UNIVERSITY OF LJUBLJANA
FACULTY OF ECONOMICS

MASTER'S THESIS

RETIREMENT SAVINGS PLAN ON STOCKS: INTER-SECTORAL COMPARISON

Ljubljana, September 2016 ROK POTOČNIK

AUTHORSHIP STATEMENT

The undersigned Rok Potočnik, a student at the University of Ljubljana, Faculty of Economics, (hereafter: FELU), author of this written final work of studies with the title Retirement savings plan on stocks: inter-sectoral comparison, prepared under supervision of prof. dr. Igor Masten

DECLARE

1. this written final work of studies to be based on the results of my own research;
2. the printed form of this written final work of studies to be identical to its electronic form;
3. the text of this written final work of studies to be language-edited and technically in adherence with the FELU's Technical Guidelines for Written Works, which means that I cited and / or quoted works and opinions of other authors in this written final work of studies in accordance with the FELU's Technical Guidelines for Written Works;
4. to be aware of the fact that plagiarism (in written or graphical form) is a criminal offence and can be prosecuted in accordance with the Criminal Code of the Republic of Slovenia;
5. to be aware of the consequences a proven plagiarism charge based on the this written final work could have for my status at the FELU in accordance with the relevant FELU Rules;
6. to have obtained all the necessary permits to use the data and works of other authors which are (in written or graphical form) referred to in this written final work of studies and to have clearly marked them;
7. to have acted in accordance with ethical principles during the preparation of this written final work of studies and to have, where necessary, obtained permission of the Ethics Committee;
8. my consent to use the electronic form of this written final work of studies for the detection of content similarity with other written works, using similarity detection software that is connected with the FELU Study Information System;
9. to transfer to the University of Ljubljana free of charge, non-exclusively, geographically and time-wise unlimited the right of saving this written final work of studies in the electronic form, the right of its reproduction, as well as the right of making this written final work of studies available to the public on the World Wide Web via the Repository of the University of Ljubljana;
10. my consent to publication of my personal data that are included in this written final work of studies and in this declaration, when this written final work of studies is published.

Ljubljana, ________________________ Author's signature: _________________________

TABLE OF CONTENTS

TABLE OF TABLES

TABLE OF FIGURES

INTRODUCTION

One of the most important topics in economics recently are demographic changes and their consequences on financial stability of pension system. Especially in developed countries population is getting older and older, and we are close to the point where public pension will not be high enough for a decent retirement. Probably, the most convenient way to solve the problem is to save money during working age. With periodical savings during working ages, it is possible to gain enough funds and hence ensure a financially safe retirement.

There are many strategies and financial instruments that enable us to invest money. We can choose between deposits, bonds, precious metals (gold, silver etc.), real estates, saving in several types of insurance instruments, and investing in stocks. It is widely known that on average, the stock returns are higher than those on any other investment (Diamond, 1999, p. 2).

It is essentially to start investing in early life and to continue with it during the whole working period, because in the long run, we are able to gather quite a lot of funds, despite fairly low periodical contributions. A sufficiently long investing period also enables us to "beat" fluctuations in equity markets.

My master's thesis coincides with implementation of life-cycle funds in Slovenia. Until recently, Slovenian pension companies and other providers of supplementary pension insurance were obliged to invest their funds almost exclusively in deposits, low risk bonds, and other very safe financial instruments. This led to quite low returns of Slovenian pension companies' funds. Life-cycle funds enable supplementary pension providers to invest higher share of their funds into stocks with the main purpose of achieving higher returns.

Nevertheless, this thesis is not applicable only to Slovenian area, but applies also to other countries, which may face a similar demographic and economic situation. Those countries are mostly developed economies.

In the following thesis it is assumed that I am the asset manager of a pension company, dealing with question to which stock sector should pension funds be invested in order to achieve the best result for the pension company's clients and for the pension company itself.

If we assume that stocks are the most profitable in long period, we have to ask ourselves, which stocks we should buy in order to achieve the highest return. In the following thesis I deal with the question, which stock sector is the most profitable and what is the amount of the expected savings at the end of the saving period for an average client of the pension company. Furthermore, I am interested in the probabilities of gaining negative average return during the entire savings period. Quite surprisingly, there are not any existing researches that would deal with stock investing in such way.

The purpose of this master's thesis is to calculate the expected wealth at the end of the savings period for an average client of the pension company. It is assumed that an average client invests 100 € monthly for the entire period of working ages, i.e. 40 years. As an asset manager, I have to calculate, to which stock sector the pension company funds should be invested. I would like to calculate the amount of money that can be expected by the pension company customers. I would also like to compare the expectations among different stock sectors.

A broad set of existing equities is classified into 10 stock sectors (financials, utilities, telecommunications, technology, oil & gas, industrials, health care, consumer services, consumer goods, and basic materials sector). With purpose of a simple comparison it is assumed that asset manager invests the entire funds into only one stock sector for the entire period. As there are 10 existing stock sectors, I have 10 different investment strategies (i.e. one investment strategy for one stock sector). For each investment strategy I calculate the amount of expected wealth for an average client. I am interested in final wealth/savings of the customer at the end of 40 years periodical monthly contributions. Furthermore, I calculate the probability of gaining a negative average return for each strategy, and then I compare the results of the sectors with each other.

The main aim of the thesis is to choose the best econometric model for time series analysis, estimate the model, simulate it 1000-times, and calculate the results (i.e. expected wealth/savings). This is done for each stock sector index. The results can be a good indicator for investment policy of the pension companies. Moreover, the results can also be a guideline for the investors (i.e. physical persons). My goal is to also gain some Matlab programming skills.

In the following thesis there are two hypotheses tested:

Hypothesis 1: There are notable differences in expected returns within sector indices.

According to the literature (Emsbo-Mattingly, Hofschire, Litvak, & Lund-Wilde, 2014), there are some sectors that are more successful in recessions and less successful in expansions, whereas for other sectors the opposite is true. However, it is highly unlikely that all 10 sectors perform the same during 40 years of the investment period.

Hypothesis 2: There is a very low (close to 0) probability to expect a negative average return on the long run periodical investments into stocks.

In line with the literature (Diamond, 1999, p. 2), we know that stocks yield 7% on average per annum; therefore I am allowed to assume that probability to achieve a negative average return during the 40 years should be low.

In line with the research problem specifics, I model the process with the Autoregressive Integrated Moving Average Model (hereinafter: ARIMA model). ARIMA models are usually used for financial time series, which exhibits trend or seasonality, and when the time series data point is dependent on its previous values. With those models, I am able to simulate time series movement many times, with the goal of minimizing effects of time series random movement (extremely positive or negative values). The average of those simulations is the expected wealth that can be anticipated by customer of a pension company.

The structure of my master's thesis is the following: the first and the second chapter present the theoretical background information of the selected topic, based on the relevant academic literature. The first chapter explains demographic trends, pension system, and life-cycle legislation in Slovenia. The second chapter outlines the main concepts of time series analysis, models that are used, and their characteristics. In the third chapter I describe research data and methodology. The fourth and fifth chapters are empirically oriented. The fourth chapter describes the steps of time series analysis, whereas fifth chapter presents results, explains them, and provide implications for the pension company investment policy. I finish the thesis with the conclusion, references, summary in the Slovene language, and appendixes.

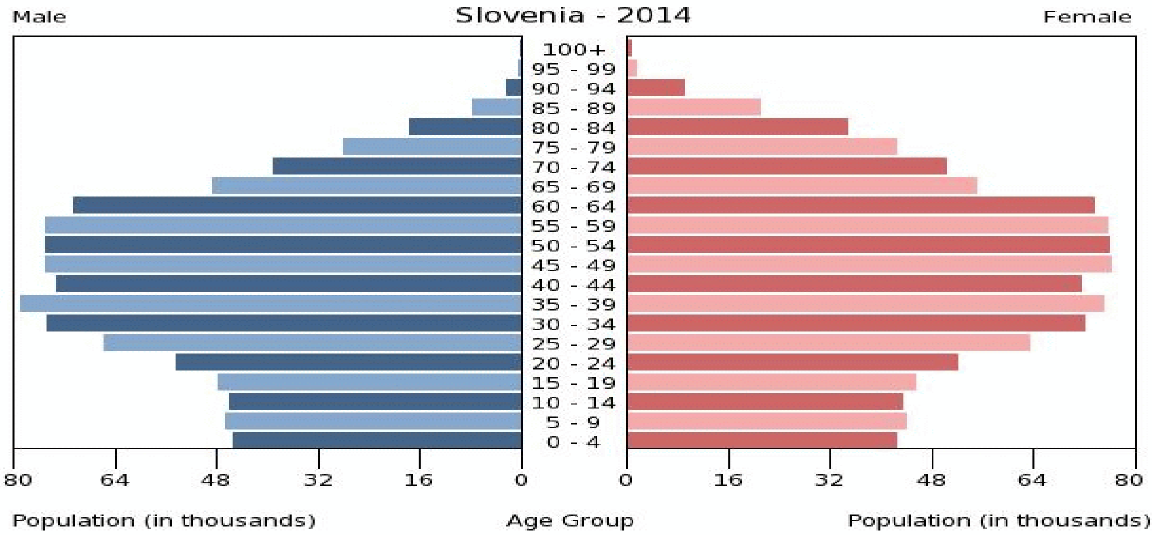
1 DEMOGRAPHY AND PENSION SYSTEM

1.1 Demographic trends in Slovenia

In the next decades, aging of the population and adverse demographic trends will put Slovenia before the challenge of ensuring a long-term financial sustainability of the system, which largely depends on demographic trends (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 5). From Figure 1 we notice that Slovenia has large generations with ages 35 to 60. Obviously, this will put pressure on the pension system in the following decades. Furthermore, extension of life expectancy also has to be considered.

According to the Ministry of labour, family, social affairs and equal opportunities (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 9), in the last decade, the stability and future promises of pension systems has become a very important issue in the Slovenian and also in the European political and public circles. Like most developed western European countries that have a long history of public pension systems, Slovenia is dealing with severe changes in demography, economy and socio-political stability. Changes emphasize the problems of the current pay-as-you-go system, such as weak financial stability, weak economic growth, lack of transparency, and failure to fulfil the equivalence principle. The problem is also lack of motivation among Slovenians to extend their active period.

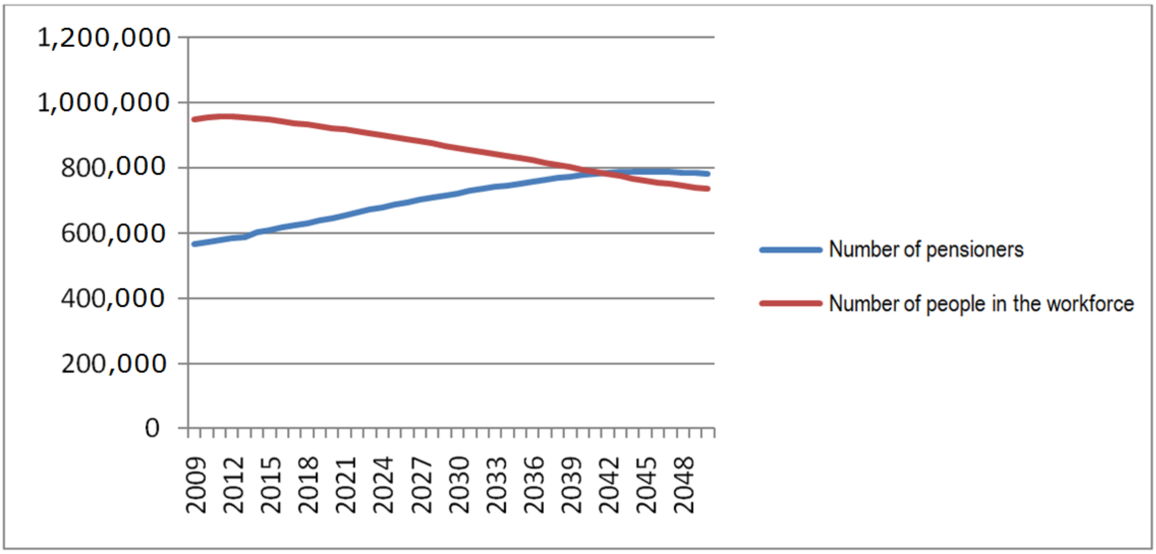
Figure 1: Slovenian population distribution across age groups



Source: Slovenia – 2015, 2016.

In line with previously stated problems, current pension system, which is based on the principle of the intergenerational solidarity (meaning that current working population is paying for current pensioners), will not be able to maintain this principle in the future, especially for generations that will retire in the next few decades.

Figure 2: Estimates on increase in the number of pensioners and decrease in the number of employees



Source: Ministry of labour, family and social affairs, Pension and disability insurance bill (ZPIZ-2), 2010, p. 8, figure 9.

Decreasing number of the working age population and growing share of retired people creates the age/employment paradox. It means that while life expectancy is increasing, participation in the labour market is strongly declining because of the early retirement trends. Economic consequences of the steadily increasing life expectancy can be solved by extending the working ages and by involvement of elderly into economic life. The fact is that the existing pension system is expensive, despite the fact that public pensions are relatively low and even steadily becoming lower (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 5).

1.2 Slovenian pension system

Slovenian pension system has quite a long tradition and it is very well placed within the broad social security system. It is considered as one of the most important sub-systems, ensuring a secure retirement. Ministry of labour, family, social affairs and equal opportunities (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 1) believes that Slovenian pension system offered a satisfactory answer to the broad socio-economic changes.

However, the future will probably be less pleasant. As shown in Table 1, there exists a permanent deficit in the public pension fund. Deficit is permanently financed by the government budget.

Table 1: Contribution of the government budget to public pension fund from 2001 to 2014 (as % of GDP)

Year	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
% of GDP	1.93	2.51	2.46	2.4	2.37	2.19	1.86	1.79	2.19	2.36	2.77	2.88	3.44	3.17

Source: Adapted from Ministrstvo za delo, družino, socialne zadeve in enake možnosti, Učinki pokojninske reforme in nadaljnji koraki, 2014, p. 10, Table 4.

To put percents in numbers: the expenditures of the pension fund in 2014 amounted to more than 5 billon €. More than one third of the sum was covered by state budget.

Currently, the deficit in the pension scheme is covered by the Slovenian state budget. However, such financing leads to the decline of other necessary state expenditures, such as infrastructure, health care etc. State budget has no reserves and the deficit is thus financed by increased borrowing (Ministrstvo za delo, družino in socialne zadeve, 2009, pp. 10–11).

Table 2 indicates that we should not be optimistic when it comes to financial sustainability of the Slovenian public pension system in the next decades, as the state pension fund is assumed to be less and less financially independent.

Table 2: Estimates of a total balance of the state pension fund (in % of GDP) using different assumptions about retirement age

Retirement age	2020	2030	2050
60	-6.0	-8.9	-13.7
65	-1.9	-3.8	-8.1

Source: Adapted from Boris Majcen & Miroslav Verbič, Slovenian pension system in the context of upcoming demographic developments, 2008, p. 10, Table 1.

It has to be mentioned, that the average period of receiving the pension is steadily increasing, due to ageing of the population. Whereas an average male is receiving pension for 16 years and 11 months, an average female is receiving pension for more than 23 years. Moreover, Slovenia has the 2nd lowest percentage of working population in European Union (hereinafter: EU), whose age is between 55 and 64 years. An employee with an average salary during average years of service pays enough contributions only for 17 years of receiving of the pension. Although the government's goal for years of service is 40 years, the real number is significantly lower: the average Slovenian employee works only for 32 years and 7 months before he gets retired (Pokojninska reforma ZPIZ2 splošna pojasnila, 2016).

Despite the increasing deficit in the public pension fund (see Table 1), pension amounts are steadily decreasing. As shown in Table 3, share of the average pension in the average salary has declined by almost 10 percentage points during 2003 and 2012.

Table 3: Share of average pension in the average salary for Slovenia between 2003 and 2012

Year	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
Ratio in %	71.1	70.2	69.1	68.6	67.1	67.1	66.6	64.7	63.4	62.1

Source: Ministrstvo za delo, družino, socialne zadeve in enake možnosti, Učinki pokojninske reforme in nadaljnji koraki, 2014, p. 7, Table 2.

There are two possible solutions to cover the deficit without any structural changes of the pension system. First one is increasing the contributions, but this is not feasible due to the international competitiveness of the Slovenian economy, as the labour tax in Slovenia is among the highest in European Union (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 10). The second solution is the reduction of the pension rights, which is again unrealistic, since 20.3% of pensioners were already under the poverty threshold in 2012 (Intihar, 2013).

The third solution for the public pension fund is to extend the working ages and/or minimum retirement age. This measure is only partially feasible, because prolonging the working ages would lead to even higher unemployment rate of young generations.

Fourth solution is to encourage citizens to save money during working ages and spend these savings during retirement.

Combination of the third and fourth solution is probably the most realistic measure that can be implemented. The following thesis deals with the fourth solution.

With the purpose of getting a better insight into the regulation of the Slovenian pension system, I describe pension pillars that exist in Slovenia. There are three pension pillars, but only one is mandatory.

Pillar 1: The first (alternative name: public or mandatory) is related to employment and is financed by social contributions. It is a state system and is compulsory. The main characteristic that has to be mentioned is that the public pillar is financed by the pay-as-you-go principle. Employees pay 15.5% of gross wage and employers pay 8.85% of gross wage (Guardiancich, 2010, p. 2). Slovenia also has a zero pillar existing since 1999 and it is part of the first pillar. It ensures a state pension. As such, it is universal, and it is unique in the region. The zero pillar provides social safety for labour force that fall out of the public pension system. I do not go more into detail, since my master thesis is focused mainly on 2^{nd} and 3^{rd} pillar.

Pillar 2: The second pillar is based on the individual accounts. It is either voluntary for private sector employees or mandatory for particular working categories and public sector employees (Guardiancich, 2010, pp. 2–3).

Pillar 3: The third pillar includes savings of an individual in pension, and life insurance vehicles. Premiums paid to this third pillar are not subject to tax relief. That is the reason for substantial difficulties of the development of the third pension pillar (Guardiancich, 2010, p. 3).

To sum up, the public pillar has to ensure social security, whereas the second and third pillars are designed to provide or maintain the economic and social status after the retirement. This multi-pillar system allows planning of some standard of living, someone wants to posses when retired (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 12).

Table 4: There Slovenian pension pillars and their types of insurance, providers, and benefits

	1st pillar	**2nd pillar**	**3rd pillar**
Type of insurance	Compulsory pension and disability insurance	Compulsory supplementary pension insurance and voluntary supplementary pension insurance	Different types of insurances and savings
Provider	Pension and disability insurance institute of Slovenia	Compulsory: Kapitalska družba Voluntary: insurance companies, mutual pension funds, pension insurance companies	Banks, insurance companies, mutual funds
Type of income	Public pension	Pension annuity from 2nd pillar	Annuity, single withdrawal

1.3 2nd Pillar and life-cycle funds

1.3.1 General characteristics of the 2nd Pillar

The 2nd pension insurance pillar came into force in 2001. Providers of the supplementary pension insurance services may be either pension companies, banks or insurance companies (Insurance supervision agency, 2015, p. 55).

Exclusively persons with the compulsory insurance (i.e. they are employees in Republic of Slovenia and as such included in the first pillar) are allowed to enter the supplementary pension and disability scheme. The difference between the first and the second pillar is substantial, since the pensioners receive the pension from the current contributions of the insured persons. The second pillar is based on the investment principle, where people save on their personal accounts, and pension companies manage and guarantee for this property (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 7).

Supplementary insurance is divided into two parts; compulsory pension insurance and voluntary supplementary pension insurance.

Workers in especially difficult and harmful working environment are included in the compulsory supplementary pension insurance. The same applies for workers in jobs that cannot be performed successfully after a certain age (e.g. policemen, soldiers, miners etc.). From 2004 on, the compulsory pension insurance is required also for civil servants (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 30).

The purpose of the voluntary supplementary insurance is to preserve the existent financial level of life in such way, that the pension annuity (receiving from the pension company) is able to cover the loss of the individual's revenue, when retired (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 33). The voluntary supplementary insurance can be either collective or individual. In case of the collective insurance, the premium is paid by the employer, while the individual insurance premium is paid by the insured person himself (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 8).

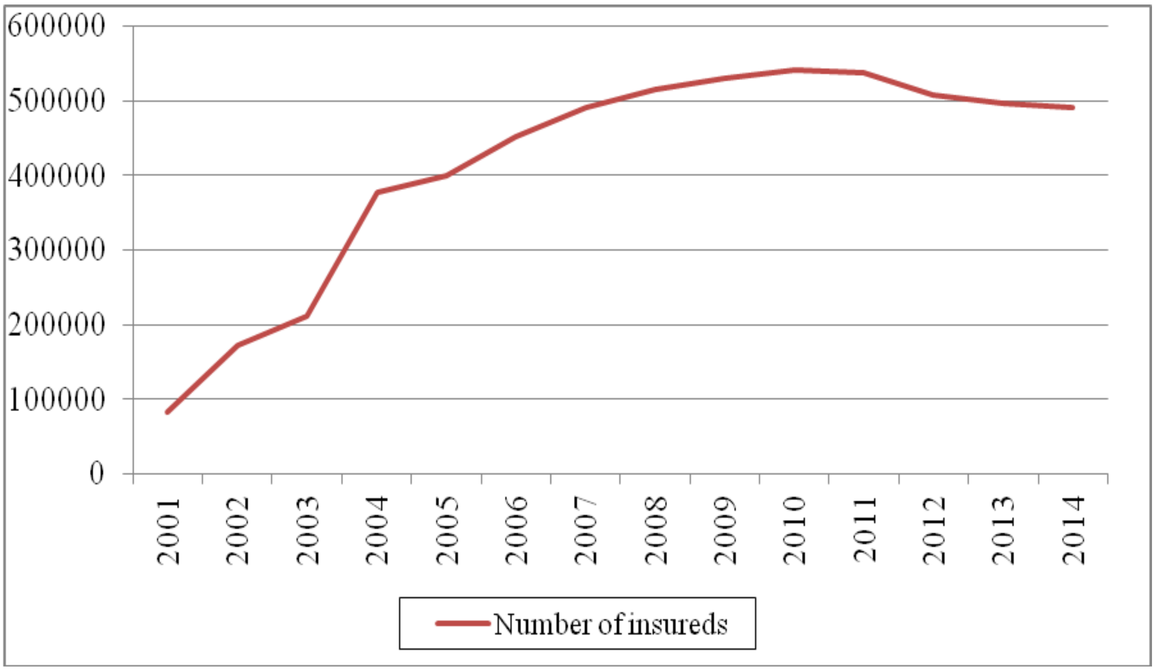
A significant advantage of the supplementary pension insurance is tax stimulation, as tax credit is offered to some amount of contributions. For example, for a company (and employee) it is better to pay monthly contributions to the employer's pension account than to increase his salary (Ekart, 2011, p. 4). The monthly contribution is limited to a maximum of 5.844% of the gross salary, but it should not be more than 234.92 € a month or 2,819.09 € a year. For instance, if the gross salary amounts to 1,559.79 € (average gross salary in Slovenia in January 2016), the maximum premium is 91.09 € per month (Plače, 2016). On this premium, there are no taxes or social contributions. By paying the premium, the company has an advantage in using the tax incentives. If the company pays the maximum premium, the employee cannot additionally contribute without being taxed (Ekart, 2011, p. 4). Premiums paid by companies can be claimed as a tax relief for corporate tax, whereas premiums paid by employees can be used to decrease personal income tax.

When it comes to investment policy, the most important feature is the minimum guaranteed yield. The minimum guaranteed yield is determined by the Ministry of finance twice per year and is relevant for a period of the next six months. It is required that the yield is at least 40% of the yield of government bonds, with maturity of more than 1 year (Ekart, 2011, p. 6). The members take investment risk for the premiums paid. The pension company that does not meet 40% requirement is required to cover the missing amount out of its reserves.

When a company wants to join the supplementary pension insurance, it must make a contract with the employee representatives, usually the trade unions. Currently, 95% of the members are included through the collective insurance by their employers. The remaining 5% of the employees pay contributions individually.

According to Ekart (2011, p. 10) the solution would be to start a true mandatory system, where it would be obligatory for youth to join the second pillar. Ekart claims that the social pension contributions should be in some way divided between the state and the individuals. This would be the only fair solution for the young generations, as it is almost certain that at the time that today's young generation will retire, the ratio between workers and pensioners will be one pensioner per one worker.

Figure 3: Number of insured in 2nd pension pillar across years



Source: Adapted from Dodatno pokojnisko zavarovanje, 2016.

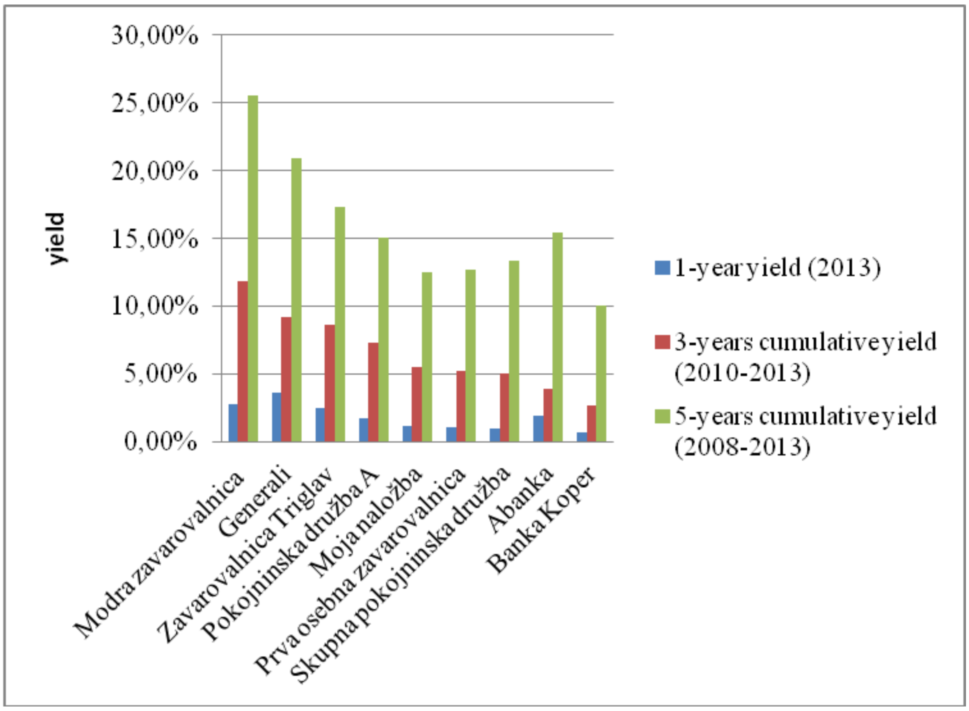
From Figure 3 we can notice that the number of the insured in the supplementary pension insurance is roughly half of a million. According to Statistical office of the Republic of Slovenia (hereinafter: SURS), there are approximately 800,000 persons in employment. This means that 3 out of 8 Slovenians are not included in the supplementary pension insurance (Aktivno prebivalstvo, Slovenija, januar 2016, 2016).

One important thing to point out is the average monthly contribution. It amounts to only 35 € and it is not enough to cover the loss from the constantly decreasing public pensions (Ministrstvo za delo, družino, socialne zadeve in enake možnosti, 2014, p. 7). For this reason, I assume 100 € monthly contribution in the empirical part of the thesis.

1.3.2 Life-cycle funds

Although the tax relief, assured by government, the expectations of supplementary pension insurance to become an alternative way of investing has not been accomplished. Insurance supervision agency believes that the reason lies in rather low returns of such investments in previous years (due to legislation regulation regarding investment strategies) and rather high returns on financial markets in the past (Insurance supervision agency, 2015, p. 55). From Figure 4 it is evident, that the returns in the past were fairly low.

Figure 4: *Yield of Slovenian pension funds*



Source: Adapted from Sredstva javnih uslužbencev za dodatno pokojnino so v dobrih rokah!, 2016.

Ministry of labour, family, social affairs and equal opportunities (Ministrstvo za delo, družino in socialne zadeve, 2009, p. 34) suggested different investment policies, with the purpose of achieving higher long-term returns on pension funds, trying to attract more employees that are out of second pillar. Ministry proposed to leave the current system of the guaranteed return unchanged, but the employees, who are willing to accept higher risk, should get an opportunity that the pension company invests in riskier financial instruments. Those instruments are usually more profitable. Consequently, supplementary pension insurance providers should offer more sub-funds, with different investment policies. The decision, whether to invest in guaranteed sub-fund or in riskier sub-fund, should depend on the preferences of the pension company client. However, a member should be allowed to move his savings to a less risky sub-fund. However, he is not allowed to switch to a riskier sub-fund.

In order to solve the problem with relatively low returns, legislator has implemented a new act. In December 2012, the new Pension and invalidity insurance act (Zakon o pokojninskem in invalidskem zavarovanju, Ur.l. RS, no. 96/12, 39/13, 99/13 - ZSVarPre-C, 101/13 - ZIPRS1415, 44/14 - ORZPIZ206, 85/14 - ZUJF-B in 95/14 - ZUJF-C, hereinafter: ZPIZ-2) was passed and published. It came into direction on 1st of January 2013, introducing several novelties in the voluntary supplementary pension insurance (Insurance supervision agency, 2015, p. 16).

The main ZPIZ-2 novelty was the life-cycle investment policy. Providers of supplementary pension insurance are allowed to offer sub-funds with higher share of stocks. Nevertheless, they must also offer the already existing minimum guaranteed sub-funds, with mainly low risk bonds and deposits as an investments (Insurance supervision agency, 2014, p. 14).

Furthermore, the ZPIZ-2 renamed voluntary supplementary pension insurance in supplementary insurance, provided through pension fund. According to Insurance supervision agency (2015, p. 16), supplementary insurance has three forms:

- Mutual pension fund, with the highest share of stocks.
- Umbrella pension fund, with the medium share of stocks.
- Long term business fund (minimum guaranteed return), with the same investment policy as before ZPIZ-2.

In the below Table 5, I present three Slovenian providers of supplementary insurance and their life-cycle investment policies.

Table 5: Three Slovenian supplementary pension insurance providers and their life-cycle policies with share of stocks across sub-funds

	Dynamic sub-fund	**Prudent sub-fund**	**Guaranteed sub-fund**
Modra zavarovalnica	<50 years 90% stocks	50-60 years 60% stocks	>60 years 10% stocks
Pokojninska družba A	<42 years 55-85% stocks	42-55 years 15-35% stocks	>55 years 10% stocks
Prva osebna zavarovalnica	<45 years 65% stocks	45-57 years 35% stocks	>57 years 5% stocks

Source: Adapted from Modra zavarovalnica, Modri dinamični podsklad: Izjava o naložbeni politiki, 2014a; Modra zavarovalnic, Modri preudarni podsklad: Izjava o naložbeni politiki, 2014b; Modra zavarovalnica, Modri zajamčeni podsklad: Izjava o naložbeni politiki, 2014c; Pokojninski načrti življenjskega cikla, 2016; Skladi življenjskega cikla, 2016.

In line with Modra zavarovalnica (Life-cycle pension fund, 2016), investing in a life-cycle pension funds has several advantages compared to other forms of saving:

- Saving in the life-cycle pension fund reduces your income tax;
- It allows monthly payments;
- It adapts to your age and your needs at different stages of life;
- It allows you to achieve higher returns than banking deposits returns;
- It guarantees the long-term security of your savings.

Practice has shown that saving in funds with stocks investments is not all that risky in the long term, as the fluctuations of economic cycles always level out over the years (Life-cycle pension fund, 2016).

The legislator stipulates (article 260 of the ZPIZ-2) that the purchase fee is allowed to be at the most 3% of each contribution. The redemption fee, stipulated on withdrawal of savings, must not exceed 1% of the saved assets (article 260 of the ZPIZ-2). Article 260 in the ZPIZ-2 provides management fee is allowed to be 1% of pension fund's assets per annum. Article 245 in the ZPIZ-2 provides stipulates that monthly contribution by each supplementary pension insurance member must be at least equal to 20 €.

Before ZPIZ-2, it was allowed to withdraw the savings after 10 years of saving (although tax penalties were high). However, since 2013 the insured has no longer an option to withdraw the saved funds after 10 years (Insurance supervision agency, 2015, p. 17). From ZPIZ-2 onwards it is also impossible to make a single withdrawal of the savings at the time of retirement, when savings amount exceeds 5,000 € (article 206 of the ZPIZ-2). The government is planning to transform the system from single withdrawals of savings to periodical monthly annuities.

1.4 3rd Pillar

The 3rd pillar is based on voluntary savings, consisting of individual savings. A considerable advantage is lack of restrictions and the fact that everyone can join the 3rd pillar. The disadvantage is non-existing tax benefits (Gornjak, 2011, p. 178).

The biggest threat of such saving is lack of individual's discipline. Not saving strictly according to the savings plan could also be problematic. It is obvious that it is mentally easier to pay the compulsory premium than voluntary premium. When considering investments in 3rd pillar, we have to be aware of the taxation on capital gains. Slovenian taxation rate on capital gains is 25% taxation rate and it is considered quite large (Remšak, 2010).

Of course, there are also benefits when investing in third pillar; it does not depend on employment status of a person. The third pillar offers several possibilities for retirement saving – savings in stocks, bonds, mutual funds, life insurances. Advantage is the ability to freely withdraw saved assets and not being obliged to wait until retirement (Remšak, 2010).

My master thesis is linked to both, 2nd and 3rd pillar. Because of assuming investment strategy with 100% of stocks (due to legislation restrictions), it cannot be directly linked to companies providing 2nd pillar supplementary insurance. However, the thesis could be linked to the companies, providing 3rd pillar investments, which have completely flexible investment policy.

2 THEORY: TIME SERIES ANALYSIS AND MODELS

2.1 Generally about time series

Horváth and Johnston (2006, p. 3) describe time series in such way: Time series can be considered as any array of time and numbers that are associated. Nevertheless, usually we think of time series as an order sequence of data points (i.e. values) of variables at equally spaced time periods. Main goal of time series models is to make sense of time series values. Economists use models in order to understand the theory that produces the time series values and to study underlying factors that affect values. The most important questions that time series models try to answer are the following:

- Where did the data come from?
- What are the statistical properties of the data?
- Are there any trends present?

The results are used to fit these models for predictive forecasting and monitoring. Time series analysis is the study of these models and is used in many cases, for example: stock market analysis, budgetary analysis, macroeconomic forecasting, inventory studies, process and quality control, utility studies, and yield projections (Horváth & Johnston, 2006, p. 3).

There are many models employed for time series analysis, however, there are three very comprehensive groups that are usually used, namely, the autoregressive models (hereinafter: AR models), the integrated (I) models, and the moving average models (hereinafter: MA models).

When combining previously mentioned models, we are able to generate new models. For example, the autoregressive moving average model (hereinafter: ARMA model) combines the AR model and the MA model. Another example is the autoregressive integrated moving average model (hereinafter: ARIMA model), which combines all three of the models, previously mentioned. The most commonly used model for time series data is the autoregressive process (Horváth & Johnston, 2006, p. 3).

Similarly, Raicharoen, Lursinsap and Sanguanbhoki (2003, pp. 741–744) define time series modelling as a dynamic research area which has attracted attentions of scientists during previous decades. The main aim of time series modelling is to collect and study past values of a time series in order to develop a proper model, which describes the structure of the time series. The chosen model is then used for forecasting, i.e. to generate future values for the series. Time series forecasting is predicting the future movement, by understanding the past movement of time series.

Historically, developments in time series theory date back to 1920s and 1930s, when Yule and Walker presented the autoregressive model. More developed and advanced theory came into practice in 1970s with a book called "Time series analysis", written by Box and Jenkins, including all steps needed to time series analysis (A brief history of time series analysis, 2016).

As stated by Greene (2002, p. 608), the reputation of Box and Jenkins methodology has raised from the fact, that fairly simple univariate time series models frequently have better forecasting performance than large complicated simultaneous-equations models (for example: two-stages least squares method).

There is another line of development of time series analysis, originating from Box-Jenkins models, dealing with non-constant variance. This class of models are called autoregressive conditional heteroscedasticity (hereinafter: ARCH models) models and generalized autoregressive conditional heteroscedasticity models (hereinafter: GARCH models). Nobel Prize, awarded to Granger and Engle in 2003, proves the significant importance of those models in economics (A brief history of time series analysis, 2016).

According to Adhikari and Agrawal (2013), during the process of choosing a proper model, we have to be aware of the principle of parsimony. Principle states that always the model with the least parameters, as it is statistically admissible, has to be chosen. Out of several suitable models, researcher should select the simplest model; however, the model has to be able to sufficiently describe the characteristics of time series. With selecting a simple econometric model, the researcher avoids the risk of overfitting (Adhikari & Agrawal, 2013, p. 16).

2.2 ARIMA

2.2.1 Generally about ARIMA

The autoregressive integrated moving average model gained popularity by Box and Jenkins in 1970. A linear combination of past values of own time series and a series of errors (also known as random shocks or innovations) forecasts future values. The ARIMA models are used in cases, where time series exhibits non-stationarity, which requires time series differencing (Hamilton, 1994).

The advantage of the ARIMA model is its flexibility to simply describe several types of time series. Moreover, it is linked to the optimal steps at Box-Jenkins methodology in econometrics modelling (Adhikari & Agrawal, 2013, p. 9).

The process is denoted by equation (1):

$$\Delta^D x_t = A_0 + A_1 \Delta^D x_{t-1} + \ldots + A_p \Delta^D x_{t-p} + v_t + \theta_1 v_{t-1} + \ldots + \theta_{t-p} v_{t-p} \tag{1}$$

Where $\Delta^D x$ represent D-th difference of time series, x_t is time series data at time t, x_{t-1} ... x_{t-p} are lagged time series values, A_0 is constant term, $A_1 \dots A_p$ are AR parameters, $\theta_1 \dots \theta_{t-p}$ are MA parameters, v_t is uncorrelated innovation process with zero mean and constant variance and $v_{t-1} \dots v_{t-p}$ are lagged innovations,

Constant. It is important to study, whether to include the constant in the model or not, as the presence of a constant causes a non-zero mean of the original time series (i.e. without differencing). Once the time series is differentiated, the constant causes a non-zero average trend in time series. It causes a non-zero average trend-in-the-trend, when first differences are used twice (in reality it happens quite rarely). Usually we do not assume that there are trends-in-trends, so the constant is removed from models with I (2). When model is I (1) the constant may or may not be present. It depends, whether time series exhibits non-zero average trend. A model without differencing generally includes a constant term (Nau, 2016a).

In line with Robert Nau, the mean and the constant are not equivalent when an econometrics model includes AR coefficients. The mean is the mean of the stationary series. The constant is the A_0 term in the ARIMA equation, which is adequate to the intercept in a regression equation.

$$Constant = mean * (1 - \Sigma \; AR \; coefficients) \qquad (2)$$

From (2) we see that the constant is equal to the mean multiplied by one, minus the sum of the AR coefficients. We should keep in mind, that when one order of differencing has been used, the mean is the trend per period. The mean is the mean of the series itself, in case if differencing is not required (Nau, 2014b, p. 15).

2.2.2 Autoregressive model

Many time series exhibit linear dependence between lagged time series values (i.e. serial autocorrelation) and thus past observations could predict current and future observations. The autoregressive process models the conditional mean of x_t as a function of past observations $x_{t-1}, x_{t-2}, \dots, x_{t-p}$. When p past values (lagged values) have effect on AR process, then we deal with an AR model with the degree p. It is denoted by AR (p) model, dependent on p past observations.

$$x_t = A_0 + A_1 x_{t-1} + \dots + A_p x_{t-p} + v_t \qquad (3)$$

Symbols at (3) are already explained at (1) equation.

2.2.2.1 AR (0) model

Autoregressive model with zero lags (hereinafter: AR (0) model) is a special type of ARIMA models (however, it is also a special part of AR models). It is rather simple, since prediction is such, that the change in the next period will be zero. It means that there is 50% probability that time series value in t+1 period will increase, and 50% probability that it will decrease, no matter how was the movement in the past. The changes are i.i.d. (i.e. independently and identically distributed) in size. The change can be explained no earlier than in t+1, because economics subjects would have already bid it up or down, if the change could have been determined in advance (Nau, 2014c). This process is called random walk without drift model:

$$x_t = x_{t-1} + v_t \tag{4}$$

Second type of AR (0) process is random walk with drift model. Constant A_0 in the equation (5) is included, when the median step is not zero. In such case, forecasts for several periods in future look like a trend line with slope A_0. Forecasts are periodically re-adjusted according to the last observed value. It is not really scientific in terms of the point forecasts we could make ("next period value will be equal to last value, plus average growth"), however it shows how much uncertainty there is in forecasting.

$$x_t = A_0 + x_{t-1} + v_t \tag{5}$$

Random walk patterns are frequently seen in historical stock prices and currencies (Nau, 2014b). Such patterns are also seen in the empirical part of my master thesis.

In the geometric random walk model, the natural logarithm of the variable is assumed to walk a random walk, usually a random walk with drift. That is, the changes in the natural logarithm from one period to the next (which are approximately the percentage changes) are assumed to be independent and identically normally distributed (Nau, 2014c, p. 7).

2.2.2.2 AR (1) model

Autoregressive model with one lag (hereinafter: AR (1) model) is the special type of ARIMA model.

$$x_t = A_0 + A_1 x_{t-1} + v_t \tag{6}$$

where A_0 is constant, A_1 is AR (1) parameter, x_t is current value of time series and x_{t-1} is value one period in past and v_t is assumed to be white noise time series with zero mean and constant variance.

This equation (6) is first order linear difference equation including x_t, which is dependent variable and x_{t-1} , which is explanatory variable. Having $|A_1| < 1$ means, that the model is stable; the further in the past a given innovation arises, the less it influences the present. The innovation term gradually dies out over time. Furthermore, that condition implies that the process has finite variance (Tsay, 2005, pp. 32–33).

When having $|A_1| > 1$, then the model blows up. Nevertheless, also for $|A_1| > 1$ it is possible to obtain a solution, but it is given as a function of the innovations from the future, which does not make any sense (Horváth & Johnston, 2006, p. 6). When having $|A_1| = 1$, we deal with random walk model, described in the previous chapter.

2.2.3 Moving average model

Moving average model is a specific type of ARIMA model. MA model (equation (7)) is another simple process that keeps only information from the q most recent periods. Basically, current value of time series is linearly regressed against current (i.e. v_t) and previously unobserved random shocks (i.e. $v_{t-1} + \ldots + v_{t-q}$).

$$x_t = A_0 + v_t + \theta_1 v_{t-1} + \ldots + \theta_{t-q} v_{t-q} \tag{7}$$

Greene (2002, pp. 609–610) points out that all moving average processes are stationary, unless they have finite MA coefficients. Consequently, whether an ARIMA (or ARMA) process is stationary or not, depends only on the autoregressive part of the model. When ARIMA (or ARMA) process includes moving average term, then ordinary/linear least squares are not consistent and researcher has to estimate by non-linear least squares (Greene, 2002, p. 622).

It has to be explained, why this thesis' empirical research does not include the MA model. As moving average error at period t has affect only on time series value x_t and q periods in the future, it has fairly low predictability strength (q is normally from 0 to 3). Contrary, autoregressive term has influence on all future periods (despite influence limits to zero in the long run).

2.2.4 Order of integration

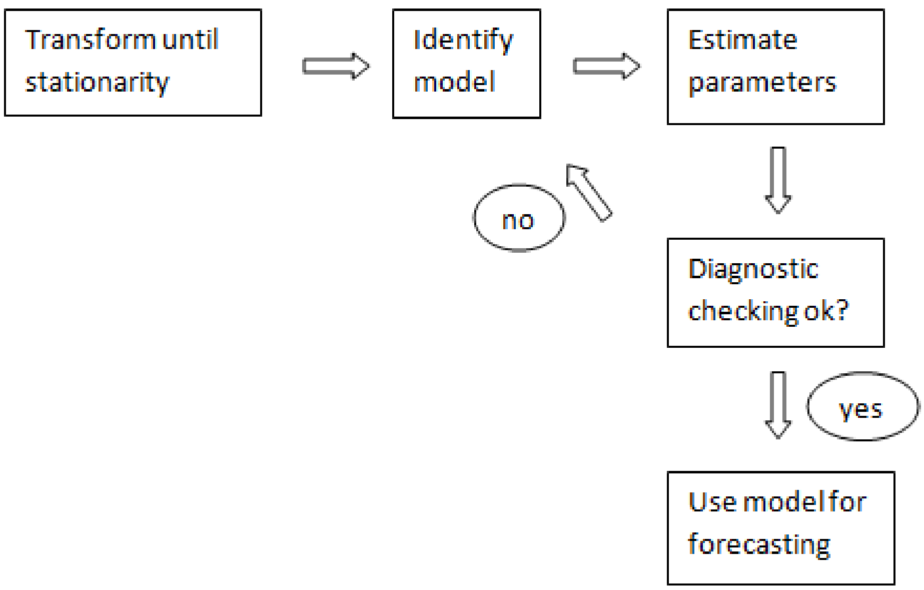
'I' term in ARIMA model stands for 'integrated', and shows how many times original time series was differenced before becoming stationary. When I is equal to 0, original time series is already without trend, so ARIMA model simplifies to ARMA model. An I (1) time series in its undifferentiated format usually exhibits constant growth or it fluctuates around without the tendency to return to mean. The majority of stock market time series and macroeconomics data are I (1). Time series that increases at ever-growing rate has to be integrated twice (i.e. I (2)). Time series, that is integrated more than twice, is extremely rare (Greene, 2002, p. 632). More on this topic is written in the next chapter.

2.2.5 Box-Jenkins methodology

The Box–Jenkins methodology, presented by statisticians George Box and Gwilym Jenkins in 1970, for modelling stochastic processes consists of the following 5 steps:

1. According to Greene (2002, p. 620), at the first step, the researcher is required to transform original time series data to obtain stationary time series (i.e. without trend, seasonality). Stationarity is reached either by taking logarithms, by first differences or both of them.
2. Select the most appropriate number for autoregressive and moving average lags. Plots of autocorrelation function (hereinafter: ACF) and partial autocorrelation function (hereinafter: PACF) of the time series data are used to define number of lags. We take a look at ACF/PACF plots, and the sequential number of lag that has significant drop of ACF/PACF indicates order of MA/AR parameter (Box, Jenkins, & Reinsel, 1994).
3. Estimate the parameters of the ARMA model, either by nonlinear least squares (when MA term is present) or by ordinary least squares (when MA term is not present) or by maximum likelihood estimation.
4. Obtain set of residuals from the previous step and examine, whether they look like white noise time series. If residuals are white noise, then we can move to the final step. If they are not white noise, then we should return to step 2 and redefine model.
5. At this step, forecasting or simulation can be carried out (Greene, 2002, p. 620).

Figure 5: Schematic illustration of Box-Jenkins methodology



2.2.5.1 Log-returns and stationarity

Why using the logarithmic returns? Most financial studies do not involve prices, but rather asset returns. Campbell, Lo, and MacKinlay (1996, p. 298) offer two main advantages of using returns. Return series is easier to deal with than raw price time series, because the return series has more appropriate statistical properties. For an average investor, return of an asset is a complete and scale-free summary of the asset performance.

Short description (8) of an asset return: r_t at time t, where p_t is the price at time t and j = t – 1:

$$r_t = (p_t - p_j) / p_j \tag{8}$$

As explained by Quantivity (Why log returns?, 2015), there are several benefits of using the log returns:

First; log-normality: if we assume, that distribution of prices is log-normal (in reality may or may not be true), then log (1 + r_t) is normally distributed, because:

$$1 + r_t = \frac{p_t}{p_j} = \exp^{\log \frac{p_t}{p_j}} \tag{9}$$

This is useful, as much statistics assumes normality.

Second; approximately raw-log equality: if returns are very low (short periods of time), the approximation (10) ensures they resemble raw returns:

$$\log\,(1 + r_t) \approx r_t\,;\; r_t \ll 1 \tag{10}$$

Third; time-additivity: consider an ordered sequence of n trades. A statistics frequently calculated from this sequence is the compounding return, which is the running return of this sequence of trades over time:

$$(1 + r_1)(1 + r_2)\ldots(1 + r_n) = \prod_1^n (1 + r_t) \tag{11}$$

This equation is quite unpleasant, as probability theory states that the product of normally distributed variables is not normal. However, the sum of normally distributed variables is normal (note: only in case when all variables are uncorrelated), which is useful when we remind log identity:

$$\log\,(1 + r_t) = \log\,\left(\frac{p_t}{p_j}\right) = \log\,(p_t) - \log\,(p_j) \tag{12}$$

Thus, compounding returns are normally distributed. Finally, this equation brings us to a simple formula for calculating compound returns:

$$\sum_i \log(1+r_t) = \log(1+r_1) + \log(1+r_2) + \ldots + \log(1+r_n) = \log(p_n) - \log(p_0) \quad (13)$$

Consequently, the compound return over n periods is the difference between initial logarithm and the last logarithm. This is especially useful when n is large. Moreover, this sum is applicable, when returns differentiate from normality, since central limit theorem proves that the average sum of the sample will converge to normality.

Fourth, mathematical ease:

$$e^x = \int e^x \, dx = \frac{d}{dx} e^x = e^x \quad (14)$$

A lot of financial mathematics assumes continuous time stochastic processes and (14) is often useful.

Fifth; numerical stability: addition of small numbers is numerically safe, but multiplying is not, because it is subject to arithmetic underflow. To solve this problem, it can be transformed into summation by logarithms.

Stationarity. A stationary time series differs from a non-stationary (i.e. trending). While stationary time series has constant mean and variance over time, non-stationary does not. We achieve stationarity (note: approximation of stationarity) with mathematical transformations. According to statistical methods, the approximation of stationarity is enough to forecast. Stationary time series is easier to interpret than a non-stationary one, because the researcher is allowed to assume that time series statistical properties will stay the same in the future (Nau, 2016b).

Determining the right combination of transformations, needed to achieve stationarity (logarithms or/and first differences) is one of the most important steps in econometric modelling.

It is pre-requested, that time series is stationary, when the researcher wants to find meaningful time series characteristics (i.e. mean, variance and correlations with other variables). If it is not, the characteristics are not correct. For illustration; in case of steadily increasing time series, the sample variance and mean constantly grow with increasing the sample size. Consequently, future sample mean and variance are always underestimated. As a result of incorrect mean and variance determination, correlation with other variables arises (Nau, 2016b). That is why we should ensure time series stationarity before progressing to other steps in econometrics modelling.

In reality, the greater part of economics and financial time series are not stationary, when expressed in original units (for example: unit of mutual fund or index value). Generally, they have trends, random walks, seasonal behaviour (e.g. quantity of beer bottles sold across months) and/or cycles (Nau, 2016b).

We have two types of stationarity; weakly stationarity (alternative name is covariance stationary) and strictly stationarity. Anyhow, financial literature usually assumes that an asset return time series has to be weakly stationary (Tsay, 2005, p. 33). According to Greene (2002, p. 612) a process is weakly stationary when it satisfies the following requirements:

Expected time series value is independent of time t: $E(x_t) = \bar{x}$ for all $t \in T$

Variance is independent of time, with positive and finite constant: $Var(x_t) = \sigma$

Covariance $cov(x_t, x_s)$ is finite function of $|t\text{-}s|$

The equations above ensure weak stationarity, whereas concept of strict stationarity could be used. It requires all joint distributions of $(x_t, \dots, x_{t-s})$ to be time invariant for any integer s (Masten, 2012).

How can we test our time series for non-stationarity?

To recall: our financial time series is probably non-stationary. We can examine stationarity either by several statistical tests or visually, by looking at sample autocorrelation function (LeSage, 1999, p. 115).

As stated by Masten (2012), some of the several tests are:

- Augmented Dickey-Fuller test (hereinafter: ADF test),
- Kwiatkowski, Phillips, Schmidt and Shin test,
- Phillips-Peron test,
- Hegy test.

Alternatively, we can plot sample autocorrelation function (hereinafter: ACF). Quick drop of ACF indicates stationary process, while gradual decline of ACF indicates non-stationary process. Another indicator for non-stationarity is positive and large (close to 1) autoregressive coefficient value (Stationarity and differencing, 2016). Using non-stationary time series data in financial models produces unreliable results and leads to poor understanding and forecasting.

The solutions for dealing with non-stationary time series are differencing and taking the (natural) logarithms. Transformations, such as logarithms, may decrease the variance of a time series. Differencing stabilizes the mean of a time series by removing changes in the level of a time series and therefore remove trend and seasonality. Differencing (or first differences) is the difference between consecutive observations (Greene, 2002, p. 649). As stated by Greene (2002, p. 631), when using differencing we say that non-stationary series is integrated of order d, if it becomes stationary after taking first differences transformation d times. It is denoted by I (d). In equation (15) I present equation for the logarithmic transformation and first differences.

$$\Delta \ln(x_t^n) = \ln(x_t^n) - ln(x_{t-1}^n) \quad for \quad t = [1,2, \ldots 513] \tag{15}$$

Most macroeconomic flows and stocks that relate to population size, such as output or employment, are I (1), meaning that first differences have to be taken once. Such series in its original (non-differenced) form constantly increases or wander around, having no tendency to return to a mean (Greene, 2002, p. 623).

A model with no orders of differencing assumes that the original series is stationary. A model with two orders of differencing assumes that the original series has a time-varying trend (Nau, 2016b).

2.2.5.2 ACF, PACF and p, q selection

According to Nau (2014b, p. 4), the next step in Box-Jenkins methodology is to determine the number of lags for AR and MA parameter that should be used in the stationary time series x_t estimation. One alternative is just trying some standard combinations of AR lags (denoted by 'p') and MA lags (denoted by 'q'). Nevertheless, systematic method for lags determination exists. It is built upon studying the ACF and PACF of time series x_t.

The autocorrelation of x_t at lag p is the correlation between x_t and x_{t-p}, i.e., it is the correlation between x_t and itself, lagged by p periods (Nau, 2014b, p. 4).

The partial autocorrelation between x_t and itself, lagged by p periods (x_{t-p}), is the autocorrelation between x_{t-p} and x_t, subtracted by a part, explained linearly by the intervening lags (Greene, 2002, p. 617).

At the beginning of this step, we have to plot ACF and PACF of stock sector time series x_t. ACF plot should be used for determination of moving average lags, whereas PACF plot should be studied for selection of autoregressive lags.

Researcher should examine, at which lag PACF plot cuts off. For example, if PACF is significantly different from zero at lag p-1, whereas at lag p (and for following lags) becomes non-significant, then researcher should consider, if autoregressive term has p lags.

Similarly holds for ACF plot, which helps to (with the same logic as above) determine the lag order for moving average term.

When having only statistically significant PACF and non-significant ACF, we should consider AR (p) model, which is sub-type of ARIMA. While facing statistical significance only in ACF plot, we should consider MA (q) model. Third option is to have serial dependence, both in autocorrelation and partial autocorrelation plot. In such case, we consider ARIMA (p, d, q) model (Nau, 2014b, p. 4).

Nau (2014b, p. 8) suggests another useful hint: we should not take into consideration isolated spikes in the PACF and ACF plots after third lag. We have to focus on the first few lags, while after that, we should examine, whether there is any systematic pattern in ACF and PACF plots. However, this hint only holds, when we are dealing with non-seasonal time series data.

In the Table 6 is presented the summary of the theory stated above.

Table 6: Summary of properties of ACF and PACF for autoregressive, moving average and ARMA models

	AR process	**MA process**	**ARMA process**
ACF	Tails off gradually	Cuts off after lag *q*	Tails off gradually
PACF	Cuts off after lag p	Tails off gradually	Tails off gradually

Source: Adapted from Robert Nau, Notes on nonseasonal ARIMA models, 2014b.

There exists another option, namely corrected Akaike information criterion (i.e. AICc) (Brockwell & Davis, 2009, p. 273). Alternatively, Bayesian information criterion and Akaike information criterion can also be used. Due to space limitation, I do not go into detail.

2.2.5.3 Parameter estimation

Model is estimated on the basis of T - p observations, where T stands for length of time series (basically, number of observations) and p is lag order. This is especially important when we deal with short time series, because p values get lost.

When residuals are normally distributed, ordinary least squares estimation is equivalent to maximum likelihood estimation (Greene, 2002, p. 588).

When ARIMA model does not include MA term, model can simply be estimated by ordinary least squares (linear least squares), using previously determined p (number of lags) and d (order of differencing) values. On the other hand, researcher can face ARIMA model with the included MA term. In such case, ordinary least squares should not be used, because the residuals are not known until the model is fitted (i.e. it is not possible to specify residuals as an independent variable) (Nau, 2016a). Consequently, non-linear least squares have to be used.

One part of estimation is to verify, whether AR and/or MA coefficients are statistically significant. When we choose correct values of p, d and q, as a part of ARIMA model, then the highest order of autoregressive and/or moving average coefficient should have a t-statistics either greater than +1.96 or lower than -1.96 and corresponding p-value lower than 0.05 (Nau, 2014b, p. 4). Those numbers are normally assumed in economic statistics.

Nau (2014a, p. 6) warns us that fitting an ARIMA (2, d, 1) model that is in reality ARIMA (1, d, 0), the coefficient estimates are not unique.

Contemporary statistical packages allow us to avoid manual differencing time series values. In line with MathWorks (Econometrics toolbox: trend-stationary vs. difference-stationary processes, 2016) we can use, in Matlab software, raw (before transformations) values and estimate an ARIMA model with previously determined order of differencing (note: also seasonal adjustment, when we have seasonality). In comparison to the alternative (manually difference time series and estimate ARMA model), we benefit from results, which are returned on original scale.

The estimation adequacy will be examined in the next step. Estimated residuals should resemble the white noise process (Greene, 2002, p. 622).

2.2.5.4 Residual diagnostics

Tsay (2005, pp. 31–44) gives an advice that estimated model should be carefully examined in order to check for possible model inadequacy. In case of no misspecifications, the residual series should resemble a white noise process. White noise is a sequence of independent and identically distributed random variables with finite mean and variance.

Residual autocorrelation. Whereas in classical regression models residual autocorrelation leads to inefficiency of ordinary least squares estimator, in all dynamic models (also in ARIMA model, as part of dynamic models) autocorrelation causes inconsistency.

In order to prove it (Masten, 2012), consider the simplicity AR (1) model (it holds in general):

$$x_t = A_0 + A_1 x_{t-1} + v_t \tag{16}$$

with autocorrelated residuals:

$$v_t = \rho v_{t-1} + \varepsilon_t \tag{17}$$

Hence, error term is correlated with the regressor, which leads to inconsistency of OLS estimator. This problem is called "omitted variable bias", which states that our model leaves out one or more important factors.

To prove this, we subtract ρx_{t-1} from x_t and rearrange to:

$$x_t = A_0 + (A_1 + \rho)x_{t-1} + A_1\rho(x_{t-2}) + \varepsilon_t \tag{18}$$

If we notice, that residual autocorrelation exists; we are allowed to assume that our model can be improved. Selecting higher autoregressive lag order is usually a solution for this problem.

MathWorks (Residual diagnostics, 2016) recommends testing residual autocorrelation with two approaches; first one is informal and consists of plots of ACF and PACF functions on residual time series. The formal approach is to conduct Ljung-Box test on residual time series to obtain critical value and the corresponding test statistics.

Residual homoscedasticity. A white noise innovation process must have a constant variance. When errors exhibit constant variance, we say that they are homoscedastic, if variance is volatile, then errors are heteroscedastic (19), meaning that they vary across time. When residuals are heteroscedastic, ordinary least squares are still unbiased, but no longer efficient, because variance is underestimated.

$$\sigma_t^2 = \gamma_0 + \gamma_1 v_{t-1}^2 + \ldots + \gamma_q v_{t-q}^2 \tag{19}$$

Again, we can evaluate this problem with several approaches; informal test is to plot ACF and PACF on squared residual time series, whereas formal testing is done with the Engle's autoregressive conditional heteroscedasticity test (ARCH test). When heteroscedasticity exists, we have to use the combination of ARIMA model and generalized autoregressive conditional heteroscedasticity model (GARCH model). ARIMA is applied for conditional mean, whereas GARCH is used for modelling conditional variance. Alternatively, homoscedasticity can be tested by Ljung-Box test (used also for autocorrelation test) applied on squared residuals (Residual diagnostics, 2016).

However, empirical economists spent much less time trying to model the exact nature of the heteroskedasticity in their data sets than the autocorrelation (Barreto & Howland, 2005).

Normality distribution. In line with Nau (2016b), we have a variety of statistical tests for normality; the Jarque-Bera test, the Shapiro-Wilk test, the Kolmogorov-Smirnov test and the Anderson-Darling test. Normality distribution of residuals (20) (Greene, 2002, p. 17) is not the most important thing to verify. Time series data, with perfectly normally distributed residuals are fairly rare, therefore it may be impossible to fit the data to the model, whose errors are normally distributed at 5% confidence level. It is more appropriate to focus on other assumptions and only check, whether there are any extreme values (i.e. outliers), that causes non-normality. In such occasion it is wise to use common sense and think about, whether the problem is systematic or not (Nau, 2016b).

$$v|X \sim N[0\,,\,\sigma^2 I] \tag{20}$$

Here a theorem has to be mentioned, which is often useful in real empirical problems: the central limit theorem. It states, that when we deal with a large dataset, normality (or non-normality) is less important. The theorem claims, that identically and independently distributed variables are approximately normally distributed (Siegrist, 2015).

When dealing with excess kurtosis (i.e. forth moment), we should consider assuming a Student's t residuals distribution (Residual diagnostics, 2016). Excess kurtosis is present, when the plot of residuals shows fatter tails.

2.2.5.5 Forecasting/simulation

If all previous steps are done correctly, then we arrive to the last step, which is forecasting or simulation of our selected model. Consider for simplicity an AR (2) model and MA (2) model.

AR (2) model:

$$x_t = A_0 + A_1 x_{t-1} + A_2 x_{t-2} + v_t \tag{21}$$

Forecast one period ahead:

$$x_{t+1} = A_0 + A_1 x_t + A_2 x_{t-1} + v_{t+1} \tag{22}$$

MA (2) model:

$$x_t = A_0 + v_t + \theta_1 v_{t-1} + \theta_2 v_{t-2} \tag{23}$$

MA (2) model (written with backshift B operator):

$$x_t = A_0 + (1 + A_1\mathrm{B} + A_2\mathrm{B}^2)v_t \tag{24}$$

Forecast one period ahead:

$$x_{t+1} = A_0 + (1 + A_1\mathrm{B} + A_2\mathrm{B}^2)v_{t+1} \tag{25}$$

As evident from (22), the AR forecasts are linear function of past data and coefficients, whereas Nau (2014a) warns us that MA forecasts (25) are a non-linear function of coefficients. That is the reason that ordinary (i.e. linear) least squares should not be used in the presence of MA term.

If the time series values are transformed manually (for example, with logarithms and first differences) before estimation, we have to undo the same transformations at this step in order to get values compared to original time series (Nau, 2014b).

Greene (2002, p. 610) claims that models with relatively small p and q values are very effective as forecasting models.

2.2.5.6 Information selection criterion

In statistical analysis it is common, that more models resemble the data to a similar degree. In such cases, information selection criterion may help us to choose the right one.

One popular criterion is the Akaike information criterion (hereinafter: AIC), introduced by Akaike in 1973. AIC is used, when the researcher starts with several candidate models that seem reasonable. Then each model is estimated and finally, AIC value is calculated. This criterion positively evaluates goodness of fit and penalizes number of parameters in model (penalizes complexity of the model). Equation for AIC is the following:

$$AIC = -2 \times log\ (L) + 2 \times K \tag{26}$$

Where K represents the number of parameters in the model, and log (L) is the maximum value of the log-likelihood function. It has to be pointed out, that the natural logarithm must be used. The multiplier 2 is included for historical reasons and it does not have any significant meaning. When we increase the number of parameters in the model (i.e. K increases), the log-likelihood also increases, but the value of -2 log (L) becomes even more negative. The model with the lowest AIC value has the best trade-off between the goodness of fit and complexity (Schwarz, 2011, pp. 25–27).

Another popular criterion function is the (Schwarz, 2011) Bayesian information criterion (hereinafter: BIC). The equation is the following:

$$BIC = - 2 \times log\ (L) + K \times ln\ (t) \tag{27}$$

We see that the first part of the formula is the same as in AIC, whereas it differs in the second part: K is the number of parameters, which is multiplied by natural logarithm of the number of observations. According to Tsay (2005, p. 42), we can infer that BIC tends to select a lower lag order selection in comparison with AIC, when sample size is not small (as we see from the (26) and (27). AIC penalizes each parameter with 2, whereas BIC with the log of number of time series observation; which is larger than 2, when model consist of at least 8 observations.

We can use AIC and BIC in time series modelling for several purposes; p and q lag order selection, and for comparison of estimated models.

There are more information criterion tests; one of them is the Hannan-Quinn information criterion. Due to space limitations, I do not go into more detail.

2.3 Stock sectors and cyclical vs. non-cyclical stocks

Before moving on to the empirical research, I would like to explain the basis of classification of stocks. There are quite a few classifications; however, I am the most interested in the classification of stocks into sectors.

Since my data is obtained from the Thomson Reuters database, it is the most appropriate that I use the Thomson Reuters sectors classification. There are 10 stock sectors; namely utilities, telecommunication, technology, industrials, health care, basic materials, financials, consumer services, consumer goods, and oil & gas sector (Thomson Reuters, 2012).

Those stock sectors are classified into two broad groups; cyclical and non-cyclical stocks. The latter can be also called defensive stocks.

A cyclical stock price is highly correlated to the economic cycles. When economy is in recession or depression, cyclical commodities producing company usually face lower profits and consequently their share price decreases. On the other hand, during expansion phase, share prices tend to increase. Good example is the stock of automobile producer. An individual is more willing to buy a new car when economy is in good shape; hence, share price probably increases (Cyclical vs. Defensive Stocks, 2016). There are 5 sectors that belong to the cyclical group of stocks: technology, industrials, basic materials, financials and consumer services sector index.

A non-cyclical (defensive) stock price has a very low correlation with the economic activity. With a little simplification we can say that company revenues and profits remain relatively stable, disregarding the economic cycle. Good example is stock of health care company. Even in recession, an individual is willing to buy its products. Defensive stock price is assumed to be less volatile (Cyclical vs. Defensive Stocks, 2016). There are 5 sectors that belong to cyclical group of stocks: utilities, telecommunications, health care, consumer goods, and oil & gas stock sector.

3 DATA AND METHODOLOGY

3.1 Data

Before starting with the empirical part, I still need to find the most appropriate data values. I use statistical database Datastream professional (hereinafter: Datastream), version 5.1., provided by Thomson Reuters. The Thomson Reuters Datastream is used, because it provides the longest time series values for the needs of the thesis.

In the empirical analysis I separate stocks into stock sectors indices. Datastream database has 10 different stock market sectors indices; financials, utilities, telecommunications, technology, oil & gas, industrials, health care, consumer services, consumer goods, and basic materials index. The number of companies in specific sector is 226, 52, 10, 106, 81, 159, 90, 130, 102, and 36, respectively.

Indices, provided by Datastream consist of 514 monthly observations (42 years and 9 months), starting on 2nd of January 1973 and ending on 1st of October 2015. However, it would be much better to deal with longer time series, but such time series values do not exist. Monthly data is the best choice, since savers are assumed to invest on monthly basis.

The series index value always starts with the number 100. Dividends are assumed to be immediately reinvested (i.e. not paid out) and included into values of indices. I also have to point out, that my data is provided by single research company and collected with the same statistical methods, in order to get the most comparable statistics.

Geographically, I focus on the United States (hereinafter: U.S) stocks, since U.S. stock market data are more complete than those for any other country (Cornell, 1999, p. 60). In addition to that, Datastream professional offers either world stock sector index values or European stock sector index values. However, I decide to deal with the U.S. stock-market data, as it provides the longest and the most comparable time series data. Comparability allows an easier analysis of several time series, whereas a longer time series span provides us with more information and we are able to select a model, which is the best fit.

Values of Datastream time series indices are nominal, meaning that inflation is not considered. Consequently, I have to subtract inflation from nominal returns in order to get real returns. As we already know, positive inflation decreases nominal returns, whereas negative inflation (i.e. deflation) increases nominal returns, since deflation allows consumer to buy more goods, with the same amount of money.

I assume investing in United States stocks and hence United States inflation has to be taken into account. Monthly consumer price index from U.S. department of Labor, Bureau of Labor statistics (Crawford, Church, & Akin, 2016) is used.

3.2 Methodology

After collecting the data, as presented in chapter 3.1, I am able to move to the core of the thesis. Before describing methodology into more detail, I would like to explain my research problem. As an asset manager of a pension company, I am interested into historical profitability of different stock market sectors. Obviously, we do not know what the future holds, hence the only possibility to prepare forecasts is to study and to understand the past time series movement. Based on the understanding of the past, I am able to make better investment strategies for company's clients.

As an asset manager, I need information about amount of savings that a customer of a pension company can expect at the end of the investment period (i.e. expected wealth). It is assumed that all pension company assets are placed in stocks. Stocks are classified into 10 different sectors. It is assumed that a pension company invests its funds into single stock market sector for the entire 40 years. I would like to know, what is the expected wealth for an average customer of a pension company for each stock sector separately. Results are compared across stock sectors. In addition to expected wealth, I am interested in probabilities that an average customer end savings period with negative average return on his contributions (i.e. he end with lower amount of savings that he invested during savings period). Knowing such probabilities, helps me to determine a successful investment strategy for a pension company fund, since customers would be really disappointed with a negative average returns. Probabilities are calculated for each stock sector separately, and compared with each other.

Savings period is assumed to last for 40 years, the same as the duration of working ages in Slovenia. Customer's financial contribution is set at 100 € monthly. Hence, 480 contributions are paid during savings period (40 years with 12 contributions each year). Monthly contribution is assumed to be in real terms (after considering inflation). All together, 48,000 € are assumed to be invested by the insured to the pension company during savings period (480 contributions, each one amount to 100 €). At the end of the savings period it is assumed that the customer withdraws his savings.

It has to be mentioned that it is not allowed to withdraw the savings by the insured before the end of 40 years of savings period. By the assumption, savers are not obliged to receive dividends, earned by pension company's stocks. Dividends are rather immediately reinvested by the pension company.

Savings in the pension fund are charged by two fees; management fee and administrative fee. Slovenian legislation states that maximum permitted management fee is 1% per annum, charged by a pension company on its pension fund. Administrative fee is set at 1% of each contribution, and it is charged by a pension company on monthly basis (Zakon o pokojninskem in invalidskem zavarovanju, Ur.l. RS, no. 96/12, 39/13, 99/13 - ZSVarPre-

C, 101/13 - ZIPRS1415, 44/14 - ORZPIZ206, 85/14 - ZUJF-B in 95/14 - ZUJF-C). Those two fees are the main source of revenue for a company and its financial stability.

Especially the management fee is of significant importance. This is, because the first contribution of insured gets charged with management fee 480 times (note: because it is invested for 40 years/480 months) with the rate of 1/12 * 1% each month. This means that out of 100 €, invested at the beginning of savings period, 40 € present revenue of a pension company.

The reason for not taking costs of each purchase into account has to be explained (from the perspective of a pension company). Simplification is that a pension company collects all contributions of its clients each month, and the complete amount of money is invested into specific stock sector. Investments can be placed into one of many sector exchange traded funds (ETFs). When doing so, brokerage fee occurs. Assumption states that pension company (or bank or insurance company) has many clients that pay 100 € premiums monthly. Hence, a large amount of money gets invested. Brokerage fee, which is charged, is relatively small in comparison to amount of money that is invested monthly by a pension company (fees are usually fixed, disregarding amount invested). Consequently, a fee does not really affect the expected wealth of average clients and thus can be neglected.

To sum up, the task is to calculate the expected final savings for an average customer of a pension company after 40 years of saving with the same investment strategy, as stated above (all contributions are invested in single stock sector index, the same amounts in real terms etc.). Furthermore, probabilities of gaining negative average return (i.e. end with less than 48,000 €) are also in my interest. Expected wealth and probability of negative average return have to be calculated for each stock sector separately. As set of stocks is divided into 10 sectors, there are 10 expected savings and also 10 probabilities for negative return (each stock sector has its own expected wealth and its own probability for negative return).

Analysis has to appropriately deal with random fluctuations in stock market prices in order to get unbiased expected wealth. Economic cycles have to be taken into account. It depends, when a customer starts with contributions to pension company funds. For example, when an average customer starts with saving at the deep recession (stock prices are almost certainly are low), he can buy a lot of stocks (with the same amount of 100 € can buy more stocks, as they are cheap) that are undervalued. Hence, he can expect to gain more than average returns in the future. Contrary, when an average client starts with saving at the expansion (stock prices are almost certainly high), he can only buy few stocks (with the same amount of 100 € he can buy less stocks, as they are expensive). Hence, he can expect to gain less than average returns in the future.

To solve the problem regarding economic cycles, I simulate the process for 1000 times. 1000 simulations are enough, that random movement in stock prices is appropriately dealt with. With calculating an average out of 1000 simulation, I can assume that expected wealth and probability of negative returns is unbiased.

Problems, mentioned above can be solved by the correct choice of econometric models. There are several different models that can be used for calculation. However, for financial time series, it is common to employ autoregressive integrated moving average model (ARIMA model). The following thesis also uses ARIMA model for analysis, simulation and forecasting. Empirical part is modelled in Matlab software, which helps us to efficiently deal with large sets of data.

Despite ARIMA models are quite easy to understand, they are still able to provide quality results. ARIMA models are appropriate to estimate time series and based on estimation they can be used for forecasting or simulation.

ARIMA modelling methodology was introduced and presented in 1970 by Box and Jenkins, hence procedure is called the Box-Jenkins methodology. The methodology starts with examination, whether time series are stationary. If they are not, transformations have to be carried out on original time series. Typical transformations are first differences or/and natural logarithms. However, a new version of Matlab allows us to estimate the model with non-stationary time series. I follow the Box-Jenkins methodology and time series are manually transformed.

With plotting partial autocorrelation function (PACF) and autocorrelation function (ACF) I choose models that could be the most accurate. The next step in methodology is the estimation of the models with Matlab software. After estimation I check, whether the residuals of the estimated models are non-autocorrelated, homoscedastic and normally distributed. Statistical significance of the parameters also plays very important part for the final choice of the model.

With the analysis, stated above I can choose the best econometrics model, which is used for simulation and calculation of the results. The most appropriate model should not be complicated, it should have statistically significant parameters and residuals should be non-autocorrelated, having stable variance across time and they should be approximately normally distributed.

When the model is defined, I can move to the core of the empirical part of the following thesis: ARIMA simulation. Matlab software offers possibility to simulate entire process for specified number of times; in this thesis 1000 times. Monthly returns are chosen randomly, out of historical returns, for each simulation separately. In simulation step I must consider both, management and administrative fee. Moreover, all calculations have to be mathematically valid.

Each simulation gives me one expected saving, hence, with 1000 simulations I get 1000 expected savings for a customer of a pension company for each stock sector. In order to get single value of expected wealth, I simply calculate an average out of 1000 expected savings. This is done for 10 stock sectors.

After that I need to calculate, what are the probabilities for a pension company customer to finish the savings period with a negative average return on his contributions. Calculus is fairly simple: I count how many simulations, out of 1000 previously calculated simulations, have expected wealth below 48,000 €. To recall, an average client is assumed to invest 48,000 € during savings period, hence savings below this amount lead to conclusion that returns are negative. Number of such simulations is divided by 1000 simulations and probability is calculated. This is done for each of 10 stock sectors. Final part of empirical analysis is commentary and explanation regarding results and implications for pension company investment strategy.

One important note, regarding the methodology is the following: Slovenia is part of the European Union and uses € as a currency. Anyhow, purchases are not made in Euros. Investments are assumed to be in U.S. stocks, and payment has to be in U.S. dollars. This does not significantly affect my results, because statistics of European central bank (Euro exchange rates USD, 2016) demonstrates that average exchange rate between Euro and dollar (since Euro implementation on 4th of January 1999 until 4th of May 2016) is 1.216 USD/EUR, which is close to exchange rate at the time when Euro was implemented. Euro dates back to the beginning of January 1999, when exchange rate was 1.179 USD/EUR. Consequently, I am able to assume, that the fluctuations in exchange rate during 40 years of savings period can be ignored, since savings period is long enough that random fluctuations of exchange rate do not significantly influence the results.

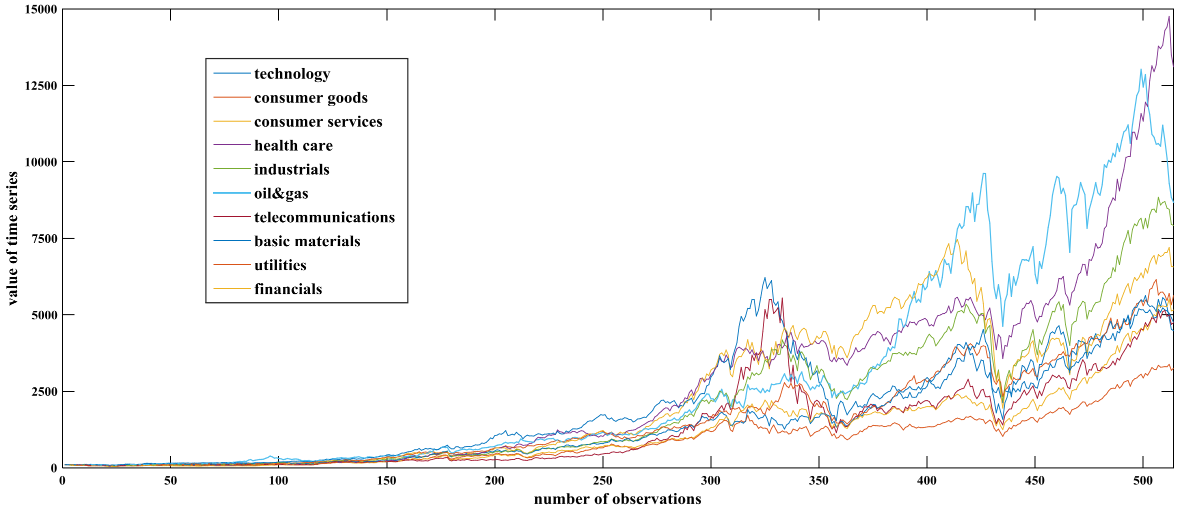
4 EMPIRICAL ANALYSIS

4.1 Stationarity and inflation

4.1.1 Stationarity

After downloading 10 stock sector indices from Datastream software I have to transform each of the indexes until it becomes stationarity. Figure 6 displays stock indices movement during observed period of 514 months (42 years and 9 months). Starting point of each index is value of 100 and itself does not mean anything. It is just an index number allowing time series comparison.

Figure 6: 10 stock sector indices on original scale, 1973-2015



Interesting detail seen in Figure 6 is that indices were half of the time (first half) very stable, with roughly linear and moderate increasing trend. Significant volatility has arisen not earlier than in last few decades. We can see huge stock market crash in the last financial crisis in 2008 (approximately 430th observation) and also crash after dot-com bubble in 2000.

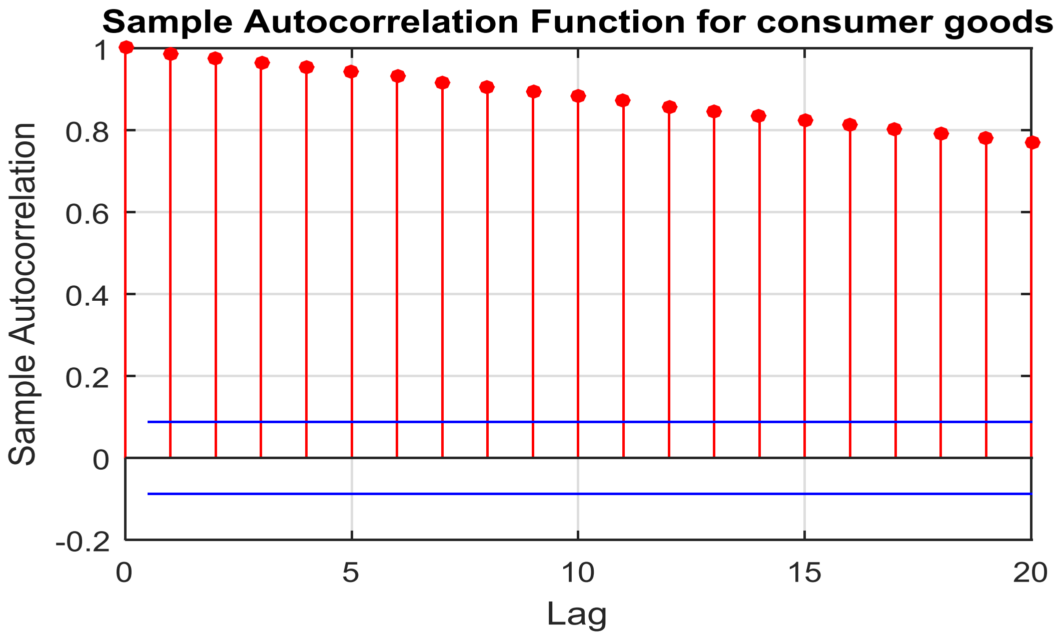
It is widely seen from the figure that sector indices exhibit upward long-term trend and this is the reason why indices are most likely non-stationary. Anyway, we have to prove non-stationarity and it can be done with augmented Dickey-Fuller test (hereinafter: ADF test).

First of all, I conduct ADF test on consumer goods index, because it has the smallest value at the end of observing period (take a look at Figure 6) and approximately the smallest values also in previous periods. As such it is has less evident upward trend and I am able to infer that consumer goods index could be stationary with the highest probability (in comparison with other indices). ADF test shows failure to reject null hypothesis, with probability value $p = 0.999$.

ADF test's null hypothesis indicates that consumer goods time series is non-stationary and by inference all other time series are almost certainly also non-stationary. Anyhow, I calculate ADF test also for other indices and probability values are nearly the same.

Another option is to visually assess whether there are any trends in time series data. Plot of autocorrelation function helps us. ACF for non-stationary time series data slowly decreases, while for stationary ACF drops off at specific lag. Figure 7 indicates that autocorrelation (red dots at the top of vertical lines) is significantly above corresponding confidence levels (blue horizontal lines) and proves that consumer goods index is non-stationarity. ACFs for other indices are similar to one, presented in Figure 7. This finding is in line with augmented Dickey-Fuller test.

Figure 7: Sample ACF for non-stationary consumer goods index



With the aim of removing of trend from each index I calculate natural logarithm (as one of the chances for time series transformation) of original index values, which stabilize the variance of time series. When performing ADF test on logarithmic consumer goods index data, test result still proves non-stationarity with probability p = 0.9983. Results for other 9 indices are fairly similar, leading to conclusion that more transformations have to be implemented.

This means that I should use another time series transformation option; first differences between consecutive observations. When taking first differences once our time series lose one value (for example: consider that we have only *two* values. We are able to calculate only *one* first difference). However, in my thesis this should not be a serious problem, since time series are sufficiently long, I calculate first differences (equation (15)) for all stock indices and ADF test reveals that finally indices have become stationary. This is presented in Table 7.

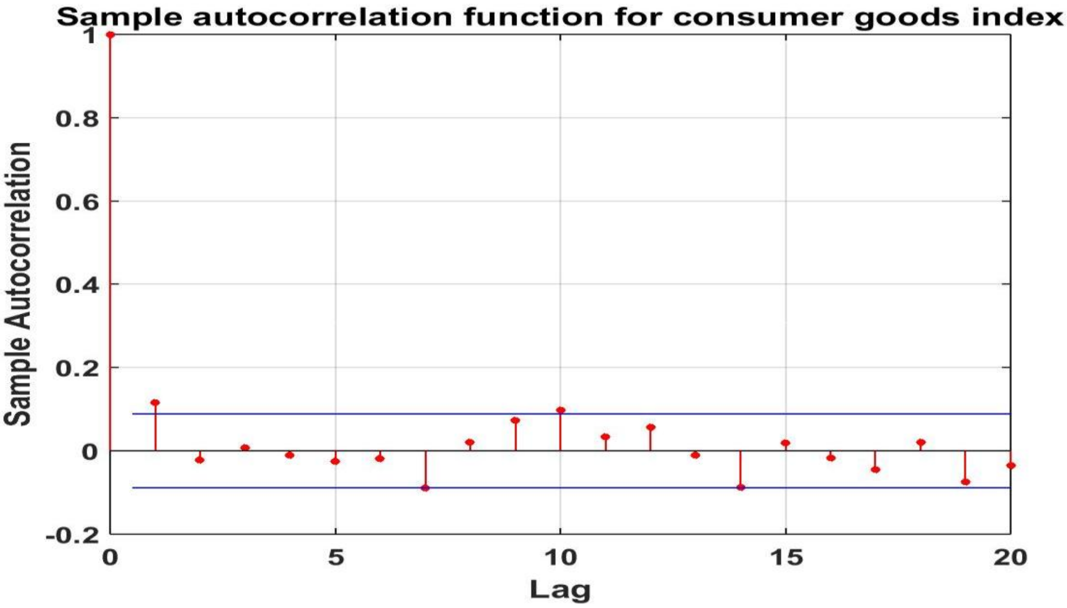
From now on, I denote $\Delta\ln(x_t^n)$, from equation (15), simply as x_t.

Table 7: P-values of ADF test, performed on the logarithmic and differentiated time series

Transformed time series	Utilities	Telecommunications	Technology	Oil & gas	Industrials	Healthcare	Financials	Consumer services	Consumer goods	Basic materials
P-value	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

As seen in Figure 8, alternative way to assess stationarity confirm ADF test, since sample ACF significantly drops off at first lag.

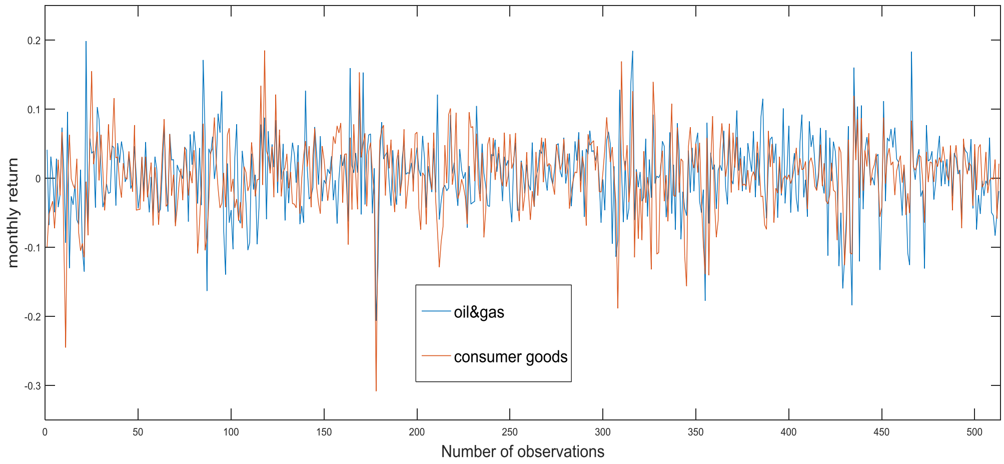
Figure 8: Sample ACF on stationary consumer goods index



Taking first differences only once (more than once is quite rare) leads to conclusion that 'I' term in ARIMA model is set to 1. All my time series have become stationary after differencing once and for such series we say that they are integrated of order 1. Financial time series with logarithmic and first differences transformations are easily to handle with, because period to period changes in data values are approximately returns (especially in cases, when changes are rather low; monthly returns are usually low).

With stationarity I achieve that mean of the process does not follow any trend/seasonality pattern (it is more or less stable). However, it does not necessarily indicate that mean of the process is zero. In order to achieve non-zero mean of time series, I include constant term in my model. Constant in once differenced time series represents average trend. For the sake of simplicity, there are only two time series after logarithmic and first differences transformation presented in the Figure 9. We can see that they are no-trending, but there are some few extremely positive or negative values (for example take a look on approximately 170th consumer goods value, which is close to - 30%). Such outliers may cause misleading statistical inference.

Figure 9: Stationary time series for oil & gas and consumer goods stock stationary time series



4.1.2 Inflation

Before determination of autoregressive and moving average terms in ARIMA model I have to take into account inflation, because it strongly affects savings amount during years/decades. Inflation has negative effects on returns (important note: it does not necessarily hold nowadays, since deflation in developed countries has become reality). Until now stock sector indices have nominal monthly returns, but inflation must not be ignored.

It is important to keep in mind that inflation index has to be determined in units comparable to stock index units, meaning that inflation index has to be defined on monthly basis and collected for the United States (note: stock indices are U.S. equities). Moreover, inflation index must be transformed with the same mathematical operations as sector indices were. In order to prevent inconsistent comparison, I take natural logarithm of inflation time series and afterwards first differences of natural logarithm. Now, I am able to deduct inflationary pressures from nominal returns.

4.2 Lag order selection

Next step in ARIMA modelling is to choose appropriate values for autoregressive (AR) and moving average (MA) parameters. AR parameter is usually characterized by letter ‘p’ and MA parameter with letter ‘q’.

This is very important stage in ARIMA modelling, because model estimates are based on results of this step. We have two options; first one (more formal) is to calculate ACF and PACF (or plot it) of stationary time series (note: in a case, when ACF and PACF do not provide precise answer, we can decide with information criterion; Akaike or Bayesian information criterion). Second option (more informal) is just to try some standard combinations of p and q.

First option is to calculate ACF and PACF functions of each time series. According to Nau (2014b, p. 8) I have to focus at first few autocorrelations and partial autocorrelations. Any isolated spikes in the ACF/PACF beyond 3rd lag are not problematic. Therefore significant partial autocorrelation at lag 8 in Figure 10 is not considered and the same holds for other significant partial autocorrelations in Figures 10 and 11.

Blue horizontal lines in Figures 10 and 11 are lower and upper significance bounds, whereas red vertical lines are partial autocorrelations at specific lag.

I start with PACF, which is calculated for each stock sector index. To recall, PACF helps us to identify AR order.

For example, take a look at Figure 10, which displays PACF for technological sector time series. We see that PACF is far within bounds already at first lag, leading to conclusion that ARIMA (0,1,q) is the best choice (note: q will be determined later by ACF). For example, take a look at Figure 11. We see that PACF for consumer goods time series looks significant at first lag, whereas it is not at the second, ARIMA (1,1,q) could be the best choice. Since PACF is not far away from bounds, ARIMA (0,1,q) might be better choice.

I repeat this procedure for the remaining 8 time series and get following results: ARIMA (0,1,q) is proposed for utilities, telecommunications, technology, industrials, health care and basic materials index. ARIMA (1,1,q) is selected for financials, consumer services, consumer goods and oil & gas index. I have to point out that none of time series have statistically significant second autoregressive lag (t-statistics for second lag varies from -1.15 for technology to +0.71 for health care time series).

Figure 10: Sample partial autocorrelation function for technology time series

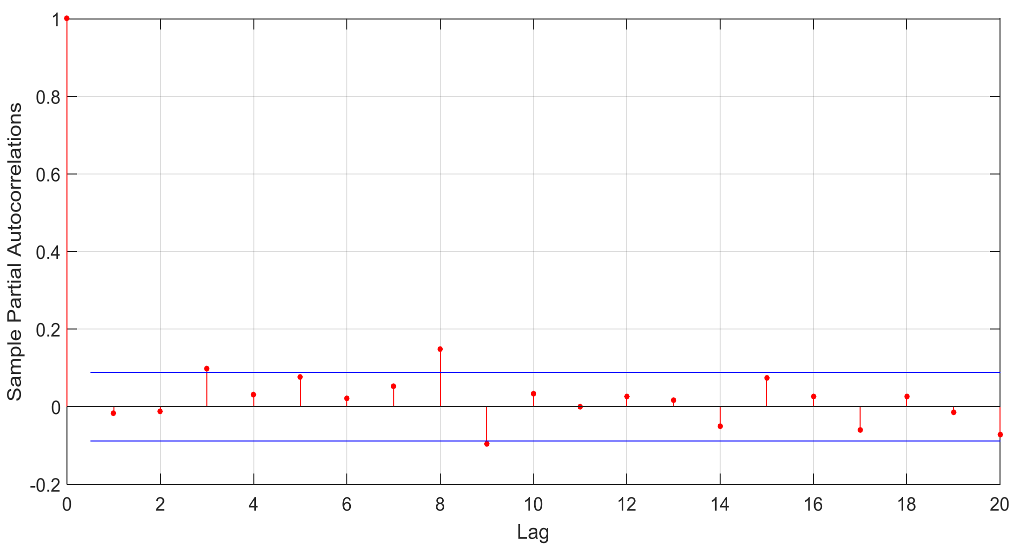
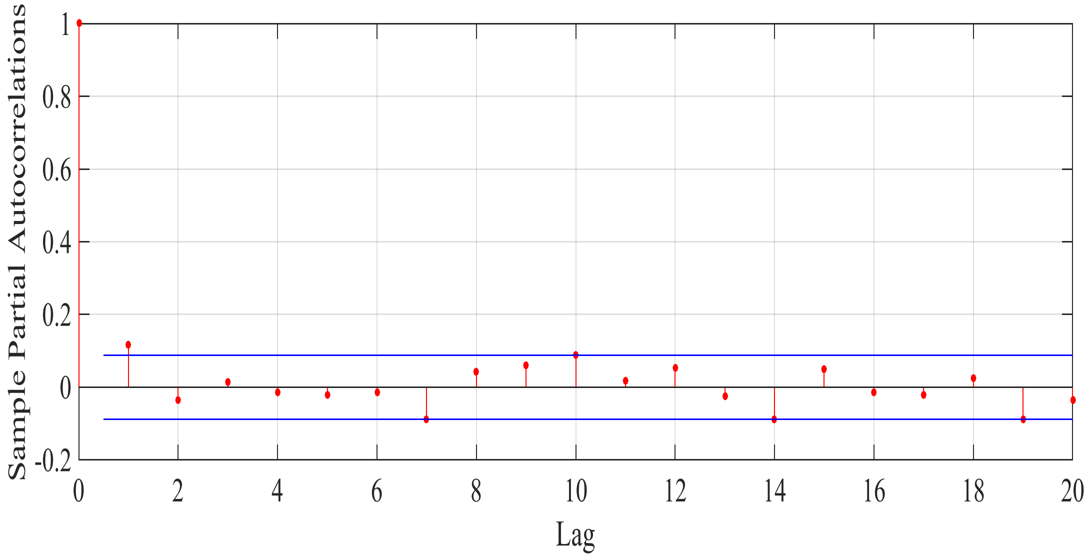


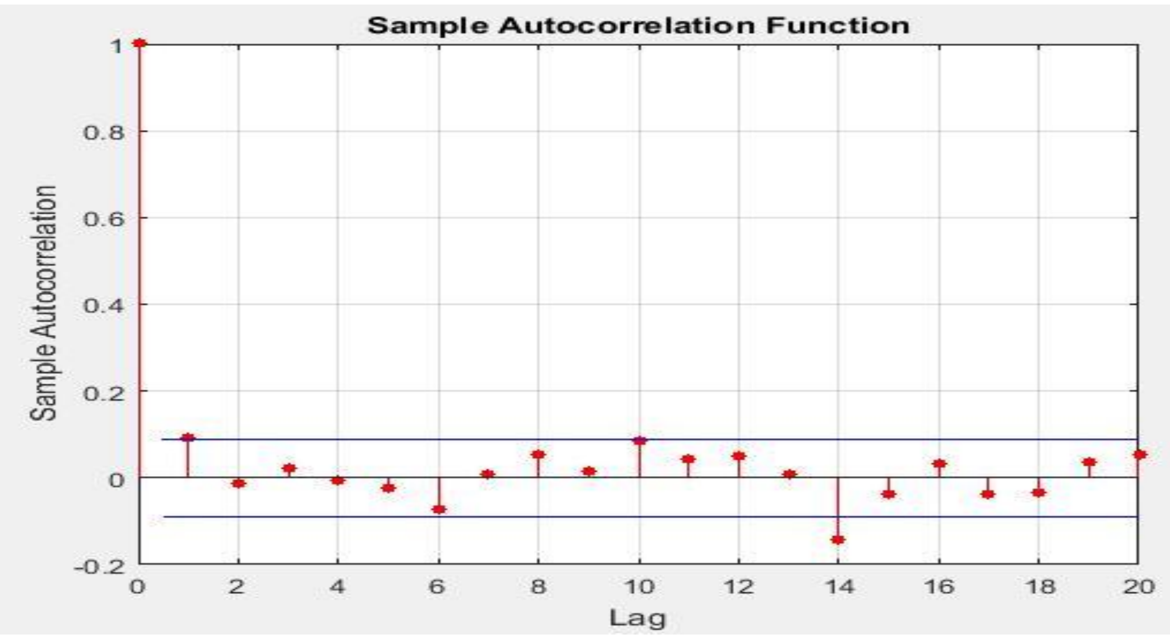
Figure 11: Sample partial autocorrelation function for consumer goods time series



I continue with ACF, which is calculated for each stock sector time series. To recall, ACF helps us to identify MA order.

The procedure is analogous to identification of AR. Take a look at Figure 12, which displays ACF for consumer services time series. We see that ACF for consumer services time series at first lag seems to be significant, whereas it is not at second lag. MA (1) may be the best choice. However, MA (0) might be correct too, since ACF is close to statistical bounds.

Figure 12: Sample autocorrelation function for consumer services



The same procedure goes for all remaining 9 stock sector indices. After doing it, I combine MA terms and AR terms and get suggested ARIMA models for each index separately, presented in Table 8. I have to point out that none of indices have statistically significant second moving average term (t-statistics varies from -1.40 for industrials index to +0.81 for basic materials index).

However, autocorrelations and partial autocorrelations for some indices, lie on the boundary lines (or really close to it). This means that lag order selection does not clearly indicate proper lag order selection. Hence, Table 8 consists of 4 indices that do not have ultimate determined lag orders.

Table 8: Sector indices and suggested models

Stock sector index	Possible models
Utilities	ARIMA (0,1,0) with constant
Telecommunications	ARIMA (0,1,0) with constant
Technology	ARIMA (0,1,0) with constant
Industrials	ARIMA (0,1,0) with constant
Health care	ARIMA (0,1,0) with constant
Financials	ARIMA (1,1,0) with constant ARIMA (1,1,1) with constant ARIMA (0,1,1) with constant ARIMA (0,1,0) with constant
Consumer services	ARIMA (1,1,0) with constant ARIMA (1,1,1) with constant ARIMA (0,1,1) with constant ARIMA (0,1,0) with constant
Consumer goods	ARIMA (1,1,0) with constant ARIMA (1,1,1) with constant ARIMA (0,1,1) with constant ARIMA (0,1,0) with constant
Basic materials	ARIMA (0,1,0) with constant
Oil & gas	ARIMA (1,1,0) with constant ARIMA (1,1,1) with constant ARIMA (0,1,1) with constant ARIMA (0,1,0) with constant

4.3 Models selection

With purpose to select the most appropriate model I calculate Akaike information criterion and Bayesian information criterion for financials, consumer services, consumer goods and oil & gas stock time series. Results are shown in Table 9.

Table 9: AIC and BIC values for financials, consumer services, consumer goods and oil & gas

Time series	ARIMA (0,1,0)	ARIMA (1,1,0)	ARIMA (0,1,1)	ARIMA (1,1,1)
Financials	-1,441 / -1,437	-1,443 / -1,434	-1,443 / -1,434	-1,441 / -1,429
Consumer services	-1,505 / -1,497	-1,507 / -1,494	-1,507 / -1,494	-1,506 / -1,489
Consumer goods	-1,537 / -1,528	-1,541 / -1,528	-1,541 / -1,528	-1,540 / -1,523
Oil & gas	-1,493 / -1,484	-1,493 / -1,481	-1,493 / -1,481	-1,491 / -1,474

Note: first value (i.e. left one) in each cell is AIC value and second value (i.e. right one) is BIC value.

From Table 9 we can see relative quality of models for a given set of data. To recall, relative quality of econometrics model is determined with AIC and BIC values. The most appropriate model for a specific time series has the lowest AIC and BIC values, however, it still has to be statistically reliable. I have to point out that AIC and BIC helps to compare candidate models only within single time series set of data. AIC and BIC do not tell anything about comparison between two or more time series data values. For example, AIC value for random walk model for consumer goods is much lower than AIC for random walk model for financials, but it does not mean that random walk in more appropriate for consumer goods than for financials.

With taking a look at values in Table 9 it is clear that 'candidate' models have similar explanatory strength. However, ARIMA (1,1,1) has slightly higher values and it is not the simplest one (it has both, autoregressive and moving average term). Therefore it cannot be the most appropriate model.

However, other three candidate models are almost in the same position, as their AIC and BIC values resemble. I do not suggest selecting ARIMA (0,1,1), because it has only one moving average term. To recall, MA (1) is affects future for only one period and has fairly low predictability strength.

In contrast to the MA term, AR term influences all future periods, despite its limit in long run is 0. Relating to Table 9 it is known that Akaike information criterion values suggest ARIMA (1,1,0) model for financials, consumer services and consumer goods index. Anyhow, for oil & gas index decision is uncertain, as AIC values are equal for random walk model and ARIMA (1,1,0). As already explained in theoretical part, Bayesian information criterion tends to select models with lower lag order. My case supports theoretical findings, because BIC suggests ARIMA (0,1,0) for financials, consumer services and oil & gas index. However, BIC values for consumer goods index are equal, hence the decision is uncertain (take a look in the table 9).

Final decision is quite complex. Acquah (2010, pp. 3–4) suggests that AIC performs well in small samples, whereas its performance does not improve when sample gets larger. In contrast, BIC tends to perform better in larger sample size. Sample size in the following thesis is 514, large enough that BIC tends to perform better than AIC. In addition to this I should try to select model with as little number of lags as it is possible to ensure statistical validity. In line with reasons stated above I decide that ARIMA (0,1,0) with constant is used for model estimation also for financials, consumer services and oil & gas sector index. BIC values for consumer goods index does not prefer random walk model over ARIMA (1,1,0). For the sake of simplicity (and comparability) I decide to model consumer goods index with ARIMA (0,1,0) with constant.

Before moving to models estimation let me revise findings from above paragraphs. Decision for 6 stock sector indices, namely utilities, telecommunications, technology, industrials, health care and basic materials sector index is clear and ARIMA (0,1,0) model with constant is unquestionably the most appropriate model. After additional analysis I decide that also remaining 4 stock sector indices, namely financials, oil & gas, consumer services and goods sector index are estimated with ARIMA (0,1,0) model with constant.

4.4 Models estimation

In this chapter I estimate the models with Matlab software. Maximum likelihood estimation is used in the models estimation. Programming code is presented in Appendix B. Furthermore, I check whether estimated parameters are statistically significant (or not). In addition to this I examine whereas estimated parameters are more statistically significant, when residuals are assumed to be either normally (i.e. Gaussian) distributed or when residuals are assumed to have Student-t distribution.

I estimate the models with two different assumptions; firstly with student-t distributed residuals and secondly, with normal or Gaussian residuals distribution. Results are present in Table 10 for t-statistics of constants for each stock sector index.

Table 10: Stock sector time series and their t-statistical significance for parameters of the estimated models assuming either Student-t or Gaussian distribution of residuals

Stock sector time series	**Residuals distribution assumption**	**t-statistics for ARIMA (0,1,0) with constant**
Utilities	Student-t Normal	t (cons) = 3.16 t (cons) = 2.34
Telecommunications	Student-t Normal	t (cons) = 2.70 t (cons) = 1.87
Technology	Student-t Normal	t (cons) = 1.88 t (cons) = 1.29
Industrials	Student-t Normal	t (cons) = 2.95 t (cons) = 1.88

table continues

Table 10: Stock sector time series and their t-statistical significance for parameters of the estimated models assuming either Student-t or Gaussian distribution of residuals (cont.)

Stock sector time series	Residuals distribution assumption	t-statistics for ARIMA (0,1,0) with constant
Health care	Student-t Normal	t (cons) = 3.62 t (cons) = 2.88
Basic materials	Student-t Normal	t (cons) = 2.27 t (cons) = 1.36
Financials	Student-t Normal	t (cons) = 3.05 t (cons) = 2.88
Consumer services	Student-t Normal	t (cons) = 2.31 t (cons) = 1.51
Consumer goods	Student-t Normal	t (cons) = 2.20 t (cons) = 1.20
Oil & gas	Student-t Normal	t (cons) = 3.11 t (cons) = 2.44

At this stage it is the most important that the values of parameters, evaluated by t-statistics, are statistically significant. Parameter is statistically significant with 95% probability (95% is usually assumed in economic statistics) when parameter's t-statistics is either less than -1.96 or higher than +1.96. This means that 95% of observations lie within +/- 1.96 standard deviations of the mean.

Table 10 shows that all 10 time series have more statistically significant (i.e. higher t-statistics in absolute value) parameters when residuals are assumed to be Student-t distributed. This result is not surprising, as for financial time series Student-t distribution is usually more appropriate than Gaussian distribution.

Table 10 proves that parameters for 9 stock sector time series are statistically significant, hence, they have to be included into econometric model. Since we are dealing with random walk model, parameter is constant/drift component in model. However, only technological time series has statistically non-significant parameter, t-statistics is 1.88, whereas it should be at least 1.96. As the difference is rather small, I neglect it and assume student-t distribution of residuals,

To summarize above results, I would like to point out that all time series are better estimated with assuming student-t distributed residuals. Furthermore, t-statistics for parameters are statistically significant and thus parameters are included in econometric models.

Table 11: Estimated econometric models for 10 stock sector time series

Stock sector time series	Econometric model
Utilities	$x_t = 0.00568 + x_{t-1} + v_t$ (0.00180)
Telecommunications	$x_t = 0.00585 + x_{t-1} + v_t$ (0.00212)
Technology	$x_t = 0.00548 + x_{t-1} + v_t$ (0.00284)
Industrials	$x_t = 0.00728 + x_{t-1} + v_t$ (0.00220)
Health care	$x_t = 0.00701 + x_{t-1} + v_t$ (0.00186)
Basic materials	$x_t = 0.00521 + x_{t-1} + v_t$ (0.00243)
Financials	$x_t = 0.00829 + x_{t-1} + v_t$ (0.00230)
Consumer services	$x_t = 0.00617 + x_{t-1} + v_t$ (0.00229)
Consumer goods	$x_t = 0.00564 + x_{t-1} + v_t$ (0.00220)
Oil & gas	$x_t = 0.00654 + x_{t-1} + v_t$ (0.00236)

where $E(v_t)$ =0 and $E(v_t v_\tau) = 0 \; for \; \tau \neq t$ holds for each stock sector time series.

Let me explain the content of Table 11. All numerical values in equations in the table above are constants, so called 'drift' in the random walk model. These constants are not the only terms that contribute to processes evolution, because residuals are also influential for time series movement. However, residuals are random, unpredictable and on average equal to zero and thus cannot be modelled. Alternatively, constant can be explained also as period to period return. In this thesis, period is set to one month, hence, constant is monthly return.

For example, take a look at utilities time series and its constant equal to 0.00568. It means that average monthly return is expected to be 0.568%. This does not mean each month such return is achieved, but on average during observing period. The same logic holds also for remaining sectors.

4.5 ARIMA (0,1,0) with constant; residuals checking

Before I move to the time series simulation it is necessary to check residuals. There are several conditions that must be met in order to get correct results. Those conditions are:

expected values of residuals have to be equal to 0 and residuals must be uncorrelated with each other. Furthermore, their variance should be statistically stable across time. At this stage it is wise to check type of residuals distribution, because residuals from financial time series are usually not normally distributed. Because I have already determined residuals distribution (it is Student-t distribution) I do not check their distribution again.

Matlab software offers a lot of possibilities for residuals checking, both, statistical and graphical tests.

In the first place, I check whether expected value of residual v_t is 0, In ARIMA (0, 1, 0) with drift, residuals are simply difference between observed time series values at time t and mean value of time series. For all 10 indices the mean of residuals is equal to 0.

Secondly, I test residuals for autocorrelation. It is really important that residuals are not autocorrelated, because autocorrelation indicates that the model could be improved. I assess autocorrelation with Ljung-Box test and with autocorrelation plot.

Test decision h=0 indicates non-autocorrelation in residuals, whereas h=1 warns us that autocorrelation is existing. For 9 stock sectors Ljung-Box test returns h=0, except for telecommunications time series, meaning that residuals are autocorrelated.

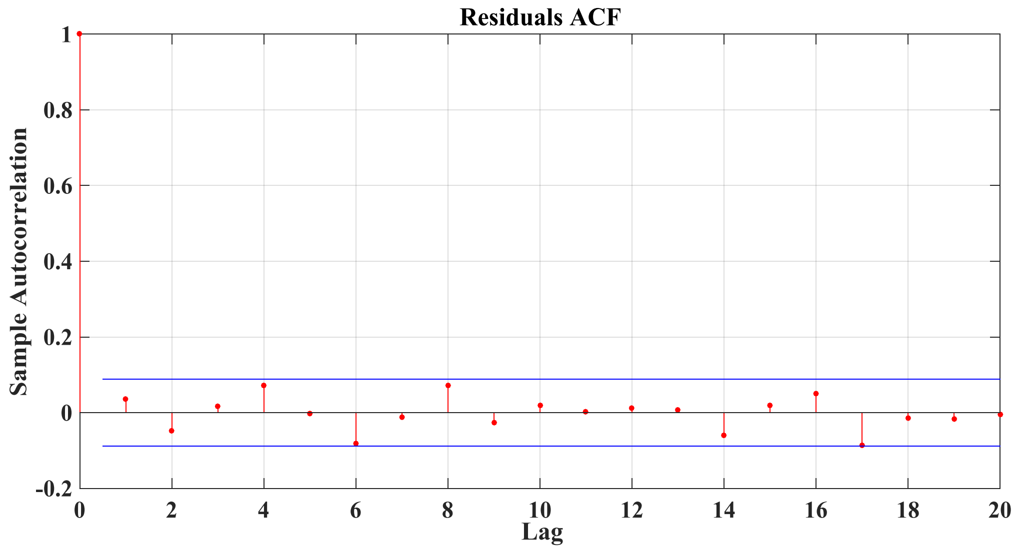
In such situation it is recommended to increase lag order. For telecommunications increasing lag order from 0 to 1 does not solve the problem. Going further with increasing lag order, I find out that AR (8) model is the only model that ensures non-autocorrelation is residuals. However, such high number of AR lag (usually the number of lags varies from 0 to 3) does not make any sense. When increasing autoregressive order does not solve the problem, than some extreme values (i.e. outliers) are most probably source of autocorrelation. Just to be sure, I plot residual autocorrelation function and only one (out of 20) lag lies outside confidence levels. Only one outlier is considered acceptable in economic statistics. Due to reasons stated above, I do not re-define the selected random walk with drift model for telecommunication stock sector time series and conclude that estimated model is acceptable.

As already mentioned above, autocorrelation can be checked also visually, by plotting residuals ACF (note: for examination of stationarity we plot ACF on original time series, while for examination of autocorrelation we plot ACF on residuals).

In Figure 13, I present plot of residuals autocorrelation function for industrial sector time series. As we can see, all 20 autocorrelations lie within statistical bounds (blue horizontal lines). The most important ACFs are at the beginning/at the left of x axis. I am able to claim that they are non-significant, leading to conclusion that residuals are not autocorrelated. To make things clear; ACF's for other indices are fairly similar.

Third thing to verify is residual homoscedasticity/heteroscedasticit. I assess homoscedasticity with Engle's ARCH (autoregressive conditional heteroscedasticity) test. However, there are two alternative possibilities to verify homoscedasticity, namely Ljung-Box test implemented on squared residuals and ACF plot of squared residuals. All three tests (namely ARCH test, Ljung-Box test and ACF plot) indicate significant conditional heteroscedasticity for all 10 stock sectors. It means that variances of the processes are not stable across time. Result is not surprising, as I deal with financial time series. Anyhow, heteroscedasticity can be a severe problem.

Figure 13: Autocorrelation plot for residuals of industrials time series



Despite existence of heteroscedasticity I do not go further with process estimation with ARCH (autoregressive conditional heteroscedasticity) model or GARCH (generalized autoregressive conditional heteroscedasticity) model. Reason for such decision lies in 1000 random simulations/iterations of entire time series process. Those simulations are able to neutralise variance volatility and provide accurate result.

If I had to model conditional variance, I would estimate variance (i.e. second moment) with GARCH model and process mean (i.e. first moment) with ARIMA model. In fact combination of ARIMA-GARCH model would be used.

5 RESULTS

5.1 ARIMA (0,1,0) simulation and expected wealth calculation

In this chapter I describe the final calculus and present the results of expected savings calculation and probabilities for an average negative return.

Before starting with simulation I have to be sure that selected models are the most accurate among other possible models. For 10 sector market indices, that are simulated by random walk with drift models, I am able to assume that those models are optimal (models are simple, but they are still able to fit to the data and satisfy statistical requirements).

I conduct 1000 iterations/simulations of estimated model in order to eliminate effects of volatility clustering (i.e. to eliminate any misleading effects of heteroscedasticity). Each time series is simulated 1000-times. With 1000 iterations I am able to assume that my expected savings at the end of saving period are unbiased. The Matlab code that is used for simulation (and for entire process) is presented in Appendix B.

I assume 40 years of monthly payments/savings/contributions. All together, 480 contributions take place in saving period. Each contribution is assumed to be 100 € in real terms. In nominal terms, the payment amount will probably increase through time (due to positive inflation expectations). Real values are also necessary, because stock sector time series values are determined in real terms (corrected for inflation). Future is uncertain, so my simulations can only be based on historical stock index movement and on assumption that the history will repeat itself.

I should not ignore the costs; 2 different fees decreases investor's savings. First fee is assumed to be 1% of each contribution and it is called administrative fee (note: in reality it is determined through negotiation with pension company. In Slovenia, administrative fee varies from 0.5% to 3%). Administrative fee causes that out of 100 € invested, only 99 € gets invested by pension company, whereas 1 € is pension company revenue. It is assumed that administrative fee is charged immediately at the payment time of client's contribution.

Second cost that occurs is management fee and it is essential in the long run for pension company financial stability and profitability. I set management fee to 1% per year and it is charged on the amount of pension fund (or equivalently 1/12 % per month). This is the highest fee allowed by Slovenian legislation and usually charged by Slovenian pension companies. It is assumed that management fee is charged each month. Hence, pension company takes 1/12 % of client's savings each month.

One important note and interesting detail: customer's first contribution (at time t=0 or at the beginning of savings period) is subject to 480 management fees in 40 years of savings (40 years have 480 months). First contribution is subject to management fee each time with one twelfth of one percent of customers savings.

$$100\ € * (1/12) * 0.01 * 480 = 40\ € \tag{29}$$

This means that client investing 100 € at time t=0 gets only 60 € at t+480, whereas remaining 40 € goes to pension company (note: for simplicity, ignore capital gains).

Simulation starts with Matlab for loop, which allows me to simultaneously iterate process for 1000 times. My raw time series data consists of 514 data values, however, one is lost through taking first differences transformation (explanation: for the sake of simplicity consider having only 2 time series values. Employing first differences transformation gives us only one number, out of two). Consequently, I deal with 513 monthly observations. Because of taking logarithmic and first differences, transformed index values are also approximations of monthly returns.

Pension investment period is assumed to last for 40 years (480 months). Each pension company customer's contribution should be randomly linked with monthly return. Evidently original index values are too long, hence I have to select 480 observations (note: observation is in fact also approximation of monthly returns) out of 513. Each payment has to be linked with specific monthly return. It is very important that selection has to be done randomly.

Because I do not want that some returns are doubled or tripled, no replacement is assumed. I have time series with 480 monthly natural returns and another time series with the same number of 100 € contributions.

Next step in modelling is to deal with those 480 randomly selected monthly returns. We have to keep in mind that we are dealing with natural returns. It means that I am allowed to sum returns. Summation has to be cumulative and it has to start from the end of time series that includes monthly returns.

Reason it the following: the first element in this time series is the first monthly return in 40 years time period. Consequently cumulative summation has to be conducted on the entire period of time, as first contribution is invested for 40 years and as such influenced with all returns that take place in 40 years (with all values of time series monthly returns). At the end of this step I have time series that has 480 cumulative natural returns. However, it has to be linked with clients' contributions and transformed in non-logarithmic units.

In order to get values in non-logarithmic units, I have to calculate exponential of each element in time series explained in above paragraph. Such operation must take place, since I deal with natural returns, instead of price returns.

In the next step I have to connect time series values that were exposed in previous paragraph with pension company customers' monthly contributions. Hence, each element of exposed time series is multiplied by number 100 (note: 100 € is monthly contribution).

However, I should not ignore the fees. Therefore, I multiply time series values calculated at previous step by 0.99. Explanation: 1% (= 0.01) of each payment is administration fee, charged by supplementary pension insurance provider. Hence, from the perspective of pension company client, only 99 € gets invested, despite 100 € is paid.

Also management fee must be taken into account. Management fee is modelled in such way, that the last contribution in savings period (at time t = 480) is subject to only one management fee (1/12 of 1%). The first contribution (at time t = 0) is subject to 480 management fees. For complete specification, please, take a look in Appendix B.

Finally I get time series that consists of 480 values. Those values represent expected amounts for each contribution at the end of savings period (at time t = 480). Those values include returns that have effect on the contributions, inflationary effects and both fees.

Simulation step is ended with summation of 480 expected amounts, calculated at above paragraph. Result represents expected wealth/savings of one process simulation at the end of savings period for a client of pension company.

With purpose to get unbiased result I simulate/iterate the whole process that is described above, for 1000 times. Each simulation has its different, randomly selected set of 480 returns, organized in random order. Matlab for loop iterates this process for 1000 times and calculates 1000 expected savings/wealths.

Calculating the mean out of 1000 previously calculated expected savings gives me the final result: expected wealth/savings at the end of investment period for an average customer of pension company that invests 100 € periodically for 40 years.

I repeat the same procedure for all 10 investing strategies. Matlab code stays almost the same for all sector indices, only raw data (historical stock sector index values) gets changed. Results are presented in the Table 12.

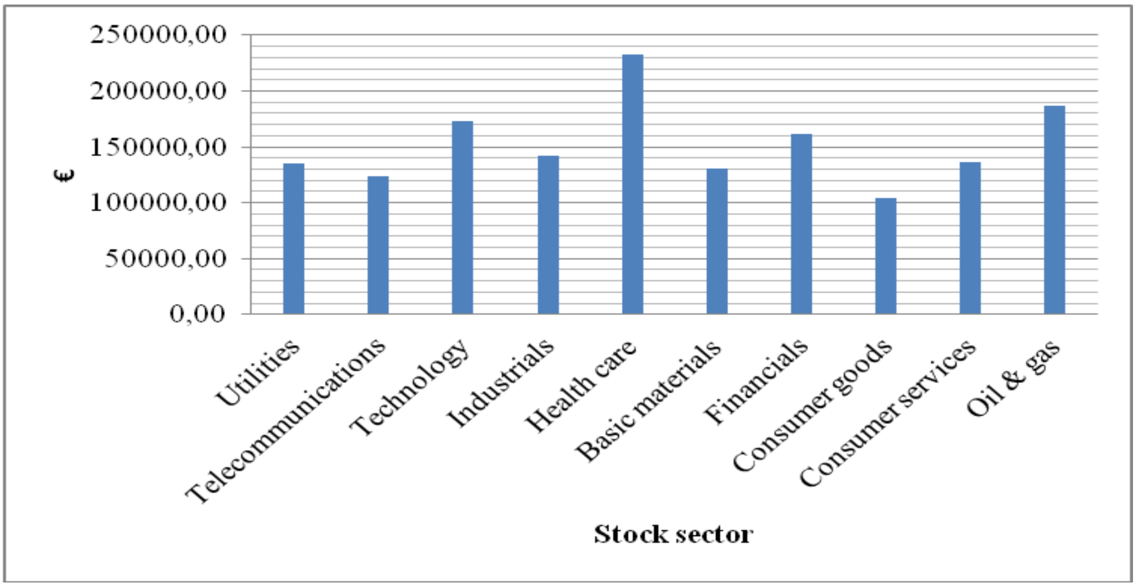
Table 12: Expected savings for 10 investment strategies after 40 years of periodical saving and corresponding average return per annum

Investment strategy	**Expected savings (in €)**	**Expected average return per annum (in %)**
Utilities	135,350	4.514
Telecommunications	123,880	4.165
Technology	173,710	4.702
Industrials	142,040	5.461
Health care	232,720	6.529
Basic materials	130,550	4.373
Financials	161,270	5.182
Consumer goods	104,480	3.479
Consumer services	136,040	4.549
Oil & gas	186,530	5.726

Table 12 displays the value of expected wealth after 40 years of periodical savings across different stock sector indices for an average pension company customer. As we can infer from Table 13, expected savings are rather diversified and vary from 104,480 €, for telecommunication index, up to 232,720 €, for health care index. Results from Table 12 are graphically presented at the Figure 14.

When we compare savings with total invested amount of 48,000 €, we can see that expected returns are significantly positive. However, it has to be emphasized, that past returns are not indicative for future returns. For example, McKinsey & Company (2016) claim that high returns in past are often followed by low returns, while low returns in the past could end with irrationally high prices.

Figure 14: Expected savings for 10 investment strategies after savings period



Although expected wealth of each sector index consists of single numerical value it does not mean that pension company customer saves exactly the same amount of money. Uncertainty in financial markets may lead to either lower or higher wealth at the end of savings period. We have to keep in mind that starting with pension saving, when stock prices are irrational high, leads to different expected savings at the end of savings period in comparison to start when stock prices are extremely low.

For this reason I calculate probability to gain negative average expected return across savings period for each of the investment strategies. Before calculating I pre-assume that probability should be zero or close to it. This statement is based on the assumption that in long run (40 years is obviously long run) investing in stocks is expected to be positive. Hence it is highly unlikely that the overall returns are negative.

Furthermore, I calculate standard deviation of expected savings for each investment strategy and present the results in Table 13.

Before calculating probability for negative average return we should remember that 1000 simulations of each stock sector time series were done. Probability for expecting negative average return is simply proportion of simulations that bring a pension company customer less than 48,000 € at the end of saving period. Probability of loss is extremely important, since nobody wants to invest money for such a long time period (40 years) and then received fewer funds that invested.

Table 13: Probability for negative average expected return when investing in different investment strategies and standard deviation of expected wealth

Investment strategy	Probability for negative return (in %)	Standard deviation of expected savings (€)
Utilities	0.1	42,914
Telecommunications	0.1	49,597
Technology	0.4	86,200
Industrials	0	73,357
Health care	0	81,918
Basic materials	0.9	66,809
Financials	0.2	71,810
Consumer goods	2.2	39,615
Consumer services	0.3	53,164
Oil & gas	0	78,566

I put in order expected savings of 1000 simulations and count how many of them ensure savings amount at least 48,000 €. Then I divide number with 1000 and get probability. For example, investing in basic materials stock index has 0.9% probability for negative return. It means that out of 1000 process iterations, 9 expected saving amounts are less than 48,000 €.

Empirical research shows that probability varies from 0% to 2.2%. The most secure sectors are industrials, health care and oil & gas sector, whereas consumer goods index performs the worst, with 2.2% probability to finish investment period with less savings than invested. The average of all 10 stock sectors (assuming equal weights) is 0.44%. It means that out of 1000 customers of pension company 4.4 customers can expect to face with negative return on their investments. I think this is reasonable low probability to conclude that saving for such a long period in stock is fairly safe and can be offered to customers of pension company.

When it comes to standard deviation of expected savings, we can see from the Table 13, that standard deviation varies from 39,615 € for consumer goods index to 86,200 € for technology index. It is interesting that consumer goods sector has to lowest standard deviation of expected savings and at the same time it has the highest chance for negative return. The average of all 10 stock sectors (assuming equal weights) is 64,395 €.

At the end of empirical research I classify stock sectors in 2 broad classes, namely cyclical and non-cyclical stocks. Cyclical stocks include technology, industrials, basic materials, financials and consumer services sector. Non-cyclical stocks are utilities, telecommunications, health care, consumer goods and oil & gas sector. In the line with theory I expect lower standard deviation for non-cyclical stocks, because such investments are suppose to be more stable.

When classification is done I calculate expected savings, average return, standard deviation and probability for negative return for both, cyclical stocks and non-cyclical stocks. Weights of economic sectors within cyclical/non-cyclical stocks are the same (i.e. each sector has the weight 0.2). Results are presented below, in the Table 14.

Table 14: Expected savings, expected average return per annum, probability for negative return on investments and standard deviation of expected savings

	Expected savings (€)	**Expected average return per annum (in %)**	**Probability for negative return (in %)**	**Standard deviation of expected savings (€)**
Cyclical stocks	148,722	4.881	0.34	70,268
Non-cyclical stocks	156,592	5.073	0.48	58,522

From Table 14 we can see that at 2, out of 3 characteristics (note: expected savings and average expected return per annum is basically the same thing), non-cyclical stocks outperform cyclical stocks. Expected savings are nearly 8,000 € higher, which is definitely not negligible difference. Standard deviation of expected savings is significantly lower for non-cyclical stocks, which allows pension company clients to have more precise expectation regarding their expected savings at the of savings period. When it comes to probability to gain negative average return over entire 40 years, cyclical stocks perform much better than non-cyclical stocks. However, probability is in both cases fairly low.

5.2 Implications for investment strategy of pension company

In the last chapter of thesis I present implications for an investment strategy of pension company funds. Implications are based on the results of master thesis.

Relating to the empirical results it is quite simple to select the most profitable stock sector to invest in. Health care index performs by far the best. Oil & gas and technology indices returns are also above average, whereas consumer goods and telecommunications indices perform worse than an average. Around average there are investment strategies with consumer services, utilities, industrials, basic materials and financial stocks.

Based on the above comment it seems that pension company should invest great part of funds in the health care industry, whereas it should avoid consumer goods stocks. However, I would not recommend such investment policy to pension company. We have to keep in mind that past returns are not indication for future returns.

Moreover, one of the most important rules in finance is to disperse the investments among several asset classes. This rule claims that the most efficient way to avoid random fluctuations in stock prices of specific sector/company is to split the savings and distribute to asset classes that are not correlated. For example: in recent years finance sector stocks returns are significantly below market average. Having all assets in finance stocks means that investor would face with severe loss of its savings. With allocation of assets in some other sector, defensive for example, he would perform much better.

According to empirical results I recommend to split pension company funds in two parts. One part should be invested in the cyclical stocks and another one should be placed to non-cyclical stocks. Amount intended for defensive stocks should be slightly larger than those for cyclical stocks, as defensive stocks have higher expected returns (note: based on the data used in the following thesis).

When it comes to allocation of assets within cyclical and non-cyclical class I suggest pension company asset manager to buy stocks of all sectors within cyclical/non-cyclical class. Based on empirical results, I advise to overweight technological and financial stocks for cyclical stocks, whereas industrials, basic materials and consumer services stocks should be underweighted. Advice is based on comparison of expected savings for specific sector with expected savings for the 5 cyclical stock sectors together. Overweighed defensive stocks (with the returns above average) should be health care and oil & gas stock, while utilities, consumer services and telecommunication stocks should be underweighted.

Anyhow, financial investments are never completely safe and thus these implications have to be only a part of successful investment strategy.

CONCLUSION

In the final chapter I conclude the thesis with the final comment on the results for both hypotheses that were tested. Moreover, I comment whether purpose and aim of the thesis were achieved.

In the first part of master thesis, theoretically oriented, I present demographic changes, financial sustainability and legislation of current Slovenian pension system. First part contains also theory about time series analysis and econometrics models that are used for analysis.

In the second part, empirically oriented, I choose the best econometrics models, estimate them and check the residuals. Moreover, I calculate expected savings, probability to gain negative average return and standard deviation for each of 10 investment strategies.

Results, presented in the Table 12 and at the Figure 14 help us to make the final decision, whether to reject or not reject the first hypothesis. Based on the Table 12 results, I claim that differences in expected profitability across sector indices are large enough not to be able to reject first hypothesis. Statement is based at the spread between the lowest average return, which is expected to be 3.479 % for strategy that invests in consumer goods stocks, whereas the highest return is expected to be 6.529 % for health care stocks. Differences between returns of investment strategies are large enough that statistical testing is unnecessary. Differences in expected returns are reflected in the expected savings across sectors, since amounts significantly differentiate (note: take a look in the Table 12). It leads to conclusion that there are notable differences in expected savings within stock indices.

Results, presented in the Tables 13 and 14 help us to evaluate second hypothesis. To recall, it states that there is very low (i.e. close to 0) probability to achieve negative average return on long term investments into stocks. Relating to the results, I claim that probabilities to achieve average negative return for a customer of a pension company are low enough not to be able to reject the second hypothesis. Explanation: expected probability for negative average return when investing in cyclical stocks is 0.34%, whereas for non-cyclical stock is 0.48% (note: take a look in the table 14). Probabilities are low enough that statistical testing is unnecessary. Consequently, I am not able to reject second hypothesis and conclude that average negative return is highly unlikely to happen.

My master thesis is a mixture of theoretical and realistic concepts. It applies theoretical time series modelling on realistic index values. I get familiar with Slovenian demographic problem, pension system and its legislation. Furthermore, I acquire knowledge about steps in time series modelling. Results can be an indication for pension company investment strategy. I gained Matlab programming skills. Hence, the main aim of thesis is achieved.

Purpose of the thesis is achieved, as I successfully calculate expected wealth for periodical investments in stocks for an average savings person and compare investment strategies with each other.

To sum up, I would like to point out that our financial safety at the retirement is also result of saving during the working ages. Despite some recessions with strong decrease of stock prices, I recommend savings in stocks, as they most likely bring positive returns. I recommend either to join pension company or to invest money at its own through numerous trading platforms. Decision should depend financial expertise of an individual.

REFERENCE LIST

1. *A brief history of time series analysis*. Retrieved February 21, 2016, http://www.statistics.su.se/english/research/time-series-analysis/a-brief-history-of-time-series-analysis-1.259451
2. Acquah, H. D. (2010, January). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of an asymmetric price relationship. *The Journal of Development and Agricultural Economics*, *2*(1), 1–6.
3. Adhikari, R., & Agrawal, R. K. (2013). *An introductory study on time series modeling and forecasting*. Saarbrücken: LAP Lambert Academic Publishing.
4. *Aktivno prebivalstvo, Slovenija, januar 2016* [Labour force, Slovenia, January 2016]. Retrieved February 22, 2016, from http://www.stat.si/StatWeb/prikazi-novico?id=5826&idp=3&headerbar=2
5. Barreto, H., & Howland, F. M. (2005). *Introductory Econometrics*. Retrieved September 15, 2015, from http://www3.wabash.edu/econometrics/EconometricsBook/chap20.htm
6. Box, G. E. P., Jenkins, G. M., & Reinsel, G. C. (1994). *Time series analysis: forecasting and control* (3rd ed.). New Jersey: Prentice Hall.
7. Brockwell, P. J., & Davis, R. A. (2009). *Time series: theory and methods* (2nd ed.). New York: Springer.
8. Campbell, J., Lo, A., & MacKinlay, A. (1996). *The econometrics of financial markets* (2nd ed.). Princeton: Princeton University Press.
9. Cornell, B. (1999). *The equity risk premium: the long run future of the stock market*. New York: John Wiley & Sons.
10. Crawford, M., Church, J., & Akin, B. (2016, April). *CPI Detailed Report: data for April 2016*. Retrieved April 15, 2016, from http://www.bls.gov/cpi/cpid1604.pdf
11. *Cyclical vs. Defensive Stocks*. Retrieved May 15, 2016, from https://www.privatebanking.societegenerale.com/en/strategy/equity-solutions/equity-nutshell/equity-nutshell-june-2013/
12. Diamond, P. (1999, September). What stock market returns to expect for the future? *Center for retirement research at Boston college Paper No. 2*. Retrieved December 15, 2015, from http://economics.mit.edu/files/637
13. *Dodatno pokojnisko zavarovanje* [Supplementary pension insurance]. Retrieved March 2, 2016, from http://www.mddsz.gov.si/si/uveljavljanje_pravic/statistika/dodatno_pokojninsko_zavarovanje/
14. *Econometrics toolbox*: *trend-stationary vs. difference-stationary processes*. Retrieved January 30, 2016, from http://www.mathworks.com/help/econ/trend-stationary-vs-difference-stationary.html
15. Ekart, M. (2011, March). *The supplementary pension insurance in Slovenia – a brief overview of the current situation*. Retrieved March 4, 2016, from http://www.ebrd.com/downloads/research/news/Slovenia.pdf

16. Emsbo-Mattingly, L., Hofschire, D., Litvak, A., & Lund-Wilde, J. (2014, September). The business cycle approach to equity sector investing. *Leadership series.* Retrieved March 24, 2016, from https://scs.fidelity.com/common/application/markets_sectors/business_cycle/Business_Cycle_Sector_Approach.pdf
17. *Euro exchange rates USD.* Retrieved May 4, 2016, from https://www.ecb.europa.eu/stats/exchange/eurofxref/html/eurofxref-graph-usd.en.html
18. Gornjak, M. (2011). Second pension pillar – opportunity for employees. In V. Dermol (ed.), *Knowledge as business opportunity* (p. 177–186). Retrieved February 23, 2016, from http://issbs.si/press/ISBN/978-961-92486-3-8/papers/ML11-20.pdf
19. Greene, W. (2002). *Econometrics analysis* (5th ed.). New Jersey: Prentice Hall.
20. Guardiancich, I. (2010, May). *Slovenia: current pension system: first assessment of reform outcomes and output.* Retrieved January 29, 2016, from http://www.ose.be/files/publication/2010/country_reports_pension/OSE_2010_CRpension_Slovenia.pdf
21. Hamilton, J. D. (1994). *Time series analysis*. Princeton: Princeton University Press.
22. Horváth, Z., & Johnston, R. (2006). *AR(1) time series process.* Retrieved March 4, 2016, from http://www.math.utah.edu/~zhorvath/ar1.pdf
23. Insurance supervision agency. (2014, June). *Report of the insurance supervision agency for 2013.* Ljubljana: Insurance supervision agency.
24. Insurance supervision agency. (2015, June). *Report of the insurance supervision agency for 2014.* Ljubljana: Insurance supervision agency.
25. Intihar, S. (2013). Mednarodni dan boja proti revščini 2013 [International Day for the Eradication of Poverty 2013]. *Statistični urad.* Retrieved January 2, 2016, from http://www.stat.si/StatWeb/glavnanavigacija/podatki/prikazi staronovico?IdNovice=5827
26. LeSage, J. P. (1999). *Applied econometrics using MATLAB.* Retrieved December 20, 2015, from http://www.spatial-econometrics.com/html/mbook.pdf
27. *Life-cycle pension fund.* Retrieved February 19, 2016, from http://www.modra-zavarovalnica.si/home/pension-funds/life-cycle-pension-fund/
28. Majcen, B., & Verbič, M. (2008). Slovenian pension system in the context of upcoming demographic developments. In R. Holzmann, L. MacKellar & J. Repanšek (eds.), *Pension reform in Southeastern Europe: linking to labor and financial market reform* (pp. 73–87). Ljubljana: Center of excellence in Finance.
29. Masten, I. (2012). *Econometrics II: Introduction to AR and VAR models* [PowerPoint slides]. Retrieved December 19, 2015, from http://studentnet.ef.uni-lj.si/Domov.aspx
30. McKinsey & Company. (2016, May). *Diminishing returns: Why investors may need to lower their expectations.* Retrieved August 15, 2016, from http://www.mckinsey.com/industries/private-equity-and-principal-investors/our-insights/why-investors-may-need-to-lower-their-sights
31. Ministrstvo za delo, družino in socialne zadeve. (2009, September 25). *Modernization of the pension system in the Republic of Slovenia – Safe age for all generations.* Retrieved December 20, 2015, from http://www.mddsz.gov.si/fileadmin/mddsz.gov.si/

pageuploads/dokumenti__pdf/word/document.modernisation.pension.system_final.doc
32. Ministrstvo za delo, družino, socialne zadeve in enake možnosti. (2014, April 18). *Učinki pokojninske reforme in nadaljnji koraki* [Effects of the pension reforms and next steps]. Retrieved March 7, 2016, from http://www.mddsz.gov.si/fileadmin/mddsz.gov.si/pageuploads/dokumenti__pdf/dpd/Analiza.PIZ.pdf
33. Ministry of labour, family and social affairs. (2010). *Pension and disability insurance bill (ZPIZ-2)*. Retrieved November 9, 2015, from http://www.mddsz.gov.si/fileadmin/mddsz.gov.si/pageuploads/dokumenti__pdf/word/zpiz_2_reasons_en.doc
34. Modra zavarovalnica. (2014a, October). *Modri dinamični podsklad: Izjava o naložbeni politiki* [The Dynamic Sub-Fund: Statement of investment Policy]. Retrieved January 19, 2016, from http://www.modra-zavarovalnica.si/fileadmin/MKPS/IoNP_Modri_DINAMICNI.pdf
35. Modra zavarovalnica. (2014b, October). *Modri preudarni podsklad: Izjava o naložbeni politiki* [The Prudent Sub-Fund: Statement of investment Policy]. Retrieved January 19, 2016, from http://www.modra-zavarovalnica.si/fileadmin/MKPS/IoNP_Modri_PREUDARNI.pdf
36. Modra zavarovalnica. (2014c, October). *Modri zajamčeni podsklad: Izjava o naložbeni politiki* [The Guaranteed Sub-Fund: Statement of investment Policy]. Retrieved January 19, 2016, from http://www.modra-zavarovalnica.si/fileadmin/MKPS/IoNP_Modri_ZAJAMCENI.pdf
37. Nau, R. (2014a). *The mathematical structure of ARIMA models*. Retrieved January 10, 2016, from http://people.duke.edu/~rnau/Mathematical_structure_of_ARIMA_models--Robert_Nau.pdf
38. Nau, R. (2014b). *Notes on nonseasonal ARIMA models*. Retrieved January 11, 2016, from http://people.duke.edu/~rnau/Notes_on_nonseasonal_ARIMA_models--Robert_Nau.pdf
39. Nau, R. (2014c). *Notes on the random walk model*. Retrieved January 6, 2016, from http://people.duke.edu/~rnau/Notes_on_the_random_walk_model--Robert_Nau.pdf
40. Nau, R. (2016a). Estimation of ARIMA models. *People duke*. Retrieved January 12, 2016, from http://people.duke.edu/~rnau/arimest.htm
41. Nau, R. (2016b). Statistical forecasting: notes on regression and time series analysis. *People duke*. Retrieved January 15, 2016, from http://people.duke.edu/~rnau/411home.htm
42. *Plače* [Earnings]. Retrieved February 22, 2016, from http://www.stat.si/StatWeb/pregled-podrocja?idp=74&headerbar=13
43. *Pokojninska reforma ZPIZ2 splošna pojasnila* [Pension Reform ZPIZ2 General Illustrations]. Retrieved March 1, 2016, from http://www.penzionist.info/index.php/novice/item/56-pokojninska-reforma-zpiz2-splo%C5%A1na-pojasnila
44. *Pokojninski načrti življenjskega cikla* [Life-cycle pension plan]. Retrieved February 19, 2016, from http://www.pokojninskad-a.si/zc/

45. Raicharoen, T., Lursinsap, C., & Sanguanbhoki, P. (2003). Application of critical support vector machine to time series prediction. *Proceedings of the IEEE 2003 International Symposium on Circuit and Systems* (pp. 741–744). Bangkok: Institute of Electrical and Electronics Engineers.
46. Remšak, B. (2010, September 14). Razlike med tremi pokojninskimi stebri [Differences between three pension pillars]. *Vzpon.* Retrieved February 5, 2016, from http://vzpon.vzajemci.com/novice/osebne_finance/ 210/razlike_med_tremi_pokojninskimi_stebri
47. *Residual diagnostics*. Retrieved January 28, 2016, from http://www.mathworks.com/ help/econ/residual-diagnostics.html?s_tid=srchtitle
48. Schwarz, C. J. (2011). *Sampling, regression, experimental design and analysis for environmental scientists, biologists, and resource managers.* Retrieved February 22, 2016, from http://people.stat.sfu.ca/~cschwarz/Stat-650/Notes/PDF/MLE-AIC.pdf
49. Siegrist, K. (2015). The central limit theorem. *Math.* Retrieved January 15, 2016, from http://www.math.uah.edu/stat/sample/CLT.html
50. *Skladi življenjskega cikla[Life cycle funds]*. Retrieved February 19, 2016, from http://www.prva.net/pokojninsko-zavarovanje/skladi-zivljenjskega-cikla/
51. *Slovenia – 2015*. Retrieved September 30, 2016, from https://www.cia.gov/library/publications/the-world-factbook/graphics/population/SI_ popgraph%202015.bmp
52. *Sredstva javnih uslužbencev za dodatno pokojnino so v dobrih rokah*![The funds of the Public employees are in save hands]. Retrieved February 19, 2016, from http://www.modra-zavarovalnica.si/podjetja/medijsko-sredisce/novice/novica/article/sredstva-javnih-usluzbencev-za-dodatno-pokojnino-so-v-dobrih-rokah/
53. *Stationarity and differencing.* Retrieved March 2, 2016, from https://www.otexts. org/fpp/8/1#fn:1
54. Thomson Reuters. (2012). *Thomson Reuters business classification methodology*. Retrieved September 15, 2015, from http://financial.thomsonreuters.com/content/ dam/openweb/documents/pdf/financial/trbc-methodology.pdf
55. Tsay, R. S. (2005). *Analysis of financial time series* (2nd ed.). New Jersey: John Wiley & Sons, Inc.
56. *Why log returns?* Retrieved December 23, 2015, from https://quantivity.wordpress. com/2011/02/21/why-log-returns/
57. Zakon o pokojninskem in invalidskem zavarovanju [Pension and invalidity insurance act]. *Uradni list RS* no. 96/12, 39/13, 99/13 - ZSVarPre-C, 101/13 - ZIPRS1415, 44/14 - ORZPIZ206, 85/14 - ZUJF-B in 95/14 - ZUJF-C.

APPENDIXES

TABLE OF APPENDIXES

APPENDIX 1: Slovenian summary

Živimo v času neverjetnih sprememb na večini področij človeškega življenja. Ena izmed njih je tudi tematika te magistrske naloge. Demografske spremembe in njihov vpliv na finančno vzdržnost pokojninskega sistema je še posebej v razvitih državah pereč problem. Sčasoma vse bolj in bolj postaja jasno, da obstoječ sistem javne pokojnine, ki izhaja iz naslova obveznega pokojninskega zavarovanja, ne bo več dolgo časa zagotavljal finančno varne starosti. V ostalih državah razvitega sveta je situacija precej podobna.

V prvem delu teoretske obravnave se ukvarjam s trenutno situacijo na področju pokojninskega varčevanja v Sloveniji. Trenutno slovenski proračun pokrije visok delež primanjkljaja v pokojninski blagajni. Posledično se deficit pokojninske blagajne akumulira v našem javnem dolgu. Opišem tudi pokojninsko ureditev pri nas in demografske trende.

Iz vsega naštetega izhaja dejstvo, da bomo morali sami dodatno varčevati v aktivni dobi in prihranke porabljati v upokojitveni dobi. Kar samo se poraja vprašanje kako varčevati, da bomo privarčevali čim več.

Slovenski zakonodajalec je problem opazil že pred časom in leta 2001 uvedel dodatno prostovoljno pokojninsko zavarovanje. S pomočjo davčnih olajšav se je poskušalo pridobiti čim več zavarovancev. Poleg davčnih olajšav je vlada določila dodatno pokojninsko zavarovanje kot obvezno za določene skupine zaposlenih; javne uslužbence, zaposlene v težavnih delovnih pogojih in za poklice, katerih izvajanje po določeni starosti postane težavno. Sistem je zaživel, ampak ne tako, kot je bilo predvideno. 300 tisoč od 800 tisoč delovno aktivnih ni vključenih v dodatnem pokojninskem varčevanju.

Razlog da je precej ljudi ostalo izven sistema je precej konservativna investicijska politika slovenskih pokojninskih družb/skladov, saj so imeli dovoljen zelo nizek delež pokojninskih skladov v delnicah. Z namenom rešitve tega problema je leta 2013 v veljavo vstopil nov pokojninski akt, ki pokojninskim družbam/skladov dovoljuje višji delež delniških investicij. Delež delnic je določen glede na starost varčevalca. Izraz za to so skladi življenjskega cikla (angl. *life-cycle funds)*.

Bistvena značilnost skladov življenjskega cikla je omogočanje posamezniku, da prilagodi varčevanje glede na naklonjenost do tveganja in glede na starost. Posamezniku so ponujeni trije skladi življenjskega cikla, med katerimi lahko izbira. Velja omeniti, da se posameznik lahko odloči za varčevanje v skladu z nižjim deležem delnic, medtem ko se ne more odločiti za varčevanje v skladu z večjim deležem delnic. Vse to z namenom, da se izogne pretiranemu izpostavljanju tveganju. Manjša novost je tudi prepoved dviga privarčevane vsote po desetih letih varčevanja. Sistem je zasnovan tako, da bo posameznik začel prejemati mesečno pokojninsko rento, ko bo izpolnil pogoje za upokojitev. Poleg tega so se določili tudi najvišji dovoljeni stroški, ki jih ponudniki dodatnega pokojninskega varčevanja smejo zaračunati.

V Sloveniji obstaja še tretji pokojninski steber, ki je povsem neobvezen in ni predmet nobenih zakonskih omejitev. Ne ponuja pa nobenih davčnih olajšav. Sem spadajo različna življenjska zavarovanja, vzajemni skladi itd.

Moja magistrska naloga je povezana z drugim in tretjim pokojninskim stebrom. Ker predpostavljam investiranje izključno v delnice, ne more biti direktno uporabna za drugi pokojninski steber, saj to ne bi bilo v skladu z zakonodajo. Lahko pa je uporabna za tretji steber. Vsekakor pa lahko služi, kot pomoč pri izbiri investicijske politike pokojninske družbe.

V drugem delu teoretske obravnave se osredotočim na analize časovnih vrst (angl. *time series*) in njihovo modeliranje. Modeliranje časovnih vrst (časovne vrste so nizi podatkov v času) je precej pogosto uporabljeno z namenom napovedovanja prihodnjega gibanja časovne vrste. Analiza delnic in napovedovanje makroekonomskih faktorjev sta dva izmed več področij, ki s pridom izkoriščata znanje o časovnih vrstah.

Poznamo različne modelov za opisovanje časovnih vrst, ki se delijo v tri obširne skupine: avtoregresijski modeli (angl. *autoregression models*), integrirani modeli in modeli premikajočega povprečja (angl. *moving average models*), ki skupaj sestavljajo integrirane avtoregresijske modele premikajočega ravnotežja (angl. *ARIMA models)*. Avtoregresijski model je uporabljen v tej magistrski nalogi. Ti modeli so precej popularni, ker so relativno preprosti, še vedno pa so sposobni dobre napovedi.

Pomembna prednost ARIMA modelov je v tem, da so sposobni pravilno obravnavati trende ali sezonska gibanja časovne vrste, kar je še posebej koristno pri ekonomskih/finančnih spremenljivkah, ki pogosto izkazujejo trend. V tem primeru moramo vključiti še konstanto v naš model. Najbolj pomembna razlika med avtoregresijskim modelom in modelom premikajočega povprečja je v tem da slednji posnema predhodno gibanje ostankov (angl. *residuals*), medtem ko avtoregresijski model obravnava gibanje predhodnih podatkov v časovni vrsti.

Na tem mestu velja omeniti metodologijo, ki sta jo zasnovala znanstvenika Box in Jenkins in se ukvarja s pravilnim postopkom modeliranja časovnih vrst. Sestavljena je iz petih korakov; modifikacijo časovne vrste v stacionarno časovno vrsto, kamor spadajo transformacije s prvimi diferencami (angl. *first differences*) in naravnimi logaritmi. Drugi korak vsebuje izbiro odlogov, ki najbolje opisujejo našo časovno vrsto. Sledi ocenjevanje parametrov časovne vrste, obravnava ostankov in preverjanje ali slednji ustrezajo belemu šumu (angl. *white noise*). Metodologija se zaključi s napovedovanjem na podlagi ugotovljenih lastnosti v prejšnjih korakih.

V nadaljevanju naloge natančneje pojasnim vsak korak v Box-Jenkins metodologiji. Začnem s pojasnilom zakaj je najbolje uporabiti logaritemske donose in opišem postopke, ki nam zagotavijo stacionarnost (logaritmi, metoda diferenciacije, statistični testi ali grafi (delne) avtokorelacijske funkcije). Sledi opis postopkov, ki nam pomagajo določiti število odlogov. Tu so nam v pomoč avtokorelacijska funkcija za določitev odlogov premikajočega povprečja in delna avktokorelacijska funkcija, ki nam pokaže število odlogov avtoregresijskega modela. Potem predstavim teoretični opis ocenjevanja parametrov, kjer velja, da smemo uporabiti metodo najmanjših kvadratov (angl. *ordinary least squares*), če imamo ničelno število odlogov pri modelu premikajočega povprečja. V kolikor pa to ne velja, pa moramo uporabiti metodo nelinearnih najmanjših kvadratov (angl. *non-linear least squares*). Pri ocenjevanju parametrov moramo preveriti statistično značilnost posameznih parametrov. V ekonomski statistiki se običajno značilnost začne ko je 'p' vrednost nižja od 0,05 oziroma t-statistika večja od +1,96 ali manjša od -1,96. V naslednjem podpoglavju se osredotočim na ocenjevanje ostankov. Ostanki modela morajo čim bolj spominjati na beli šum. V primeru, da so ostanki avtokorelirani med sabo, je naš model mogoče izboljšati. V primeru, da imajo ostanki spremenljivo varianco moramo uporabiti kombinacijo ARIMA modela z generaliziranim avtoregresijskim modelom pogojne heteroskedastičnosti (angl. *GARCH model*).

V naslednjem podpoglavju pa opišem teste za informacijske kriterije, ki nam pridejo v pomoč, ko več različnih modelov 'preživi' sito analize. Opišem Akaike-jev in Bayesovski informacijski kriterij, katerih vrednosti nam povesta, kateri izmed modelov ima najboljše razmerje med preprostostjo in napovedno močjo.

Teoretični del sklenem s kvalifikacijo vseh desetih sektorjev delnic v dve večji skupini; ciklični in neciklični sektorji. Industrija, tehnologija, materiali, finance in sektor cikličnih potrošnih dobrin spadajo pod ciklične sektorje. Pod neciklične sektorje spadajo sektorji javnih dobrin, telekomunikacij, zdravstva, energetike in sektor necikličnih potrošnih dobrin.

Sledi vmesno poglavje, kjer opišem podatke, ki so uporabljeni v empiričnem delu magistrske naloge. Podatki so mesečne vrednosti ameriških ekonomskih panožnih indeksov in so pridobljeni iz Thomson Reuters podatkovne baze. Thomson Reuters delnice razvrsti v 10 različnih sektorjev. Ameriške indekse delnic uporabim, ker nam ponudijo najdaljše časovne vrste. Zbrani so na mesečni osnovi, z začetkom januarja leta 1973 in koncem oktobra leta 2015. Vse skupaj je 514 mesečnih vrednosti. Inflacija ni upoštevana v vrednostih indeksov. Ker imam na voljo ameriške delnice, upoštevam inflacijski indeks Združenih držav Amerike, ki zniža nominalne donose indeksov. Dividendna donosnost delnic je v indeksih upoštevana; predpostavljeno je, da se dividende takoj investirajo nazaj v nakup delnic posameznega sektorja.

Nato razložim metodologijo, ki je uporabljena v empiričnem delu. Postavim se v vlogo svetovalca upravljavca premoženja pokojninske družbe. Rad bi izvedel podatek o pričakovani vrednosti premoženja na koncu varčevalnega obdobja za povprečnega varčevalca pokojninske družbe. Predpostavljam da povprečni varčevalec mesečno prispeva v pokojninski sklad realnih 100 € (če bo inflacija pozitivna, bo mesečno vplačilo nominalno naraščalo). Varčevalec prispeva celotno delovno dobo, torej 40 let in posledično 480-krat vplača 100 €. Zanima me, koliko je pričakovana vrednost premoženja za varčevalca, če pokojninska družba investira vsa mesečna vplačila po določeni investicijski strategiji. Imam 10 investicijskih strategij, ki se med sabo razlikujejo glede na to v kateri delniški sektor so sredstva varčevalca investirana. 10 strategij je posledica 10-ih sektorjev, v katere so razdeljene delnice. Za lažjo primerjavo donosnosti delniških sektorjev je predpostavljeno, da pokojninska družba vsa sredstva naloži v delnice enega sektorja (znotraj posamezne investicijske strategije).

Pomemben podatek, kar se tiče metodologije raziskave je tisočkratna simulacija (angl. *simulation)*, z namenom dobiti čim bolj realno pričakovano vrednost premoženja (angl. *expected savings/wealth*). Ker naredim tisoč ponovitev, vsa morebitna ekstremna gibanja procesa izgubijo (oziroma se jim zmanjša) vpliv na končno pričakovano vrednost. Simulacije procesa so narejene s pomočjo Matlab programa. Potrebno je poudariti, da se upošteva vse stroške, ki jih pokojninska družba zaračuna varčevalcu. Vsaka mesečna premija je soočena z 1% vstopnimi stroški. To pomeni, da se dejansko investira samo 99 €. Poleg tega varčevalcu niža pričakovano vrednost premoženja tudi upravljaljska provizija, ki se izračunava mesečno; 1/12 od 1% vrednosti prihrankov varčevalca vsak mesec pobere pokojninska družba.

Investicije so narejene v ameriške delnice, torej so izvedene z ameriški dolarji. Glede na to da je dolgoročno razmerje ameriškega dolarja in Eura precej stabilno, sem se odločil, da zanemarim valutno razmerje pri izračunu vrednosti premoženja. Dividende se takoj investirajo nazaj v nakup delnic in se ne izplačajo varčevalcem.

V empiričnem delu magistrske naloge najprej preverim časovne vrste delniških indeksov, ki morajo biti stacionarni (vrednosti naj ne naraščajo/padajo skozi čas). Ker niso, najprej izračunam naravni logaritem od vsake vrednosti indeksa. Ker to ni dovolj za stacionarnost, izračunam prve razlike (diference) med sosednjima vrednostnima v indeksu. Z Dickey-Fullerjevim testom dokažem, da so transformacije zadoščale za dosego stacionarnosti. Tudi avtokorelacijska funkcija kaže na stacionarnost. Velja si zapomniti, da so vrednosti časovne vrste, po teh transformacijah, približki naravnih donosov.

Sledi pravilna vključitev inflacijskega indeksa v časovne vrste, katerih vrednosti niso popravljene za inflacijo. Ker se ukvarjam z ameriškimi delnicami moram uporabiti ameriški inflacijski indeks, na katerem izvedem enake transformacije, kot v prejšnjem odstavku. Nato samo odštejem transformirane vrednosti inflacijskega indeksa od transformiranih vrednosti delniških indeksov.

Sledi izbira avtoregresijskega odloga in odloga premikajoče sredine za vsak delniški indeks. Avtoregresijski odlog je določen z delno avtokorelacijsko funkcijo, medtem ko avtokorelacijska funkcija določi število avtoregresijskih odlogov.

Na tem mestu je pomembno opomniti, zakaj ne uporabim modelov premikajočega povprečja. Slednji lahko napovedujejo samo toliko period vnaprej (oziroma vsebujejo informacije) kot je število odlogov. V primerjavi s avtoregresijskim modelom ima relativno nizko napovedno moč.

S avtokorelacijsko funkcijo ugotovim, da je za 6 sektorskih indeksov model naključnega hoda (angl. *random walk model with drift* ali ARIMA (0,1,0) model) najbolj primeren. Za finančni, energetski sektor in sektorja cikličnih oziroma necikličnih potrošnih dobrin pa avtokorelacijska funkcija ne pokaže dokončne izbire.

Pri dokončni izbiri modela naključnega hoda tudi za zgoraj navedene 4 indekse mi je pomagal izračun Akaike-jevega in Bayes-ovega informacijskega kriterija (AIC in BIC). Ker je BIC z daljšanjem časovne vrste vedno bolj konsistenten ga upoštevam nekoliko bolj kot AIC. To pomeni, da bom vse časovne vrste ocenil z modelom naključnega hoda.

V naslednjem poglavju ocenim ekonometrične modele in preverim če so parametri statistično značilni. Poleg tega ugotovim ali so parametri bolj statistično značilni, ko predpostavljam normalno razporeditev ostankov ali ko predpostavljam študentsko-t razporeditev ostankov. Matlab mi pomaga oceniti modele in rezultati pokažejo, da so koeficienti statistično bolj značilni, ko predpostavljam študentsko-t razporeditev ostankov. Rezultati tudi pokažejo, da bom za simulacijo uporabil ARIMA (0,1,0) model s konstanto, ki predstavlja povprečni pričakovani mesečni donos delniškega indeksa.

Ko so ekonometrični modeli ocenjeni je potrebno preverit, če ostanki teh modelov predstavljajo beli šum. To pomeni, da mora biti njihova pričakovana vrednost enaka nič, med sabo ne smejo biti korelirani in njihova varianca mora biti čim bolj stabilna skozi čas. Ko preverim pričakovano vrednost, ugotovim da je enaka nič. Morebitno avtokorelacijo ostankov preverim z Ljung-Box testom, ki za 9 indeksov pokaže odsotnost avtokorelacije, medtem ko je avtokorelacija za telekomunikacijski sektorski indeks posledica nekaj ekstremnih vrednosti v samem indeksu. Avtokorelacija je lahko preverjena tudi z grafom avtokorelacijske funcije ostankov modela. Stabilnost variance ostankov preverim z testo avtoregresijske pogojne heteroskedastičnosti, ki pokaže nestabilnost variance ostankov, kar ni dobro. Ker bom naredil tisoč simulacij modelov in bo to zadoščalo za omilitev posledic nestabilne variance, ne grem naprej v modeliranje GARCH modela (posplošen avtoregresijski model pogojne heteroskedastičnosti).

Sledi poglavje, kjer prestavim rezultate magistrske naloge. Še prej natančno opišem relativno zapleten proces tisočih simulacij modela. V računanju pričakovane vrednosti premoženja za povprečnega varčevalca pri pokojninski družbi je treba upoštevati precej stvari. Vedeti je treba, da bo končni znesek zmanjšan za vstopne stroške vsakega vplačila. Vstopni strošek je postavljen pri 1% od vplačila, torej varčevalec prispeva 100 €, le 99 € pa se vloži v nakup delnic. Poleg tega je potrebno upoštevati še upravljalsko provizijo, ki znaša na letni ravni 1% od vrednosti portfelja varčevalca, oziroma 1/12 od 1% mesečno.

Vsaka časovna vrsta (po transformacijah) je sestavljena iz 513 mesečnih donosov. Na tem mestu je potrebno ponoviti investicijsko strategijo, ki pravi, da pokojninski sklad kupuje delnice izključno enega sektorja delnic. Sektorjev je 10, torej imam tudi 10 investicijskih strategij. Ker je moje varčevalno obdobje dolgo 40 let oziroma 480 mesecev, s pomočjo naključne funkcije v Matlabu izberem 480 vrednosti iz časovne vrste in jo pripišem posameznemu 100 evrskemu vplačilu. V naslednjem koraku moram določiti skupni donos, ki se je zgodil na posameznem vplačilu. Zavedati se je potrebno, da na vplačilo iz prvega meseca varčevanja vplivajo vsi donosi, ki so se zgodili po v naslednjih 40 letih. Ker imam v časovni vrsti naravne donose, smem le-te kumulativno sešteti.

Sledi izračun pričakovane vrednosti posameznega vplačila na koncu varčevalnega obdobja. Upoštevati je potrebno upravljalski stroške in strošek posameznega vplačila. Poleg tega je izjemno pomembno naranvne donose pretvorit v absolutne vrednosti. To naredim z eksponiranjem vsakega posameznega elementa časovne vrste kumulativnih donosov. Vse skupaj pomnožim s 100 in dobim časovno vrsto s 480 elementi, ki predstavljajo pričakovano vrednost vsakega posameznega vplačila na koncu varčevanja.

Pričakovana vrednost posamezne ponovitve privarčevanega portfelja na koncu 40-letne periode je preprosto seštevek zgoraj navedene časovne vrste s 480 elementi. Dobljena vsota predstavlja eno simulacijo ekonometričnega modela.

Ker pa želim narediti 1000 ponovitev procesa, zgoraj opisan proces ponovim 1000-krat. To naredim s pomočjo Matlab for zanke, ki vedno znova ponovi zgoraj opisan postopek (naključno izbere donose, jih pripiše mesečnim vplačilom in jim pripiše donose). S tem dobim 1000 ponovitev modela in izračun povprečja iz vseh 1000-ih simulacij pričakovanih vrednosti portfelja nam pokaže vsoto, ki jo povprečni varčevalec v pokojninski družbi lahko pričakuje, če 40 let periodično vplačuje po 100 € mesečno. Torej, dobil sem pričakovano premoženje za eno investicijsko strategijo.

Ker pa imam 10 investicijskih strategij, postopek ponovim še za ostalih 9. Matlab koda ostane identična, z izjemo podatkov, ki vstopajo v končni izračun. Torej, ko računam vrednost portfelja pri strategiji z investiranjem v delnice tehnološkega sektorja, moram uporabljati pretekle vrednosti delnic tehnološkega sektorja.

Ko dobim izračune za vseh 10 investicijskih strategij, z namenom lažje primerjave, izračunam še pričakovan povprečen letni donos posamezne strategije. Iz rezultatov je razvidno, da so razlike v donosnostih med posameznimi sektorji delnice precej velike, saj povprečnemu varčevalcu investiranje v delnice zdravstva prinese kar 232.720 € (oziroma 6,529 % letno), medtem ko investiranje v delnice necikličnih potrošnih dobrin (angl. *consumer goods)* 104.480 € (oziroma 3,479 % letno). Zdravstvenemu sektorju sledijo s precej nižjimi donosi sektorji energije, tehnologije in financ. Še nižje in precej podobne pričakovane donose imajo sektorji industrije, cikličnih potrošnih dobrin (angl. *consumer services),* materialov, javnih storitev in telekomunikacij.

Naj poudarim, da je varčevalec prispeval 48.000 € tekom dobe varčevanja. Če to primerjamo s pričakovanim izkupičkom na koncu, vidimo, da se varčevanje v delnicah izplača (tudi če varčujemo v najmanj donosnem sektorju). Kakorkoli že, treba je poudariti da pretekli donosi niso nujno indikacija prihodnosti, saj obdobjem nadpovprečnih donosov pogosto sledijo leta padajočih cen delnic.

Ne glede na to, da so pričakovane vrednosti donosov občutno višje od 0, je potrebno preverit kakšna je verjetnost, da se varčevalec sooči z negativnim povprečnim donosom. To pomeni, da bo na koncu varčevanja imel v portfelju manj od investiranih 48.000 €.

Vemo, da so finančni trgi lahko zelo nepredvidljivi, zato izračunam še verjetnost negativnega povprečnega donosa za varčevalca, za vsako investicijsko strategijo posebej. Poleg tega predstavim še standardni odklon (angl. *standard deviation)* pričakovane vrednosti premoženja za posamezno strategijo.

Izračuni potrdijo predvidevanja, saj imajo 3 investicijske strategije ničelno verjetnost negativnega povprečnega donosa. To so sektor zdravstva, energetike in industrije. Tudi ostali sektorji imajo zelo nizke verjetnosti, še najslabše se izkaže investicijska strategija z vlaganjem v delnice necikličnih potrošnih dobrin, ki pa tudi ni zares problematična s 2,2 % verjetnostjo negativnega donosa. Menim, da so donosnosti dovolj nizke, da lahko trdim, da je varčevanje v delnicah na tako dolgi rok varno in primerno.

Pri izračunu standardnega odklona pričakovane vrednosti portfelja ugotovim, da so standardni odkloni kar visoki in tudi pozitivno korelirani z vrednostjo pričakovane vrednosti portfelja. Najnižji standardni odklon ima investicijska strategija vlaganja v sektor necikličnih potrošnih dobrin (39.615 €), ki ima obenem tudi najnižjo pričakovano vrednost portfelja. Najvišji standardni odklon ima investicija v sektor tehnologije (86.200 €), ki je sicer na tretjem mestu po pričakovani vrednosti prihrankov. Relativno visoke vrednosti standardnih odkloni dokazujejo težko napovedljivo in volatilno naravo delniških investicij.

Za konec računskega dela magistrske naloge pa vseh 10 sektorjev delnic razdelim v dve širši skupini, in sicer, ciklične in neciklične delnice. Med ciklične delnice spadajo tiste, ki so v povprečju nadpovprečno donosne v obdobjih ekonomske ekspanzije, medtem ko se za neciklične delnice pričakuje relativno višji donos, od cikličnih delnic, v časih recesije. Sektorji, ki spadajo pod ciklične so industrija, tehnologija, materiali, finance in sektor cikličnih potrošnih dobrin. Sektorji, ki spadajo pod neciklične so sektor javnih dobrin telekomunikacije, zdravstvo, energetika, in sektor necikličnih potrošnih dobrin.

Za obe skupini izračunam iste parametre, kot za vse posamezne investicijske strategije. Vsaka skupina vsebuje 5 sektorjev in izračunam njihovo povprečje. Vsak sektor ima enako utež v končnem izračunu. Pričakovana vrednost portfelja ob investiranju v neciklične delnice je skoraj za 8.000 € višja, kot v primeru cikličnih delnic, ki so tudi bolj tvegane, saj imajo občutno višji standardni odklon. Slednji znaša 70.268 € za ciklične delnice in 58.522

€ za neciklične. Rezultat ni presenetljiv, saj ciklične delnice veljajo za bolj volatilne. Če pogledamo verjetnost negativnega donosa ugotovimo, da je slednja višja za neciklične delnice. Vrednost pa je nižja od polovice odstotka in tako ne predstavlja večje nevarnosti za varčevalca.

Glede na rezultate bi bila najbolj očitna investicijska strategija kupiti čim več delnic zdravstvenega sektorja, medtem ko bi se bilo potrebno izogibati delnicam necikličnih potrošnih dobrin. Vseeno pa upravljavcu premoženja ne bi svetoval take investicijske politike, saj eno izmed najbolj osnovnih pravil v financah govori o pomembnosti razpršitve premoženja med čim več finančnih inštrumentov. Svetoval bi nekoliko nadpovprečen delež delnic sektorjev, ki dosegajo nadpovprečne donose in nekoliko podpovprečne delež manj donosnih sektorjev. Prav gotovo pa svetujem nakup delnic vseh sektorjev, tudi najmanj uspešnih, saj nizka cenovna vrednotenja kažejo na možnost nadpovprečnih donosov.

V zaključku ugotovim, da na podlagi empiričnih rezultatov, ne morem zavrniti obeh hipotez. Prva hipoteza govori o tem, da so donosnosti med investiranjem v posamezne indekse različni. Tabela številka 13 nam pokaže kako različne so pričakovane vrednosti premoženja za varčevalca in njihove pričakovane povprečne letne donosnosti. Vrednosti se gibljejo od 3,479 %, za investicijsko stategijo vlaganja v sektor necikličnih potrošnjih dobrih, do 6,529 % za investicijsko strategijo vlaganja v zdravstveni sector. Na podlagi rezultatov iz Tabel št.14 in 15 ugotovim, da ne morem zavrniti druge hipoteze. Verjetnosti negativnega pričakovanega povprečnega donosa se gibljejo od 0 % do 2,2 %. Torej, varčevalec lahko upravičeno pričakuje pozitiven povprečni donos na svoje prispevke.

Namen magistrske naloge je izpolnjen, saj sem uspešno izračunal pričakovano vrednost premoženja za vsako izmed 10-ih investicijskih strategij posebej.

APPENDIX 2: List of abbreviations

ACF	Autocorrelation function
ADF	Augmented Dickey-Fuller test
AIC	Akaike information criterion
AICc	Corrected Akaike information criterion
AR	Autoregressive (model)
ARCH	Autoregressive conditional heteroskedasticity (model)
ARIMA	Autoregressive integrated moving average (model)
ARMA	Autoregressive moving average (model)
BIC	Bayesian information criterion
ETF	Exchange traded funds
EU	European Union
EUR	The euro
GARCH	Generalized autoregressive conditional heteroskedasticity (model)
GDP	Gross Domestic Product
MA	Moving average (model)
PACF	Partial autocorrelation function
SURS	Statistical Office of the Republic of Slovenia
U.S.	United States
USD	United States dollar
ZPIZ-2	Pension and disability insurance bill

APPENDIX 3: Matlab code for ARIMA (0,1,0)

```
% set key parameters
Lags=0;
Nsim=1000;  % 1000 simulations
d=1;  % order of time series differencing
x;    % x is logarithmic sector stock market index (vector),
      % corrected for inflation index

% estimate the model
Mdl=arima(lags,d,0);
Mdl.distribution='t'; % residuals distribution
EstMdl = estimate(Mdl,x);
v=x-mean(x); % v is vector of residuals

% iterate model 1000 times
Nsim=1000;   % number of simulations
payment=100;   % monthly savings
ones(480,1);
fee=ones(480,1).*((1/12)/100);    % monthly management fee
fee=cumsum(fee,'reverse');  %cumulative management fee
fee1=1-fee;
expwealth = NaN(1,Nsim);  % vector of expected wealth

for ii=1:Nsim
    index=datasample(x,480,'Replace',false);
    B=cumsum(index,'reverse');
    c=exp(B);
    0.99.*(payment.*(c));  % administration fee 1% on each contribution
    g=(0.99.*(payment.*(c))).*(fee1);  % management fee
    sum(g);  % expected wealth of single iteration
    expwealth(ii)=sum(g);
    end
expwealth
mean(expwealth)  % average of 1000 simulations
```