Charles University in Prague

Faculty of Social Sciences
Institute of Economic Studies

MASTER'S THESIS

Impact of Sports Results on Czech Stock Market

Author: **Bc. Michal Urban**

Supervisor: **PhDr. František Čech**

Academic Year: **2016/2017**

Declaration of Authorship

The author hereby declares that he compiled this thesis independently, using only the listed resources and literature, and the thesis has not been used to obtain a different or the same degree.

The author grants to Charles University permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, May 15, 2017

Signature

Acknowledgments

I would like to thank my supervisor PhDr. František Čech for supervising this thesis and my family for the continuous support during my studies.

Abstract

This thesis analyses the impact of sports results on Czech stock market. As sports results we used the matches played by Czech national football and ice hockey teams in main international competitions. As a proxy for Czech stock market we use PX index, the official index of Prague Stock Exchange. In our thesis, we applied ARMA-GARCH-t model to study the impact of results. We found only negative effect after losses of Czech national football team. The effect was statistically significant when we controlled for importance of the matches and was stronger for elimination games. There was no significant effect after wins and draws. We also did not find any significant effect after matches of Czech national ice hockey team.

Keywords Stock Market, Sport, Behavioral economics

Author's e-mail Mmichal.urban@gmail.com
Supervisor's e-mail fero.cech@gmail.com

Abstrakt

Tato práce analyzuje vliv výsledků sportu na český akciový trh. Jako sportovní výsledky jsme použili zápasy českých národních fotbalových a hokejových týmů v hlavních mezinárodních soutěží. Jako proxy pro český akciový trh jsme použili index PX, oficiální index Burzy cenných papírů Praha. V naší práci jsme použili ARMA-GARCH-t model pro analýzu vlivu sportovních výsledků. Našli jsme pouze negativní efekt po prohrách českého národního fotbalového týmu. Efekt byl statisticky významný, když jsme brali v úvahu důležitost zápasů a efekt byl silnější pro zápasy ve vyřazovacích kolech. Nenašli jsme žádný efekt po vítězstvích a remízách. Také jsme nenašli žádný statisticky významný efekt po zápasech českého národního hokejového týmu.

Klíčová slova Burza, Sport, Behaviorální ekonomie

E-mail autora Mmichal.urban@gmail.com
E-mail vedoucího práce fero.cech@gmail.com

Contents

List of Tables

List of Figures

Acronyms

ACF autocorrelation function

ADF Augmented Dickey-Fuller

AR autoregressive

ARCH autoregressive conditional heteroskedasticity

ARMA autoregressive moving average

GARCH general autoregressive conditional heteroskedasticity

KPSS Kwiatkowski–Phillips–Schmidt–Shin

MA moving average

LM Lagrange multiplier

OLS ordinary least squares

PACF partial autocorrelation function

PX index Prague Stock Exchange index

Master’s Thesis Proposal

Author	Bc. Michal Urban
Supervisor	PhDr. František Čech
Proposed topic	Impact of Sports Results on Czech Stock Market

Motivation:

One of the fundamental assumptions of economic theory is that decision-making of economic agents is rational. This assumption is questioned by behavioral economics, that suggests that economic agents’ emotions and mood can make them decide irrationally. The problem is that emotions and mood are difficult to measure. Since the last decade, many papers were written about modelling a link between mood and stock market returns.

Since football results of national teams have some suitable attributes, Edmans et al. (2007) used them as variable that is able to describe mood. They drive mood in a substantial and unambiguous way, they impact large proportion of the population and the effect is correlated across the majority within a country. Edmans et al. found significant market decline after losses and increase after wins of football national teams. The impact of other sports results was much weaker, there was only evidence of small market decline after losses.

The motivation of this thesis is to adopt this idea and apply it on Czech data. Czech nation is known as nation that is very enthusiastic about football and ice hockey matches and in both sports is also very successful, especially in ice hockey.

Hypotheses:

Hypothesis #1: Sports results have statistically significant impact on Prague Stock Exchange

Hypothesis #2: Negative effect after loss is larger than positive effect after win

Hypothesis #3: Elimination rounds have larger impact than group games

Methodology:

To test these hypotheses, we model Prague Stock Exchange index using GARCH model which was introduced by Engle (2001). As Prague Stock Exchange index will be used PX index (for data until March 2006 the PX 50) and as sports results we will use results of Czech football and ice hockey national teams results in main competitions. For football games we will use matches from European Championships, World championship and qualifications for these competitions. For ice hockey games we will use matches from World championships and Olympic Games.

Since financial series are usually non-stationary we will most likely use them in logarithmic differenced form. The stationarity of original data will be tested with Augmented Dickey Fuller test (ADF) and KPSS test. We use Schwarz, Hannan Quinn and Akaike info criteria to determine how many AR and MA we are going to use. To test whether there is some conditional heteroskedasticity we use ARCH-LM test. Then we will use suitable GARCH model as most of the stock market indices suffer from conditional heteroskedasticity. When we decide which model best fits our data, we add to this model the variables of our interest.

Expected Contribution:

This thesis will try to extend existing literature of behavioral economics about impact of investors' mood changes on stock markets. It will use sports matches as proxy for mood changes and it will be the first study written on this topic with using data just from Czech stock market.

The results from this thesis could be used as one of the ways to predict behaviour of stock markets and for further research on this topic.

Outline

1. Introduction
2. Literature Review:
3. Data, methodology and models
4. Empirical results
5. Discussion
6. Conclusion

Core bibliography

Berument, M, & Ceylan, N 2013, "Soccer and stock market risk: Empirical evidence from the Istanbul stock exchange", Psychological Reports, 112, 3, pp. 763-770.

Edmans, A, García, D, & Norli, Ø 2007, "Sports Sentiment and Stock Returns", Journal Of Finance, 62, 4, pp. 1967-1998

Engle, Robert. 2001. "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics."Journal of Economic Perspectives, 15(4): 157-168.

Hirshleifer, D 2001, "Investor Psychology and Asset Pricing", Journal Of Finance, 56, 4, pp. 1533-1597.

Huang, D, Jiang, F, Tu, J, & Zhou, G 2015, "Investor Sentiment Aligned: A Powerful Predictor of Stock Returns", Review Of Financial Studies, 28, 3, pp. 791-837.

Stracca, L 2004, Behavioral finance and asset prices: Where do we stand?, Journal Of Economic Psychology, 25, 3, pp. 373-405.

Author

Supervisor

Chapter 1

Introduction

The one of the fundamental assumptions of economic theory is that economic agents always behave rationally. As empirical evidence suggests, this assumption is not always realistic. One of the explanations of economical biases is brought by behavioral economics. It suggests that emotions and mood can make economic agents decide irrationally. Investors that are in good mood tend to evaluate future prospects more optimistically than people that are in bad mood.

The problem is that mood and emotions are difficult to measure. However, in the last two decades large number of papers documented links between mood and stock returns. Hirshleifer & Shumway (2003) examined the relationship between stock market returns and the sunshine that is narrowly connected with good mood. Kamstra *et al.* (2000) documented that the lack off sleep can affect negatively the stock market returns on the following days.

There are many papers that document links between sports and mood of the people. Hirt *et al.* (1992) investigated the effect of game outcome on sports fans' expectations of the team's and their own future performance. They found that fans' mood and self-esteem were affected by game outcome. Wann & McGeorge (1994) found that spectators exhibit negative affective reactions after their team's loss and positive reactions after their team's wins and that these effects are moderated by level of identification with that team. The effect was stronger after difficult wins than after easy wins.

Edmans *et al.* (2007) used sports results as a mood variable that can describe mood of the investors as they have some suitable attributes. They drive mood in a substantial and unambiguous way, they impact large proportion of the population and the effect is correlated across the majority within a country. Edmans *et al.* (2007) found strong evidence of loss effect after international football matches on the next working days in stock markets. The effect of wins is much weaker what is in line with economic theory as investors are affected more with bad news than the good news. They also found that the effects were stronger after elimination games. Edmans *et al.* (2007) also adapted this idea for other sports. They found statistically significant loss effect after cricket, basketball, and rugby international matches, for ice hockey they found no significant effect. The win effect was statistically insignificant for all these sports.

This idea is also adapted by this thesis. As sports results we use matches played by Czech national football and ice hockey teams in main international competitions and qualifiers for these competitions. These sports are very popular in the Czech Republic and the national teams are also relatively successful in these international competitions. The Czech nation is also very enthusiastic about these sports, so the effect of results on the mood of Czech investors is expected to be large. As a proxy for Czech stock market we use Prague Stock Exchange index (PX index), the official index of Prague Stock Exchange.

As most of financial market data suffer usually from autoregressive conditional heteroskedasticity (ARCH), we use to model these data ARMA-GARCH model. The ARCH model family was introduced by Engle (1982). This model family is able to deal with this heteroskedasticity by modelling not just mean equation but also the variance equation. The GARCH model is the generalization of the ARCH model introduced by Bollerslev (1986).

In this thesis we want to examine whether sports results have statistically significant impact on PX index. We expect that after losses there should be abnormal decline in the index and after wins there should be abnormal increase. The negative effect after losses is expected to be stronger than positive effect after wins. We also test whether the effect is stronger after elimination matches and more important matches. We also want to examine whether there is some significant effect after draws.

This thesis is structured in the following way. In chapter 2 is summarized literature which was written about this topic. In chapter 3 are described used methodology, data, and econometric models. In chapter 4 are presented the empirical results. In chapter 5 we discuss these empirical results and in chapter 6 we conclude this whole thesis.

Chapter 2

Literature Review

This chapter is divided into three parts. The first part sums up literature which was written about connection between mood of investors and stock returns. In the second part is summed up literature that describes how sports can affect mood of people and in the third part is summarized literature written about impact of sports results on stock markets.

2.1 Mood and stock market returns

A lot of papers have shown that investors do not always behave rationally as economic theory predicts. Sudden changes in mood are one of the explanations of irrational behavior. People that are in good mood tend to evaluate future prospects more optimistically than people that are in bad mood.

There are two basic approaches how to connect mood to stock market returns. The first one is event approach and the second one is using continuous variable that describes mood. The main advantage of event approach is that the changes in mood are sudden and hence give large signal-to-noise ratio in returns. The disadvantage of event approach is usually small number of events which reduce the statistical power of this approach. (Edmans *et al.* 2007)

It was shown in psychological experiments that weather has significant effect on human behavior. Saunders & Jr. (1993) studied the relation between local New York City weather and daily changes in indexes of listed stocks in New York City. They discovered that the cloudiness in New York City is significantly correlated with major stock indices in New York City. Hirshleifer & Shumway (2003) also examined the relationship between stock market returns and the weather. They found that sunshine is highly correlated with daily stock returns.

They did not find the correlation of daily stock returns with the other weather conditions such as rain or snow.

Kamstra *et al.* (2000) studied how the lack off sleep that is narrowly connected with bad mood can affect the stock market returns. They found that the weekend effect after daylight saving time change is statistically and economically significant in compare to normal weekend effect.

Cohen-Charash *et al.* (2013) examined whether press reports on collective mood of investors can influence stock prices. They searched for use of words that describes emotions in newspaper reports and analyzed how much they were correlated with next working day returns. They found that pleasant emotional words were followed by increase in NASDAQ prices and unpleasant mood words predicted decrease in NASDAQ prices.

2.2 Mood and sports

Hirt *et al.* (1992) investigated the effect of game outcome on sports fans' expectations of the team's and their own future performance. For this purpose they gave questionnaires to the fans of university basketball men's team after the team's games where the fans should rate the team performance in the game, next season performance, and their own prediction of their performance in various tasks. Hirt *et al.* (1992) found that fans' mood and self-esteem were affected by game outcome, the predictions of the team and their own performance were significantly more optimistic after the wins of the team that they were supporting.

Wann & McGeorge (1994) tested the hypotheses that spectators exhibit negative affective reactions after their team's loss and positive reactions after their team's wins and that these effects are moderated by level of identification with that team. Their hypotheses were confirmed in their study, fans with low identification with a team exhibited significantly less intense negative postgame effect on their emotions after the team's defeat than the fans with high identification with a team. It was also confirmed for positive reactions after the wins of team that they support, the effect was stronger after difficult wins than after easy wins.

Schweitzer & Zillmann (1992) examined the influence of sport results on estimates of the likelihood that a feared event will happen. Specifically, they tested how college football results influence a perception of the likelihood of a war in the Middle East. Their finding showed that postgame emotions are capable of influencing perception of the likelihood of the war. Their hypotheses that watching the victory of a supported team makes the war appear less likely and less threatening and watching a loss makes it appear more likely and more threatening were confirmed.

Rainey *et al.* (2011) tested how fan expectations of team performance, fan investments in the team, and identification with the team predict fans' end of season disappointment after unsuccessful season of team that they support. They found that fans that invested more in the team and were more identified with the team were much more disappointed after the unsuccessful season.

2.3 Sports and stock market returns

Edmans *et al.* (2007) introduced a new mood variable, international sports results, which can influence the effect of investor sentiment on stock returns. They argued that a mood variable must satisfy three key characteristics to affect investor sentiment.

- Drive mood in a substantial and unambiguous way
- Impact large proportion of the population
- The effect is correlated across the majority within a country

They believed that international football results satisfy all these three characteristics. As the international football results they used the results from the World Cups and main continental cups such as European Championship or Copa America including the qualifications of these football competitions. They collected data from competitions that were held from January 1973 through December 2004. Their sample consisted of 1162 matches, which were relevant "mood events", played by 39 different countries. To measure the impact of the results on stock prices, they used the return of the stock market index on the first trading day following the match.

To model these data, they used GARCH(1,1) model with panel corrected standard errors. They were controlling the model for the day of week effect using dummy variables of the days and they also included compounded daily U.S. dollar return to control for correlation with world market.

Edmans *et al.* (2007) found the strong evidence of negative loss effect after football matches that was also economically significant. The excess returns associated with a football loss exceeded 7%. The positive effect on the stock after wins of international football teams was much weaker. They also found out that the loss effect is stronger for more important games and for smaller stocks.

Edmans *et al.* (2007) also adapted this idea for other team sports, namely rugby, basketball, ice hockey, and cricket. They found statistically significant loss effect after cricket, basketball, and rugby international matches, for ice hockey they found no significant effect. The win effect was statistically insignificant for all these sports.

Berument *et al.* (2006) assessed the effect of football wins of Turkish major football teams (Besiktas, Fenerbahce, and Galatasaray) in Winner's cup on Turkish stock market returns. They used GARCH in mean specification to assess the effect of risk on return. They found out that only wins of Besiktas against foreign rivals in Winner's Cup had significant positive effect, the same effect was not present for the other teams.

Boido & Fasano (2007) analyzed the effect of the results of three Italian teams that are quoted on Italian stock market (Juventus, AS Roma, and Juventus). They found out that average price/return ratio following the wins is higher than after unsuccessful matches. They also analyzed impact of tied matches and it appeared that Italian investors dislikes tied matches.

Chang *et al.* (2012) examined the effect of american football matches on local stocks. They compared the stock returns of stocks that are geographically near to the NFL teams that played against each other. They found that the local team's loss led to lower next day returns for local stocks in compare to stock of winning team. The effect was stronger for a surprising loss or a critical game loss.

Berument & Ceylan (2012) further claimed that the mood changes affect not only stock market returns but also the return-volatility relationship. They assumed that investors become more risk averse after a loss and less risk averse after a win. They used international matches of football clubs from Turkey, Chile, Spain, and the United Kingdom. As a model they used EGARCH. The results supported their presumptions. The effect of losses of relatively more successful teams from Spain and the United Kingdom was that market return decreased and investors became more risk averse. For relatively less successful teams from Chile and Turkey, the effect of wins was that stock market returns increased and investor became less risk averse.

Tsounis *et al.* (2014) examined effects of football results in European international matches of four football clubs, namely FC Porto, Benfica Lisabon, Juventus FC and Ajax Amsterdam. The effects on stock markets are quite different amongst these clubs. For Benfica and Ajax, they found positive effects of ties on stock returns, for Juventus they found negative effects after ties and losses. They reported no effect for FC Porto club.

Kaplanski & Levy (2010) examined market returns during World Cups on the U.S. market, which at first glance should not be affected by football outcomes as football is not very popular in the USA. They believed that the U.S. market should also be affected by soccer championships as many foreign investors invest in U.S. market because the U.S. market is very liquid and with low transaction cost in compare to other markets and asymmetry between loss and win effect of foreign investors. According to their hypothesis there should be decline during championships. They found that average return on the U.S. market over the World Cup's period is -2.58% and average return for other periods with same length is +1.21% so there is quite significant difference.

Chapter 3

Methodology, data and models

This chapter is divided into three parts. In the first part is described theoretical background of the methodology used in this thesis. In the second part are described data and variables used in our econometric analysis, and the third part presents the models that are used for testing our hypotheses which are also stated in this section.

3.1 Methodology

The core framework of applied econometrics is the least squares model family. It is natural because most econometricians are interested in measuring how much one variable will change in response to a change in another variable. The basic least squares model assumes homoskedasticity what means that the expected value of squared error is the same for every observation.

Unfortunately, real data from financial time series usually do not satisfy this assumption as the variance of errors is usually not constant over time or we can say that errors suffer from heteroskedasticity. In financial time series we can usually see some periods with higher volatility, which can be described as "volatility clustering". In presence of heteroskedasticity, the coefficients of least squares regression are still unbiased but the estimated standard errors are smaller and hence confidence intervals are narrower what gives us false sense of precision.

One approach how to deal with heteroskedasticity is using robust standard errors. They can give quite good estimates of standard errors when sample size is large but work poorly when the sample size is smaller. The other approach is to treat heteroskedasticity as a variance to be modelled. It can be done using

ARCH and GARCH models (Engle 2001). These models are used in this thesis and described in the following subchapters.

3.1.1 ARCH model

The ARCH model was introduced by Engle (1982). Since then the model was often used by many applied econometricians and the basic model was also widely extended and created quite large ARCH family framework.

If we consider random variable y_t that is drawn from conditional density function $f(y_t|y_{t-1})$ then the forecast of today's value is simply $E(y_t|y_{t-1})$. The variance of this forecast is given by $Var(y_t|y_{t-1})$. This expression suggests that the conditional variance depends on the past information and may be a random variable. Under classical assumptions, the conditional variance does not depend on the y_t. Engle (1982) proposed a class of models where the variance does depend upon past observations.

Consider simple AR(1) model without constant:

$$y_t = \gamma y_{t-1} + \upsilon_t \tag{3.1}$$

Where υ is white noise with $Var(\upsilon) = \sigma^2$. Then the conditional and unconditional means and variances of this model are following:

$$E(y_t) = 0 \tag{3.2}$$

$$E(y_t|y_{t-1}) = \gamma y_{t-1} \tag{3.3}$$

$$Var(y_t) = \frac{\sigma^2}{1-\gamma^2} \tag{3.4}$$

$$Var(y_t|y_{t-1}) = \sigma^2 \tag{3.5}$$

From (3.2) we can see that if we do not have any known observation, our best forecast of y_t is 0. If we know the observation from time $t-1$, our best prediction of y_t is γy_{t-1}. We can also see that conditional variance is smaller than the unconditional variance if $|\gamma| < 1$, which is the stationarity condition for autoregressive models. We can also notice that in this model conditional variance does not depend on the last observation and is expected to be constant over time. As variance is not usually constant over time in real financial data, more general class of models seems desirable.

The standard approach how to deal with heteroskedasticity is to use an exogenous variable x_t which predicts the variance. The example of this approach is the following model:

$$\begin{aligned} y_t &= \upsilon_t x_{t-1} \\ Var(\upsilon) &= \sigma^2 \end{aligned} \tag{3.6}$$

Conditional variance of this model is $\sigma^2 x_{t-1}^2$. This approach is not very useful in financial time series as it requires the variable that causes changes in variance, rather than recognizing that conditional mean and variance jointly evolve over time.

Model that allows the conditional variance evolve over time is the following bilinear model:

$$\begin{aligned} y_t &= \upsilon_t y_{t-1} \\ Var(\upsilon) &= \sigma^2 \end{aligned} \tag{3.7}$$

Conditional variance of this model is $\sigma^2 y_{t-1}^2$. The weakness of this model is that the unconditional variance of this model is zero or infinity. Engle (1982) suggested slight generalizations to avoid this problem. The suggested modified model is:

$$\begin{aligned} y_t &= \upsilon_t \sqrt{h_t} \\ Var(\upsilon) &= 1 \end{aligned} \tag{3.8}$$

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 \\ \epsilon_t &= \upsilon_t \sqrt{h_t} \end{aligned} \tag{3.9}$$

This is an example of autoregressive conditional heteroskedasticity (ARCH) model. The model is composed from a mean equation (3.8) and a variance equation (3.9). If we add an assumption of normality, we can rewrite the mean equation using the information set available in time t denoted by ψ_t:

$$y_t|\psi_{t-1} \sim N(0, h_t) \tag{3.10}$$

The variance equation can be generalized in the following way:

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 = h(\epsilon_{t-1}, \epsilon_{t-2}, ..., \epsilon_{t-p}, \alpha) \tag{3.11}$$

The p in this equation is the order of the ARCH process and α is a vector of unknown parameters. In (3.9) we had the simplest ARCH(1) process.

The ARCH model is estimated using maximum likelihood method. Using assumption of normality, likelihood function of this model for time t is:

$$L_t = \frac{1}{(2\pi h_t)} exp\left(-\frac{y_t^2}{2h_t}\right) \tag{3.12}$$

From (3.12) we can easily get the log likelihood function:

$$l_t = \frac{1}{2}log(2\pi) + \frac{1}{2}log(h_t) + \left(-\frac{y_t^2}{2h_t}\right) \tag{3.13}$$

To estimate the unknown parameter α, we search for maximum of this log likelihood function. We differentiate the log likelihood function with respect to α and equate to zero:

$$\frac{\partial l_t}{\partial \alpha} = \frac{1}{2h_t}\frac{\partial l_t}{\partial \alpha}\left(\frac{y_t^2}{h_t} - 1\right) = 0 \tag{3.14}$$

3.1.2 Testing for ARCH disturbances

Before we make any estimation of ARCH model, we need to be sure that the disturbances in our data are conditionally heteroskedastic. If the disturbances are homoskedastic, OLS is the more appropriate procedure to model our data. Engle (1982) proposed ARCH-LM test procedure to test whether there is some ARCH effect present in the data. The null hypothesis of this test is that first p lags are equal to zero. The alternative hypothesis is autocorrelation in the squared residuals.

$$H_0 : \alpha_1 = \alpha_2, \ldots, \alpha_p = 0 \tag{3.15}$$

To test this hypothesis we run simple OLS regression and save the residuals (ϵ_t).

$$y_t = \beta_0 + \epsilon_t \tag{3.16}$$

Then we regress saved squared residuals ($\hat{\epsilon_t}$) on intercept and p lags.

$$\hat{\epsilon_t} = \alpha_0 + \sum_{i=1}^{p} \alpha_i \hat{\epsilon_t}^2 + \upsilon_t \tag{3.17}$$

Then we test TR^2 as a χ^2_p where T is number of observations and R^2 is R^2 of the equation (3.17).

3.1.3 GARCH model

The problem with basic ARCH models is that sometimes we need a large number of squared lagged residuals to specify the model correctly. Bollerslev (1986) introduced general autoregressive conditional heteroskedasticity (GARCH) model, what is an extension of ARCH model, that allows more flexible lag structure than ARCH models. The extension of the ARCH process to the GARCH is similar to the extension of simple time series AR process to the general ARMA process.

The GARCH(p,q) process is described by the following set of equations:

$$\epsilon_t | \psi_{t-1} \sim N(0, h_t) \tag{3.18}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon^2_{t-i} + \sum_{i=1}^{p} \beta_i h_{t-i} \tag{3.19}$$

where

$$p \geq 0, q > 0, \alpha_0 > 0$$

$$\alpha_i \geq 0, i = 1, ..., q$$

$$\beta_i \geq 0, i = 1, ..., p$$

ϵ_t is a real-valued discrete time stochastic process, ψ_t is the information set available in time t. For $p = 0$ we have the ARCH(q) process. For $p = 0$ and $q = 0$ ϵ_t is simple white noise. From (3.19) we can see that the GARCH process assumes that the best prediction for next period variance is weighted average of long-term variance (α_0), new information about the variance from last q periods (α_i) and the variance predicted for the last p periods (β_i).

Similarly as AR can be seen as infinite MA(∞), it can be proven that

GARCH(1,1) may be seen as an infinite dimensional ARCH process:

$$\begin{aligned} h_t &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1} = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1(\alpha_0 + \alpha_1\epsilon_{t-2}^2 + \beta_i h_{t-2}) = \\ &= \alpha_0(1+\beta_1) + \alpha_1(\epsilon_{t-1}^2 + \beta_1\epsilon_{t-2}^2) + \beta_i h_{t-2} = \ldots = \\ &= \alpha_0(1+\beta_1+\beta_1^2+\ldots) + \alpha_1(\epsilon_{t-1}^2 + \beta_1\epsilon_{t-2}^2 + \beta_1\epsilon_{t-3}^3 + \ldots) = \\ &= \frac{\alpha_0}{1-\beta_1} + \sum_{i=1}^{\infty}\beta_1^{i-1}\epsilon_{t-i}^2 = \gamma_0 + \sum_{i=1}^{\infty}\gamma_i\epsilon_{t-i}^2 \end{aligned} \tag{3.20}$$

GARCH(p,q) process is stationary if the sum of all coefficients α_i and β_i is strictly smaller than one.

$$\sum_{i=1}^{q}\alpha_i + \sum_{i=1}^{q}\beta_i < 1 \tag{3.21}$$

To determine number of dimensions of GARCH process, we can use ACF and PACF functions which are described in 3.1.6. In most financial applications the lower order models are used such as GARCH(1,1) or GARCH(2,1).

3.1.4 GARCH-t model

Errors of financial time series are not usually conditionally normally distributed what is one of assumptions of simple GARCH model. They are usually leptokurtic what means that the distribution of the errors is heavy-tailed. The probability of extreme values of returns is larger than normal distribution would suggest.

To deal with this issue, Bollerslev (1987) introduced an extension of GARCH model which assumes that errors are conditionally t-distributed because the student's-t distribution better explains leptokurtic distribution of time series than normal distribution. The GARCH-t model can be described by the following set of equations.

$$y_t = \epsilon_t \tag{3.22}$$

$$\epsilon_t = \upsilon_t\sqrt{h_t}, \epsilon_t \sim f_\nu$$

$$h_t = \alpha_0 \sum_{i=1}^{q}\alpha_i\epsilon_{t-i}^2 + \sum_{i=1}^{p}\beta_i h_{t-i} \tag{3.23}$$

Where ν denotes the number of degrees of freedom of t-distribution. The lower number of degrees of freedom t-distribution is, the heavier tails of the distribution are. The student's-t distribution converges to normal distribution for infinite number of degrees of freedom.

3.1.5 ARMA-GARCH model

The common way how to build GARCH models is by removing linear dependencies in the data by ARMA model. The ARMA(r,s)-GARCH(p,q) model can be described by following set of equations:

$$y_t = \phi_0 + \sum_{i=1}^{r} \phi_i y_{t-i} + \sum_{i=1}^{s} \theta_i \epsilon_{t-i} + \epsilon_t \tag{3.24}$$

$$\epsilon_t = \upsilon_t \sqrt{h_t}$$

$$h_t = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \tag{3.25}$$

3.1.6 Stationarity and unit root testing

When we model time series data, we need to have our data stationary. If our time series is not stationary, the persistence of shocks will be infinite for our series. We can also have problem with spurious regression what means that when we use two trending variables they will have high coefficient of determination even if they are unrelated. Also the standard assumptions for asymptotic analysis will not be valid if the variables in the regression are not stationary, so the usual t-ratios does not follow t-distribution and we cannot test validly our parameters with usual tests.

There are two basic types of stationary process of the series. The first is a strictly stationary process. The process is said to be strictly stationary if the following equation holds for every m:

$$P\{y_{t_1} \leq b_1, \ldots, y_{t_n} \leq b_n\} = P\{y_{t_1+m} \leq b_1, \ldots, y_{t_n+m} \leq b_n\} \tag{3.26}$$

The other type of stationary process is a covariance (weakly) stationary process. We say that the process is covariance (weakly) stationary if following set of equation holds.

$$E(y_t) = \mu, \quad t = 1, 2, \ldots, \infty \tag{3.27}$$

$$E(y_t - \mu)(y_t - \mu) = \sigma^2 < \infty \tag{3.28}$$

$$E(y_{t_1} - \mu)(y_{t_2} - \mu) = \gamma_{t_2 - t_1} \forall t_1, t_2 \tag{3.29}$$

So the time series is covariance stationary if it has constant mean, variance is constant and finite, and covariances depend on the difference between t_1 and t_2. The covariances (γ_s) are known as autocovariances. Autocovariances normalised by dividing by the variance are called autocorrelations (τ_s).

$$\tau_s = \frac{\gamma_s}{\gamma_0} \; s = 0, 1, 2, \ldots \tag{3.30}$$

The series of autocorrelations is called autocorrelation function (ACF). The other important function is partial autocorrelation function (PACF). It measures the correlation between an observation k periods ago and the current observation with controlling for observations in lags that lie between them. So the partial autocorrelation denoted by τ_{ss} measures correlation between y_t and y_{t-k} after removing effects of $y_{t-k+1}, \ldots, y_{t-1}$.

At first lag the PACF is equal to ACF ($\tau_{11} = \tau_1$), at second lag $\tau_{22} = \frac{\tau_2 - \tau_1^2}{1-\tau_1^2}$. For following lags is formula much more complex.

The model which is frequently used to describe non-stationary series is the random walk model with drift. It can be described by the following equation.

$$y_t = \mu + y_{t-1} + u_t \tag{3.31}$$

Where μ is the drift and u_t is the error term. The model may be further generalized to simple AR(1) model using coefficient ϕ.

$$y_t = \mu + \phi y_{t-1} + u_t \tag{3.32}$$

This process is stationary for $|\phi| < 1$, in this case the shocks gradually die away. If $\phi = 1$, the model is the same as the random walk model (3.31) and shocks persist and never die away, we say that the time series have unit root. For $\phi > 1$ the process is explosive and shocks become more influential over time. This case is not common in financial series, so it is not further considered. The case $\phi \leq -1$ would be also non-stationary but it is very unrealistic so also this case is not further considered.

To identify whether we have a non-stationary series, we can use various methods. One of the methods is simple look on the plots of the series. Most of the time it is obvious from the plots that our series has not constant mean or variance over time. The other informal method is to look on ACF function of the series. If there is high autocorrelation that does not die out for higher lags, we probably have non-stationary series.

One of the formal methods how to recognize that our series is non-stationary, is Dickey-Fuller test which was introduced by Dickey & Fuller (1979). The test uses random walk model and the null hypothesis of this test is that time series has unit root ($\phi = 1$), the alternative hypothesis is that the time series is stationary ($\phi < 1$). The usual way how to test this hypothesis is subtract y_{t-1} from the equation and test whether coefficient of y_{t-1} is equal to 0. The procedure of the test can be described by following set of equations.

$$y_t = \phi y_{t-1} + u_t \tag{3.33}$$

$$\Delta y_t = \psi y_{t-1} + u_t \tag{3.34}$$

$$H_0 : \phi = 1 \ or \ \psi = 0 \tag{3.35}$$

$$H_A : \phi < 1 \ or \ \psi < 0 \tag{3.36}$$

The test statistic is defined as:

$$test\ statistic = \frac{\hat{\psi}}{SE(\widehat{\psi})} \tag{3.37}$$

This test statistic does not follow t-distribution under the null hypothesis so we cannot use critical values for t-distribution. Dickey & Fuller (1979) derived critical values using Monte Carlo simulation. The null hypothesis of unit root is rejected when the test statistic is lower than the critical value.

The DF test is valid only when the u_t is white noise. If there will be some autocorrelation in dependent variable Δy_t, the u_t is not white noise and the DF is not valid. To deal with this problem, we can "augment" the test using p lags of the dependent variable:

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta y_t + u_t \tag{3.38}$$

The test statistic and critical values for ADF test are the same as for DF test. We just need to determine how many lags we should use. This can be done using information criteria.

The second formal method how to identify whether we have non-stationary time series, is to use the KPSS test which was introduced by Kwiatkowski *et al.*

(1992). They derive their test with using the following model.

$$y_t = \boldsymbol{\beta' D_t} + \mu_t + u_t$$
$$\mu = \mu_{t-1} + \epsilon_t, \epsilon_t \sim WN(0, \sigma_\epsilon^2) \quad (3.39)$$

Where $\boldsymbol{D_t}$ contains deterministic components, u_t is stationary process which may be heteroskedastic, and μ_t is a random walk with variance σ_ϵ^2. The null hypothesis of this test is formulated as $H_0 : \sigma_\epsilon^2 = 0$, which implies that μ_t is a constant. The alternative hypothesis is formulated as $H_A : \sigma_\epsilon^2 > 0$ (Zivot & Wang 2006). In other words, the null hypothesis of this test is that our time series is stationary, the alternative hypothesis is that time series contains one unit root.

3.1.7 Hypotheses testing

For testing our hypotheses we use Student's t-test. First we compute t-statistics ratio which can be described by following equation:

$$t_{\hat{\beta}_i} = \frac{\hat{\beta}_i - a}{SE_{\hat{\beta}_i}} \quad (3.40)$$

Where $\hat{\beta}_i$ is computed coefficient, a is the number that we test our coefficient against, in our hypotheses we always test against $a = 0$. $SE_{\hat{\beta}_i}$ is the standard error of the coefficient.

If we want to test whether our variables are significantly different from zero, we need to reject hypothesis that our coefficients are zero. To test this hypothesis, we compare t-statistics ratio of the coefficient with two-sided critical value of t-distribution (denoted by c) on five percent significance level. For higher degrees of freedom the critical value of t-distribution is almost equal to normal distribution critical value so we can use the normal distribution critical value. If the absolute value of t-statistics is lower than the critical value, which is for normal distribution 1.96, we cannot reject the hypothesis that our coefficient is zero what means that our coefficient is not statistically significant.

$$|t_{\hat{\beta}_i}| < c = 1.96 \quad (3.41)$$

Then we also find appropriate p-values to determine how strong our evidence is.

If we want to test whether our coefficient is negative or positive, we need to use one-sided critical values instead of two-sided. The critical value for five percent confidence interval for normal distribution is 1.645. In other words, we want to test whether our coefficient is significantly smaller or larger than zero. For testing whether our coefficient is negative we use (3.42), for testing whether our coefficient is positive we use (3.43). If these inequalities hold, then we cannot reject the hypothesis that our coefficients are zero, so our hypotheses will not be supported.

$$t_{\hat{\beta}_i} > -c = -1.645 \tag{3.42}$$

$$t_{\hat{\beta}_i} < c = 1.645 \tag{3.43}$$

3.2 Data

In this part, we present data that are used in this thesis.

3.2.1 PX index

As a proxy for Czech stock market index is used Prague Stock Exchange index (PX index). We collected daily data of PX index in the period from 1.8.1996 to 24.6.2016. The data were downloaded from the official web site of Prague Stock Exchange.

PX index is the official price index of Prague Stock Exchange. It was first calculated on 20.3.2006 when it replaced PX 50 and PX-D indices. PX index took over data from PX 50 index so for period before creation of PX index we use the data of PX 50 index.

PX index is calculated using the following formula[1].

$$PX(t) = 1000 \frac{\sum_{i=1}^{N(t)} q_i p_i(t) FF_i RF_i}{Start\ cap.} AF(t) \tag{3.44}$$

Startcap. is the market capitalization of the index on launch date, $AF(t)$ is the adjustment factor at time t, q_i denotes the number of securities of the i-th index, $p_i(t)$ denotes the price quotation of the i-th index issue at time t, FF_i denotes the free float factor, RF_i denotes the representation factor, $N(t)$ the number of index issues at time t.

[1]Source: https://www.pse.cz/en/indices/description-of-indices/px-index/

PX index data description is in the table 3.1. In the first line is PX index data in level form, in the second line are logarithmic returns of PX index. They are defined by the following equation.

$$R_t = \log\left(\frac{PX_t}{PX_{t-1}}\right) = \log(PX_t) - \log(PX_{t-1}) \tag{3.45}$$

The logarithmic returns are included because PX index in level form is not stationary and we use them in further analysis.

Table 3.1: PX index data description

Variable	Mean	Standard deviation	Min	Max	Number of observations
PX_t	888.3	393.8	316	1936.1	5140
$ldPX$	0.00014	0.014	-0.162	0.124	5139

Source: author's computations.

3.2.2 Sports results

We collected results of Czech national teams from main football and ice hockey competitions from 1996 through 2016. The data include football matches from football European Championships, European Championships qualifications, World Cup, and World Cup qualifications. The ice hockey data include matches from World Championships and Olympic Games.

We divided the football matches into three groups according to phase which they were played in. In the first group are matches played in qualificitions, in the second group are matches from group stages of the main tournaments, and in the last group are matches from elimination phases. The Czech national football team played 133 competitive matches in this period. In the following table are summarized results of Czech national football team in each groups. There are also added means and standard deviations of logarithmic returns of PX index from the following working day.

Table 3.2: Football results

	Wins			Ties			Losses		
Games	N	Mean	SD	N	Mean	SD	N	Mean	SD
All	81	0.0002	0.016	17	-0.009	0.02	35	0.003	0.022
Qualifying	65	0.001	0.018	15	-0.011	0.021	20	0.006	0.028
Group stage	9	-0.004	0.013	2	0.004	0.007	10	0.0007	0.012
Eliminiation	7	-0.004	0.007	0	0	0	5	-0.005	0.004

Source: author's computations.

We divided the hockey games only into two groups because there are no qualifications to major tournaments. The first group of games is consisted from games played in group stage and the second group is consisted of games played in elimination games.

The problem with ice hockey games is that often are played two games before one working day. If it happens, we use only the second match as it should influence the mood much more than the previous one on the following working day.

The Czech national hockey team played 207 matches during this period but because of the previously mentioned problem we can use only 138 matches as our mood events. The used results with mean and standard deviation of followed logarithmic returns of PX index are summarized in the following table.

Table 3.3: Ice hockey results

	Wins			Ties			Losses		
Games	N	Mean	SD	N	Mean	SD	N	Mean	SD
All	87	0.001	0.016	8	0.0007	0.01	43	-0.0005	0.016
Group stage	64	0.001	0.018	6	0.004	0.01	25	0.0019	0.014
Eliminiation	23	0.0008	0.009	2	-0.009	0.007	18	-0.004	0.018

Source: author's computations.

3.2.3 Proxies for importance of matches

As the literature suggested that more important matches have larger influence on mood of spectators, we need to find some proxy that can describe the importance of matches.

For football matches we decided to use as proxy for importance of matches the attendance of the match as we expect that the more important matches are much more attractive for watching than the less attractive matches.

Also another possible proxy would be television viewership. We decided not to use this proxy in our analysis because of the fact that not all matches were broadcasted in television, also the data of television viewership are not consistent through the time, and they are also not accurate. Nevertheless the data of television viewership still bring some information so we used them to evaluate suitability of our used proxy. We downloaded data of television viewership from official site[2] of Czech Television what is the public television broadcaster in the Czech Republic which brodcasts most of the matches of Czech football national team. To compare with attendances of matches we used only data since year 2010 because the older data were calculated differently. The correlation coefficient of these two variables is 0.5313 what shows quite large correlation between them so the attendance of the match seems to be quiet good proxy for importance of the match for football.

Table 3.4: Football attendance

Variable	n	Min	Max	Average	Standard deviation
Attendance	133	900	73611	21964.01	15767.42
Viewership	35	291000	1244000	603314.29	213965.47

Source: author's computations.

In our models we use attendance divided by 10000 because otherwise the coefficient of this variable would be very small if we used the variables in level form. The t-ratios should not be affected by this transformation.

For ice hockey matches we decided to use as proxy for importance of matches the strength of an opponent as we presume that matches against top countries are much more important for fans than matches against teams that are not as good in ice hockey. We consider the attendance as not appropriate for ice hockey matches because it is more influenced more by popularity of ice hockey in country where the championship or Olympic games are held than the importance of the matches. With television viewership for ice hockey matches we have the same problem as we have with television viewership for football matches.

To measure strength of the opponents, we divided teams in three groups. The first group consists of teams that are with Czech team considered as long-term top 7 teams in ice hockey and only with one exception[3] won every medals

[2]Source: www.ceskatelevize.cz/vse-o-ct/sledovanost-a-spokojenost/topy-sledovanost/
[3]Switzerland won silver medal in 2013

in last fifty years in ice hockey championships and Olympic Games. The other group is consisted of teams that are regular participants of world championships and occasionally beat the top teams. The rest of the teams are in the third group. The division of teams with assigned index is summarized in the following table:

Table 3.5: Ice hockey teams division

First group	Second group	Third group
index 3	index 2	index 1
Sweden	Norway	Italy
Finland	Germany	Kazakhstan
Canada	Belarus	France
Russia	Switzerland	Japan
United States	Latvia	Slovenia
Slovakia	Austria	
	Denmark	

Source: author's computations.

3.3 Models and hypotheses

3.3.1 Model A

The first model is the simple ARMA-GARCH model with sports results as exogeneous variables. Following Edmans *et al.* (2007), we include dummies for days of the weak to capture weekday effects.

$$R_t = \phi_0 + \sum_{i=1}^{r} \phi_i R_{t-i} + \sum_{i=1}^{m} \theta_i \epsilon_{t-i} + \gamma_1 win_{st} + \gamma_2 loss_{st} + \gamma_3 draw_{st} + \sum_{i=1}^{4} \rho_i D_i + \epsilon_t \tag{3.46}$$

$$\begin{aligned} \epsilon_t &= \upsilon_t \sqrt{h_t}, \epsilon_t \sim f_\nu \\ h_t &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned} \tag{3.47}$$

In the mean equation (3.46) R_t is the return of PX index in time t, $\phi, \theta, \gamma, \rho$ are the coefficients, ϵ_t is the error term in time t, q is the order of AR process, m is the order of MA process. Win, loss and draw are dummy variables for sports results in time t, s is the index of the sport. D_i are dummy variables for working days, we exclude friday to avoid the dummy trap. In the variance equation (3.47) α, β are the coefficients of GARCH variables, q and p are the orders of GARCH process.

With this model we test whether our variables are significant without taking into account the importance of the matches. For testing the significance of our variables, we test the following hypotheses:

Table 3.6: Model A hypotheses

Variable tested	H_0	H_A
Win	$\gamma_1 = 0$	$\gamma_1 > 0$
Loss	$\gamma_2 = 0$	$\gamma_2 < 0$
Draw	$\gamma_3 = 0$	$\gamma_3 \neq 0$

The null hypotheses for each test is that coefficient of the variable is equal to 0. As we want to test whether there is some positive effect after wins, the alternative hypothesis for wins is that coefficient of wins is greater than 0. Similarly for losses the alternative hypothesis is that coefficient of losses is less than zero. For draws we are interested whether they are significant hence the alternative hypothesis is that coefficient of draws is not equal to zero.

3.3.2 Model B

The second model is the simple extension of the first model. We use the same variables as in the first model, the difference is that we also include dummy variables for elimination games. We include only wins and losses because the elimination games have always winner.

$$R_t = \phi_0 + \sum_{i=1}^{r} \phi_i R_{t-i} + \sum_{i=1}^{m} \theta_i \epsilon_{t-i} + \gamma_1 win_{st} + \gamma_2 loss_{st} + \\ + \gamma_3 draw_{st} + \gamma_4 winel_{st} + \gamma_5 lossel_{st} + \sum_{i=1}^{4} \rho_i D_i + \epsilon_t \quad (3.48)$$

$$\epsilon_t = v_t \sqrt{h_t}, \epsilon_t \sim f_\nu \\ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \quad (3.49)$$

With this model we test whether the loss and win effects are larger for elimination games. For testing the significance of our variables, we test the following hypotheses:

Table 3.7: Model B hypotheses

Variable tested	H_0	H_A
Win	$\gamma_1 = 0$	$\gamma_1 > 0$
Loss	$\gamma_2 = 0$	$\gamma_2 < 0$
Draw	$\gamma_3 = 0$	$\gamma_3 \neq 0$
Winel	$\gamma_4 = 0$	$\gamma_4 > 0$
Lossel	$\gamma_5 = 0$	$\gamma_5 < 0$

3.3.3 Model C

The third model is very similar to the first model, the only difference is that we use instead of dummy variables for results our proxies for importance of the matches.

$$R_t = \phi_0 + \sum_{i=1}^{r} \phi_i R_{t-i} + \sum_{i=1}^{m} \theta_i \epsilon_{t-i} + \gamma_1 win_{st} imp_{st} + \\ + \gamma_2 loss_{st} imp_{st} + \gamma_3 draw_{st} imp_{st} + \sum_{i=1}^{4} \rho_i D_i + \epsilon_t \tag{3.50}$$

$$\epsilon_t = \upsilon_t \sqrt{h_t}, \epsilon_t \sim f_\nu \\ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{3.51}$$

With this model we want to examine whether taking into account importance of the matches improves our results. The hypotheses are the same as for the first model.

3.3.4 Model D

The last model is combination of the second and the third model. We add to the third model the elimination games dummies multiplied by relevant proxy variable for importance of matches.

$$R_t = \phi_0 + \sum_{i=1}^{r} \phi_i R_{t-i} + \sum_{i=1}^{m} \theta_i \epsilon_{t-i} + \gamma_1 win_{st} imp_{st} + \gamma_2 loss_{st} imp_{st} + \\ + \gamma_3 draw_{st} imp_{st} + \gamma_4 winel_{st} imp_{st} + \gamma_5 lossel_{st} imp_{st} + \sum_{i=1}^{4} \rho_i D_i + \epsilon_t \tag{3.52}$$

$$\epsilon_t = \upsilon_t \sqrt{h_t}, \epsilon_t \sim f_\nu$$
$$h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \tag{3.53}$$

The hypotheses are the same as for the second model.

Chapter 4

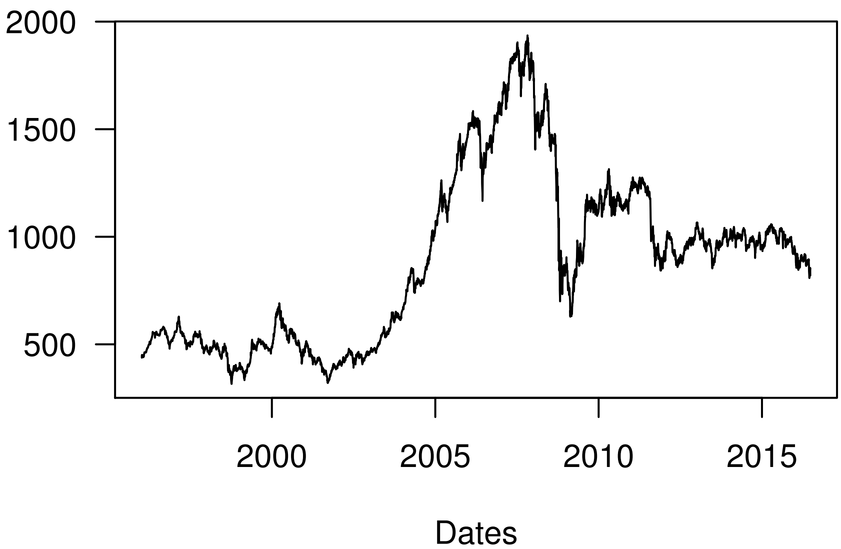
Empirical part

In this chapter, we present empirical results of our models. In the first part of this chapter is initial analysis where we check whether our time series is stationary, determine order of ARMA and GARCH processes. In the second part, we estimate our model using R software and test our hypotheses.

4.1 Initial analysis

4.1.1 Stationarity

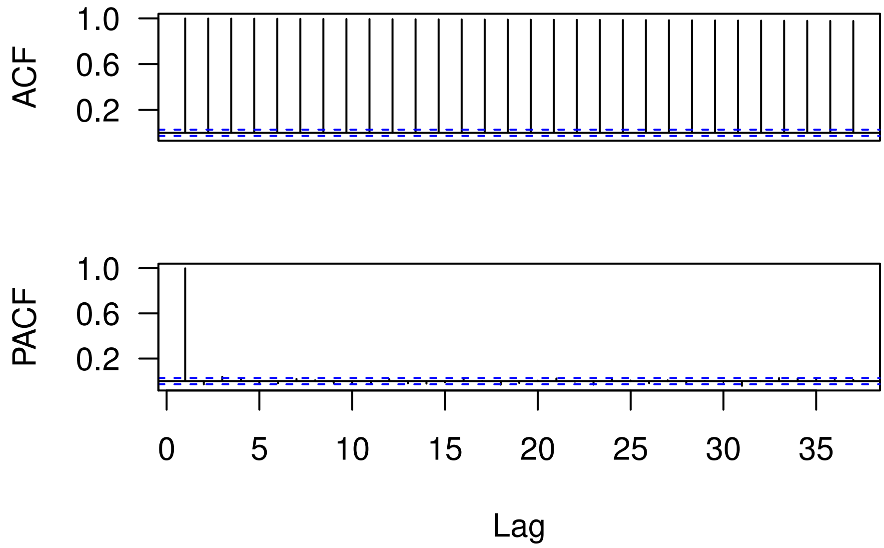
Figure 4.1: Plot of PX index



Source: author's computations.

Before we estimate our models, we have to make sure that our time series is stationary. It can be easily seen from Figure 4.1 that our series is not stationary in level form. This is also confirmed by ADF test that does not reject the hypothesis of unit root with p-value 0.86 and by KPSS test that strongly rejects the hypothesis of stationarity with p-value smaller than 0.01. From the plots of ACF and PACF (Figure 4.2), we can also see that there the series is strongly autocorrelated so it is definitely not stationary.

Figure 4.2: ACF and PACF of PX index



Source: author's computations.

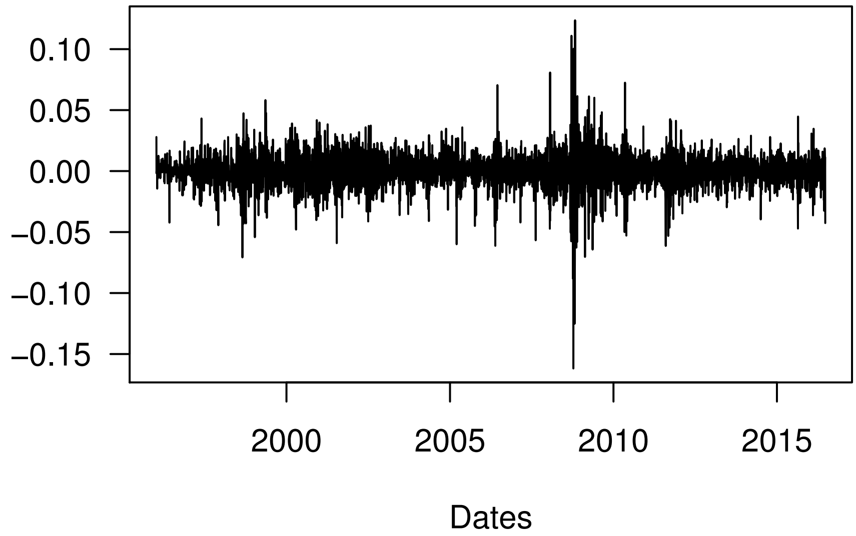
To deal with the problem with non-stationarity, we use logarithmic returns of the PX index. They are defined as:

$$R_t = \log\left(\frac{PX_t}{PX_{t-1}}\right) = \log(PX_t) - \log(PX_{t-1}) \tag{4.1}$$

Where PX_t is value of PX index in time t.

The plot of logarithmic returns is on the Figure 4.3. The series seems stationary at the first sight, there are just some volatility clusters present but these can be explained as an ARCH effect. With using ADF test, we are able to strongly reject the hypothesis of unit root even at 1% confidence interval. KPSS test also does not reject null hypothesis of stationarity even at 10% confidence level. On the plots of ACF and PACF, which are presented in Figure 4.5, we can see that autocovariances die away in larger lags.

Figure 4.3: Plot of logarithmic returns of PX index



Source: author's computations.

4.1.2 Distribution of returns

From the histogram of logarithmic returns, which is presented in Figure 4.4 (the black line shows how the histogram would look like if we had normally distributed errors), we can see that distribution of the returns is leptokurtic so they are not normally distributed. There are much more extreme values and values near to the mean than normal distribution suggests. To deal with this problem, we will use student's-t distributed errors instead of normal distributed errors.

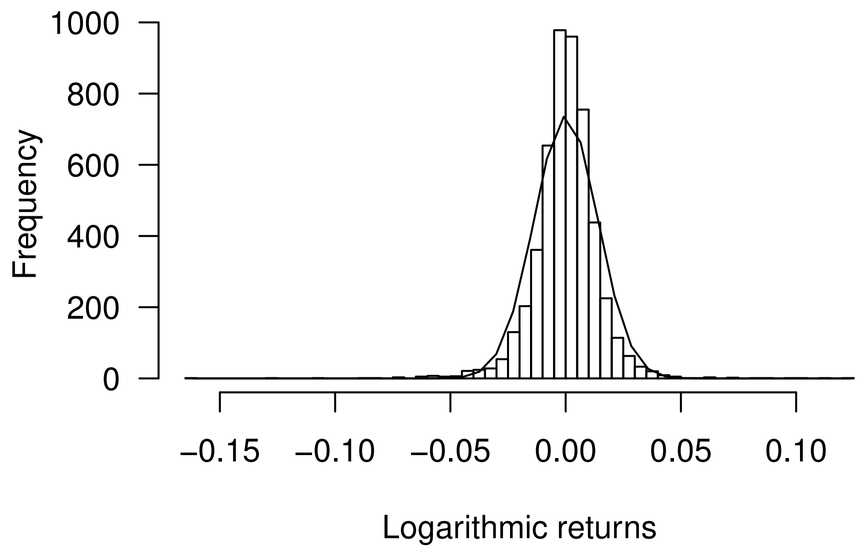
4.1.3 ARMA analysis

To determine number of orders of ARMA process, we use ACF and PACF functions shown in Figure 4.5. We can see that the first lags of these functions are highly significant. The second lags are still significant but very weakly. It seems that ARMA(0,1) best explains our data. To check whether our choice is right, we run simple ARMA model and save the residuals.

$$R_t = \alpha + \theta_1 \epsilon_{t-1} + \epsilon_t \tag{4.2}$$

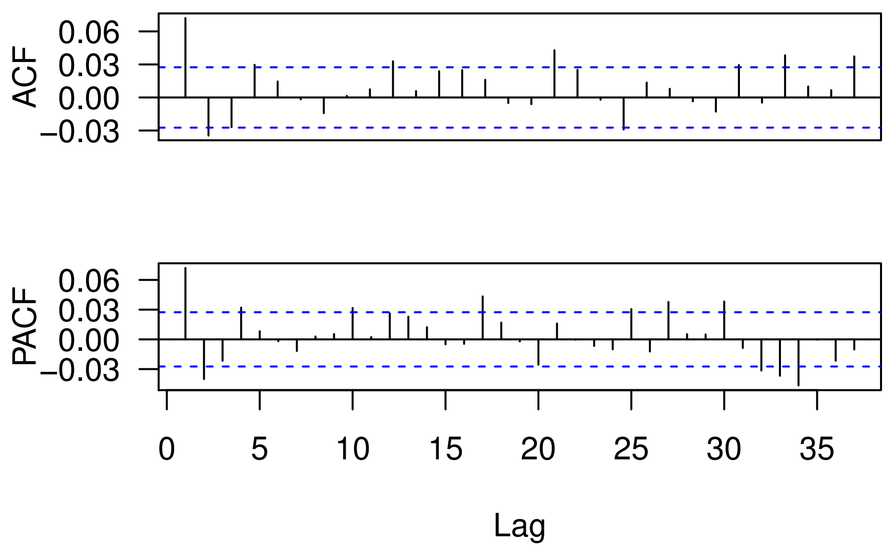
The estimated coefficients are summarized in the following table.

Figure 4.4: Histogram of returns of PX index



Source: author's computations.

Figure 4.5: ACF and PACF of returns of PX index



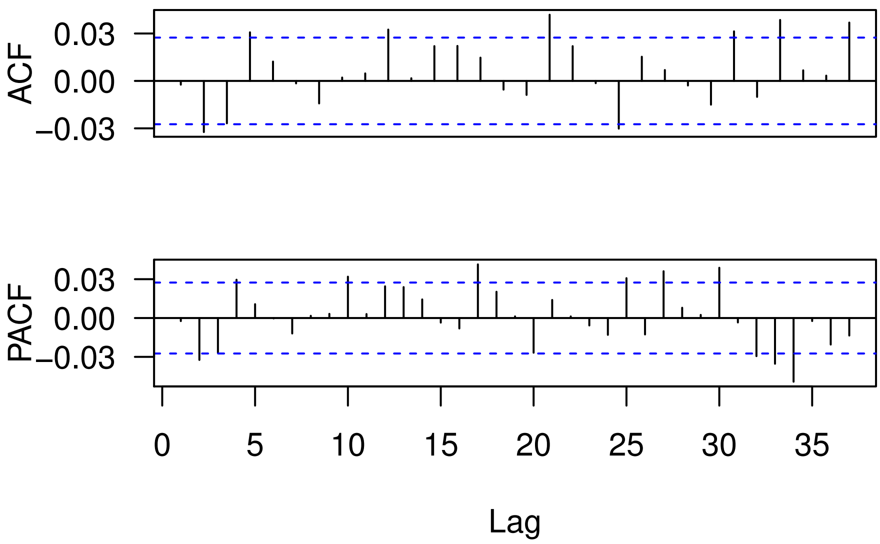
Source: author's computations.

Table 4.1: ARMA model

Variable	coefficient	standard error	t-ratio	p-value
$constant(\alpha)$	0.0002	0.0001	0.471	0.637
$MA_1(\theta_1)$	0.0775	0.0144	5.38	0.00

The coefficient of first lag of MA process is strongly significant, the mean is not significant. The ACF and PACF functions (see Figure 4.6) of residuals have just very weakly significant third lag. By trying other models, we did not find any model that would fit better than this model so we decided to use ARMA(0,1) for our further analysis.

Figure 4.6: ACF and PACF of residuals of ARMA(0,1) model



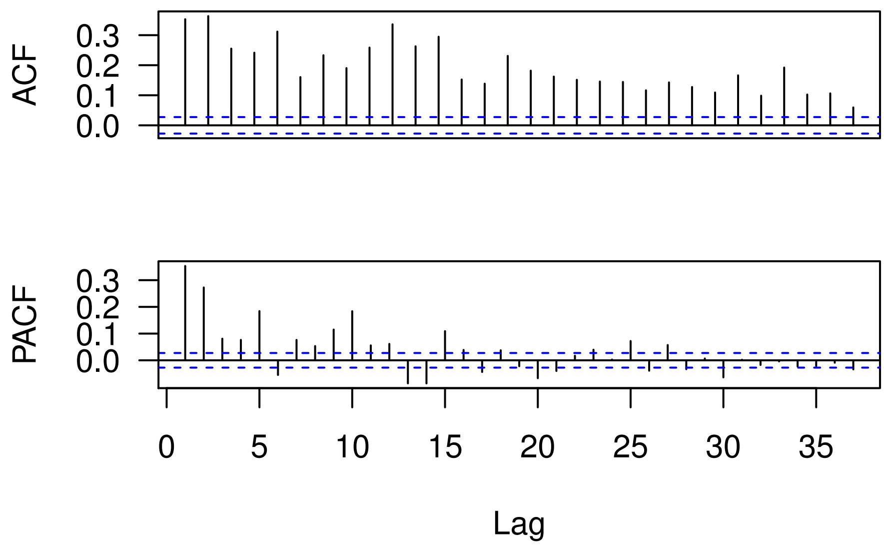
Source: author's computations.

4.1.4 GARCH analysis

In subsection 4.1.1 we saw on the plot of logarithmic returns couple of volatility clusters what suggests that there is probably some ARCH effect present. To check whether this effect is truly present, we use ARCH-LM test and check ACF and PACF functions of squared residuals of our ARMA model. ARCH-LM test strongly rejects the hypothesis of no ARCH effect and we also see very high autocorrelation in ACF and PACF functions (Figure 4.7).

Because ARCH effect is present, we should use some model from ARCH family. After fitting various models, we decided to use GARCH(1,1). More

Figure 4.7: ACF and PACF of squared residuals of ARMA(0,1) model



Source: author's computations.

specifically, we use ARMA-GARCH-t model. Our models can be described by the following equations:

$$R_t = \phi_0 + \theta_1 \epsilon_{t-1} + \gamma variables + \sum_{i=1}^{4} \rho_i D_i + \epsilon_t \tag{4.3}$$

$$\begin{aligned} \epsilon_t &= v_t \sqrt{h_t}, \epsilon_t \sim f_\nu \\ h_t &= \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned} \tag{4.4}$$

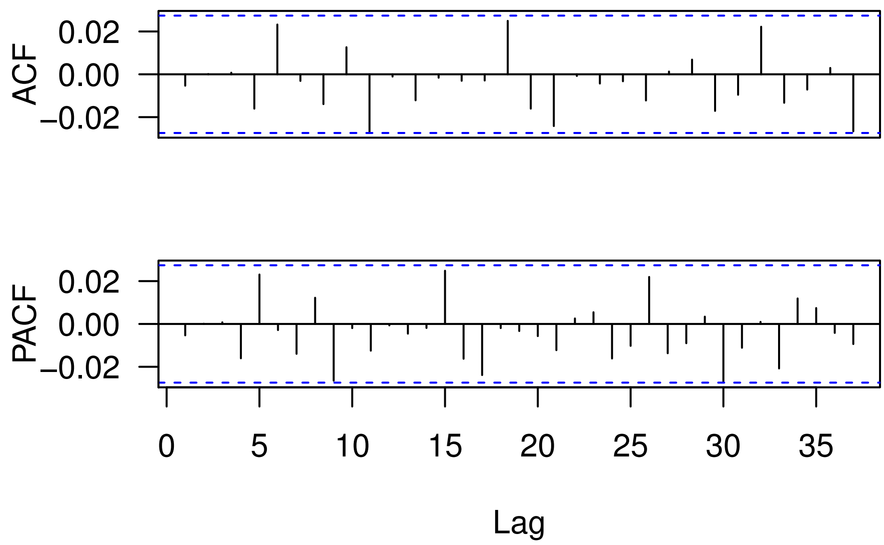
The estimated coefficients of the GARCH model are summarized in the following table.

Table 4.2: GARCH model

Variable	coefficient	standard error	t-ratio	p-value
ϕ_0	0.001	0.000	3.815	0.000
θ_1	0.072	0.016	4.368	0.000
α_0	0.000	0.000	0.978	0.328
α_1	0.124	0.019	6.424	0.000
β_1	0.86	0.028	31.23	0.000
ν	7.347	0.743	9.883	0.000

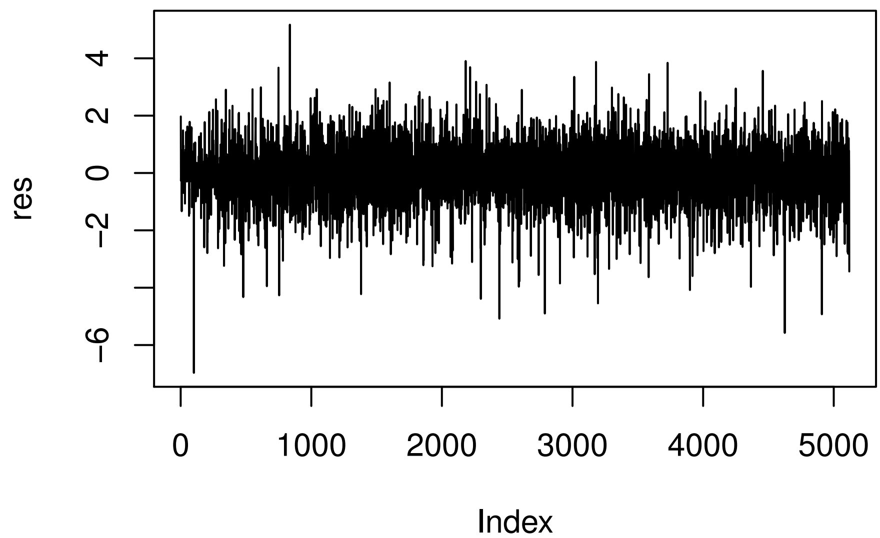
We can see that coefficient of MA process and coefficients of GARCH process are very strongly significant. The constant in mean equation is also strongly significant and also the coefficient of shape of the distribution. The constant in variance equation is not significant.

Figure 4.8: ACF and PACF of squared residuals of GARCH model



Source: author's computations.

Figure 4.9: Plot of squared residuals of GARCH model



Source: author's computations.

When we check squared residuals from this model, we can see on ACF and PACF functions (see Figure 4.8) that all lags of these functions are not significant, so the ARCH is no longer present. This is also confirmed by ARCH-LM test that does not reject the null hypothesis that ARCH effect is present with p-value larger than 0.3. From the plot of residuals from this model (Figure 4.9), we can see that the clusters of high volatility are not present so it seems that our model explains our data well and we will use it in our further analysis.

4.2 Empirical results

4.2.1 Model A

The model A is the simplest model where we use simple dummy variables for wins, losses and draws.

$$R_t = \phi_0 + \theta_1\epsilon_{t-1} + \gamma_1 win_{st} + \gamma_2 loss_{st} + \gamma_3 draw_{st} + \sum_{i=1}^{4}\rho_i D_i + \epsilon_t \qquad (4.5)$$

$$\begin{aligned} \epsilon_t &= v_t\sqrt{h_t}, \epsilon_t \sim f_\nu \\ h_t &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned} \qquad (4.6)$$

The empirical results of our model A for football are presented in the following table.

Table 4.3: Model A for football results

Variable	Coefficient	Value	standard error	t-ratio	p-value
constant	ϕ_0	0.0007	0.0003	2.3581	0.0184
ϵ_{t-1}	θ_1	0.072	0.0164	4.3878	0.0000
win	γ_1	0.0005	0.0012	0.4215	0.6734
loss	γ_2	-0.0021	0.0016	-1.3572	0.1747
draw	γ_3	-0.0025	0.0022	-1.1545	0.2483
monday	ρ_1	0.0002	0.0004	0.4739	0.6356
tuesday	ρ_2	-0.0004	0.0004	-1.0143	0.3105
wednesday	ρ_3	0.0001	0.0004	0.1629	0.8706
thursday	ρ_4	0.0001	0.0004	0.2417	0.809
constant	α_0	0.0000	0.0000	0.9751	0.3295
ϵ_{t-1}^2	α_1	0.1236	0.0193	6.4066	0.0000
h_t	β_1	0.8604	0.0276	31.2159	0.0000
shape	ν	7.3361	0.74	9.9138	0.0000

From the results of the A model for football we can see that only coefficients of MA and GARCH processes are significant. The coefficient of win dummy variable is positive as we expected but very strongly statistically insignificant. With one-sided test we also cannot reject the null hypothesis that coefficient is equal to zero with p-value equal to 0.34 so our hypothesis that there is positive effect after wins is rejected.

The coefficient of loss dummy variable is negative as we expected and is also highly insignificant but the p-value is not as high as for win dummy variable. With one-sided test the p-value is equal to 0.087 so we can reject the null hypothesis that coefficient is equal to zero on ten percent significance level but we cannot reject the null hypothesis on five percent significance level so our hypothesis that there is negative effect after losses is rejected. The coefficient of draw dummy variable is also negative and not significant but again not so strongly insignificant as dummy variable for wins. The coefficients of weekdays are all insignificant.

In the following table are summed up results of our hypotheses. All of them are for this model rejected as we are not able to reject the null hypotheses at five percent significance level.

Table 4.4: Model A hypotheses results for football

Variable tested	H_0	H_A	p-value	Result	Our hypothesis
Win	$\gamma_1 = 0$	$\gamma_1 > 0$	0.3367	H_0	Rejected
Loss	$\gamma_2 = 0$	$\gamma_2 < 0$	0.0874	H_0	Rejected
Draw	$\gamma_3 = 0$	$\gamma_3 \neq 0$	0.2483	H_0	Rejected

The empirical results of model A for ice hockey are presented in the following table.

Table 4.5: Model A for ice hockey results

Variable	Coefficient	Value	standard error	t-ratio	p-value
constant	ϕ_0	0.0007	0.0003	2.3194	0.0204
ϵ_{t-1}	θ_1	0.0719	0.0164	4.38	0.0000
win	γ_1	0.0003	0.0014	0.1966	0.8441
loss	γ_2	0.0001	0.0015	0.091	0.9275
draw	γ_3	0.0000	0.003	0.0055	0.9956
monday	ρ_1	0.0001	0.0004	0.3625	0.7169
tuesday	ρ_2	-0.0004	0.0004	-0.9983	0.3181
wednesday	ρ_3	0.0001	0.0004	0.1616	0.8716
thursday	ρ_4	0.0001	0.0004	0.229	0.8189
constant	α_0	0.0000	0.0000	0.9773	0.3284
ϵ^2_{t-1}	α_1	0.1236	0.0192	6.4256	0.0000
h_{t-1}	β_1	0.8604	0.0275	31.3085	0.0000
shape	ν	7.3364	0.7443	9.8568	0.0000

From this table we can see again that only coefficients of MA and GARCH processes are significant. The coefficient of wins is positive as we expected but very strongly insignificant. With one-sided test we also cannot reject the null hypothesis that coefficient is equal to zero with p-value equal to 0.42 so our hypothesis that there is positive effect after wins is strongly rejected.

The coefficient of losses is also positive what is in contrary to what we expected. With one-sided test we have p-value equal to 0.54 so we cannot reject the null hypothesis that coefficient is equal to zero. The coefficient of draws is also strongly insignificant, p-value is almost equal to 1 so we cannot reject the null hypothesis that coefficient is equal to zero.

Similarly as for the football, all our hypotheses were rejected for model A as we were not able to reject any null hypothesis that coefficient is equal to zero. The results of our hypotheses are described in the following table.

Table 4.6: Model A hypotheses results for ice hockey

Variable tested	H_0	H_A	p-value	Result	Our hypothesis
Win	$\gamma_1 = 0$	$\gamma_1 > 0$	0.422	H_0	Rejected
Loss	$\gamma_2 = 0$	$\gamma_2 < 0$	0.5363	H_0	Rejected
Draw	$\gamma_3 = 0$	$\gamma_3 \neq 0$	0.9956	H_0	Rejected

4.2.2 Model B

In model B we use simple dummy variables for results and we also add dummy variables for wins and losses in elimination games.

$$R_t = \phi_0 + \theta_1 \epsilon_{t-1} + \gamma_1 win_{st} + \gamma_2 loss_{st} + \gamma_3 draw_{st} + \\ + \gamma_4 winel_{st} + \gamma_5 lossel_{st} + \sum_{i=1}^{4} \rho_i D_i + \epsilon_t \quad (4.7)$$

$$\epsilon_t = \upsilon_t \sqrt{h_t}, \epsilon_t \sim f_\nu \\ h_t = \alpha_0 + \alpha_1 \epsilon_{t-1}^2 + \beta_1 h_{t-1} \quad (4.8)$$

The empirical results of model B for football are presented in the following table.

Table 4.7: Model B for football results

Variable	Coefficient	Value	standard error	t-ratio	p-value
constant	ϕ_0	0.0007	0.0003	2.389	0.0169
ϵ_{t-1}	θ_1	0.0716	0.0165	4.3433	0.0000
win	γ_1	0.0011	0.0012	0.8992	0.3685
loss	γ_2	-0.0013	0.002	-0.6463	0.5181
draw	γ_3	-0.0025	0.0022	-1.1635	0.2446
winel	γ_4	-0.0035	0.0024	-1.4192	0.1558
lossel	γ_5	-0.0048	0.0014	-3.3626	0.0008
monday	ρ_1	0.0002	0.0004	0.4541	0.6497
tuesday	ρ_2	-0.0005	0.0004	-1.0414	0.2977
wednesday	ρ_3	0.0000	0.0004	0.106	0.9156
thursday	ρ_4	0.0001	0.0004	0.212	0.8321
constant	α_0	0.0000	0.0000	0.9743	0.3299
ϵ_{t-1}^2	α_1	0.1239	0.0193	6.407	0.0000
h_{t-1}	β_1	0.8603	0.0276	31.1947	0.0000
shape	ν	7.319	0.7388	9.9072	0.0000

From the results we can see that not only MA and GARCH coefficients are significant in this model but coefficient of dummy variable for losses in elimination games is also significant.

The coefficient of dummy variable for wins is positive as we expected but insignificant. With one-sided hypothesis we cannot reject the null hypothesis that coefficient is equal to zero even at 10 percent confidence level as we have p-value equal to 0.184. The coefficient of loss dummy variable is negative as we expected but we also cannot reject the null hypothesis with one-sided test with p-value equal to 0.259. The coefficient of draw variable is again negative and insignificant. The coefficient of wins in elimination games is surprisingly negative so with one-sided test we cannot strongly reject the null hypothesis of coefficient being equal to zero with p-value equal to 0.922. As we mentioned before, the only significant variable is the dummy variable for losses in elimination games and it is also negative as we expected. With one-sided test we can reject the null hypothesis that the coefficient is equal to 0 even at one percent confidence level so our hypothesis that the coefficient of dummy variable for losses in elimination games is negative is not rejected. The effect is also statistical significant as difference between normal return and return is expected to be 48 basis points.

Our results of hypotheses from model B for football are summed up in the following table. We see that four of our hypotheses were rejected but the most important hypothesis was not rejected by our test. According to the results from this model, we can state that the losses from elimination games have significant negative impact on stock market returns in Czech stock market. The other results of games are not significant.

Table 4.8: Model B hypotheses results for football

Variable tested	H_0	H_A	p-value	Result	Our hypothesis
Win	$\gamma_1 = 0$	$\gamma_1 > 0$	0.184	H_0	Rejected
Loss	$\gamma_2 = 0$	$\gamma_2 < 0$	0.259	H_0	Rejected
Draw	$\gamma_3 = 0$	$\gamma_3 \neq 0$	0.244	H_0	Rejected
Winel	$\gamma_1 = 0$	$\gamma_4 > 0$	0.922	H_0	Rejected
Lossel	$\gamma_2 = 0$	$\gamma_5 < 0$	0.0004	H_A	Not rejected

The empirical results of model B for ice hockey results are presented in the following table.

Table 4.9: Model B for ice hockey results

Variable	Coefficient	Value	standard error	t-ratio	p-value
constant	ϕ_0	0.0007	0.0003	2.344	0.0191
ϵ_{t-1}	θ_1	0.072	0.0164	4.38	0.0000
win	γ_1	0.0003	0.0016	0.1811	0.8563
loss	γ_2	0.0009	0.0023	0.3846	0.7005
draw	γ_3	0.0004	0.003	0.1256	0.9001
winel	γ_4	-0.0018	0.0031	-0.5876	0.5568
lossel	γ_5	0.0001	0.0024	0.0218	0.9826
monday	ρ_1	0.0001	0.0004	0.3479	0.7279
tuesday	ρ_2	-0.0004	0.0004	-1.0197	0.3079
wednesday	ρ_3	0.0001	0.0004	0.1631	0.8705
thursday	ρ_4	0.0001	0.0004	0.2338	0.8151
constant	α_0	0.0000	0.0000	0.9737	0.3302
ϵ^2_{t-1}	α_1	0.1234	0.0194	6.3662	0.0000
h_{t-1}	β_1	0.8607	0.0276	31.1729	0.0000
shape	ν	7.3141	0.7406	9.8752	0.0000

From the results we can see that in the model B for ice hockey are again only MA and GARCH coefficients significant. The coefficient of wins is positive as we expected but very strongly insignificant. With one-sided test we also cannot reject the null hypothesis that coefficient is equal to zero with p-value equal to 0.43. The coefficient of losses is also positive what is in contrary to what we expected. With one-sided test we have p-value equal to 0.65 so we cannot reject the null hypothesis that coefficient is equal to zero. The coefficient of draws is also strongly insignificant, p-value is equal to 0.9. Even dummy variable of coefficients for wins and losses in elimination games are strongly insignificant and have the opposite signs in respect to what we expected. The p-values of one-sided tests are 0.721 and 0.509 respectively so we cannot reject the null hypotheses that coefficients are equal to zero.

Similarly as in the model A for ice hockey, all our hypotheses were rejected for model B as we were not able to reject any null hypothesis that coefficient is equal to zero. The results of our hypotheses are described in the following table.

Table 4.10: Model B hypotheses results for ice hockey

Variable tested	H_0	H_A	p-value	Result	Our hypothesis
Win	$\gamma_1 = 0$	$\gamma_1 > 0$	0.428	H_0	Rejected
Loss	$\gamma_2 = 0$	$\gamma_2 < 0$	0.65	H_0	Rejected
Draw	$\gamma_3 = 0$	$\gamma_3 \neq 0$	0.9	H_0	Rejected
Winel	$\gamma_1 = 0$	$\gamma_4 > 0$	0.721	H_0	Rejected
Lossel	$\gamma_2 = 0$	$\gamma_5 < 0$	0.509	H_0	Rejected

4.2.3 Model C

In model C we use our importance proxy variables instead of simple dummy variables.

$$R_t = \phi_0 + \theta_1\epsilon_{t-1} + \gamma_1 win_{st}imp_{st} + \gamma_2 loss_{st}imp_{st} + \gamma_3 draw_{st}imp_{st} + \sum_{i=1}^{4}\rho_i D_i + \epsilon_t \tag{4.9}$$

$$\begin{aligned} \epsilon_t &= \upsilon_t\sqrt{h_t}, \epsilon_t \sim f_\nu \\ h_t &= \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1} \end{aligned} \tag{4.10}$$

The results of model C for football results are presented in Table 4.11.

From the results we can see that the variable for losses is significant in this model and negative as we expected, we can reject that coefficient is equal to zero at five percent confidence level. With one-sided test we have p-value equal to 0.009 so we can reject this null hypothesis even at one percent confidence level. The coefficient of loss effect is smaller in compare to the coefficient of losses in elimination games in model B. This could be explained by the property of our proxy variable as it is not always equal to 1. The expected decline after losses in more important matches that have attendance over thirty thousands spectators is expected to be more than 15 basis points. On the contrary losses in less important matches re expected to have much lower impact. The win variable is in this model insignificant and surprisingly negative. With one-sided test we have p-value equal to 0.832 so we strongly cannot reject the null hypothesis. Draw variable is again negative and insignificant.

Table 4.11: Model C for football results

Variable	Coefficient	Value	standard error	t-ratio	p-value
constant	ϕ_0	0.0007	0.0003	2.3619	0.0182
ϵ_{t-1}	θ_1	0.0719	0.0164	4.3746	0.0000
winimp	γ_1	-0.0004	0.0004	-0.9614	0.3364
lossimp	γ_2	-0.0005	0.0002	-2.3501	0.0188
drawimp	γ_3	-0.0009	0.001	-0.8525	0.3939
monday	ρ_1	0.0002	0.0004	0.5687	0.5695
tuesday	ρ_2	-0.0004	0.0004	-1.0244	0.3056
wednesday	ρ_3	0.0001	0.0004	0.1896	0.8496
thursday	ρ_4	0.0001	0.0004	0.2958	0.7674
constant	α_0	0.0000	0.0000	0.9739	0.3301
ϵ_{t-1}^2	α_1	0.1238	0.0193	6.4127	0.0000
h_{t-1}	β_1	0.8604	0.0275	31.3017	0.0000
shape	ν	7.3324	0.741	9.8948	0.0000

The results of our hypotheses are summed up in the following table. Only the hypothesis that there is negative effect after losses was not rejected, the other two our hypotheses were rejected for this model.

Table 4.12: Model C hypotheses results for football

Variable tested	H_0	H_A	p-value	Result	Our hypothesis
Winimp	$\gamma_1 = 0$	$\gamma_1 > 0$	0.832	H_0	Rejected
Lossimp	$\gamma_2 = 0$	$\gamma_2 < 0$	0.009	H_A	Not rejected
Drawimp	$\gamma_3 = 0$	$\gamma_3 \neq 0$	0.394	H_0	Rejected

The results of model C for ice hockey results are presented in the following table.

Table 4.13: Model C for ice hockey results

Variable	Coefficient	Value	standard error	t-ratio	p-value
constant	ϕ_0	0.0007	0.0003	2.3429	0.0191
ϵ_{t-1}	θ_1	0.0719	0.0164	4.3768	0.0000
winimp	γ_1	0.0000	0.0006	0.0424	0.9662
lossimp	γ_2	-0.0001	0.0005	-0.1871	0.8516
drawimp	γ_3	0.0004	0.0013	0.3234	0.7464
monday	ρ_1	0.0002	0.0004	0.3794	0.7044
tuesday	ρ_2	-0.0004	0.0004	-1.0000	0.3173
wednesday	ρ_3	0.0001	0.0004	0.165	0.8689
thursday	ρ_4	0.0001	0.0004	0.2204	0.8256
constant	α_0	0.0000	0.0000	0.9753	0.3294
ϵ^2_{t-1}	α_1	0.1236	0.0193	6.4088	0.0000
h_{t-1}	β_1	0.8605	0.0275	31.245	0.0000
shape	ν	7.3339	0.7405	9.9038	0.0000

From the results of model C for ice hockey we can see that again only coefficients of MA and GARCH variables are significant. The coefficient of variable for wins is positive but very strongly insignificant, with one-sided test we have p-value equal to 0.483 so we cannot reject the null hypothesis. The coefficient of variable for losses is negative as we expected but also very strongly insignificant, with one-sided test we have p-value equal to 0.426. The draw variable is positive and also strongly insignificant.

Similarly as in previous models for ice hockey, we reject all our hypotheses. The results of hypotheses are summarized in following table.

Table 4.14: Model C hypotheses results for ice hockey

Variable tested	H_0	H_A	p-value	Result	Our hypothesis
Winimp	$\gamma_1 = 0$	$\gamma_1 > 0$	0.483	H_0	Rejected
Lossimp	$\gamma_2 = 0$	$\gamma_2 < 0$	0.426	H_0	Rejected
Drawimp	$\gamma_3 = 0$	$\gamma_3 \neq 0$	0.746	H_0	Rejected

4.2.4 Model D

The last model D is the combination of model B and model C. We use our proxy variables instead of simple dummy variables and we also include the variables for elimination games.

$$R_t = \phi_0 + \theta_1\epsilon_{t-1} + \gamma_1 win_{st}imp_{st} + \gamma_2 loss_{st}imp_{st} + \gamma_3 draw_{st}imp_{st} + \\ + \gamma_4 winel_{st}imp_{st} + \gamma_5 lossel_{st}imp_{st} + \sum_{i=1}^{4}\rho_i D_i + \epsilon_t \quad (4.11)$$

$$\epsilon_t = \upsilon_t\sqrt{h_t}, \epsilon_t \sim f_\nu \\ h_t = \alpha_0 + \alpha_1\epsilon_{t-1}^2 + \beta_1 h_{t-1} \quad (4.12)$$

Table 4.15: Model D for football results

Variable	Coefficient	Value	standard error	t-ratio	p-value
constant	ϕ_0	0.0007	0.0003	2.3675	0.0179
ϵ_{t-1}	θ_1	0.0715	0.0165	4.3431	0.0000
winimp	γ_1	-0.0001	0.0005	-0.1372	0.8909
lossimp	γ_2	-0.0004	0.0004	-1.1019	0.2705
drawimp	γ_3	-0.0009	0.001	-0.8568	0.3916
winelimp	γ_4	-0.0002	0.0005	-0.4517	0.6515
losselimp	γ_5	-0.0013	0.0004	-3.0135	0.0026
monday	ρ_1	0.0002	0.0004	0.5651	0.572
tuesday	ρ_2	-0.0004	0.0004	-1.0254	0.3052
wednesday	ρ_3	0.0001	0.0004	0.1575	0.8749
thursday	ρ_4	0.0001	0.0004	0.291	0.7711
constant	α_0	0.0000	0.0000	0.9703	0.3319
ϵ_{t-1}^2	α_1	0.124	0.0195	6.3701	0.0000
h_{t-1}	β_1	0.8602	0.0277	31.1073	0.0000
shape	ν	7.3273	0.7421	9.8731	0.0000

Similarly as in model B, not only the coefficients of MA and GARCH variables are significant, also the coefficient of losses in elimination games is strongly significant.

The coefficient of dummy variable for wins is negative but strongly insignificant. With one-sided hypothesis we cannot reject the null hypothesis that coefficient is equal to zero with p-value equal to 0.554. The coefficient of loss dummy variable is negative as we expected but also insignificant. We cannot reject the null hypothesis with one-sided test with p-value equal to 0.135. The coefficient of draw variable is again negative and insignificant. The coefficient of wins in elimination games is surprisingly negative so with one-sided test we cannot strongly reject the null hypothesis of coefficient being equal to zero with p-value equal to 0.674. The only significant variable is the dummy variable for losses in elimination games and it is also negative as we expected. With one-sided test we have p-value equal to 0.001 so we can reject the null hypothesis that the coefficient is equal to 0 even at one percent confidence level. The expected decline after losses in more important elimination matches that have attendance over thirty thousands spectators is expected to be more than 40 basis points what is definitely economically significant.

Similarly as in model B, the four of our hypotheses were rejected but the most important hypothesis was not rejected by our test. The results of our hypotheses are summed up in the following table.

Table 4.16: Model D hypotheses results for football

Variable tested	H_0	H_A	p-value	Result	Our hypothesis
Winimp	$\gamma_1 = 0$	$\gamma_1 > 0$	0.554	H_0	Rejected
Lossimp	$\gamma_2 = 0$	$\gamma_2 < 0$	0.135	H_0	Rejected
Drawimp	$\gamma_3 = 0$	$\gamma_3 \neq 0$	0.392	H_0	Rejected
Winelimp	$\gamma_1 = 0$	$\gamma_1 > 0$	0.674	H_0	Rejected
Losselimp	$\gamma_2 = 0$	$\gamma_2 < 0$	0.001	H_A	Not rejected

The last model is model D for ice hockey. The results of this model are described in Table 4.17.

Similarly as in previous models for ice hockey, only coefficients of MA and GARCH variables are significant.

The coefficient of variable for wins is negative what we did not expected and strongly insignificant. With one-sided test we have p-value equal to 0.533. The coefficient for losses has also opposite sign to what we expected and is also insignificant. The p-value of one-sided test is 0.532. The draw variable is positive and insignificant. Wins and losses in elimination games have also opposite signs to what we expected. The p-values of one-sided tests are 0.636 and 0.595 respectively.

Table 4.17: Model D for ice hockey results

Variable	Coefficient	Value	standard error	t-ratio	p-value
constant	ϕ_0	0.0007	0.0003	2.354	0.0186
ϵ_{t-1}	θ_1	0.0721	0.0164	4.3862	0.0000
winimp	γ_1	-0.0001	0.0007	-0.0833	0.9336
lossimp	γ_2	0.0001	0.0008	0.0799	0.9363
drawimp	γ_3	0.0004	0.0014	0.2711	0.7863
winelimp	γ_4	-0.0004	0.0011	-0.3484	0.7276
losselimp	γ_5	0.0002	0.0008	0.2417	0.809
monday	ρ_1	0.0002	0.0004	0.3751	0.7076
tuesday	ρ_2	-0.0004	0.0004	-1.0077	0.3136
wednesday	ρ_3	0.0001	0.0004	0.1704	0.8647
thursday	ρ_4	0.0001	0.0004	0.2246	0.8223
constant	α_0	0.0000	0.0000	0.9688	0.3326
ϵ_{t-1}^2	α_1	0.1234	0.0195	6.318	0.0000
h_{t-1}	β_1	0.8607	0.0277	31.0218	0.0000
shape	ν	7.3408	0.7469	9.8284	0.0000

Similarly as in all previous models for ice hockey, we reject all our hypotheses of this model. The results of hypotheses are summarized in following table.

Table 4.18: Model D hypotheses results for ice hockey

Variable tested	H_0	H_A	p-value	Result	Our hypothesis
Winimp	$\gamma_1 = 0$	$\gamma_1 > 0$	0.533	H_0	Rejected
Lossimp	$\gamma_2 = 0$	$\gamma_2 < 0$	0.532	H_0	Rejected
Drawimp	$\gamma_3 = 0$	$\gamma_3 \neq 0$	0.786	H_0	Rejected
Winelimp	$\gamma_1 = 0$	$\gamma_1 > 0$	0.636	H_0	Rejected
Losselimp	$\gamma_2 = 0$	$\gamma_2 < 0$	0.595	H_0	Rejected

4.2.5 Summary of results

From the above results we can see that there is some negative effect on Czech stock market after losses of Czech national football team. The effect is significant when we control for importance of the matches and in elimination games what is in line with previous literature. The coefficient of losses are also economically significant. In contrast with previous literature, we did not find any positive effect on Czech stock market after wins. Also the effect after draws is in all models insignificant. The summary of our hypothesis for football is in the following table.

Table 4.19: Summary of hypotheses for football

Variable tested	Model A	Model B	Model C	Model D
Win	Rejected	Rejected	Rejected	Rejected
Loss	Rejected	Not rejected	Rejected	Rejected
Draw	Rejected	Rejected	Rejected	Rejected
Winel		Rejected		Rejected
Lossel		Not rejected		Not rejected

Similarly as Edmans *et al.* (2007), we did not find any significant effect on Czech stock market after matches of Czech national ice hockey team.

Chapter 5

Discussion

In this chapter, we discuss our empirical results that were presented in the previous chapter. We documented only negative effect on Czech stock market after losses of Czech national football team. We did not find any positive effect after wins and after draws. We also were not able to find any effect after games of Czech national ice hockey team.

In following subchapters, we present some possible causes why we did not find those effects and why we were able to find only negative effect after losses in football games.

5.1 Loss effect versus win effect

There are two straightforward explanation why the negative effect after losses is stronger than positive effect after wins.

The first one comes from economic theory where agents usually much more react to bad news than good news. The second one is that wins usually result only in advancing to the next stage and losses lead to elimination of national team from the competition.

Fans also usually overestimate the strength of their favorite teams. For example, there were survey in England whether they believe they beat Brazil in the 2002 World Cup quarter final. The 86% believed that the English national team would beat Brazilian national team, in contrast bookmakers assigned only 42% probability of the win of English team. Brazilian team won that match and then also won the whole World Cup. (Edmans *et al.* 2007) So the losses are usually more surprising for them than wins and therefore frustration after losses is larger than satisfaction after wins.

Another possible explanation why the positive effect after wins is insignificant is that Czech nation is very well known for drinking. It is quiet usual that Czech fans celebrate after wins with lots of drinks and they can feel bad in the next day because of it so the positive effect can be neutralize by that.

The insignificance of draws is not surprising as sometimes can be draw taken as a good result and sometimes as a bad result.

5.2 PX index issues

One possible explanation why some our hypotheses were not supported is that PX index is not suitable for explaining behavior of Czech investors on Czech stock market. The composition of PX index (see Table 5.1) is quite unusual in compare to other stock indices.

The PX index is composed from 13 companies, more than 60% of PX index is composed by financial institutions and almost 60% of PX index is composed by only three companies (Erste Group Bank, Komerční banka, ČEZ) what makes PX index quite undiversified. The stock indices of other countries are usually much more diversified as they are composed from more companies from different sectors and the share of individual companies are not as large as in PX index.

Table 5.1: Current PX composition

Name	Sector	Index Portion
Erste Group Bank	Banking	21.7%
Komerční banka	Banking	19.75%
ČEZ	Electric Utilities	18.25%
Moneta Money Bank	Banking	9.89%
VIG	Insurance	8.93%
O2 Č.R.	Telecommunications	6.23%
Unipetrol	Oil & Gas	3.59%
Pegas Nonwovens	Personal Products	2.75%
Philip Morris ČR	Food, Beverage & Tobacco	2.73%
CETV	Media	2.49%
Stock	Food, Beverage & Tobacco	2.26%
Fortuna	Leisure & Gambling	0.8%
Kofola ČS	Food, Beverage & Tobacco	0.62%

Source: https://www.pse.cz/

The solution to this problem could be creating own index that is more varied than PX index and with more suitable companies.

There is also question how large is proportion of real investors that can be influenced by mood variables and how large proportion of investments is done by various algorithms that are naturally not influenced by emotions. The proportion of algorithmic trading is raising in the time and because we use data from recent years, the proportion of algorithmic trading is in our data larger than in previous studies. Therefore the effect of mood variables is naturally weaker in our data.

5.3 Distribution of matches

Another explanation can be not ideal distribution of games for our causes. One of the problems with distribution of games is that large proportion of games are played on Fridays and Saturdays, for these games we used returns on Mondays. This could be a problem because the mood of people is probably less influenced by these games because there is more than one day difference between match day and working day.

In chapter 3 we also mentioned problem that at the ice hockey World Championships there are often played two matches between two working days. In our models, we use only the second matches in that case but mood can be influenced also by the first matches. Similar problem can also occur when there are played two matches during two working days, the mood of investors is probably still influenced by the match that is played two days ago. This problem can be one of the explanations why the performance of our models are much worse for ice hockey than for football because in football is the pause between matches longer.

To solve this problem, we suggest to use for ice hockey also result from matches that were played two days ago as explanatory variable not just the match result from previous day.

5.4 Low number of observations

Another explanation can be low number of competitive matches played by national teams. This is the disadvantage of event approach as the number of events is always limited and it reduces the statistical power of our tests. The average number of competitive matches is around 7 per year for football and 10 for ice hockey. Because of the previously mentioned problem for ice hockey, we cannot use all of them and the average number of ice hockey matches per year is also just around 7. As it is usually around 250 working per year, it means that approximately only 3% of working days are expected to be affected in our models what is quite low proportion. The number for elimination games is naturally much lower so there is a question whether our results for elimination games are even usable as they can be easily affected by some outliers.

There is no simple solution if we want to use just competitive matches of Czech national teams. One possible solution is to add friendly matches but it would not increase the number of matches a lot and the results from them are usually not important for the fans so they should not effect the mood at all.

Probably better solution would be to use matches of local clubs. They play much more matches during the year than national teams. From Czech football clubs would be Sparta Prague natural choice. It is the most successful team in Czech republic with the largest number of fans.

But the problem with using local clubs is that not all football fans like them but also many fans do not like them. Especially fans of their city rival Slavia Prague will have good mood after losses of Sparta Prague and not bad mood as model expects so the assumption that the effect is correlated across the majority within a country would not be fulfilled. However, the results still could be better than for national teams because the identification with local teams is for many people stronger than with national team.

This idea was used by Berument *et al.* (2006), they adapted it on three major Istanbul teams (Besiktas, Fenerbahce, and Galatasaray), they found significant effect only after wins of Besiktas. From Czech football could be used apart from Sparta also previously mentioned Slavia Prague, and also Viktoria Pilsen. As Czech hockey teams we would recommend Sparta Prague and Kometa Brno as they have the most fans in the Czech republic.

5.5 Misspecification of model

As we expected that the more important matches have larger effect on stock market, we used our own proxy variables for importance of matches. There is no simple and right way how to measure importance of matches but there are many possibilities, we just chose one of them for each sport and they could be not appropriate proxy variables for importance of matches. The other possible proxy variable for importance of the matches could be television viewership why we did not use this variables we explained in chapter 3.

In these times there are very popular social media like Twitter or Facebook so the number of reactions on these social medias could be taken as good proxy for importance of matches for fans as they definitely react more to matches that are important to them. However, it is quite hard to distinguish which reactions should be calculated because there are many not official sites and their count differs over time. Also these social medias cannot be used for older data as these social medias were created in recent years.

As another proxy variable for importance of a match could be taken number of newspaper articles that were written about the certain match. The more important match would have definitely much more articles written than the less important matches. But there is again problem how to specify correct number of articles, which newspapers to choose as relevant sources.

Edmans *et al.* (2007) used the Elo rating to calculate probability that the team will win the match. Usually the outcomes from more balanced matches are more important for fans. Similarly could be used pre-match betting odds published by betting offices to calculate the probability of various outcomes from the matches. Especially unexpected losses probably affect mood of fans to a large extent. The volume of bets on the matches could be also good proxy for importance of matches as people usually bet more on important matches but these would probably be hard to collect as it is private information of each betting office.

The other explanation could be that the relationship between outcomes of sports matches and stock market returns simply does not exist. Gerlach (2011) believed that low returns on the U.S. market during the football World Cup documented by Kaplanski & Levy (2010) were not related to football as there were similarly low returns on the U.S. market four weeks before World Cup started and there were no similar effect for continental championships, which according to their theory should have a similar effect. They also examined whether outcome of international football matches also affects returns of bordering countries stock indices as they believed that the abnormal declines can be caused by some regional causes that are not related to football. They found that the decline in stock market is not significantly larger in country whose national team lost the match in previous day than in stock market of the bordering country.

It is also possible that we excluded some variables that should be included in the model. The period which we examined is quite long so there could be some structural breaks that could change behavior of investors on the Czech stock market such as financial crisis if 2007-2008. As mentioned before, the structure of investors is also changing over time, nowadays large portion of trade on Czech stock market is done by algorithmic trading and it is also much easier to invest in foreign stock markets as the world is much more connected. The legislative and taxes also evolve over time so the volume of trades can be also influenced by these changes. Our model is same for the whole period so it does not cover these possible changes that may affect the Czech stock market.

Chapter 6

Conclusion

In this thesis, we analyzed the impact of sports results on Czech stock market. As sports results we used the matches played by Czech national football and ice hockey teams in main competitions. As Czech stock market index we used PX index, the official index of Prague Stock Exchange. The main idea is that the mood of Czech investors is affected by the outcomes from the matches of Czech national teams and investors that are in good mood tend to evaluate future prospects more optimistically and invest more. On the contrary, people with bad mood tend to evaluate future prospects less optimistically and invest less. Because our data suffered from autoregressive conditional heteroskedasticity, we used ARMA-GARCH to model our data.

We used four different models to find whether sports results are statistically significant and whether more important matches and elimination matches have larger impact than normal matches. As proxy for importance of football matches we used the attendance of the certain as we expect that more important matches have larger attendance. As proxy for importance of ice hockey matches we used the index of strength of the opponent.

We found negative effect after losses of Czech national football team. The effect was statistically significant when we used our proxy for importance of matches and effect was much stronger for elimination matches. We documented no significant effect after wins of Czech national football team. This can be explained by asymmetry between perception of good and bad news, and also by the fact that losses in elimination games result in elimination from the tournament. We did not find any effect after draws.

Similarly as Edmans *et al.* (2007), we did not find any significant effect after matches of Czech national ice hockey team. The one possible reason could be structure of ice hockey tournaments which are played in just two weeks so we have to drop some observations as there are often played two matches between two working days.

The other reason why we did not find significant effect for some variables could be the limited number of observation as there are not many competitions that are played by national teams. To deal with this problem, we suggest to use football clubs instead of the national teams in further research. They play much more matches during year and for many fans is identification with football clubs much greater than with national teams. The disadvantage of this approach is that there is often big rivalry between fans of the clubs so for fans of the rivalry team the losses would have positive impact instead of negative.

Bibliography

Berument, H. & N. B. Ceylan (2012): "Effects of soccer on stock markets: The return–volatility relationship." *Social Science Journal* **49(3)**: pp. 368 – 374.

Berument, H., N. B. Ceylan, & E. Gozpinar (2006): "Performance of soccer on the stock market: Evidence from turkey." *Social Science Journal* **43(4)**: pp. 695 – 699.

Boido, C. & A. Fasano (2007): "Football and mood in italian stock exchange." *ICFAI Journal of Behavioral Finance* **4(4)**: pp. 32 – 50.

Bollerslev, T. (1986): "Generalized autoregressive conditional heteroskedasticity." *Journal of Econometrics* **31(3)**: pp. 307–327.

Bollerslev, T. (1987): "A conditionally heteroskedastic time series model for speculative prices and rates of return." *Review of Economics & Statistics* **69(3)**: p. 542.

Chang, S.-C., S.-S. Chen, R. K. Chou, & Y.-H. Lin (2012): "Local sports sentiment and returns of locally headquartered stocks: A firm-level analysis." *Journal of Empirical Finance* **19(3)**: pp. 309 – 318.

Cohen-Charash, Y., C. A. Scherbaum, J. D. Kammeyer-Mueller, & B. M. Staw (2013): "Mood and the market: Can press reports of investors' mood predict stock prices?." *PLoS ONE* **8(8)**: pp. 1 – 15.

Dickey, D. A. & W. A. Fuller (1979): "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American Statistical Association* **74(366)**: pp. 427–431.

Edmans, A., D. García, & Ø. Norli (2007): "Sports sentiment and stock returns." *Journal of Finance* **62(4)**: pp. 1967 – 1998.

ENGLE, R. (2001): "GARCH 101: The Use of ARCH/GARCH Models in Applied Econometrics." *Journal of Economic Perspectives* **15(4)**: pp. 157 – 168.

ENGLE, R. F. (1982): "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation." *Econometrica* **50(4)**: pp. 987 – 1007.

GERLACH, J. R. (2011): "International sports and investor sentiment: do national team matches really affect stock market returns?." *Applied Financial Economics* **21(12)**: pp. 863 – 880.

HIRSHLEIFER, D. & T. SHUMWAY (2003): "Good day sunshine: Stock returns and the weather." *The Journal of Finance* **58(3)**: pp. 1009–1032.

HIRT, E. R., D. ZILLMANN, G. A. ERICKSON, & C. KENNEDY (1992): "Costs and benefits of allegiance: Changes in fans' self-ascribed competencies after team victory versus defeat." *Journal of Personality & Social Psychology* **63(5)**: pp. 724 – 738.

KAMSTRA, M. J., L. A. KRAMER, & M. D. LEVI (2000): "Losing Sleep at the Market: The Daylight-Savings Anomaly." *American Economic Review* **90(4)**: pp. 1000–1005.

KAPLANSKI, G. & H. LEVY (2010): "Exploitable predictable irrationality: The fifa world cup effect on the u.s. stock market." *Journal of Financial & Quantitative Analysis* **45(2)**: pp. 535 – 553.

KWIATKOWSKI, D., P. C. PHILLIPS, P. SCHMIDT, & Y. SHIN (1992): "Testing the null hypothesis of stationarity against the alternative of a unit root." *Journal of Econometrics* **54(1)**: pp. 159 – 178.

RAINEY, D. W., J. H. YOST, & J. LARSEN (2011): "Disappointment theory and disappointment among football fans." *Journal of Sport Behavior* **34(2)**: pp. 175 – 187.

SAUNDERS, E. M. & JR. (1993): "Stock prices and the wall street weather." *American Economic Review* **83(5)**: p. 1337.

SCHWEITZER, K. & D. ZILLMANN (1992): "Perception of threatening events in the emotional aftermath of a televised college football game." *Journal of Broadcasting and Electronic Media* **36(1)**: p. 75.

TSOUNIS, N., A. VLAHVEI, & C. FLOROS" (2014): "International conference on applied economics, icoae 2014 football and stock returns: New evidence." *Procedia Economics and Finance* **14**: pp. 201 – 209.

WANN, D. L. & K. K. MCGEORGE (1994): "Relationships between spectator identification and spectators' perceptions of influence, spectators' emotions, and competition outcome." *Journal of Sport and Exercise Psychology* **16(4)**: pp. 347 – 364.

ZIVOT, E. & J. WANG (2006): "Unit root tests." *Modeling Financial Time Series with S-PLUS* pp. 111–139.