer, New York.
- Pearce, Douglas K. (2005). *Professional Forecasts of Interest Rates and Exchange Rates: Evidence from the Wall Street Journal's Panel of Economists*. URL: www4.ncsu.edu/unity/users/d/dkpearce/www/wsjspaper.pdf (visited on 04/11/2016).
- Pearson, Karl (1906). *A Mathematical Theory of Random Migration*. Dulau and Co., London.
- Piazzesi, M. (2002). "An Econometric Model of the Yield Curve with Macroeconomic Jump Effects". In: *Journal of Monetary Economics* 50, pp. 745–787. URL: web.stanford.edu/~piazzesi/AP.pdf.
- Pooter, Michiel De (2007). *Examining the Nelson-Siegel Class of Term Structure Models: In-sample fit versus out-of-sample forecasting performance*. Econometric Institute and Tinbergen Institute Erasmus University Rotterdam, The Netherlands.
- Reppa, Zoltán (2009). "A joint macroeconomic-yield curve model for Hungary". In: *MNB Working Papers* 1.
- Rojas, Raul (2013). *Neural Networks: A Systematic Introduction*. Springer Science and Business Media, New York.
- Rycof, Chris H. (2005). *Introduction to Random Walks and Diffusion*. URL: <http://ocw.mit.edu/courses/mathematics/18-366-random-walks-and-diffusion-fall-2006/lecture-notes/lec01.pdf> (visited on 03/09/2016).
- Samuelson, Paul (1965). "Proof That Properly Anticipated Prices Fluctuate Randomly". In: *Industrial Management Review - MIT* 6:2.

- Schumacher, J.M. (2009). *Syllabus Financial Models*. Tilburg University, Tilburg, The Netherlands.
- Sherwood, J., R. Derakhshani, and T. Guess (2008). *Region 5 Conference, 2008 IEEE*. Chap. A Comparative Study of Linear and Nonlinear Data-Driven Surrogate Models of Human Joints, pp. 1–6.
- SHOOK, ANDREW P. (2013). *Yield Curve Spread Trades: Opportunities and Applications*. URL: www.cmegroup.com/education/files/yield-curve-spread-trades.pdf (visited on 04/13/2016).
- Sims, Christopher A, James H. Stock, and Mark W. Watson (1990). "Inference in Linear Time Series Models with Some Unit Roots". In: *Econometrica* 58, pp. 113–144.
- Singh, Mahipal (2011). *Security Analysis with Investment and Portfolio Management*. Gyan Publishing House, Gyan Kunj.
- Stock, James H. and Mark W. Watson (2001). *Vector Autoregressions*. URL: faculty.washington.edu/ezivot/econ584/stck_watson_var.pdf (visited on 04/10/2016).
- Suryanarayana, T.M.V and P. B Mistry (2016). *Vector Autoregressive Model*. Springer.
- Tadjudje Kamgaing, Joseph (2005). *Competing Neural Networks as Models for Non Stationary Financial Time Series*. URL: kluedo.ub.uni-kl.de/files/1613/dfin.pdf.
- T.Chai and R. R. Draxler (2014). "Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature". In: *Geosci. Model Dev. Discuss* 7. URL: www.geosci-model-dev.net/7/1247/2014/gmd-7-1247-2014.pdf.
- Tetko, Igor V., David J. Livingstone, and Alexander I. Luik (1995). "Neural Network Studies 1. - Comparison of overfitting and overtraining". In: *J. Chem. Inf. Sci.* 35, pp. 826–833. URL: pubs.acs.org/doi/pdf/10.1021/ci00027a006.
- Vasicek, Oldrich (1977). "An equilibrium characterization of the term structure". In: *Journal of Financial Economics* 5, pp. 177–188.
- Vela, Daniel (2013). "Forecasting Latin-American Yield Curves: An Artificial Neural Network Approach". In: *Borradores de economía* No. 761. URL: www.banrep.gov.co/en/borrador-761.
- Venturini, Andrea (2011). "TIME SERIES OUTLIER DETECTION: A NEW NON PARAMETRIC METHODOLOGY". In: *Statistica* LXXI. URL: <http://rivista-statistica.unibo.it/article/viewFile/3617/2968>.
- Virili, F. (2000). "Nonstationarity and data preprocessing for neural network predictions of an economic time series". In: *Neural Networks, 2000. IJCNN 2000, Proceedings of the IEEE-INNS-ENNS International Joint Conference on* 5.
- Vries, Bert de and Jose C. Principe (1990). "A Theory for Neural Networks with Time Delays". In: *NIPS-3 Proceedings of the 1990 conference on Advances in neural information processing systems* 3, pp. 162–168. URL: papers.nips.cc/paper/356-a-theory-for-neural-networks-with-time-delays.pdf.
- Waibel, Alexander et al. (1989). "Phoneme Recognition Using Time-Delay Neural Networks". In: *IEEE Transactions on Acoustics, Speech and Signal Processing* 37, pp. 328–339. URL: www.cs.toronto.edu/~fritz/absps/waibelTDNN.pdf.

- Willmott, C. J., K. Matsuura, and S. M. Robeson (2009). "Ambiguities inherent in sums-of-squares-based error statistics". In: *Atmos. Environ.* 43, 749–752.
- Yu, Hao and Bogdan M. Wilamowski (2010). *Levenberg–Marquardt Training*. URL: www.eng.auburn.edu/~wilambm/pap/2011/K10149_C012.pdf (visited on 04/01/2016).
- Yu, Lean, Shouyang Wang, and Kin Keung Lai (2007). "Data Preparation in Neural Network Data Analysis". In: *International Series in Operations Research and Management Science* 107.