Charles University in Prague

Faculty of Social Sciences

Institute of Economic Studies

Master Thesis

2012 **Vyacheslav Lypko**

Charles University in Prague

Faculty of Social Sciences

Institute of Economic Studies

Vyacheslav Lypko

Modelling Dynamics of Correlations between Stock Markets with High-frequency Data

Master Thesis

Prague 2012

Author: **Vyacheslav Lypko**
Supervisor: **PhDr. Josef Baruník, Ph.D.**
Academic Year: **2012**

Bibliographic record

LYPKO, Vyacheslav. *Modelling Dynamics of Correlations between Stock Markets with High-frequency Data* . Prague, 2012. 59 pages. Master Thesis (M.A.) Charles University, Faculty of Social Sciences, Institute of Economic Studies.

Thesis supervisor PhDr. Jozef Baruník, Ph.D.

Abstract

In this thesis we focus on modelling correlation between selected stock markets using high-frequency data. We use time-series of returns of following indices: FTSE, DAX PX and S&P, and Gold and Oil commodity futures. In the first part of our empirical work we compute daily realized correlations between returns of subject instruments and discuss the dynamics of it. We then compute unconditional correlations based on average daily realized correlations and using feedforward neural network (FFNN) to assess how well the FFNN approximates realized correlations. We also forecast daily realized correlations of FTSE:DAX and S&P:Oil pairs using heterogeneous autoregressive model (HAR), autoregressive model of order p (AR(p)) and nonlinear autoregressive neural network (NARNET) and compare performance of these models.

Abstrakt

Tato prace je zamerena na modelovani korelaci mezi vybranymi akciovymi trhy a komoditami s pouzitim vysokofrekvencnich dat. Nasledujici casove rady jsou pouzite pro ucely teto analyzy: FTSE, DAX, PX, S&P, Gold commodity futures a Oil commodity futures. V prvni casti teto prace denni realizovane korelace jsou vypocitane a jejich dynamika je diskutovana. Dal jsou vypocitane korelace pomoci neuronove site (feed forward neural network, nebo FFNN). Tyto korelace jsou porovane s prumernymi dennimi realizovanymi korelacemi. V posledni casti teto prace jsou vypocitane prognozy dennich realizovanych korelaci pomoci HAR modelu, AR(p) modelu a dynamicke neuronove site NARNET.

Klíčová slova

Korelace, dynamika korelaci, vysokofrekvencni data, realizovane korelace, neuronove site

Keywords

Correlation, dynamics of correlations, high-frequency data, realized correlations, neural networks

Prohlášení

1. Prohlašuji, že jsem předkládanou práci zpracoval samostatně a použil jen uvedené prameny a literaturu.
2. Prohlašuji, že práce nebyla využita k získání jiného titulu.
3. Souhlasím s tím, aby práce byla zpřístupněna pro studijní a výzkumné účely.

V Praze dne Vyacheslav Lypko

Declaration

1. I hereby declare that I have compiled this master thesis independently using only the listed literature and resources.
2. I hereby declare that this master thesis was not used for acquiring a degree different from that at the Institute of Economic Studies.
3. I hereby consent for this thesis to be made available for study and research purposes.

Prague, Vyacheslav Lypko

Acknowledgements

I would like to thank my supervisor, PhDr. Jozef Baruník, Ph.D., for his endless support, patience, helpful and highly professional comments. His continuous willingness to advice on any kind of problem on the shortest possible notice helped me the most during my research.

Institute of Economic Studies
Master Thesis

Proposed Topic:

Modelling dynamics of correlations between stock markets with high-frequency data
Modelování dynamiky korelací finančních trhů pomocí vysokofrekvenčních dat

Topic Characteristics:

My thesis will focus on relatively new and advanced field of econometrics – artificial neural networks and their application in quantitative financial analysis using high frequency data. After studying several relevant works and consulting with my supervisor, I've decided to focus my research on discovering possible correlations between top 3 (5) world's financial markets. The subject markets are still to be selected. We have high frequency data available for the following ones: Germany (DAX), Japan (Nikkei 225), Czech Republic (PX index), Shanghai (SSE composite index), United States (SP500), Hungary (Budapest SE BUX index), Poland (WIG20 index), England (London, FTSE), United States (NYSE index), Hong Kong (Hang Seng indices).

The principal reason why neural networks receive increased attention in last decade is that today's driving forces of quantitative finance analysis, financial decisions making and policies valuation are precise calculation of the odds and risks in contemporary financial system. Another reason is that recent financial crisis showed the weaknesses of standard time-series models.

Artificial neural networks are computational structures that allow accurate forecasts without any specific assumptions about the distribution or characteristics of the observed variables. Thus proposed methodology allows for more advanced and sophisticated modelling of underlying processes.

Why it is important and beneficial to study the correlations between financial markets? Let's imagine the portfolio management problem: the more financial instruments are available, the more options how to diversify specific risks away there are. However, presence of significant correlations between given instruments (in our case these are financial markets indices), both positive and negative, it will affect the investment decision. The positive correlation between two instruments reduces the benefits of including them in the portfolio whereas negative correlation increases those benefits.
This topic is also interesting in terms of getting deeper understanding of the methodology and mathematical and econometric apparatus of the neural networks for the possible further research of this methodology.

Hypotheses:

1. Research/modeling dynamics of correlations between stock markets with high-frequency data
2. Neural network allows to model processes underlying the correlations of mentioned financial markets indexes better than other selected econometric models,
3. Neural Networks are able to model the dynamics of correlation better than comparable models

Methodology:

In my research I'm going to apply the neural network methodology of econometric analysis and modeling based on the high frequency data in order to support/reject proposed hypotheses. Since there are several types of models of that class (feed forward networks, multilayer feed forward networks, jump connections networks, smooth-transition regime switching networks etc.), I'll first try to identify one that will be able to model given processes better. I will also apply further econometric methods/models and compare their outcomes with those of neural network in order to support/reject the hypothesis that neural network is capable of modeling the underlying processes better.

Outline:

A. Theoretical foundations of neural networks
B. Foundations of portfolio and risk management and the role of correlations identification
C. Data description: information about selected financial markets, description of the data being analyzed
D. Selection of the appropriate neural network model type
E. Application of the selected neural network model type for correlations analysis
F. Application of other econometric models on the same problem
G. Comparison of obtained results
H. Conclusions and further suggestions

Core Bibliography:

1. Back, T. (1996), "Evolutionary Algorithms in Theory and Practice", Oxford: Oxford University press.
2. Beltratti, Andrea, Sergio Margarita, and Pietro Terna (1996), "Neural Networks for Economic and Financial Modelling". Boston: International Thomson Computer Press.
3. Campbell, John Y., Andrew W. Lo, and A. Craig MacKinlay (1997), The Econometrics of Financial Markets. Princeton, NJ: Princeton University Press.
4. Clark, Todd E., and Michael W. McCracken (2001), "Tests of Forecast Accuracy and Encompassing for Nested Models," Journal of Econometrics 105: 85–110.
5. Dayhoff, Judith E., and James M. DeLeo (2001), "Artificial Neural Networks: Opening the Black Box." Cancer 91: 1615–1635.
6. Diebold, Francis X., and Roberto Mariano (1995), "Comparing Predictive Accuracy," Journal of Business and Economic Statistics, 3: 253–263.
7. Engle, Robert, and Victor Ng (1993), "Measuring the Impact of News on Volatility," Journal of Finance 48: 1749–1778.
8. Franses, Philip Hans, and Dick van Dijk (2000), Non-linear Time Series Models in Empirical Finance. Cambridge, UK: Cambridge University Press.
9. Genberg, Hans (2003), "Foreign Versus Domestic Factors as Sources of Macroeconomics Fluctuations in Hong Kong." HKIMRWorking Paper 17/2003.
10. Paul D. McNelis (2005), "Neural Networks in Finance: Gaining predictive edge in the market", Elsevier Academic Press, Inc.

Author

Supervisor

Contents

List of abbreviations

AIC	Akaike information criterion
ADF	Augmented Dickey-Fuller tests
ANN	Artificial neural network
ANFIS	Adaptive Neuro-Fuzzy Inference System
AORD	All Ordinaries, Australian stock market index
AR	Autoregressive
ARCH	Autoregressive conditional heteroscedasticity model
CV	Coefficient of variation
DAG	Directed acyclic graphs
DAX	Index of Deutsche Börse
DJIA	Dow Jones Industrial Average
ECM	Error correction model
EHM	Efficient market hypothesis
EMU	European Monetary Union
EU	European Union
FFNN	Feedforward neural network
FTSE	Financial Times Stock Exchange 100
FX	Foreign exchange
GARCH	Generalized Autoregressive conditional heteroscedasticity
HAR	Heteroscedastic auto-regressive model
HARCH	Heterogeneous generalized autoregressive conditional heteroscedasticity
HF-data	High-frequency data
HMH	Heterogeneous market hypothesis
IC	Information criterion
IID	Independent and identically distributed
JB	Jarque-Bera test
LM	Levenberg-Marquardt
MA	Moving average
MLP	Multi-layer perceptron
MZ	Mincer-Zarnowitz regression
NARNET	Nonlinear autoregressive neural network
NYSE	New York Stock Exchange
OLS	Ordinary least squares model
PACF	Partial autocorrelation function
PX	Prague Stock Exchange 50
RBF	Radial basis function
RC	Realized correlation
RCOV	Realized covariance
RMSE	Root mean squared error
RV	Realized variance
RVOL	Realized volatility
SM	Stock market
SMH	Stock market holography
S&P	Standard & Poor's 500
SW-ARCH	Switching autoregressive conditional heteroscedasticity model
TAR	Threshold autoregressive model

TASE	Tel Aviv Stock Exchange
UK	United Kingdom
US	United States of America
VaR	Value at risk
VAR	Vector autoregression

1. Introduction

One of the main problems in quantitative finance is modelling of volatility and correlation of returns of financial assets. The concept of statistical correlation is considered, for example, in financial decision making, risk management, portfolio analysis or derivative financial instruments pricing. Research on correlation of returns of financial assets and underlying indices, including stock markets indices, was developing rapidly over the past decades. Financial econometricians and researchers have contributed to a development of advanced methods that are extensively used in modern financial time-series analysis. These models can be generally divided into two broad groups: "mainstream" models of Autoregressive Moving Average (ARMA) and Autoregressive Conditional Heteroscedasticity (ARCH, used for volatility modelling) families, and alternative models, such as an artificial neural networks. Each family of models has its strong and weak sides. Even though artificial neural networks are somewhat more powerful computational structures capable of approximating any nonlinear function with finite number of discontinuities (Beale et al., 2012), ARMA / ARCH models are used in practice more often since they require less effort to setup and apply. In our empirical analysis we utilize artificial neural networks as one possible alternative for correlation modelling.

Another reason of increased attention of financial decision makers to advanced and robust econometric techniques is a series of financial crises, starting with recent global financial crisis and followed by the current EU debt crisis. In turbulent markets, decision makers tend to search for more reliable models to support them.

During the past decade, interconnectedness between national financial markets increased significantly. Different segments of financial market experience different degree of interdependence. Overall, big markets, such as American, British or Japanese, react relatively fast to change in one of them, especially when there are problems in financial system. There also was a big deal of deregulation of financial markets in the US starting 1970's. Integration of European financial markets increased with the establishment of European Monetary Union. For example, paper by Baele et al. (2004) investigates degree of integration of main sectors of financial market of the EU, such as money market, government bonds market, corporate bond market, bank credit market and equity market. Authors find out that degree of integration of above mentioned segments increased significantly after the establishment of EMU. They argue that only 20% of local returns variance was explained by aggregate European and US shocks, whereas this proportion increased to 40% by 2000. These factors, along with other processes, have led to the increased vulnerability of national stock markets and higher probability of financial contagion, i.e. cross-border spread of financial crises. Based on stated above it becomes clear that importance of proper estimation of correlation can't be overestimated.

In portfolio analysis and risk management, correlation is one of the central concepts. Decision making process takes place in the risk-return space, where risk of an asset or portfolio can be measured by a variance of its returns. In the compendium of the risk-return calculus portfolio variance is defined as follows:

$$\sigma_P^2 = \sum_{i=1}^{N} \theta_i^2 \sigma_i^2 + 2\sum_{i=1}^{N} \sum_{j=i+1}^{N} \theta_i \theta_j \sigma_i \sigma_j \rho_{ij} \tag{1.1}$$

where $\theta_i, i = 1,...,N$ is a share of *i-th* asset in the portfolio, σ_i^2 is a variance of a risky asset *i*, and ρ_{ij} correlation coefficient between rates of return of assets *i* and *j*. As it can be seen from the formula, high positive correlation of assets included in the portfolio increases its variance and reduces diversification benefits. Negative correlation, on the other hand, reduces the variance and allows for better portfolio diversification.

Another important measure in the risk management is so-called value at risk (VaR). VaR measure aims at providing a single number expressing the total risk of a portfolio comprising financial assets. Basically, VaR expresses expected maximum loss given the target horizon of investment and confidence level. VaR of the portfolio can be expressed in terms of VaRs of individual assets included in it:

$$VaR_P = \alpha_c \sigma_P W_P^0 = \sqrt{\sum_{i=1}^{N} (VaR_i)^2 + 2\sum_{i=1}^{N} \sum_{j=i+1}^{N} VaR_i VaR_j \rho_{ij}} \tag{1.2}$$

where W_P^0 is a portfolio's initial value, σ_P is a volatility of the portfolio, α_c is a certain cut-off rate, $VaR_i, i = 1,...,N$ is a value at risk of an asset *i* and ρ_{ij} is the correlation structure among assets included in the portfolio. Again, it can be seen that depending on the correlation degree and direction, VaR of the portfolio can be increased or reduced. These two examples mentioned above give the reader an idea of possible applications and importance of the correlation measure in portfolio and risk management.

In this thesis we focus on modelling of correlations between selected stock markets and commodities between 2008 and 2011. We have included commodities in our analysis because in times of crisis it is natural for investors to seek a safe place for their funds. One of the choices for them is commodities, such as gold or crude oil. Partially because of it we can observe sharp increases in price of mentioned commodities when there is a risk of or already ongoing crisis. In the second part of this thesis we study the possibility of forecasting realized correlations and try to find models that suit best for that problem. We utilize dynamic neural network and Heterogeneous Autoregressive model (HAR) for these purposes. We also use Autoregressive model of order *p* (AR(p)) to see how it compares to more complex techniques.

We begin with a brief analysis of the existing works studying stock markets correlation to understand its possible underlying determinants. There were several attempts to analyze the stock markets correlation and model it using econometric models. Attention to stock market correlation increased after stock market crash in October, 1987. It was observed, that correlation increases when markets experience common adverse shock. King and Wadhwani (1990) developed the concept of stock market contagion, or crisis spillover, when shock in major market, e.g. US, spreads to other stock markets. This topic was further explored by Yang and Bessler (2006) in their empirical study investigating contagion among seven international markets around the 1987's stock markets crash. They build their work on Vector Autoregressive Analysis (VAR) conducting data-determined historical decompositions to provide day-by-day picture of price fluctuation transmission. They conclude that the crisis originated in the US spilled over to main European and Japanese markets. King et al. (1994) argues that majority of stock markets co-movements are determined by unobservable factors, i.e. an investor sentiment etc. Ammer and Mei (1996), on the other hand, find out that equity risk premium rather than fundamental factors explain most of the co-movements in national indices.

Now turning to the existing literature on time-varying (dynamics) properties of stock markets correlation, Longin and Solnik (1995) discover that correlations and covariances are not stable over time. Authors use a bivariate GARCH model to capture the structure of conditional covariance. They show that conditional correlations may be influenced by such factors as short-term interest rates and dividend yields. Ramchand and Susmel (1998) also show the time dependence of correlations using the SW-ARCH model. Bodart and Reding (1999) in their study examine the impact of FX-rate variance on international stock markets correlation. According to their findings, correlation of bond and stock markets increase with decreasing FX-rate variance.

Some authors make use of graphing techniques to investigate stock market correlations and their dynamics. Groenen and Franses (2000), using just mentioned technique, observe three clusters of stock markets with high correlations, namely US, Europe and Asia. Another paper by Bessler and Yang (2003) investigates the dynamic structure of interdependencies of nine major stock markets using an Error-Correction Model (ECM) and Directed Acyclic Graphs (DAG) approaches. They conclude, for example, that the US market is the only market that has a consistently strong impact on price movements in other major stock markets in the longer-run. Another recent attempt to assess the structure and dynamics of stock markets through correlations between them was undertaken by Kenett et al. (2010). In their study named "Dynamics of Stock Market Correlations" authors apply so called Stock Market Holography (SMH) method complemented by the eigenvector entropy measure to quantify changes in information in market. Their analysis is based on data collected from New York Stock Exchange (NYSE) and Tel Aviv Stock Exchange (TASE). Authors conclude that there is a significant variation over time of the amount of information in the stock market. It also can be said that SMH may be a useful first-hand quantitative tool. However, it is good only for ex post analysis.

Based on stated above we can see that there is a large number of factors that can determine the existence and degree of correlations between the stock markets as well as plethora of quantitative tools to be used to analyze it. Another important recent trend in financial econometric research is increasing availability of intraday, or high-frequency, transaction data. With the development of computer technologies and trading systems, large amounts of high-frequency data on financial assets prices became available. Sampling frequencies of intraday data achieve levels of 5 minutes or even real-time. Due to the different structure of the high-frequency data, as compared to "classical" daily observations, mainstream models are not able to fully utilize information contained in this data, which leads to inaccurate estimates. To address this problem range of alternative estimators, or so-called realized measures of volatility and correlation, were developed. Realized measures are relatively new econometric techniques. First attempts to describe properties and advantages of realized volatility estimator was undertaken, for example, in works of Andersen et al. (1998) or Andersen et al. (1999). The rigorous theoretical background, however, was developed by mid 2000's. Probably the most influential work was presented by Andersen et al. (2003) where authors provide robust theoretical framework for modeling and forecasting realized volatility. This work also had significant implications for the realized correlation estimation method. For the forecasting of realized volatility or correlation so-called Heterogeneous Autoregressive (HAR) model can be used. We utilize the HAR model in our empirical analysis for realized correlation forecasting and compare its performance to performance of artificial neural network and simple AR(p) model. We describe realized measures and HAR model in more detail in Chapter 3 of this thesis.

Another aspect that makes this thesis worth attention is the fact that we apply the neural networks on high-frequency data to model the stock markets correlations. Artificial neural networks have received a lot of attention in past decade. After failures of several major players in financial markets caused by opportunism, supervision and regulator systems faults, decision makers are in a constant search for more robust quantitative methods to support them. The internet bubble, for example, showed that conventional econometric models have substantial shortages in precision and robustness (McNelis, 2005). Majority of conventional models try to explain underlying patterns using linear relationships and assumptions and rely heavily on assumptions about analyzed data distribution. However, most of the empirical evidence does not support the assumption about linear relationships in financial data. Furthermore, financial data is usually non-normally distributed and has so-called fat tails. That makes non-linear method of financial quantitative analysis more appropriate. The artificial neural networks are computational structures that allow for accurate analysis and forecasts without any specific assumptions about the linearity of underlying relations or statistical distribution of observed variables. We analyze existing literature regarding application of neural networks in the quantitative finance in more detail in Chapter 4 of this thesis. However, important observation is that prior our empirical work artificial neural networks have never been applied to forecast correlations of stock markets. Majority of existing papers describe application of neural networks to forecasting of stock markets themselves.

We have set several aims for our research. Firstly, we model correlations of selected stock markets and commodities and analyze changes of conditional daily realized correlations over the given period of time. We also compare estimates of unconditional correlations obtained using feedforward neural network to average daily realized correlations to see how well the FFN approximates RC. We also forecast selected time-series of realized correlations.

We start this thesis by introducing basics of time-series analysis and classical time-series analysis techniques in the Chapter 2. We also introduce Autoregressive model of order p – AR(p) that we use in our empirical research. In the Chapter 3 we motivate use of realized measures in our empirical work. We provide theoretical foundations of realized measures and introduce the computation of realized correlation measure. Chapter 4 is devoted to artificial neural networks. In this chapter we establish theory of artificial neurons – building blocks of any neural network. It is followed by description of feedforward neural network (FFNN) and nonlinear autoregressive neural network (NARNET) that we utilize in our empirical research. We present our empirical findings in the Chapter 5 of this thesis. Chapter 6 concludes.

2. Theoretical Background of Classical Time-Series Analysis Techniques

As it was discussed in the introduction to this thesis, correlation is one of the central concepts in quantitative finance. There are a lot of econometric models and computational techniques that academics and practitioners can use for correlation modelling. These models range from classical correlation measure to most advances contemporary techniques, such as artificial neural networks. Since the main focus of this thesis is to model correlations of selected stock markets and selected commodities, in this chapter we present econometric models and techniques that we utilize in our analysis.

In the first section we describe basic statistic concepts of stationarity, integration of time-series and correlation. We then turn to the linear autoregressive model of order *p* (AR(p)) which we use in the second part of our empirical research devoted to forecasting of realized correlation that we describe in Chapter 5 of this thesis.

2.1 Basic Concepts

Fist important concept used in financial time-series analysis is so-called (weak) stationarity. The time-series y_t is said to be stationary if:

1. it fluctuates around a constant long-run mean, e.g. $E(y_t) = \mu$;
2. y_t has a finite variance which is not dependent upon time, e.g. $Var(y_t) = \sigma^2$;
3. covariance between two values of y_t depends only on the difference apart in time, e.g. $Cov(y_t, y_{t-n}) = \chi(n)$.

Property of stationarity is important for the forecasting tasks to assure that the forecasted time-series won't diverge to infinity. Since the majority of financial time-series do not reveal stationarity, initial data should be transformed in a certain way in order to allow for estimation of model and forecasting of assumed process. One way how to transform the data is integration. Time-series *y(t)* is integrated of order one if:

$$y^*_t = y_t - y_{t-1} \tag{2.1}$$

Integration, or first differencing, has a drawback because it causes the partial loss of information of initial time-series. On the other hand, if integration is applied on financial asset prices or stock market index values, integration transfers the problem to the asset or index nominal returns analysis:

$$r_t = P_t - P_{t-1} \tag{2.2}$$

Due to the stationarity issues, major part of analysis of financial time-series is focused on the returns volatility or correlations estimation, modelling and forecasting. The detailed description of concepts of stationarity and integration can be found in any academic book regarding financial time-series analysis (for example, Tsay, 2005).

Correlation coefficient measures the degree and direction of statistical relationship between two random variables. For two variables X and Y it is defined as follows:

$$\rho_{X,Y} = corr(X,Y) = \frac{\operatorname{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \tag{2.3}$$

where $\mu_i, i = X, Y$ is an expected value of a random variable, and $\sigma_i = \sqrt{\sigma_i^2}, i = X, Y$ is a volatility.

2.2 Normality

There are several assumptions imposed on the residuals in the classical econometric models. One of them is that residuals come from a Gaussian or normal distribution. One of the most often used tests for residuals normality is the Jarque-Bera test. It was proposed by the Jarque-Bera (1980). Test statistics is based on an assumption that normally distributed data has a skewness of zero and the kurtosis of three. The tests statistics $JB(\hat{\varepsilon})$ has a $\chi^2(2)$ distribution and can be defined as follows (McNelis, 2005):

$$JB(\hat{\varepsilon}) = \frac{T-k}{6}\left(s(\hat{\varepsilon})^2 + .25(kr(\hat{\varepsilon})-3)^2\right) \tag{2.4}$$

where T is the total number of observations, k is the number of parameters in the model, $\hat{\varepsilon}$ is the estimated vector of residuals, and $s(\hat{\varepsilon})$ and $kr(\hat{\varepsilon})$ are skewness and kurtosis respectively. Skewness and kurtosis are computed as follows:

$$s(\hat{\varepsilon}) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t - \overline{\hat{\varepsilon}}}{\sigma_{\hat{\varepsilon}}}\right)^3 \tag{2.5}$$

$$kr(\hat{\varepsilon}) = \frac{1}{T}\sum_{t=1}^{T}\left(\frac{\hat{\varepsilon}_t - \overline{\hat{\varepsilon}}}{\sigma_{\hat{\varepsilon}}}\right)^4 \qquad (2.6)$$

where $\overline{\hat{\varepsilon}}$ is the estimated mean and $\sigma_{\hat{\varepsilon}}$ is the estimated standard deviation of the residuals vector.

2.3 *Stationarity*

When stationarity of a time-series is considered, it is necessary to test if the unit roots are present. Presence of the unit roots can lead to the problem of a spurious regression. If the subject time-series contains a unit root, stationarity cannot be assumed, and estimated relationship may not reflect the true process. One way how to address this problem is differencing of the subject time-series and testing each order for the unit roots.

One of the most often used methods for unit roots testing is the Dickey-Fuller test. It was proposed by Dickey and Fuller (1979). The null hypothesis is that the subject time-series contains a unit root against an alternative that the time-series is stationary. The test is based on the autoregressive process underlying a change in the subject time-series $\{y_t\}$. For theoretical foundations of DF-test, reader may refer to McNelis (2005). The test statistics for DF-test is constructed as follows:

$$DF = \frac{\sum_{t=1}^{T} r_{t-1}\varepsilon_t}{\hat{\sigma}_{\varepsilon}\sum_{t=1}^{T} r_{t-1}^2} \qquad (2.7)$$

To verify the null hypothesis the t-test is usually constructed. Augmented Dickey-Fuller test extends the test presented above and allows for inclusion of the constant term and trend terms in the regression described above. Main drawback of the ADF-test is that it tends to over reject the null hypothesis.

2.4 Linear Models

Linear time-series models provide a simple framework for the analysis of stock market returns. Returns are said to be described by a linear process if they can be modeled as follows:

$$r_t = \mu + \sum_{i=0}^{\infty} \alpha_i \varepsilon_{t-i} , \tag{2.8}$$

where μ is the mean of the process, α_i are the weights with $\alpha_0 = 1$, and $\{\varepsilon_t\}$ is the white noise process. The white noise process is defined as follows:

$$\varepsilon_t = u_t, \text{ where } u_t \sim IID(0, \sigma^2) \tag{2.9}$$

As it was already discussed above, the stationarity is a crucial in order to estimate and model the process underlying development of financial time-series. If we assume the r_t process to be weakly stationary, mean and variance of that process can be computed as follows:

$$E(r_t) = \mu , \quad \text{and} \quad Var(r_t) = \sigma^2(\varepsilon)\sum_{i=0}^{\infty} \alpha_i^2 , \tag{2.10}$$

where $\sigma^2(\varepsilon)$ is the variance of $\{\varepsilon_t\}$.

2.4.1 AR Model

Autoregressive model is one of the simplest techniques for time-series analysis that uses lagged values of r_t to model the underlying process. The econometric literature defines the p-order process as follows:

$$r_t = \phi_0 + \sum_{i=1}^{p} \phi_i r_{t-i} + \varepsilon_t \tag{2.11}$$

where $\{\varepsilon_t\}$ is an $IID(0, \sigma^2)$ white noise and p is a non-negative integer. The series of white noise $\{\varepsilon_t\}$ is the source of randomness in the model. Regarding the stationarity, AR(p) process is stationary if:

$$\sum_{i=1}^{p} \phi_i < 1, \tag{2.12}$$

otherwise the process is said to be explosive. The given process, if assumed to be stationary, can be described by mean:

$$E(r_t) = \frac{\phi_0}{1 - \phi_1 - ... - \phi_p}, \quad (2.13)$$

and can be alternatively defined as follows (without intercept):

$$\phi(L) r_t = \varepsilon_t, \quad (2.14)$$

where

$$\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p. \quad (2.15)$$

When determining the order of AR-process in practice, two main approaches can be applied based on the order specification procedures described in mainstream time-series analysis literature. The first one is based on so-called partial autocorrelation function (PACF) to determine the optimal number of lags. The second one uses different information criteria. All of the available ICs are likelihood based. An example of IC used to set the order p of AR-process is Akaike information criterion (Akaike, 1973). It is defined as follows:

$$AIC = \frac{1}{T}(2n - 2\ln(L)), \quad (2.16)$$

where T is the sample size, n is the number of parameters and L is the likelihood function for the estimated model. The logarithm of likelihood function measures the goodness of fit of estimated AR-model to data while former term in the eq.1.11 penalizes a potential model for the number of parameters used. Some researchers even use and AIC to determine the neural network architecture (for example, Panchal et al. (2010)).

3 Realized measures

Because we work with high-frequency data in our analysis, it makes sense to utilize so-called realized measures techniques to model correlations. In following chapter we introduce the realized measures – statistical measures and models designed especially to be applied on high-frequency data.

As it is argued in Yu et al. (2009), main problem of classical parametric models, if applied on the high-frequency data, is that microstructure noise contained in the data of frequencies higher than 15 minutes sampling do not allows for as good asymptotic results of estimators, used in the parametric models, as when working with data of lower frequencies. As a result, in-sample fitting and out-of-sample forecasts are not precise because parametric models do not fully utilize the information contained in the data of high frequencies.

Motivation behind using realized measures is quite strong. On one hand, the construction of realized volatility, for example, is trivial: it is just sum of squared intra-period high-frequency returns (Andersen et al., (1999)). On the other hand, Andersen and Bollerslev (1998) show that realized volatility computed on high-frequency data is effectively an error-free volatility measure under usual diffusion assumption. Since it is error-free, volatility estimated on high-frequency data can be treated as observable. Problem with classical models for volatility modelling, such as GARCH, is the fact that volatility is not observable directly, thus classical volatility models describe latent volatility. Let's formalize return innovation as follows:

$$r_t = \sigma_t \cdot z_t \tag{3.1}$$

where z_t denotes an independent zero mean and unit variance stochastic process, and σ_t stands for latent volatility evolving according to the selected model. Even if the model for σ_t^2 is correctly specified and squared innovation provides unbiased estimate for the latent volatility factor (i.e. $E_{t-1}(r_t^2) = E_{t-1}(\sigma_t^2 \cdot z_t^2) = \sigma_t^2$), squared innovation may yield very noisy estimate due to the idiosyncratic error term z_t^2. Poor forecasting performance of volatility models, using r_t^2 as a measure for ex-post volatility, is an inevitable consequence of noise present in the process that generates returns. These arguments show a need for a fundamentally different approach to volatility modelling.

In theory, with an increase of observation frequency form daily to infinitesimal, volatility measure based on cumulative intra-period squared returns converges to the true measure of latent volatility. Even though in practice there are certain limitations due to the availability and quality of data, Andersen and Bollerslev (1998) show that realized volatility measure provide significant reduction of noise and improved temporal stability as compared to measures based on daily returns. In other words, if we would have infinitely many data points, we would actually observe a true process. We can say, that with high-frequency data we can observe volatility. In fact, based on most influential work of Andersen et al. (2003), where general framework for integration of high-frequency intraday transaction data into the measurement, modelling and forecasting of returns volatilities and distributions of lower frequencies was presented, we expect that the neural network will not outperform HAR model based on realized measures significantly in forecasting of realized correlations.

We start the Chapter 3 by discussing main problems one can face when using high-frequency data. We then present foundations of construction of realized measures, namely realized volatility and realized correlation. We finish this chapter by introducing Heterogeneous Autoregressive model (HAR) which we use for forecasting realized correlation.

3.1 Main Problems with High-Frequency Data

High-frequency data is intraday tick-by-tick transaction level data on returns of selected stock market indices. For the purposes of our analysis we use data with sampling frequency of 5 minutes. With the development of trading systems and transactions tracking systems, large sample of data on stocks, indices etc. became available. One good thing about using high-frequency data for correlations modelling, except of the fact that it allows to observe volatility, is that this approach reduces probability of random co-movement of subject stock market indices that does not reflect any statistical dependency (if we look at the daily movements – indices may move the same direction as a result of different independent events). In our research we assumed that using the high-frequency data reduces a probability of such random co-movement.

Other problems related to high-frequency sampling are as follows: large number of data points requiring additional "preparation work" on samples that later will be used in the analysis; "dirty" data with irregular spacing; heavy tailed returns series; long memory behavior. Another problem with computing correlations between national stock market that relates to all sampling frequencies is different trading hours. In our work we have undertaken initial data preparation to adjust for these differences and compute correlation only on those parts of sample that overlap among subject stock and commodity markets.

3.2 Construction of Realized Measures

Realized measures is the family of statistical measures, such as realized correlation or realized volatility, developed especially to utilize high frequency data. The way how realized measures are constructed is presented below. We start with describing construction of realized volatility (realized variance). Let's consider $p_{i,t}$ to be a logarithmic price of an asset *i* at time *t*. Let *m* denote the number of observed data points per one trading session, and *n* to denote a fraction of trading session with given sampling frequency. This way *n = 1/m.* In other words, *n* can be viewed as a time step given the sampling frequency. Let *T* to be a number of days in the sample. Based on it, the total number of data points will *m* x *T.* The realized variance on day *t* is then constructed as a sum of squared returns, computed in each step during the day:

$$RV_{i,t} = \sum_{j=1}^{m} r_{i,t-1+jn}^2 , \tag{3.2}$$

Important property of realized variance is that it is asymptotically unbiased and consistent estimator of the integrated variance given certain assumptions.

Let's consider jump-diffusion continuous-time no-arbitrage price process, which is defined as follows:

$$dp_t = \mu_t dt + \sigma_t dW_t + c_t dJ_t , \tag{3.3}$$

where μ_t is a predictable component of a price change, W_t is a Wiener process, J_t is a constant-intensity Poisson process and jump magnitude is determined by $c_t \sim N(0,\sigma^2)$. Quadratic variation of the above defined price process if formalized as follows:

$$QV_t = \int_0^t \sigma_s^2 ds + \sum_{s=1}^{t} J_s^2 \tag{3.4}$$

The asymptotic result for realized variance if then as follows:

$$RV_t \to \int_0^t \sigma_s^2 ds, n \to 0, (m \to \infty) \text{, and } \frac{r_t}{\sqrt{RV_t}} \sim N(0,1) \tag{3.5}$$

The proof this classical property of RV can be found, for example, in Protter (2004).
Realized volatility is defined as follows:

$$RVOL_{i,t} = \sqrt{RV_{i,t}} \tag{3.6}$$

Realized covariance of two assets x and y on day *t* is defined as follows:

$$RCov_{t,(x,y)} = \sum_{j=1}^{m} r_{x,t-1+jn} r_{y,t-1+jn} \tag{3.7}$$

$RCov_{t,(x,y)}$ will converge in probability to the covariance matrix of two assets X and Y, thus it is a consistent estimator.

Finally, realized correlation on day *t* is then computed as follows:

$$RC_t = \frac{RCov_{t,(x,y)}}{\sqrt{RV_{t,x}}\sqrt{RV_{t,y}}} \tag{3.8}$$

3.3 Heterogeneous Autoregressive Model

In our empirical work, except of modelling correlations of selected stock markets and commodities, we also focus on forecasting realized correlations of selected pairs of indices and commodities. We compute daily realized correlations based on intraday high-frequency data and perform out-of-sample forecast of this time-series assuming autoregressive process behind RC development. We use several models for our analysis, namely simple Autoregressive model of order *p* (AR(p)), dynamic Nonlinear Autoregressive Neural Network (NARNET) and Heterogeneous Autoregressive model (HAR) to compare performance of these models and comment on which one addresses issue of forecasting realized daily correlation better. We apply a HAR setup which is described below to model assumed autoregressive process of development of realized correlations.

Before heterogeneous models such as GARCH or HAR were introduced, the idea of efficient market prevailed in academic field. So-called Efficient Market Hypothesis, or EHM, was developed by Eugene Fama in late 1960's – early 1970's. According to EHM, all available information is translated into prices quickly enough so there is no possibility of making profits in the long run and prices are at equilibrium level. This statement, among other things, relies on rather strong assumption of homogeneity of market participants. In this case it means homogeneity of their expectations, perception and incorporation of available information into their decisions, and similar investment horizons.

At this point it is necessary to note that efficient market hypothesis was challenged both by academics and financial markets themselves. Recent financial crisis of 2008-2009 is a good example of it.

Relevant to evolution of heterogeneous models is the work of Muller et al. (1993), where authors challenge the assumption of homogeneity. According to them it is more realistic to assume market participants to differ in their expectations and investment horizons if different types of investors are considered, such as daily traders, medium-sized funds and large institutional investors. Sources of heterogeneity may be different preferences regarding risk and liquidation of opened positions. Based on this argumentation authors propose Heterogeneous Market Hypothesis ("HMH") framework where they try to address issues of heterogeneity of market agents.

HMH was further utilized in the proposition of Heterogeneous Generalized Autoregressive Conditional Heteroscedasticity model ("HARCH"). This type of models provides a framework for capturing above mentioned heterogeneity of market participants. For the formalization of the HARCH framework reader can refer to Dacorogna et al. (1997).

Finally, Corsi (2007), based on works discussed above, proposed an additive cascade model of volatility components that are defined over different periods of time. Suggested model utilizes simple AR-process and is based on realized measures.

The model for realized variance can be formalized as follows:

$$RV_t = \beta_0 + \beta_1 RV_{t-1} + \beta_2 RV_{t-1}^{(5)} + \beta_2 RV_{t-1}^{(22)} + \varepsilon_t \quad (3.9)$$

where RV_t is a variance realized during the day *t*, $RV_{t-1}^{(5)}$ is a *(t-1)* lagged average realized variance for 5 days and $RV_{t-1}^{(22)}$ is a is a *(t-1)* lagged average realized variance for 22 days. Although the setup of the model is rather simple and the model lacks true long memory properties, Corsi argues that based on empirical results model performs relatively well in reflecting main empirical features of financial time-series, i.e. fat tails, self–similarity and long memory. Moreover, author also reports that HAR model performs extremely well in forecasting tasks.

In our analysis we use HAR model for out-of-sample forecasting of realized correlation. The model we use is similar to one used for RV_t and is defined as follows:

$$RC_t = \beta_0 + \beta_1 RC_{t-1} + \beta_2 RC_{t-1}^{(5)} + \beta_2 RC_{t-1}^{(22)} + \varepsilon_t \quad (3.10)$$

This setup allows to model autoregressive process of realized correlation using lagged values that effectively contain information about the process lagged up to one month.

4 Neural networks

As it was discussed in the introduction to this thesis, our research is focused on modelling correlations based on high-frequency data and forecasting of realized daily correlations of selected stock markets and commodities.

In the first part of our empirical work we model correlations of selected stock markets and commodities using intraday 5 minutes data on returns. We model conditional correlations using daily realized correlation measure. We also estimate unconditional correlations using average daily realized correlations over the given periods of time. We also apply feedforward neural network on high frequency data to obtain estimated unconditional correlations for the given period of time. We then compare unconditional estimates obtained using neural network to the average realized daily correlations to see if the network is capable of approximating the realized correlations process.

In the second part of our empirical work we forecast time-series of realized daily correlations we computed in the first part of our empirical research, assuming certain autoregressive process behind the RC development. The task is to apply several selected models, namely Autoregressive model of order *p* (AR(p)), Heterogeneous Autoregressive model (HAR) and dynamic Nonlinear Autoregressive Neural Network (NARNET) to modelling of realized daily correlations and assess (i) relative performance of mentioned models, and (ii) overall goodness of obtained forecasts. Detailed description of data used and explanation of application of selected methods to modelling and forecasting tasks are presented in the Chapter 5 of this thesis.

We introduce theoretical foundations of artificial neural networks in Chapter 4 of this thesis. We start with a revision of existing literature on application of artificial neural networks in stock markets analysis. Following section presents theoretical foundations of a building block of any neural network – artificial neuron, where we describe biological concepts of human neuron, transfer functions and connection of neurons in the network. Section that follows describes architecture of neural networks we used in our research. The last section is devoted to network learning algorithms and backpropagation method.

4.1 Application of Neural Networks in Quantitative Finance

Even though artificial neural networks have been used in quantitative finance for relatively short time, there is already a vast pool of literature regarding application of neural networks in stock markets analysis. In general, existing literature on neural networks in stock markets analysis can be divided into two categories. First category of works suggests that there is no gain from using artificial neural networks instead of classical econometric models. Second category of works regards neural networks as powerful computational mechanisms and authors report significant improvement of results if NN is used.

One of the first works regarding application of neural networks to stock market analysis was prepared by McCluskey (1993). Author applied feedforward and recurrent neural networks to forecasting S&P 500 (S&P) stock market index. Even though in the introduction to his research author claims that recurrent NN may improve forecasting performance significantly, empirical results of his work do not suggest any significant improvement in forecasting performance. Another attempt to use neural networks in stock market prediction was made by Kulkarni (1996) where S&P 500 (S&P) stock index one-week ahead prediction was obtained. The forecast was based on lagged values of S&P itself along with past values of long and short term interest rates. Author use feedforward neural network and reports that trained NN performs very well in prediction task even when applied on time-series with lots of unexpected rises and falls. Artificial neural networks were applied by Pan et al. (2005) to forecast Australian stock market index AORD. Authors utilize multi-layer feedforward neural networks to find best specification of NN for prediction purposes. As input information authors use six daily returns of AORD, one daily return of S&P and day of the week. Having that approach to the problem they report up to 80% directional prediction correctness. Quah (2007) applies radial basis function (RBF) neural network to selection of Dow Jones Industrial Average (DJIA) equities problem. He uses General Growing and Pruning Radial Basis (GGAP-RBF) neural network and compares its performance to less complicated Multi-layer Perceptron (MLP) network and Adaptive Nero-Fuzzy Inference Systems (ANFIS). Author concludes that there is a potential for good performance of neural networks in equities selection problem and improvement of results of investment strategy, but in his work RBF did not outperformed MLP and ANFIS techniques.

Turning to the training methods of neural networks used in practice, paper by Rumelhart et al. (1986) argues that backpropagation is the most used learning method in stocks and stock markets analysis. Several authors, namely Jovina and Akhtar (1996), Brownstone (1996), Quah and Sirinvasan (1999), Oh and Kim (2002), Salim (2011) used backpropagation algorithm and achieve relatively good performance of the networks. Another works by Kohzadi et al. (1996), McNelis et al. (1998) and Yim (2002) suggest that artificial neural networks perform better than conventional time-series analysis methods.

One very important feature of artificial neural network is its generalization ability. Beale et al. (2012) argues that properly trained multi-layer neural network in general gives reasonable results when inputs that network has not seen are presented to it. New inputs lead to as good result as one obtained during the network's training if this input is the same as ones used for training. This feature is important because it allows training the network on representative set of inputs, rather than on all available subsets of inputs. Generalization is also crucial for avoiding overfitting. Network may perform very well on the training data set, when error is forced to be on the low level, but such a network won't perform well with new data sets. Thus it is also important for application of NN to forecasting of financial time-series. One of the attempts to improve generalization ability of the network was undertaken by Hochreiter et al. (1997) by introducing an algorithm for location of a flat minimum of an error function. Authors argued that this algorithm improved performance of the network significantly.

Recurrent neural networks have drawn a lot of attention recently. Beale et al. (2012) argues that delayed inputs in the recurrent network make it more efficient in forecasting tasks. Lee Giles et al. (2001) used recurrent neural network for noisy FX time-series prediction and reported 47.1% error rate of next day prediction, and reduction of error rate down to 40% if intervals with low system confidence were removed. Other authors that utilized recurrent neural networks in their work are Zimmermann et al. (1996), Kohara et al. (1997), Kim and Chun (1998). Authors conclude that recurrent networks have good potential in in-sample fitting and forecasting tasks.

Another important issue associated with use of neural networks is a selection of number of hidden layers and neurons. The problem is loss of degrees of freedom with increased number of layer due to larger number of parameters to be estimated (McNelis, 2005). Use of networks with more than one hidden layer was discussed, for example, in works of Hornik et al. (1989) and Dayhoff and DeLeo (2001). Authors conclude that a general function approximation theorem has been proven for three-layer neural networks and that two-layer network with trainable parameters (weights and bias) can approximate any nonlinear function. Beale et al. (2012) makes the same conclusions regarding approximation capabilities of two-layer neural networks.

Based on reviewed literature and to our best knowledge, we conclude that neural networks have never been applied to forecasting of realized correlations. We use dynamic recurrent neural network in the second part of our empirical work for realized correlations forecasting, which is described in the Chapter 5 of this thesis.

4.2 Artificial Neuron - Building Block of Neural Network

Artificial neural networks are computational structures that are based on the concept of functioning of biological nervous system. This concept is built on the mathematical apparatus developed to simulate the behavior of neurons of human brain and nervous system. NN use relatively simple mathematical operations for solving nonlinear stochastic problems. The biological nervous system functions as follows: it consists of interconnected neurons, e.g. individual cells characterized by structural properties that represent very fast transmission of electrical signals among neurons in the system (Churchland and Sejnowski, 1992). Each neuron acts as independent parallel computational centre that reacts in a unique way to a received signal and sends transformed signals to neighboring neurons in the network. The key feature of that network is that weights of incoming signals from other neurons are not uniform, e.g. some connections are more important, thus have higher priority in information input. Such interconnectedness and non-uniformity of signal weights is the fundamental idea behind the artificial neural networks.

For example, the biological neuron can be approximated by relatively simple mathematical concept of perceptron. This concept is based on summing function and can be represented based on Rojas (1996) as follows:

$$\sum_{i=1}^{k} w_i x_i = \xi , \tag{4.1}$$

where x_i are inputs and w_i are respective weights. The perceptron then compares the sum ξ with the respective threshold θ (defined separately for each particular case).

Let's say, that if value of ξ exceeds that of θ, the perceptron sends out the signal of 1. In other case it sends the signal of 0. That signal is the received by neighboring perceptrons in the hypothetical network, which again assign the weights to each input and send out further signals. That way the initial inputs are transformed with neural network into the output signal.

In other words, like the linear an nonlinear methods (such as (G)ARCH models), the neural network relates a set of input variables $\{x_i\}, i = 1, ..., k$ to a set of one or more output variables $\{y_j\}, j = 1, ..., k$. The difference between a neural network and the other approximation methods is that NN makes use of one or more hidden layers with one or more neurons, in which the input variables are squashed or transformed by a special function, known as logistic or log-sigmoid transformation (McNelis, 2005). Hidden layers approach represents very efficient way to model nonlinear stochastic processes.

Let's examine simple artificial neural network. According to Haykin (1994), so-called single neuron (node) is an information-processing unit that is fundamental to the neural network functioning. The following figure shows the model of a neuron which forms the basis for designing artificial neural networks:

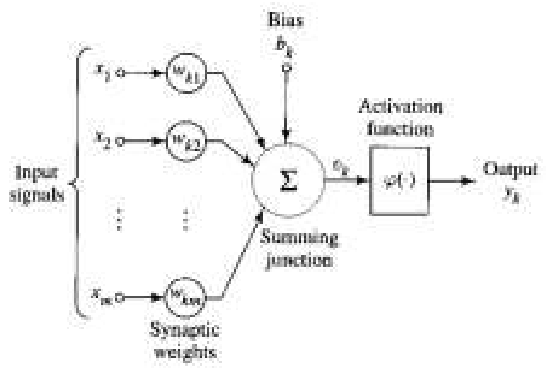


Fig. 4.1: Simple neuron model

There are three basic elements of neuronal model:

1. Set of synapses or connection links, each of which is characterized by a weight or strengths. Each input (or signal) x_i connected to a subject neuron (in this case it is a neuron k) is multiplied by the synaptic weight w_{ki}. The k subscript denotes the subject neuron. The weights in artificial neuron can have negative as well as positive values.
2. Adder, for summing the input signals (represented on the figure as summing junction).

3. Activation or squashing function that limits the amplitude of the output of a neuron to some finite value.

The external bias term b_k, also included in the model, lowers or increases the net input of the activation function.

In mathematical terms the neuron k can be defined as following pair of equations. Note, that it basically similar to the perceptron concept described above in this section:

$$u_k = \sum_{i=1}^{m} w_{ki} x_i \;, \tag{4.2}$$

and

$$y_k = \varphi(u_k + b_k)\,, \tag{4.3}$$

where $\{ x_i \}$ are inputs; $\{ w_{ki} \}$are synaptic weights of neuron k; u_k is the linear combiner output respective to the input signals; b_k is the bias term; $\varphi(\cdot)$ is the activation (or squashing) function; and y_k is the output of neuron. The external bias term b_k has an effect of applying an affine transformation on the output u_k depending on whether it is negative or positive. The term $u_k + b_k$ can be also regarded as local induced field or activation potential υ_k.

One of the most important steps when designing an artificial neural network is the selection of activation or squashing function. Along with number of hidden layers neurons in the system, it will define the way inputs are transformed into the outputs of neurons. First one, the simplest one, is so-called threshold function defined as follows:

$$\varphi(\upsilon) = \begin{cases} 1, \upsilon > 0 \\ 0, \upsilon \le 0 \end{cases} \tag{4.4}$$

Such a neuron was presented by McCulloch and Pits (1943) and it has all-or-none property: the output of neuron takes the value of 1 if the local induced field is nonnegative and 0 otherwise. Figure 3.2 shows the plot of such activation function:

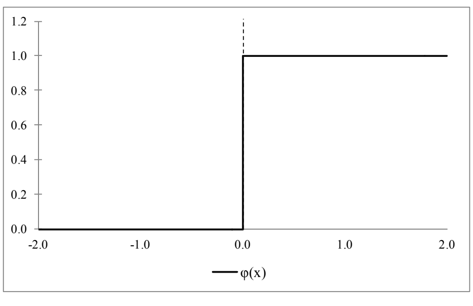


Fig. 4.2: Threshold activation function

Second type of potential activation function is so-called pricewise-linear function. It is defined as follows:

$$\varphi(\upsilon) = \begin{cases} 1,\_\ \upsilon \geq +\frac{1}{2} \\ \upsilon,\_\ -\frac{1}{2} \leq \upsilon + 1/2 \leq +\frac{1}{2} \\ 0,\_\ \upsilon \leq -\frac{1}{2} \end{cases}, \tag{4.5}$$

This form of an activation function can be seen as a rough approximation of a nonlinear amplifier. The following figure shows the graph of pricewise-linear function:

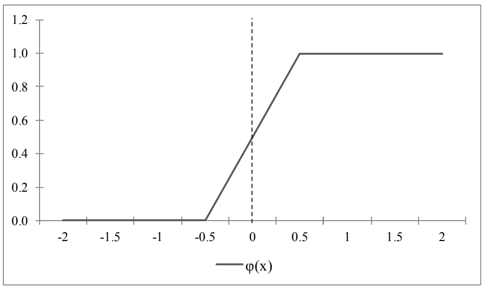


Fig. 4.3: Pricewise-linear activation function

The last but not least type of activation function is nonlinear sigmoid function. That particular function is one of the most often used in application of artificial neural networks. It is defined as strictly increasing function possessing linear and nonlinear characteristics. An example of sigmoid function can be so-called logistic function. It can be defined as:

$$\varphi(\upsilon) = \frac{1}{1+\exp(-a\upsilon)}, \tag{4.6}$$

where *a* is the slope coefficient. By changing the slope coefficient we can change the shape of sigmoid function. If *a* approaches infinity in limit, sigmoid function approaches the simple threshold activation function. The sigmoid function has several important properties. Firstly, it assumes the continuous range of values, whereas threshold function assumes 1 or 0. Secondly, the sigmoid function is differentiable. In practice, the selection of an activation function depends on assumed underlying stochastic process and assumed distribution of examined random variables. The following figure illustrates the sigmoid function with different parameter *a*.

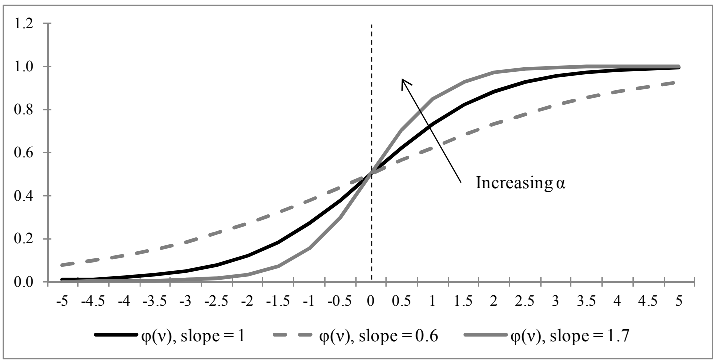


Fig. 4.4: Sigmoid activation function

Another frequently used activation function is so-called tansig or tanh function (McNelis, 2005). It is defined as:

$$\varphi(\upsilon) = \frac{e^{\upsilon} - e^{-\upsilon}}{e^{\upsilon} + e^{-\upsilon}}, \qquad (4.7)$$

This function squashes the linear combinations within the interval of [-1;1] rather than [0;1] in the sigmoid function. The behavior of such function is illustrated on the following figure:

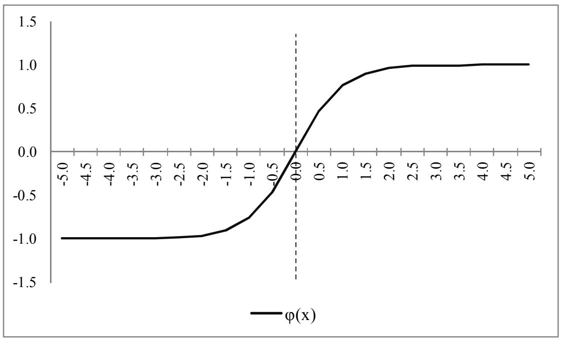


Fig. 4.5: Tansig or tanh activation function

It is necessary to note, that there are much more types of activation functions than just mentioned above, e.g. Gaussian cumulative function, hyperbolic tangent function etc. In practice sigmoid function is used most often. The reason for that is sigmoid function’s “threshold like” behavior which is considered to be appropriate for describing behavior of economic variables. A good example can be a reaction of savings, on the changes in interest rate. Small deviations from the somewhat stable level of interest rate as at particular moment of time most likely won’t affect savings level significantly. However, further increase in interest rate will probably cause changes in savings level in the economy. On the other hand, there is certain level of savings that cannot be exceeded regardless even larger increase in interest rate. One can easily imagine that this behavior of saving is similar to the behavior of the sigmoid function. Kuan and White (1994) argue that this threshold feature is a fundamental characteristic of nonlinear response of the neural network. In particular, they characterize that feature as the "tendency of certain types of neurons to be quiescent of modest levels of input activity, and to become active only after the input activity passes a certain threshold, while beyond this, increases in input activity have little further effect" [Kuan and White (1994), p. 2].

Logsig or tansig function should be selected based on the distribution of subject variables. As it can be seen from the functions formulations, or even graphs, main difference between them is the speed of saturation. In other words, logsig function will responds faster to the same inputs than tansig one. For the purposes of our analysis we have utilized both mentioned activation functions and compared performance of the neural network with each of them.

4.3 Neural Networks Architecture

The way neurons and inputs are organized in the artificial neural network is called the network architecture. This organization structure of a network is also linked with its learning algorithm that is used to train the network.

We start with describing a feedforward neural network that we applied in the first part of our empirical analysis for modelling correlations among selected stock market indices and commodity futures. This kind of network is considered to be a static one meaning it is used for in-sample fitting. For the second part of this thesis, i.e. for the forecasting of realized correlations, we selected nonlinear autoregressive neural network (NARNET). It is considered to be a recurrent dynamic network. We describe the NARNET later in this chapter.

Having described main building blocks of neural network in the previous subsection it is necessary to generalize the way neural network works. As it was stated above, neural network can be seen as a computational structure capable of approximating unknown nonlinear process given inputs and targets, and after the network is trained. One advantage of NN if compared to classical econometric models described previously in this thesis is the fact that neural network does not rely on assumptions about the distribution of input variable and errors, such as normality, homoscedastic errors etc.

4.3.1 Feedforward Neural Network

The simplest type of a neural network is a single-layer feedforward network. Picture below represents the feedforward network with one hidden layer containing 2 neurons, three inputs $\{x_i\}, i = 1, 2, 3$ and one output y (McNelis, 2005).

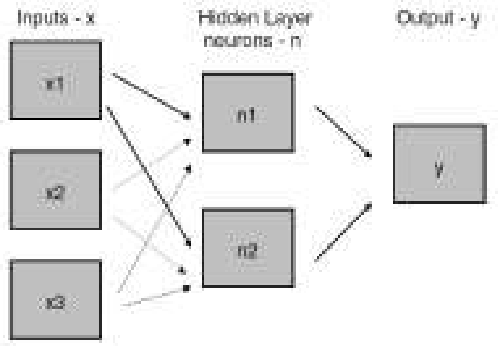


Fig. 4.6: Feedforward neural network

We can see the parallel processing of inputs inside the network. McNelis (2005) refers to the connectors between input variables as input neurons and to the neurons in hidden layer and connectors between the hidden layer neurons and output variable as output neurons. Inputs are summed with neurons weights and bias in order to create so-called net input. Net input is then processed by each neuron and network's estimate of the target y is returned. Estimate value is then compared to the true output value and error is calculated. To obtain most accurate approximation of relationship between inputs and output, error function is constructed and minimized. The process of minimizing the error function is called network training. Weights along with bias are free parameters of the network that are adjusted during the training. Training algorithms and methods will be discussed further in this thesis.

Next picture represents single-layer neural network as it is constructed in the Matlab software. We can observe inputs, weights w and bias b. We can also observe the flow of signals in the model – through the hidden layer, with sigmoid function neurons, to the output. There is another layer present, called output layer, which is based on the linear transfer function neuron. In majority of cases, this layer is used just to speed up the processing of signals by linear transformation of hidden layer's output, thus hidden layer is of particular interest.

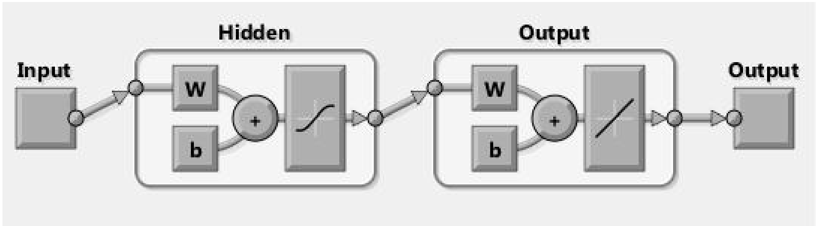


Fig. 4.7: Single-layer feedforward neural network (Matlab)

Single layer feedforward neural network can be represented mathematically as follows (McNelis, 2005):

$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i*} \omega_{k,i} x_{i,t} \tag{4.8}$$

$$N_{k,t} = L(n_{k,t}) = \frac{1}{1+e^{-n_{k,t}}} \tag{4.9}$$

$$y_t = \gamma_0 + \sum_{k=1}^{k*} \gamma_k N_{k,t} \tag{4.10}$$

where $L(n_{k,t})$ represents the logsigmoid activation function formalized as $\frac{1}{1+e^{-n_{k,t}}}$. There are $i*$ input variables $\{x\}$ in the system, and $k*$ neurons. Variable $n_{k,t}$, or net input, is formed by linear combination of input variables as at time *t*, $\{x_{i,t}\}, i = 1,...,i*$, with set of input weights $\omega_{k,i}, i = 1,...,i*$, and constant term, or bias, $\omega_{k,0}$. This net input is then transferred using logsigmoid function $L(n_{k,t})$ and becomes a neuron $N_{k,t}$ at time *t*. Hyperbolic tangent, or tansig, function (4.7) can be used as an alternative to the logsig to transfer the net input. The set of $k*$ neurons is then put in a linear combination with vector of coefficients $\{\gamma_k\}, k = 1,...,k*$ and a constant term γ_0 to form a forecast $\hat{y}_t$ at time *t*. Feedforward neural network combined with a sigmoid activation function is also known as multi-layer perceptron or MLP network. It is a basic network architecture that is often used by econometric researchers as a starting specification of the network.

Certain econometric problems may require more complex structure of the network. Single-layer neural network can be extended by adding one or more hidden layers. Multilayer networks are very powerful approximators. As it is argued in Beale et al. (2012), network of two layers, with sigmoid transfer function in the first layer and linear in second, is capable, after it is properly trained, of approximating any nonlinear function with finite number of discontinuities. Multilayered neural network is formalized as follows given k^* neurons in the first hidden layer and l^* neurons in the second hidden layer and assuming logsigmoid transfer function (McNelis, 2005):

$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i^*} \omega_{k,i} x_{i,t} \tag{4.8}$$

$$N_{k,t} = \frac{1}{1+e^{-n_{k,t}}} \tag{4.9}$$

$$p_{l,t} = p_{l,0} + \sum_{k=1}^{k^*} p_{l,k} N_{k,t} \tag{4.11}$$

$$P_{l,t} = \frac{1}{1+e^{-p_{l,t}}} \tag{4.12}$$

$$y_t = \gamma_0 + \sum_{l=1}^{l^*} \gamma_l P_{l,t} \tag{4.13}$$

We utilize the MLP network architecture with two hidden layers with ten neurons each for modelling correlations in the first part of our empirical work presented in the Chapter 5 of this thesis.

4.3.2 Nonlinear Autoregressive Neural Network

Another type of network we use in our analysis is a nonlinear autoregressive neural network, or NARNET. It is dynamic recurrent network based on Elman recurrent network architecture. In this type of network neurons depend not only on external inputs, but also on their own lagged values. Following chart depicts general structure of Elman recurrent network as presented in McNelis (2005):

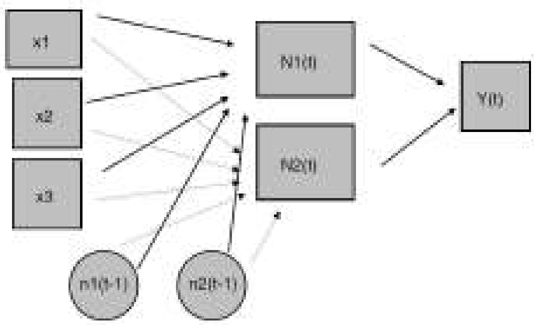


Fig. 4.8: Elman recurrent network

Elman network builds “memory” in the evolution of neurons. This type of network is similar to moving-average (MA) process that is often used in financial time-series analysis. In the MA process, dependent variable y is a function of independent variables

x as well as current and lagged values of a random shock ε . Equations formalizing the a q-th order MA process are as follows:

$$y_t = \beta_0 + \sum_{i=1}^{i*} \beta_i x_{i,t} + \varepsilon_t + \sum_{j=1}^{q} \nu_j \hat{\varepsilon}_{t-j} \qquad (4.14)$$

$$\hat{\varepsilon}_{t-j} = y_{t-j} - \hat{y}_{t-j} \qquad (4.15)$$

In a similar fashion, the Elman recurrent network makes use of current as well as delayed observations of unobservable unsquashed neurons in a hidden layer. In a same as for the MA model, it is necessary to use multi-step estimation procedure for the Elman recurrent network (McNelis, 2005). First step is to initialize vector of lagged neurons with lagged neurons proxies from a simple feedforward network. Next step is to estimate their coefficients and recalculate the vector of lagged neurons. Parameter values are recalculated in a recursive manner. The process goes on until the convergence occurs.

It is necessary to note that recurrent network has a time dimension. It means that it can be applied only on time-series data as opposed to feedforward network that can be used as well for panel data. System of equations formalizing recurrent neural network is presented below:

$$n_{k,t} = \omega_{k,0} + \sum_{i=1}^{i*} \omega_{k,i} x_{i,t} + \sum_{k=1}^{k*} \phi_k n_{k,t-1} \qquad (4.16)$$

$$N_{k,t} = \frac{1}{1+e^{-n_{k,t}}} \qquad (4.9)$$

$$y_t = \gamma_0 + \sum_{k=1}^{k*} \gamma_k N_{k,t} \qquad (4.10)$$

In practice, Elman's type of the network is often used for forecasting foreign exchange and stock markets returns using intraday high-frequency data. Structure of the network changes over time thanks to lagged neurons being feed back into current ones. However, lagged neurons are feed back before they are transformed by sigmoid function. So the network does not use lagged neurons from the output level. This process allows for memory capturing in financial markets.

Following scheme depicts an example of NARNET architecture with five delays, two hidden layer containing ten neurons with sigmoid function each as presented in the Matlab software:

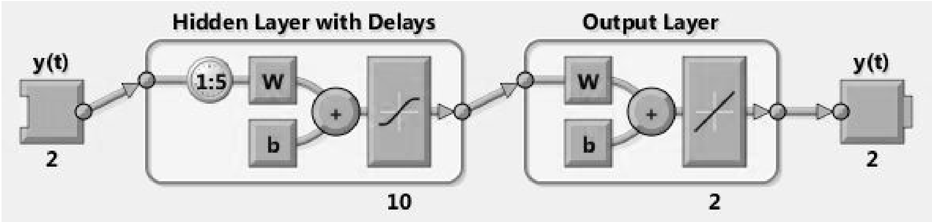


Fig. 4.9: Nonlinear autoregressive network (Matlab)

In our analysis we utilize nonlinear autoregressive neural network for forecasting realized correlations. Data and procedures used during this exercise are described in the fifth chapter of this thesis.

4.4 Network learning

Distinctive feature of a neural network is that it learns from data it is applied on about the environment. Learning is done in a way of iterative adjusting free parameters of the networks – its weights and biases. Aim of learning, or adjustments iterations, is to make network "understand" underlying process in the presented data. Each iteration increases network's understanding of the environment. Every time external information is presented to the networks and free parameters are adjusted, internal structure of the network is changed and so is its reaction to the new information.

There are different ways, or algorithms, how the learning of a network can be done. It should be noted that the network learning can be actually viewed as a nonlinear estimation problem. As it was already discussed earlier in this chapter, the network's output is compared to the actual values of a target and error is computed after each iteration. The aim is to minimize the error function. Nonlinear estimation means that there can be multiple locally optimal solutions none of which yielding the global minimum of an error function. When estimating the network one should start with some initial values of weights. Basically, these initial values are some kind of guess that can move final optimal solution closer to or further from the global optimum. This problem can be illustrated by the following picture taken from the McNelis (2005).

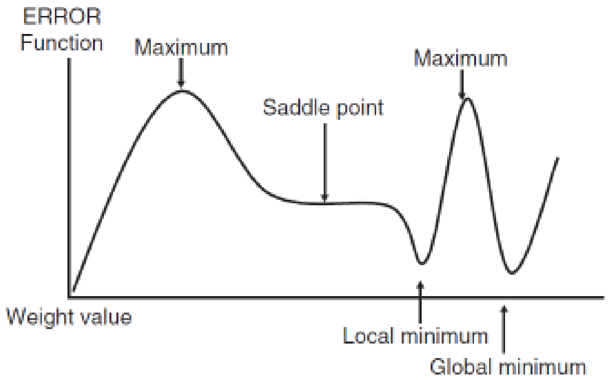


Fig. 4.10: Weight values and error function

Clearly, when estimating the network one can easily get stuck at some locally optimal point. Unfortunately, there is no single way how to avoid locally optimal solutions. In our work when estimating the network we set initial weights randomly and ran estimation several times to see how the results will differ. If the difference is marginal we assumed that the result is best solution network can provide.

To formalize the learning process we describe main principles on the example of a single layer network. To find the optimal set of coefficients $\Omega = \{\omega_{k,i}, \gamma_k\}$ the loss function Ψ is minimized (McNelis, 2005). The loss function is defined as a sum of squared differences between the actual values of the target y_t network's output $\hat{y}_t$. The whole problem can be defined as follows:

$$\min_{\Omega} \Psi(\Omega) = \sum_{t=1}^{T} (y_t - \hat{y}_t)^2 \tag{4.11}$$

The network's output is a function of the inputs x_t and the set of weights and bias Ω:

$$\hat{y}_t = f(x_t; \Omega) \tag{4.12}$$

Generally, there are three main ways how to minimize a nonlinear function $\Psi(\Omega)$: local gradient-based search; stochastic search; and evolutionary stochastic, or genetic algorithms, search. In our work we have used the first method, i.e. local gradient-based search. It is based on first- and second-order derivatives of Ψ with respect to network parameters included in the set Ω. After every iteration of network learning, values of parameters are adjusted and this process continues until certain stopping criteria are met. For technical details of all three methods of the network learning reader may refer to McNelis (2005).

5 Empirical findings

In the following chapter we will utilize theory established previously in this thesis regarding classical econometric models and techniques, realized measures and artificial neural networks to analyze correlations between selected stock markets approximated by the stock market indices, namely FTSE, DAX and PX, and between S&P 500 stock market index and commodities futures, namely gold and crude oil. We assess the dynamics of the correlations by computing and analyzing daily realized correlations. Based on these estimates of conditional correlations we see how the correlations have been changing over the given horizon. We then utilize these estimates of daily realized correlations to get estimates of unconditional correlations for each individual year from interval between 2008 and 2011. We do so to compare these unconditional estimates with ones obtained using feedforward neural network. Based on literature discussed in the Chapter 4 of this thesis, we expect neural network to perform very well in in-sample fitting and forecasting tasks. The aim here is to see how good the NNs are in reality and understand if it is useful to utilize them in further research.

We start with the description of data we use in our empirical research and present main statistical characteristics of examined time-series. Second section of this chapter deals with modelling correlations between stock markets and commodities specified above. We assess dynamics of these correlations based on development of daily realized correlations. We then compare estimates of unconditional correlations based on average daily correlations and ones obtained using FFNN to see how weel the network is able to approximate the average daily RC. In the third section of this chapter we present results of forecasting realized correlations or pairs *DAX:FTSE* and *S&P:Oil* using AR(p) model, HAR model and dynamic recurrent neural network NARNET.

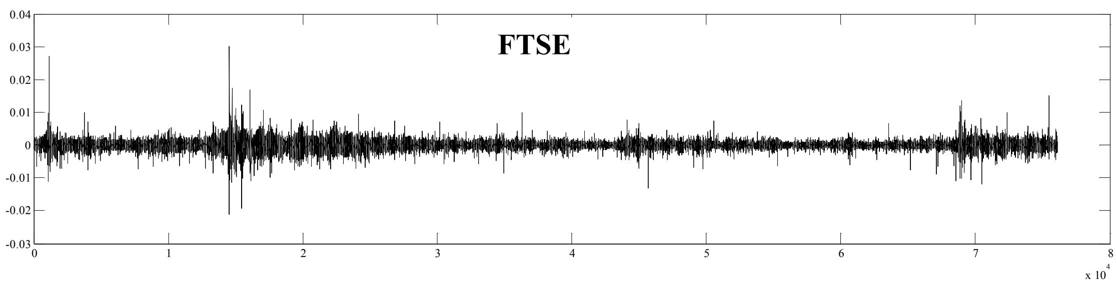
5.1 Data description

For the purposes of our analysis we have selected three stock markets – British, German and Czech represented by Financial Times Stock Exchange 100 (FTSE), Deutsche Börse (DAX) and Prague Stock Exchange 50 (PX) indices respectively. There were several reasons why these stock markets were selected. Firstly, overlapping trading hours that could capture many effects in cross-country equity returns correlation, starting from the co-movement of markets as a reaction to common news, markets contagion and ease of trading with another market participant at different location (Flavin et al., 2001). Adjustments to the initial data sets of FTSE, DAX and PX indices were made to correctly reflect parts with overlapping hours. Secondly, the size of the stock markets and level of concentration can affect cross-border correlations according to the mentioned paper. Authors argue that more liquid markets can have stronger correlations than markets with low liquidity. Thus we expect stronger correlations between DAX and FTSE than between DAX and PX and FTSE and PX. General conclusion by Flavin et al. (2001) is that geographical variables that are used to explain trade in goods, such as distance between markets, common border etc., are also applicable to the financial assets markets.

As an extension to modelling correlations between selected stock markets we also include commodities, represented by gold and crude oil futures traded on NYSE, in our analysis and model the correlations between them and a stock market represented by Standard & Poor's 500 (S&P) index. As it has been discussed earlier in this thesis, during the time of crisis investors tend to seek a safe place for their funds and may turn to commodities when there is a problem in the stock markets.

Our subject data samples comprise intraday logarithmic returns of specified indices and commodities collected on 5 minutes basis. Since the dynamics of the correlation was of our particular interest, especially its changes during the crisis, i.e. in 2008-2009, and right after it, selected time span of analyzed data for stock indices is 2 January, 2008 – 9 December, 2011; for commodities and S&P it is 1 February, 2008 – 29 November, 2011. Data was sourced from the Tick Data. After adjustments for differences in trading hours and holidays, stock markets indices sample, i.e. FTSE, DAX and PX, contains 76,085 observations. Sample with observations of gold, oil and S&P returns contains 62,832 observations.

Following figures plot subject time-series used for stock markets correlation analysis:



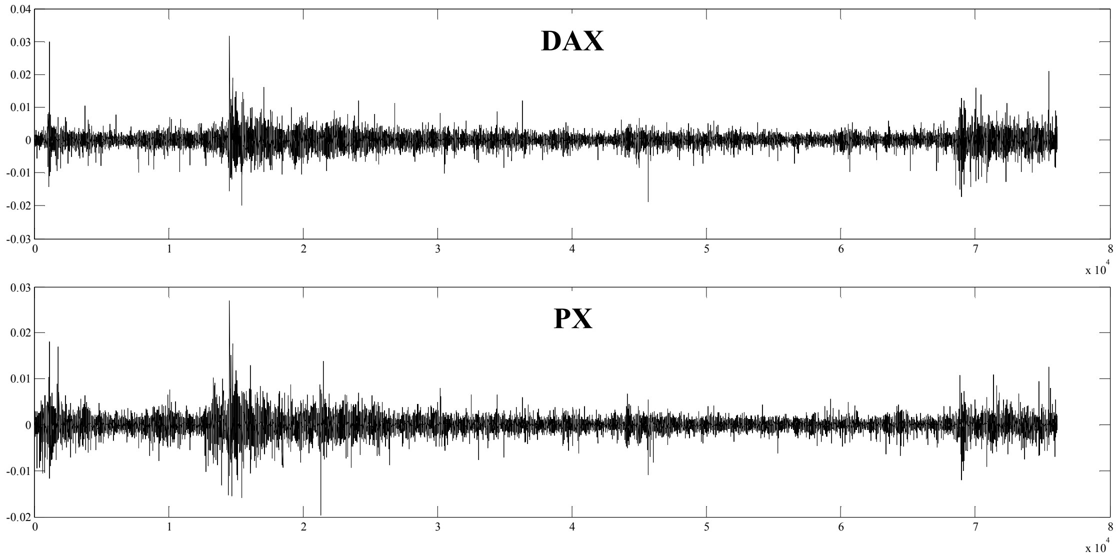
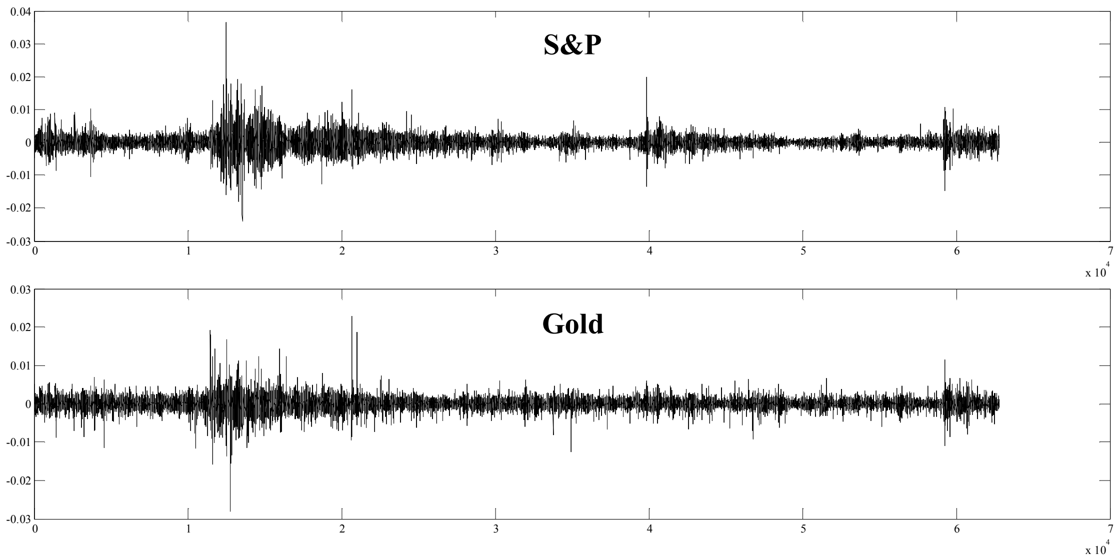


Fig. 5.1: FTSE, DAX and PX 5 minutes logarithmic returns

As it can be observed from the graphs presented above, there are clear clusters of relatively high volatility of returns in the end of 2008 – beginning 2009 and during 2011. Period between 2008 and 2011was characterized by relatively lower returns volatility. Worsening Euro zone debt crisis partially caused increased volatility in the end of 2010 – beginning 2011.

Following figures plot gold, crude oil and S&P time-series used for commodities and stock market index correlation analysis:



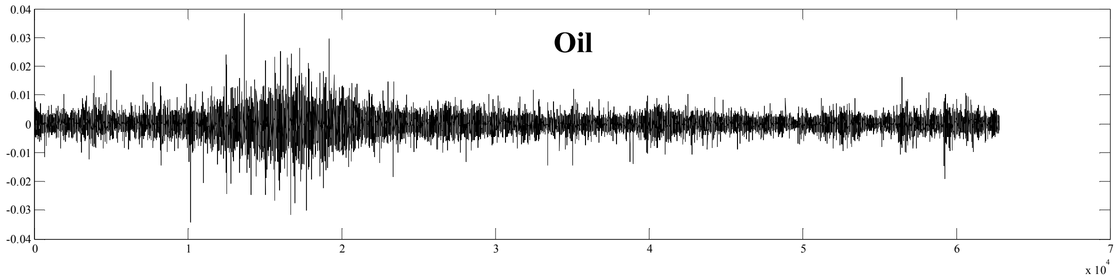


Fig. 5.2: S&P, Gold futures and Crude Oil futures 5 minutes logarithmic returns

From the charts above we can clearly observe clusters of increased volatility during the second half of 2008, especially for the oil futures returns, when spread of the US financial crisis hit the rest of the world. As in the case of three European indices mentioned above, during the period between second half of 2009 till late 2010 was characterized by relatively lower volatility of returns of subject instruments.

Working with such a data is a quite challenging task for any type of econometric model, either solving in-sample fit or out-of-sample forecast problem. In tables below we summarize descriptive statistics for subject time-series. CV stands for the Coefficient of Variation defined as a ratio of the standard deviation to mean.

Index / futures	Mean	Median	Minimum	Maximum
FTSE	0.00000	0.00000	-0.02109	0.030277
DAX	0.00000	0.00000	-0.02003	0.031769
PX	-0.00002	0.00000	-0.01967	0.027016
S&P	0.00000	0.00001	-0.02389	0.036665
Gold	0.00000	0.00000	-0.02809	0.022964
Oil	0.00001	0.00000	-0.03418	0.038403

Table 5.1: Summary statistics (1)

Index / futures	Std.dev	CV	Skewness	Kurtosis
FTSE	0.00118	242.570	0.22440	21.128
DAX	0.00140	338.750	0.07963	18.147
PX	0.00108	58.036	-0.16798	25.371
S&P	0.00159	768.530	0.38592	18.794
Gold	0.00120	1495.900	-0.19104	22.702
Oil	0.00239	339.390	-0.06853	12.278

Table 5.2: Summary statistics (2)

As it can be observed from the table 5.2, selected time-series possess feature of high excess kurtosis. This supports one of the stylized facts about financial time-series regarding the fat, or heavy, tails in probability distribution. In statistical literature it is argued that high kurtosis reflects infrequent extreme deviations which are specific to financial time-series. These deviations can be easily observed from the graphs of returns of selected indices and futures presented above. Once again, such data makes our task of modelling and forecasting correlations rather challenging.

Based on the summary of descriptive statistics we can see that the Oil future has suffered largest loss (-3.4%) among selected indices and commodities given the sampling frequency. In terms of largest positive returns among stock market indices S&P experienced largest increase of 3.6%. However, in terms of positive returns crude oil futures has outperformed S&P index with returns of 3.8%. It should be noted, that we are discussing the 5 minutes data and that picture of daily returns can be different. For the reader's reference, charts depicting returns distribution of the subject time-series with fitted normal distribution can be found in the Appendix 1.

Following table summarizes results of Jarque-Bera and Augmented Dickey-Fuller tests performed for all subject time-series of logarithmic returns.

Index / futures	**J-B test**	***p-value***	**ADF test**	***p-value***
FTSE	1,415,756.37	*0.000*	-160.11	*0.000*
DAX	1,044,118.53	*0.000*	-160.57	*0.000*
PX	2,040,929.52	*0.000*	-139.10	*0.000*
S&P	926,258.72	*0.000*	-145.26	*0.000*
Gold	1,349,683.06	*0.000*	-147.25	*0.000*
Oil	394,686.60	*0.000*	-144.15	*0.000*

Table 5.3: JB and ADF tests

Based on test statistics and p-values of the JB-test we can reject the null hypothesis about normal distribution of returns. This result is expected since we work with the financial time-series and it is a well known fact that such data is not usually distributed normally. Regarding the stationarity of subject time-series, based on the results of ADF-test we can reject the null hypothesis that there is a unit root present in the subject time-series. As it was discussed in the Chapter 2, ADF test has a tendency to over-reject H_0 even if there is a unit root. But based on the extremely low p-value of test statistics we will assume subject time-series of returns to be stationary.

5.2 Correlations Modelling and Dynamics

In the first part of our empirical analysis we analyze the correlations between three selected stock markets represented by FTSE, DAX and PX stock market indices. We use data for the period starting January. 2008 till December, 2011. We also analyze correlations between two commodities, namely gold and crude oil, represented by commodity futures and the US stock market index S&P 500 (S&P) for the period starting January, 2008 till November, 2011. As was discussed previously in this thesis there are certain limitations associated with computation of correlations among selected instruments (different trading hours, holidays etc.). That is why, given the available data, correlations were modelled for two groups of indices and commodities:

1) First group includes pairs: *FTSE:DAX, FTSE:PX* and *DAX:PX;*
2) Second group includes pairs: *Gold:Oil, Gold:S&P, Oil:S&P.*

Our analysis is based on high-frequency intraday data with sampling frequency of 5 minutes. We start with computing daily realized correlations for pairs specified above. Looking at the daily realized, or conditional, correlations is one way how to assess the dynamics of correlations.

We then compute averages of daily realized correlations for each year in considered range between 2008 and 2011 to get an estimate of unconditional correlations for each period. We also compute standard errors of these estimates to see how these unconditional correlations for each period have been changing and if year-on-year changes in these unconditional correlations are statistically significant. We then compute unconditional correlations for each period using multilayer feedforward neural network (FFNN). If estimates of unconditional correlations obtained from the FFNN lay in the interval of standard errors of average daily realized correlations it can be said that the neural network approximates these unconditional estimates.

Chart below illustrate the daily realized correlation estimates for pairs specified above. We start with the first group of indices. Daily realized correlations were computed for the range from January, 2008 till December, 2011. Number of observations in each data set is 954.

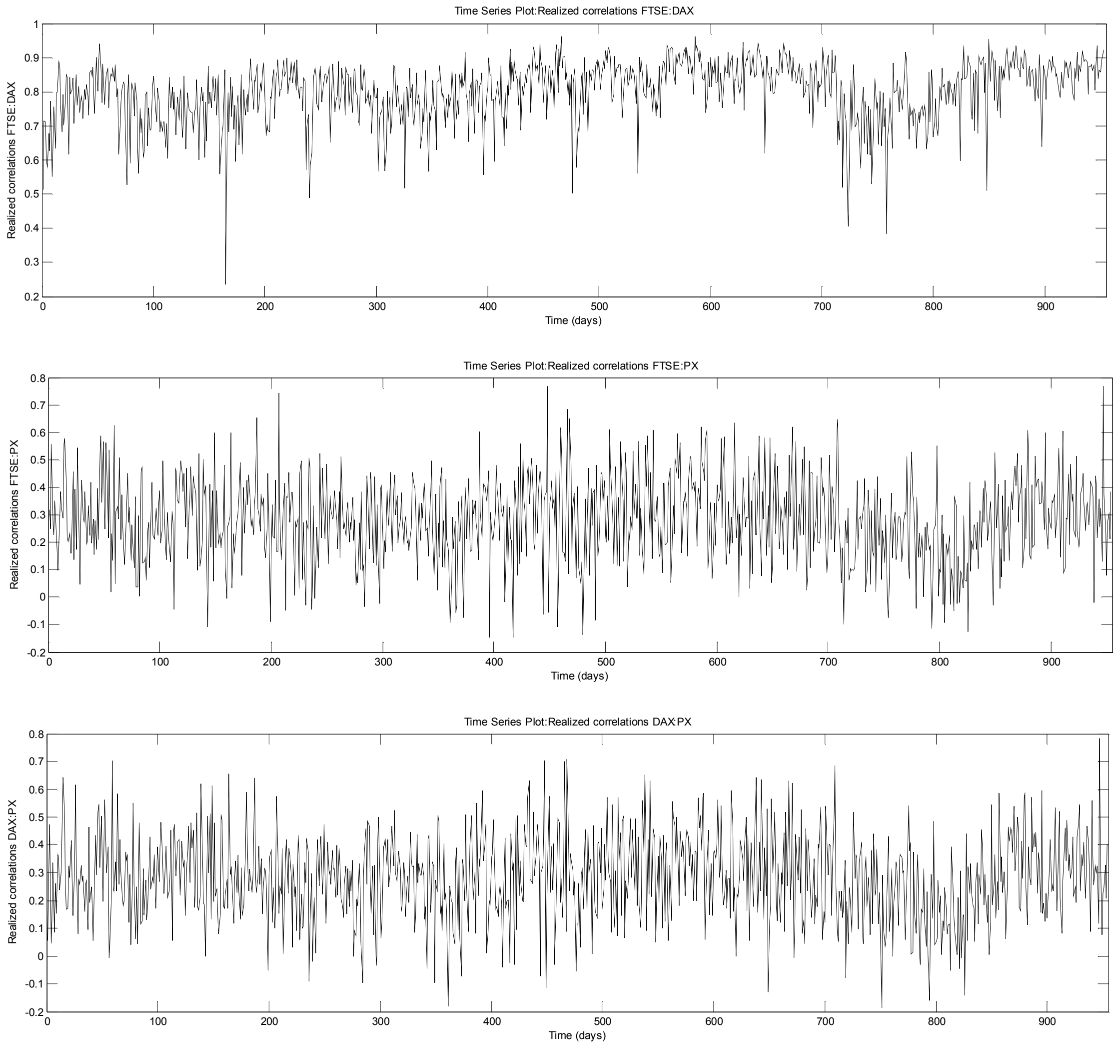


Fig. 5.3: Daily realized correlations for pairs *FTSE:DAX*, *FTSE:PX* and *DAX:PX* computed for the period starting 2 January, 2008 – 9 December, 2011.

From the charts above we can make first conclusions regarding the correlations between pairs *FTSE:DAX*, *FTSE:PX* and *DAX:PX*. Firstly, we can see that FTSE and DAX indices had much stronger correlation over the given period as compared to other two pairs in the group. Secondly, *FTSE:DAX* realized daily correlations experienced several significant decreases, namely between days 100 and 200 (year 2008), 400 and 500 (year 2009), and 700 and 800 (end of 2010 – beginning of 2011). Tables below summarize main statistics of subject time-series of daily realized correlations.

Pair	Mean	Median	Minimum	Maximum
FTSE:DAX	0.80119	0.81611	0.23583	0.962480
FTSE:PX	0.28046	0.28235	-0.14833	0.769050
DAX:PX	0.28032	0.28058	-0.18625	0.784650

Table 5.4: Summary statistics of daily realized correlations, indices (1)

Pair	Std.dev	CV	Skewness	Kurtosis
FTSE:DAX	0.08893	0.111	-1.25460	2.855
FTSE:PX	0.15434	0.550	-0.00513	-0.052
DAX:PX	0.15494	0.553	0.04656	-0.023

Table 5.5: Summary statistics of daily realized correlations, indices (2)

From the statistics presented in tables above it is clear that strongest correlation was between the UK and German stock market indices. Mean daily realized correlation between FTSE and DAX indices was on the level of 0.8 over the given horizon. Even though the standard deviations for this pair were relatively low (around 0.088), over the given period correlation of this pair experienced deepest decline down to 0.24 in August, 2008 and highest increase up to 0.96 in 2009. Overall, it can be said that these indices remained positively correlated during the whole examined period even though daily correlations experienced several significant upside and downside movements. On the other hand, correlations among *FTSE:PX* and *DAX:PX* pairs were significantly lower. For both pairs over the given period, mean daily realized correlation was on the level of 0.28. Both pairs have shown relatively high standard errors of 0.154 and both reached negative values of around -0.15 and -0.19 respectively. Lower correlations between FTSE, DAX and Prague stock exchange index can be explained by several factors. One of them may be the fact that UK and German markets are more developed then Czech one in terms of liquidity and depth of the market, i.e. number of transactions and volumes. This result is in line with results obtained by Flavin et al. (2001) where authors, using gravity model, concluded that size and liquidity of the stock markets can explain significant portion of their co-movements.

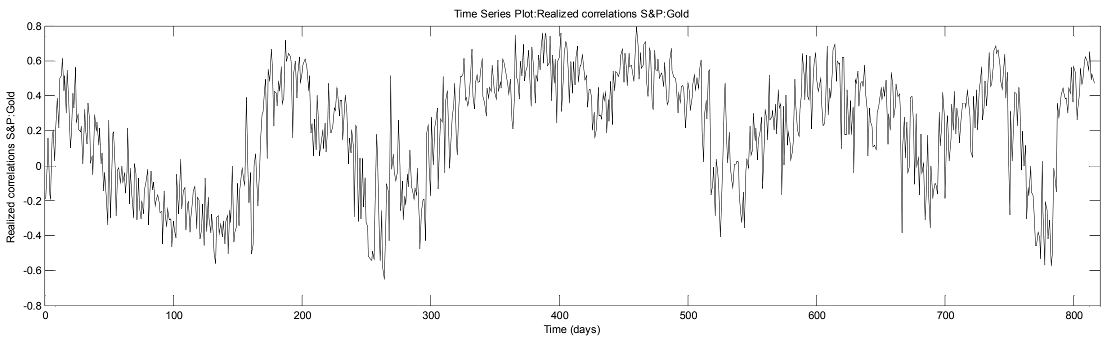
We use the time-series of daily realized correlations between FTSE and DAX in the second part of our empirical research where we forecast selected series of daily realized correlations using AR(p) model, HAR model and dynamic neural network NARNET. Since *FTSE:DAX* daily realized correlations are of our particular interest, let’s have a look at the year by year split of summary statistics of it.

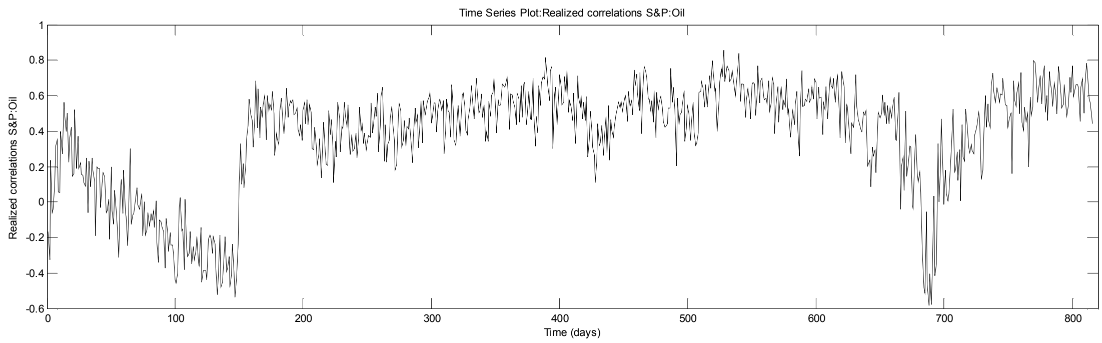
FTSE:DAX	2008	2009	2010	2011
Mean	0.76398	0.79206	0.84201	0.806830
Median	0.78440	0.79871	0.85635	0.830710
Minimum	0.23583	0.50390	0.40525	0.383460
Maximum	0.93951	0.96248	0.96188	0.955140
Std.dev	0.09161	0.08034	0.07615	0.089250
CV	0.11991	0.10143	0.09044	0.110620
Skewness	-1.33780	-0.91604	-2.18530	-1.144200
Kurtosis	3.96630	1.16180	7.85150	1.911500

Table 5.6: *FTSE:DAX* daily realized correlations, summary statistics split by years

Several notes on the numbers presented above regarding daily realized correlations of *FTSE:DAX* pair. Firstly, we can observe a slight increase in correlation from mean of circa 0.76 in 2008 up to 0.84 in 2010, taking in consideration decreasing standard deviations. Secondly, upper bound of daily realized correlations remained above 0.93 in all four years whilst lower bound has been changing from 0.24 in 2008 up to 0.38 in 2011.

Having discussed development of changes in daily realized correlations among indices from the first group, let's now turn to the daily realized correlations among S&P index and selected commodities futures. The daily realized correlations were computed for the period starting February, 2008 and ending November, 2011. Number of observations in each set is 816. Following chart illustrate subject time-series of daily realized correlations.





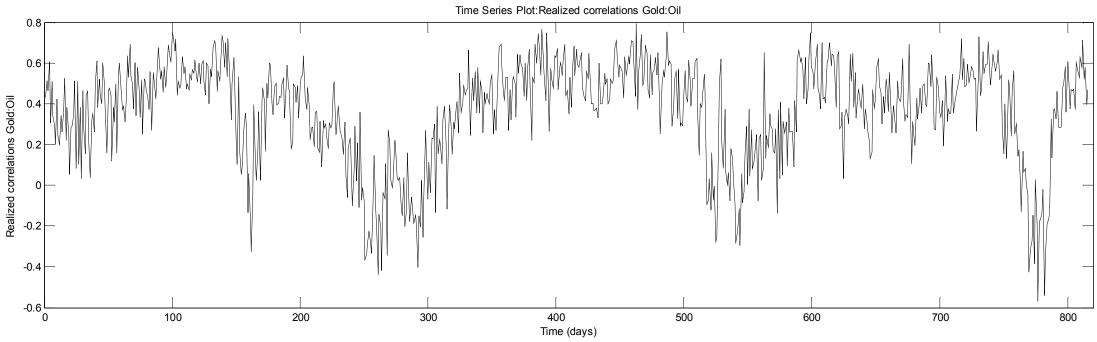


Fig. 5.4: Daily realized correlations for pairs *S&P:Gold*, *S&P:Oil* and *Gold:Oil* computed for the period starting 1 February, 2008 – 29 November, 2011.

From the charts above we can see how daily realized correlations were developing over the given period of time. It can be seen that daily realized correlations of all three pairs reached negative values during 2008 and between 2010 and 2011. To provide exact numbers, following tables summarize main statistics for subject time-series.

Pair	Mean	Median	Minimum	Maximum
S&P:Gold	0.20622	0.27282	-0.64906	0.790760
S&P:Oil	0.37473	0.46810	-0.58393	0.855250
Gold:Oil	0.36061	0.41106	-0.57146	0.792000

Table 5.7: Summary statistics of daily realized correlations, S&P and commodities (1)

Pair	Std.dev	CV	Skewness	Kurtosis
S&P:Gold	0.32365	1.569	-0.48878	-0.772
S&P:Oil	0.29979	0.800	-1.18780	0.732
Gold:Oil	0.24764	0.687	-1.01530	0.707

Table 5.7:Summary statistics of daily realized correlations, S&P and commodities (1)

We can see from the tables with summary statistics that daily correlations changed a lot between 2008 and 2011. We can say so based on mean value and minimum / maximum values of subject realized correlations. They also reveal significant standard errors. Interesting observation is that dynamics of daily realized correlations among S&P and two commodities is rather similar. There was a steady decrease of correlations from the beginning of 2008 followed by a sharp increase somewhere in August, 2008. Another deep decline, common for both index-commodity pairs, occurred in 2011.

We have selected *S&P:Oil* pair to use in the second part of our empirical work. Table below provides split of main statistics for this pair by individual years.

FTSE:DAX	2008	2009	2010	2011
Mean	0.07985	0.48179	0.54822	0.397130
Median	0.07942	0.49098	0.56149	0.486470
Minimum	-0.53686	0.10756	0.08752	-0.583930
Maximum	0.68548	0.81610	0.85525	0.796470
Std.dev	0.32270	0.13786	0.13839	0.296310
CV	4.04140	0.28614	0.25243	0.746130
Skewness	0.01778	-0.22713	-0.56193	-1.394200
Kurtosis	-1.18500	-0.47610	0.19254	1.799900

Table 5.8: *S&P:Oil* daily realized correlations, summary statistics split by years

Change in mean daily realized correlation between 2008 and 2009 should be noted. Increase in 2009 from 0.08 to 0.48 was followed by further increase in 2010 up to 0.55. Standard errors in 2008 were significantly higher than those in 2009 and 2010.

We now turn to the estimating unconditional correlations for each year by averaging daily realized correlations described above and by applying FFNN on high-frequency data. We use feedforward neural network with two hidden layers containing ten neurons each. Theoretical foundations and mathematical apparatus of feedforward neural network were presented in the Chapter 4 of this thesis. Neurons that we use are based on sigmoid transfer function. Actually, we employ both logsig and tansig transfer functions to see which one performs better on each individual data set. We also include bias in each neuron. Backpropagation method is used with Levenberg-Marquardt learning function, described in the Chapter 4 of this thesis. We use mean squared error function (MSE) as a performance function in our network.

Correlations for specified pairs of indices and commodities were calculated as follows: one index or commodity from the specified pairs was used as an input to the network and second one as a target to which network's output is compared. We solve in-sample fitting problem and use observations of both input and target form the same time step t. Main idea behind this approach is that the network should model, or approximate, the relation between two variables, in our case between two indices or commodities. The output of the network represents fitted values to each observation of the target based on the input values. Basically, output of a network is a result of network's "understanding of the process" underlying development of a target given the behaviour of inputs. In our case this process is a relation between indices and commodities included in each pair. Network's understanding of the process improves after each epoch of learning. As it was discussed in the fourth chapter of this thesis, learning occurs in an iterative way. After each step output of the network is compared to the target and error is calculated. Consequently, free parameters of the network, namely weights and biases, are adjusted to minimize the error. This process continues until convergence takes place. Typical behaviour of the error function during the learning, validation and testing is presented by the picture below:

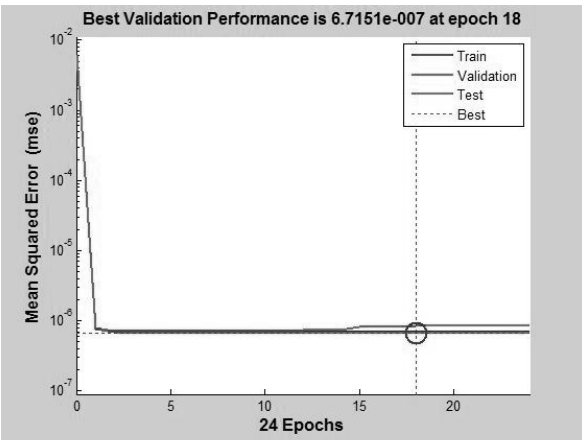


Fig. 5.5: Training, validation and testing performance, *FTSE:DAX* pair

Now, it is necessary to explain how exactly we calculate correlations for subject pairs of indices and commodities. As it was stated above, output of the network represents network's estimate of the target variable. We obtained estimates of the target variable for different time periods. We started with using the full sample of subject observations, i.e. data from the interval 1.2008-12.2011. We then calculated estimates of the target variables for each individual year. It should be noted that it is quite challenging to comment on the raw output from the neural network thus certain transformation of the output is needed. For the purposes of our analysis we have calculated coefficient of determination, or R^2, for each set of estimates, i.e. estimates for the whole period of time and individual years. Please note that square root taken from the determination coefficient is basically a correlation coefficient. This way unconditional correlations estimated using feedforward neural network that we report in this section are calculated based on goodness of fit of network's outputs. Since we work with intraday high-frequency data and it is not feasible to calculate daily correlations using NN and then report some kind of average correlation for the period with deviations due to small number of observations inside the day, we estimate the correlation for the period of time corresponding to the period of data set used. We then compare these estimates to the average daily realized correlations.. We discuss results obtained using realized correlation measure and their comparison to NN's outputs later in this section.

As it was discussed in the Chapter 4, one of the problems with NN training is the problem of local minimum of the error function. This issue can be addressed in different ways. In our work we assure that initial weights values do not lead to locally optimal solution by setting the initial weights randomly and running training, validation and testing of the network 5 to 10 times. If the difference in network's outputs is marginal, we accept such an outcome. Please note, that by using the same setup of the network several time with randomly defined initial weights, one will never get to the exactly same output (if only by chance). That is why marginal difference in networks output is acceptable.

Following tables report estimated unconditional correlations for indices pairs for each period using the FFNN. We also report average daily realized correlations and their standard deviations to see if FFNN's estimates fall into the interval of average daily RC standard errors.

Part of sample	**Neural network in-sample fit R^2**	**Unconditional correlation estimated using FFNN**	**Average realized daily correlation**	**Standard deviation**
		Tansig transfer function		
Full sample	0.66640	**0.81633**	0.801300	0.08893
2008	0.64420	**0.80262**	0.764000	0.09161
2009	0.63200	**0.79498**	0.792100	0.08034
2010	0.68310	**0.82650**	0.842000	0.07615
2011	0.72410	**0.85094**	0.806800	0.08903
		Logsig transfer function		
Full sample	0.66580	**0.81597**	0.801300	0.08893
2008	0.64070	**0.80044**	0.764000	0.09161
2009	0.63720	**0.79825**	0.792100	0.08034
2010	0.72730	**0.85282**	0.842000	0.07615
2011	0.72580	**0.85194**	0.806800	0.08903

Table 5.9: *FTSE:DAX* pair estimated unconditional correlations

Part of sample	**Neural network in-sample fit R^2**	**Unconditional correlation estimated using FFNN**	**Average realized daily correlation**	**Standard deviation**
		Tansig transfer function		
Full sample	0.11140	**0.33377**	0.280500	0.15434
2008	0.12320	**0.35100**	0.284200	0.15101
2009	0.07950	**0.28196**	0.273000	0.15395
2010	0.15350	**0.39179**	0.307100	0.15700
2011	0.11050	**0.33242**	0.256000	0.15168
		Logsig transfer function		
Full sample	0.11170	**0.33422**	0.280500	0.15434
2008	0.11680	**0.34176**	0.284200	0.15101
2009	0.08210	**0.28653**	0.273000	0.15395
2010	0.14850	**0.38536**	0.307100	0.15700
2011	0.10690	**0.32696**	0.256000	0.15168

Table 5.10: *FTSE:PX* pair estimated unconditional correlations

Part of sample	**Neural network in-sample fit R^2**	**Unconditional correlation estimated using FFNN**	**Average realized daily correlation**	**Standard deviation**
		Tansig transfer function		
Full sample	0.09190	**0.30315**	0.280300	0.15494
2008	0.10680	**0.32680**	0.285200	0.14405
2009	0.09050	**0.30083**	0.272400	0.15870
2010	0.14050	**0.37483**	0.306000	0.15614
2011	0.08910	**0.29850**	0.256300	0.15722
		Logsig transfer function		
Full sample	0.09140	**0.30232**	0.280300	0.15494
2008	0.10600	**0.32558**	0.285200	0.14405
2009	0.09020	**0.30033**	0.272400	0.15870
2010	0.13910	**0.37296**	0.306000	0.15614
2011	0.12230	**0.34971**	0.256300	0.15722

Table 5.11: *DAX:PX* pair estimated correlations

From the tables above we can see that the estimates of unconditional correlations obtained using the specified FFNN are in the interval of standard errors of average daily realized correlations using both logsig and tansig transfer functions. We can conclude that for purely index pairs the FFNN approximates unconditional correlations based on daily realized measures relatively well. We do the same check for S&P and commodities pairs. Tables below summarize the results.

Part of sample	**Neural network in-sample fit R^2**	**Unconditional correlation estimated using FFNN**	**Average realized daily correlation**	**Standard deviation**
		Tansig transfer function		
Full sample	0.02920	**0.17088**	0.206200	0.32365
2008	0.03530	**0.18788**	0.026300	0.32154
2009	0.05800	**0.24083**	0.256700	0.33306
2010	0.10000	**0.31623**	0.333200	0.24418
2011	0.03840	**0.19596**	0.211500	0.30377
		Logsig transfer function		
Full sample	0.02250	**0.15000**	0.206200	0.32365
2008	0.03340	**0.18276**	0.026300	0.32154
2009	0.05390	**0.23216**	0.256700	0.33306
2010	0.09940	**0.31528**	0.333200	0.24418
2011	0.04560	**0.21354**	0.211500	0.30377

Table 5.12: *S&P:Gold* pair estimated unconditional correlations

Part of sample	**Neural network in-sample fit R^2**	**Unconditional correlation estimated using FFNN**	**Average realized daily correlation**	**Standard deviation**
		Tansig transfer function		
Full sample	0.15000	0.38730	0.374700	0.29979
2008	0.09540	0.30887	0.079800	0.32270
2009	0.19320	0.43955	0.481800	0.13786
2010	0.34890	0.59068	0.547700	0.13815
2011	0.21470	0.46336	0.397100	0.29631
		Logsig transfer function		
Full sample	0.14940	0.38652	0.374700	0.29979
2008	0.09360	0.30594	0.079800	0.32270
2009	0.19160	0.43772	0.481800	0.13786
2010	0.34380	0.58634	0.547700	0.13815
2011	0.21590	0.46465	0.397100	0.29631

Table 5.13: *S&P:Oil* pair estimated unconditional correlations

Part of sample	**Neural network in-sample fit R^2**	**Unconditional correlation estimated using FFNN**	**Average realized daily correlation**	**Standard deviation**
		Tansig transfer function		
Full sample	0.08490	0.29138	0.360600	0.24764
2008	0.12310	0.35086	0.419000	0.18262
2009	0.04720	0.21726	0.286700	0.28483
2010	0.12850	0.35847	0.376800	0.23882
2011	0.07220	0.26870	0.359000	0.25865
		Logsig transfer function		
Full sample	0.08620	0.29360	0.360600	0.24764
2008	0.12400	0.35214	0.419000	0.18262
2009	0.05130	0.22650	0.286700	0.28483
2010	0.13010	0.36069	0.376800	0.23882
2011	0.06790	0.26058	0.359000	0.25865

Table 5.14: *Gold:Oil* pair estimated unconditional correlations

From the tables reported above it is clearly seen that as in case of purely index pairs, estimated of unconditional correlation obtained using the FFNN are in the interval of average daily realized correlations, thus FFNN approximates these unconditional correlations well. Based on these results we can say that specified neural network performs well in approximating given process and it is a powerful quantitative tool that can be used for the correlations modelling.

5.3 Forecasting Realized Correlation

Second part of our empirical research is devoted to forecasting selected time-series of daily realized correlations. Based on the results obtained in the previous section we've selected two pairs, one from each group, with strongest daily realized correlations – *FTSE:DAX* and *S&P:Oil*. There are two aims we've set for the forecasting exercise:

1. analyze possibility of forecasting time-series of daily realized correlations, and
2. assess which of selected forecasting techniques, namely AR(p) model, HAR model and nonlinear autoregressive neural network (NARNET) perform better in forecasting task.

Importance of correlation measure in different areas of finance and risk management was discussed in the introduction to this thesis. Since realized correlation is actually observed, or ”true”, correlation it makes sense to analyze it and forecast it to apply further in financial decision making. Time-series of daily realized correlations for mentioned pairs were obtained by calculating daily realized correlations for the period between 2008 and 2011 using realized correlation measure discussed in detail in the Chapter 3. We judge forecasting performance of selected techniques using following statistics: Root Mean Squared Error (RMSE), QLIKE loss function, Mincer-Zarnowitz regression (MZ).

Descriptions of subject time-series, summary statistics and their plots were already presented in the previous section of this chapter. First observation that should be made based on summary statistics and charts of subject time series is that both time-series were not stable over the examined period. There is significant difference between mean and median values of *S&P:Oil* pair. Furthermore, pair *S&P:Oil* revealed negative daily realized correlations during 2008 and at the end of 2010 – beginning 2011. This may support our assumption established earlier, that investors may escape form stock markets to safer commodities during the crisis. Realized correlation between FTSE and DAX has never reached negative values.

Before we start with forecasting subject time-series, we also report results of JB and ADF tests to see if subject series are stationary and how far they are from being normally distributed.

Pair	**J-B test**	**p-value**	**ADF test**
FTSE:DAX	573.76	0.000	-10.74
S&P:Oil	210.10	0.000	-4.27

Table 5.15: JB and ADF tests

From the statistics in table 5.15 it is clear that both time-series are not normally distributed and presence of the unit root cannot be rejected, so we cannot assume stationarity. However, we did not make any transformation to the considered series since we are interested in forecasting realized correlations as they are. Let’s start with description of our approach to the forecasting problem. As it was mentioned earlier, we focus on forecasting selected series of daily realized correlations using autoregressive model of order *p* (AR(p)), heterogeneous autoregressive model (HAR) and nonlinear autoregressive neural network (NARNET). Specification of AR(p) model was provided in the Chapter 2. HAR model that we use in this section is presented by equation 3.8.

NARNET model was specified in the Chapter 4. We assume certain autoregressive process behind the development of realized correlations. We estimate that process using specified techniques.

We forecast one day-ahead value of *FTSE:DAX* and *S&P:Oil* realized correlations using rolling window forecast approach. Rolling window means that out of total T observations first n are selected. Than specified AR(p) and HAR models are estimated, or NARNET is trained, on these n observations. Next step is to obtain forecast of $(n+1)$ value. The size of the window, i.e. n, is kept during the whole exercise. We then take actual observations $t=2,...,(n+1)$and estimate or train our models and get an estimate of $(n+2)$ value. This process goes on till the forecast of last observation T is obtained. We use windows of different size to see how increase in training sample, i.e. information that is available to the models, improves forecasting performance.

Our selected models were setup as follows. Firstly, for AR(p) model to define number of lags p, we estimated number of significant lags for each year from 2008 to 2011. For different years number of significant lags was between 1 and 3. Since the training window covers all four years we use all three specifications of AR(p) model for the forecasting exercise to account for different significant lags. Setup of the HAR model has been provided in the Chapter 3 (equation 3.8). As for the NARNET model, selection of number of lagged values is rather subjective. We started with 2 lagged values and estimated models with number of autoregressive lags all the way up to 7. Till lag 5 RMSE of NARNET forecast was declining. However, starting with lag 6 RMSE started to increase for forecasts performed for subject time-series. That is why we use NARNET with 5 lags in our work.

To define the size of window we decided to use two sizes – one that represent approximately 90% of the full sample, and second that represents 95%. It means that for FTSE:DAX pair window sizes were 850 and 900 observations out of total 954. For S&P:Oil pair these were 725 and 770 out of 816.

For the purposes of comparison of forecasts obtained using different models we've selected RMSE and QLIKE loss functions. Mentioned loss functions are defined as follows:

$$RMSE = \sqrt{N^{-1}\sum_{i=1}^{N} e_t^2} \quad (5.1)$$

$$QLIKE = N^{-1}\sum_{i=1}^{N} (\ln fcast_t + corr_t / fcast_t) \quad (5.2)$$

where e_t is the difference between target series *"corr"* and forecasted values *"fcast"*. The better the forecast is, the lower should be the value of a loss function.

Table below summarizes statistics based on which we compare the forecasting performance of selected models. We start with *FTSE:DAX* pair and training window of 850 observations.

Model	**RMSE**	**QLIKE**
NARNET	0.05699	87.20052
AR1	0.05974	87.22154
AR2	0.05606	87.18966
AR3	0.05266	87.16155
HAR	0.05008	87.14105

Table 5.16: *FTSE:DAX* one day-ahead forecast performance, window size 850

It can be seen from the table that on sample of 850 observations NARNET did not outperformed AR(3) and HAR models in terms of yielding lower values of loss functions. Moreover, HAR model forecast of one day-ahead value of realized correlation is best among selected models. Among AR(p) models, AR(3) specification performed best in terms of statistics reported above.

Next table present results of Mincer-Zarnowitz regression (MZ-regression). MZ-regression is defined as follows:

$$y_{n+h} = \alpha + \beta y_{n+h(n)} + e_{n+h(n)} \tag{5.3}$$

where y_{n+h} is a target series being forecasted, $y_{n+h(n)}$ is an (n+h) step forecast based on n previous observations of the target series and $e_{n+h(n)}$ is an error term. MZ-regression basically tells us how good the forecast is by regressing actual values of the target on forecasts. If the forecast is statistically good, beta coefficient should be close to one and significant.

Model	**MZ-regression, beta coeeficient**	**Significance**
NARNET	0.18568	-
AR3	0.98688	***
HAR	0.99295	***

Table 5.17: *FTSE:DAX* one day-ahead forecast tests, window size 850

Based on the values of MZ-regression we can see that forecast obtained using NARNET is not significant. On the other hand, results for HAR model are relatively good with beta of MZ-regression close to 1 and significant. We report MZ-regression results for AR(3) model only because this specification performed best among selected AR(p) specifications. Chart below illustrates forecasts obtained using NARNET and HAR model.

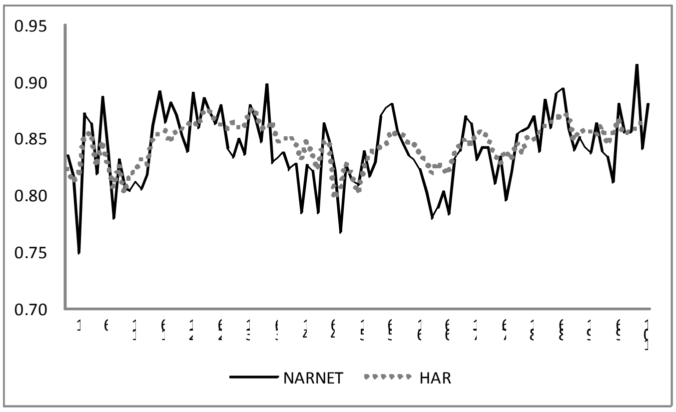


Fig. 5.6: *FTSE:DAX* daily realized correlations forecasts, window size 850

Following tables present statistics and tests for the same *FSTE:DAX* pair with larger training window of 900 observations.

Model	RMSE	QLIKE
NARNET	0.04295	45.08104
AR1	0.05164	45.11223
AR2	0.04689	45.09424
AR3	0.04431	45.08516
HAR	0.03957	45.06999

Table 5.18: *FTSE:DAX* one day-ahead forecast performance, window size 900

Model	MZ-regression, beta coeeficient	Significance
NARNET	0.61524	**
AR3	0.63571	*
HAR	0.71777	***

Table 5.19: *FTSE:DAX* one day-ahead forecast tests, window size 900

On the training sample comprising 900 observations NARNET outperformed AR(3) model in terms of loss functions values. Value of MZ-regression beta for NARNET trained on 900 observations is significantly better than one obtained on 850 data points. However, NARNET still was not capable of outperforming simple HAR model.

Following chart plots forecasts obtained using NARNET and HAR model to illustrate differences between two forecasts.

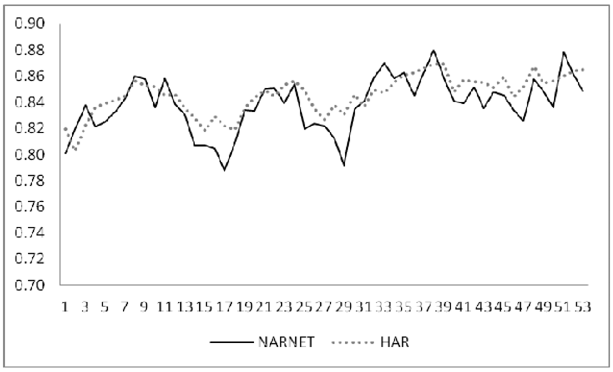


Fig. 5.7: *FTSE:DAX* daily realized correlations forecasts, window size 900

We now turn to the results of forecasting daily realized correlations of *S&P:Oil* pair calculated over the period between 2008 and 2011. As before, we used two sizes of training window, namely 725 and 770 observations out of total 816. Since the time-series of daily realized correlations for this pair is much more "complicated" for the models we've selected for our analysis in terms of larger jumps during the considered period, we expect worse overall forecasting performance of the models than in case of *FTSE:DAX* pair. We start with presenting results we've obtained using training window comprising 725 observations. We use the same specifications of AR(p), HAR and NARNET models and rolling window forecast approach as in case of *FTSE:DAX* pair. Tables below present values of coefficient of determination, RMSE and QLIKE statistics for forecasts obtained using selected models.

Model	**RMSE**	**QLIKE**
NARNET	0.13446	40.38853
AR1	0.14367	41.76373
AR2	0.13510	40.50729
AR3	0.12964	40.05818
HAR	0.12736	39.74229

Table 5.20: *S&P:Oil* one day-ahead forecast performance, window size 725

Based on the values of RMSE statistics we can conclude that among all models HAR performed best once again. AR(3) model performed better than NARNET and other specifications of AR(p). Next table summarizes results of Mincer-Zarnowitz regression for AR(3) model, HAR model and the NARNET.

Model	MZ-regression, beta coeeficient	Significance
NARNET	0.50040	***
AR3	0.54659	***
HAR	0.58413	***

Table 5.21: S&P:Oil one day-ahead forecast tests, window size 725

Results of MZ-regression suggest that forecast obtained using HAR model was again better than those of AR(3) and NARNET models. Following chart illustrates forecasts obtained using HAR and NARNET.

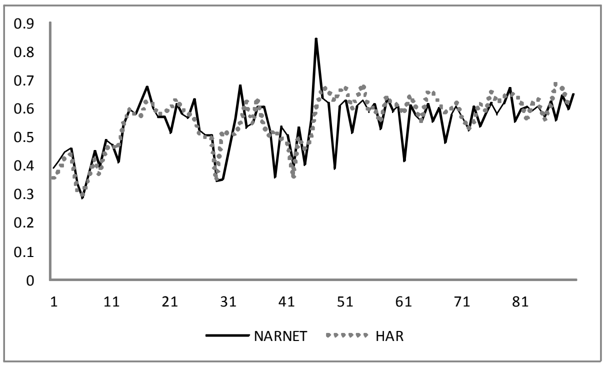


Fig. 5.8: *S&P:Oil* daily realized correlations forecasts, window size 725

Following tables present values of RMSE, QLIKE loss functions and MZ-regression results for selected models and training window of 770 observations for S&P:Oil pair.

Model	RMSE	QLIKE
NARNET	0.10882	24.30812
AR1	0.11963	24.34759
AR2	0.11417	24.17612
AR3	0.10760	24.07544
HAR	0.10336	24.01367

Table 5.22: *S&P:Oil* one day-ahead forecast performance, window size 770

Model	MZ-regression, beta coeeficient	Significance
NARNET	-0.38375	-
AR3	-0.22280	-
HAR	-0.29803	-

Table 5.22: S&P:Oil one day-ahead forecast tests, window size 770

Following chart plots forecasts of subject realized daily correlations obtained using NARNET and HAR models.

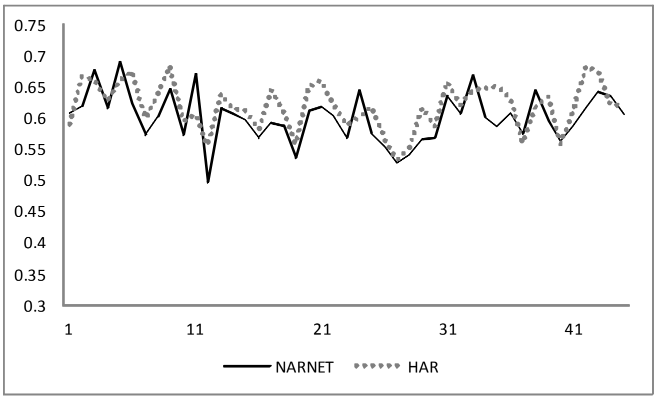


Fig. 5.9: *S&P:Oil* daily realized correlations forecasts, window size 770

Based on the results presented above and aims that were set for this part of our empirical research we can make several conclusions. Firstly, HAR model performed best among selected models in terms of RMSE and QLIKE loss functions and Mincer-Zanowitz regression results. Despite its simple composition, HAR is rather powerful model for forecasting realized correlations. Secondly, even though the specified nonlinear autoregressive neural network did not outperformed HAR model in forecasting realized correlations if loss functions and MZ-regression are considered, it is still a powerful computational structure with good potential in forecasting exercises.

6 Summary

In this thesis we have addressed several issues related to the modelling of correlations between stock markets and commodities, namely FTSE, DAX, PX, S&P, Gold futures and Oil futures, and analyzing their dynamics. We have based our analysis on the high-frequency data with a sampling frequency of 5 minutes. We started by computing daily realized correlations and analyzing their development over the period between 2008 and 2011. We have significant fluctuations in subject daily realized correlations. Among purely indices pairs, strongest correlation was registered for the FTSE:DAX pair of indices. Even though fluctuating a lot during 2008 and 2010-2011, these indices remained positively correlated. Other pairs experienced decreases to negative values over the given period. For index-commodity pairs we've seen several decreases to negative correlations. It may support the hypothesis that investors may escape from risky stock markets during the crisis and turn to the commodities instead. These empirical results may have implications, for example, for the portfolio management where correlation is one of the key concepts and diversification possibilities are important. We also assessed how well the feedforward neural network performs in approximating unconditional correlations based on average daily realized correlations. Our results suggest that FFNN is very power computational structure with a good potential in in-sample fitting.

In the second part of our empirical research we've forecasted time-series of FTSE:DAX and S&P:Oil daily realized correlations. We used HAR model, AR(p) model and nonlinear autoregressive neural network NARNET. Based on the result discussed in the Chapter 5, even though HAR model performed best in forecasting subject time-series, NARNET is also a good choice for realized correlations forecasting.

References

1. Ait-Sahalia, Y., Yu, J. (2009). *High Frequency Market Microstructure Noise Estimates and Liquidity Measures*. Institute of Mathematical Statistics.

2. Akaike, H. (1973). *Information theory and an extension of the maximum likelihood principle*. In Petrov, B. N., Csaki, F. (eds.). 2nd International Symposium on Information Theory, 267-281. Akademia Kiado, Budapest.

3. Ammer, J. And Mai, J. (1996). *Measuring International Economic Linkages with Stock Market Data.* Journal of Finance, Vol. 51, 1743-1763.

4. Andersen, T. G., Bollerslev, T. (1998). *Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts*. International Economic Review, 39(4), pp. 885-905.

5. Andersen, T.G., Bollerslev, T., Diebold, F.X., Labys, P. (1999). *(Understanding, Optimizing, Using and Forecasting) Realized Volatility and Correlation*. Manuscript, Northwestern University, Duke University and University of Pennsylvania.

6. Andersen, T. G., Bollerslev, T., Diebold, F.X., Labys, P. (2003). *Modeling and Forecasting Realized Volatility*. Econometrica, 71(2), 579-625.

7. Kulkarni, A.S., (1996). *Application of Neural Networks to Stock Markets Prediction*.

8. Baele, L., Ferrando, A., Hordahl, P., Krylova, E., Monnet, C. (2004). *Measuring Financial Integration in the Euro Area*. European Central Bank, Occasional Paper Series, No. 14.

9. Beale, M.H., Hagan, M., T., Demuth, H., B., (2012). *Neural Network Toolbox. User's Guide*. Matlab, Math Works.

10. Bessler, D.A., Yang, J., (2003). *Structure of interdependence in international stock markets*. Journal of International Money and Finance, Vol. 22, 261–287.

11. Bodart, V., Reding, P. (1999). *Exchange Rate Regime, Volatility and International Correlations on Bond and Stock markets*. Journal of International Money and Finance, Vol. 18, 133-151.

12. Brownstone, D. (1996). *Using percentage accuracy to measure neural network predictions in stock market movements*. Neurocomputing, vol. 10, no. 3, s. 237-250.

13. Churchland, P. S., Sejnowski, T. J. (1992). *The computational brain*. MIT Press, Cambridge (Massachusetts). ISBN 0-262-03188-4.

14. Corsi, F. (2007). *A Simple Approximate Long Memory Model of Realized Volatility*. Journal of Financial Econometrics (2009), 7 (2), pp. 174 196.

15. Dacorogna, M. M., Muller, U. A., Davé, D. D., Olsen, R. B., Pictet, O. V., von Weizsacker, J. E. (1997). *Volatilities of Different Time Resolutions. Analyzing the Dynamics of Market Components*. Journal of Empirical Finance 4 (1997), pp. 219 239.

16. Dayhoff, J., E., DeLeo, J.M. (2001). *Artificial Neural Networks: Opening the Black Box*. Cancer 91: 1615–1635.

17. Dickey, D.A., Fuller, W.A. (1979). *Distribution of the Estimators for Autoregressive Time Series With a Unit Root*. Journal of the American Statistical Association 74: 427–431.

18. Flavin, T.J., Hurley, M.J., Rousseau, F., (2001). *Explaining Stock Market Correlation: A Gravity Model Approach*. National University of Ireland, Maynooth.

19. Groenen, P.J.F., Franses, P.H. (2000). *Visualizing Time-varying Correlations across Stock markets*. Journal of Empirical Finance, Vol. 7, 155-172.

20. Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. Saddle River, NJ: Prentice-Hall.

21. Hochreiter, S., Schmidhuber, J. (1997). *Flat minima*. Neural Computation, vol. 9, s. 1-42.

22. Hornik, K., Stinchcombe, M., White, H. (1989). *Multilayer feedforward network are universal approximators*. Neural Networks, vol. 2, no. 5, s. 359-366.

23. Jarque, C.M., Bera, A.K. (1980). *Efficient Tests for Normality, Homoskedasticity, and Serial Independence of Regression Residuals*. Economics Letters 6: 255–259.

24. Jovina, R., Akhtar, J. (1996). *Backpropagation and Recurrent Neural Networks in Financial Analysis of Multiple Stock Market Returns*. Proceedings of the 29th Annual HawaiiInternational Conference on System Sciences.

25. Kenett, D.Y., Shapira, Y., Madi, A., Bransburg-Zabary, S., Gur-Gershgoren, G., Ben-Jacob, E. (2010). *Dynamics of Stock Market Correlations*. Czech Economic Review 4, 330-340.

26. Kim, S. H., Chun, S. H. (1998). *Graded forecasting using an array of bipolar predictions: application of probabilistic neural networks to a stock market index*. International Journal of Forecasting, vol. 14, no. 3, s. 323-337.

27. King, M.A. and Wadhwani, S. (1990). *Transmission of Volatility between Stock Markets*. Review of Financial Studies, Vol. 3, 5-33.

28. King, M.A., Sentana, E., Wadhwani, S. (1994). *Volatility and Links between National Stock Markets*. Econometrica, Vol. 62, 901-933.

29. Kohara, K. et al. (1997). *Stock price prediction using prior knowledge and neural networks*. International Journal of Intelligent Systems in Accounting, Finance, and Management, vol. 6, no.1, s. 11-22.

30. Kohzadi, N. et al. (1996). *A comparison of artificial neural network and time series models for forecasting commodity prices*. Neurocomputing, vol. 10, no. 2, s. 169-181.

31. Kuan, Chung-Ming, and White, H. (1994). *Artifical Neural Networks: An Econometric Perspective*. Econometric Reviews 13: 1–91.

32. Lee Giles, C., Lawrence, S., Tsoi, A. C. (2001). *Noisy Time Series Prediction using a Recurrent Neural Network and Grammatical Inference.* Machine Learning, Volume 44, Number 1/2, July/August, pp. 161–183

33. Lim, G. C., McNelis, P. D. (1998). *The effect of the Nikkei and the S&P on the all ordinaries: a comparison of three models*. International Journal of Finance & Economics, vol. 3, no. 3, s. 217-228.

34. Longin, F. and Solnik, B. (1995). *Is the Correlation in International Equity Returns Constant: 1960-1990?*. Journal of International Money and Finance, Vol. 14, 3-26.

35. McCluskey, P.C., (1993). *Feedforward and Recurrent Neural Networks and Genetic Programs for Stock Market and Time Series Forecasting*. Submitted in partial ful llment of the requirements for the Degree of Master of Science in the Department of Computer Science at Brown University.

36. McCulloch, W. S., Pitts, W. H. (1943). *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics, vol. 5, s. 115-133.

37. McNelis, P. D. (2005). *Neural Networks in Finance: Gaining Predictive Edge in the Market*. Elsevier Academic Press, California. ISBN 0-12-485967-4.

38. Müller, U. A., Dacorogna, M. M., Dave, R. D., Pictet, O. V., Olsen, R. B. and Ward, J. R. R. (1993). *Fractals and Intrinsic Time: A Challenge to Econometricians*. Opening lecture of the 39th International Conference of the Applied Econometrics Association (AEA), Real Time Econometrics — Sub monthly Time Series, 14 15 Oct 1993, in Luxembourg, and at the 4th International PASE Workshop, 22 26 Nov 1993, in Ascona (Switzerland).

39. Oh, K. J., Kim, K. (2002). *Analyzing stock market tick data using piecewise nonlinear model*. Expert Systems with Applications, vol. 22, no. 3, s. 249-255.

40. Pan, H., Tilakaratne, C., Yearwood, J. (2005). *Predicting Australian Stock Market Index Using Neural Networks Exploiting Dynamical Swings and Intermarket Influences*. Journal of Research and Practice in Information Technology, Vol. 37, No. 1, February 2005

41. Panchal, G. et al. (2010). *Searching most efficient neural network architecture using Akaike´s information criterion*. International Journal of Computer Applications, vol. 1, no. 5, 41-44.

42. Protter,P. (2004). *Stochastic integration and differential equations*. Springer-Verlag, Berlin.

43. Quah, T., Srinivasan, B. (1999). *Improving returns on stock investment through neural network selection*. Expert Systems with Applications, vol. 17, no. 4, s. 295-301.

44. Quah, T., (2007). *Using Neural Network for DJIA Stock Selection*. Advance online publication.

45. Ramchand, L. and Susmel, R. (1998). *Volatility and Cross Correlation across Major Stock markets*. Journal of Empirical Finance, Vol. 5, 397-416.

46. Rojas, R. (1996). *Neural networks: a systematic introduction*. Springer, Berlin. ISBN 978-3-54060-505-8.

47. Rumelhart, D. E., Hinton, G. E., Williams, R. J. (1986). *Learning internal representations by error propagation*. In Rumelhart, D. E., McClelland, J. L. (eds.). Parallel distributed processing: exploration in the microstructure of cognition. MIT Press, Cambridge (Massachusetts). ISBN 978-0-26248-120-4. s. 318-362.

48. Salim, L. (2011). *A Comparative Study of Backpropagation Algorithms in Financial Prediction*. International Journal of Computer Science, Engineering and Applications (IJCSEA) Vol.1, No.4.

49. Tsay, R.S. (2005). *Analysis of Financial Time Series*. Second edition. Wiley Interscience.

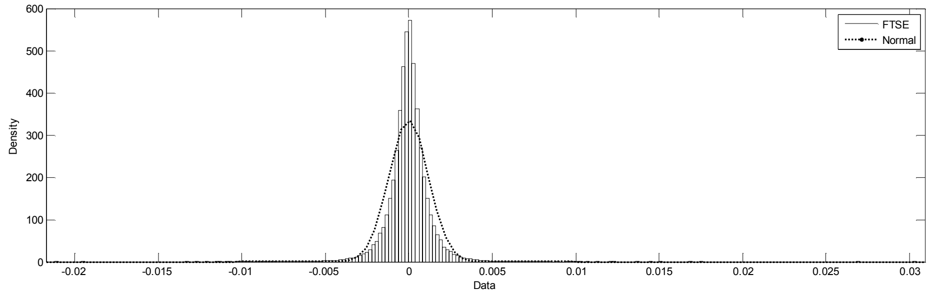
50. Yang, J., Bessler, D.A. (2006). *Contagion around the October 1987 stock market crash*. European Journal of Operational Research, Vol. 184, 291-310.

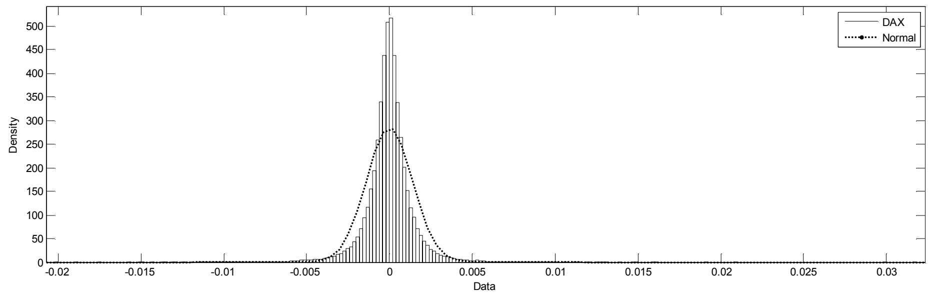
51. Yim, J. (2002). *A comparison of neural networks with time series models for forecasting returns on a stock market index*. In Hendtlass, T., Ali, M. (eds.). IEA/AIE 2002, LNAI 2358. Springer, London. ISBN 3-540-43781-9. s. 25-35.

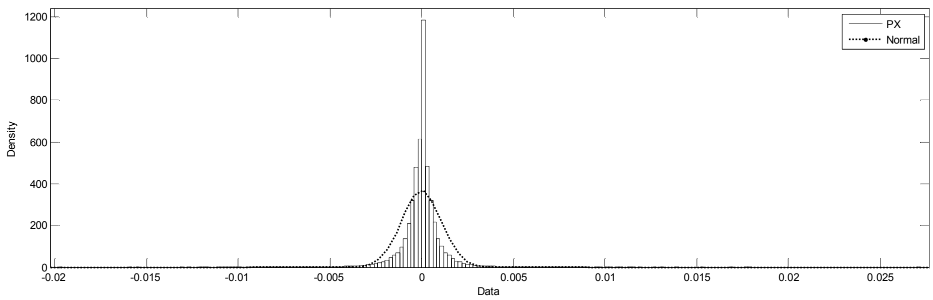
52. Zimmermann, H.G., Neuneier, R., Grothmann, R. (1996). *Active Portfolio-Management based on Error Correction Neural Networks*. Siemens AG

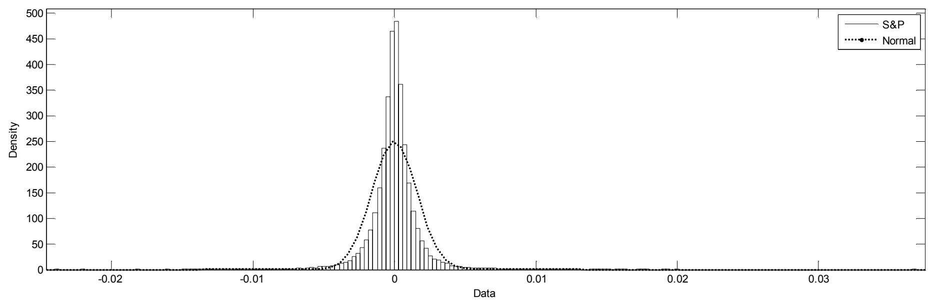
Appendix

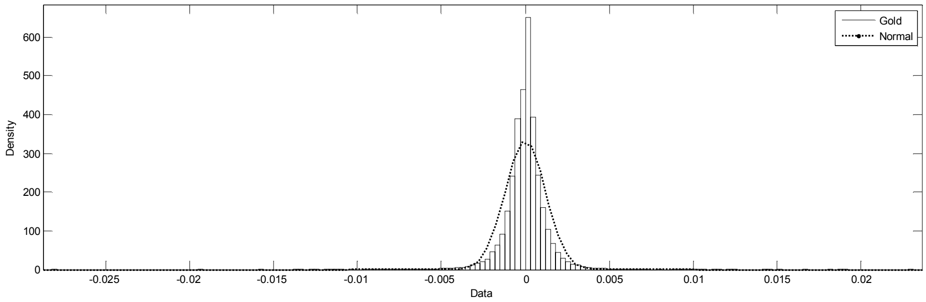
Appendix 1: Probability distributions of subject time-series analyzed in section, 5 minutes sampling frequency

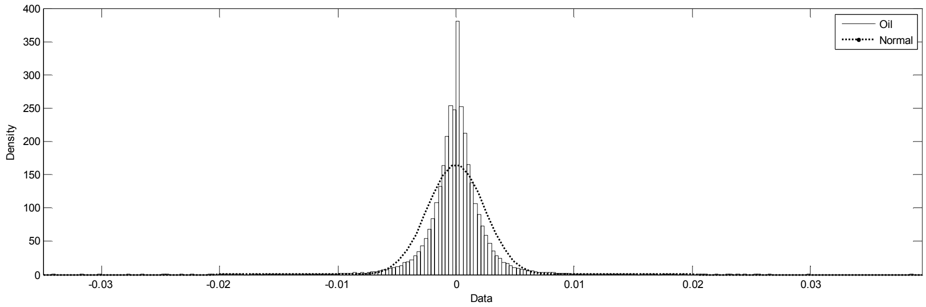









Gold
Normal
Density
Data
0
100
200
300
400
500
600
-0.025
-0.02
-0.015
-0.01
-0.005
0
0.005
0.01
0.015
0.02


Oil
Normal
Density
Data
0
50
100
150
200
250
300
350
400
-0.03
-0.02
-0.01
0
0.01
0.02
0.03