CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Bachelor thesis

2018 **Jonáš Kratochvíl**

CHARLES UNIVERSITY

FACULTY OF SOCIAL SCIENCES

Institute of Economic Studies

Jonáš Kratochvíl

Efficiency of Prague Stock Exchange Market using Markov Chains

Bachelor thesis

Prague 2018

Author: Jonáš Kratochvíl

Supervisor: RNDr. Michal Červinka, Ph.D.

Academic Year: 2017/2018

Bibliographic note

KRATOCHVÍL, Jonáš. *Efficiency of Prague Stock Exchange Market using Markov Chains.* Prague 2018. 65 pp. Bachelor thesis (Bc.) Charles University, Faculty of Social Sciences, Institute of Economic Studies. Thesis supervisor RNDr. Michal Červinka, Ph.D.

Abstract

The main intention of this thesis is to analyze the weak form efficiency of Prague Stock Exchange. We conduct our empirical analysis on daily, weekly and monthly return data of the PX index collected in time period 1994-2017. The theory of Markov chains is employed to decide whether the index returns follow a random walk, the evidence of weak form efficiency. Bayesian Information Criterion is used to establish the optimal order of the Markov chain, which is in turn tested against the order 0 by Likelihood ratio criterion. The model assumptions of time homogeneity, irreducibility and aperiodicity of transition probability matrix are validated. We reject the weak form efficiency for daily index returns and establish its optimal Markov chain order to be 1. The weak form efficiency is not rejected for weekly and monthly index returns so is the assumption of time homogeneity for the whole time period 1994-2017. We propose further analysis of daily returns for time period 2006-2017, which exploits the fact of the weak form inefficiency. Discussion of results and related literature is provided as well as the presentation of all contemplated methods.

Keywords

efficient market hypothesis, random walk hypothesis, Markov chain, probability transition matrix, time homogeneity, likelihood ratio criterion, Bayesian Information Criterion, maximum likelihood estimator

Abstrakt

Hlavním cílem této práce je analýza slabé efektivity Burzy cenných papírů v Praze. Naše empirická analýza zkoumá denní, týdenní a měsíční data z časového období 1994-2017. K testování hypotézy náhodné procházky burzovního indexu PX, která je indikátorem slabé formy efektivity je použita teorie Markovských řetězců. K určení optimálního řádu Markovského řetězce používáme metodu Bayesova informačního kritéria. Tento optimální řád je posléze testován proti řádu 0 za použití metody poměrů nejvěryhodnějších odhadů. Předpoklady modelu časové homogenity, ireducibility a aperiodicity jsou ověřené. Na základě našich výsledků zamítneme slabou efektivitu trhu pro denní výdělky indexu PX a ustanovíme jejich optimální řád 1. Slabou efektivitu nezamítneme pro týdenní a měsíční data, stejně jako předpoklad časové homogenity pro celé časové období 1994-2017. V závěru práce navrhujeme technickou analýzu, která využívá neefektivity trhu pro denní data. Práce také zahrnuje diskuzi výsledků a srovnání s již publikovanou literaturou na dané téma.

Klíčová slova

hypotéza efektivního trhu, hypotéza náhodné procházky, Markovské řetězce, pravděpodobnostní matice přechodu, časová homogenita, Bayesovo informační kriterium, odhad největší věrohodnosti

Declaration of Authorship

I hereby proclaim that I wrote my bachelor thesis on my own under the leadership of my supervisor and that the references include all resources and literature I have used.

I grant a permission to reproduce and to distribute copies of this thesis document in whole or in part.

Prague, 30 April 2018

Signature

Acknowledgment

I am grateful especially to my thesis supervisor RNDr. Michal Červinka, Ph.D. for his time and constructive remarks that greatly improved my thesis. I would also like to thank my family and friends, who have been an indispensable support throughout my entire studies.

Bachelor thesis proposal

Author Jonáš Kratochvíl

Supervisor RNDr. Michal Červinka, Ph.D.

Topic Efficiency of Prague Stock Exchange Market using Markov Chains

The main focus of our Bachelor thesis will be on testing the hypothesis whether the Prague Stock Exchange performs efficiently. In order to test this hypothesis, we will discuss the link between market efficiency and a random walk behavior of a Prague Stock Exchange Index: PX. In a specific situation, the above can be said to be equivalent (Fama, 1965), therefore, rather than answering the question about the market efficiency we will focus on a problem whether the PX index does indeed follow a random walk. If the later transpires to be truthful a direct implication of the market efficiency will be made and consequently our hypothesis would not be rejected.
We believe that the results of our thesis can be both academically and practically beneficial. We have reasons to believe that the methodology of Markov Chains, which will be our main tool in the research, is somehow unfairly neglected among students and economists in general. The fact that the theory of Markov Chains is not covered in any undergraduate course at our institute motivates us to devote a section of our thesis to present the key ideas and definitions of this concept as clearly as possible. Our goal by doing so will be to perhaps motivate future students to consider such method as an alternative to more common ones and provide them with a basic introduction to Markov Chain theory.

Contribution

To our best knowledge, there has not yet been conducted a research on the efficiency of the Prague Stock Exchange using the approach of Markov Chains. Therefore an independent result about the efficiency will be provided and compared with other outcomes, which were previously obtained using different methods and approaches. This can have interesting theoretical consequences, especially if the result derived using our method will vary from another. The comparison section of our thesis will be devoted to such discussion and will leave a space for analysis of the consistency of results across different approaches.

In the case that our hypothesis is rejected, the biggest practical contribution, especially for investors and financial advisers will the fact that the Prague Stock Exchange is not efficient. In such a case it implies that it is possible to systematically beat the market and earn revenues in a long run. Fama (1970) defines that in order for a market to perform efficiently, all information influencing the price of an asset must be reflected in a price of this asset. Therefore the consequence of an inefficient market is the existence of insider information that can be exploited to outperform the market and yield revenues for investors.

Methodology

For our thesis, we will use the approach of Markov Chains. We will model the behaviour of the PX index as Higher Order Markov Chain. The null hypothesis will state that the Markov Chain is of order zero, which means that previous states are independent of each other, thus the price of a share in a time t+1 is independent on a price in time t. The alternative hypothesis is that the index is modeled as a Markov Chain of higher order than 0, which implies the inefficiency of the Prague Stock Exchange.

Bayesian or Akaike Information Criterion will be used to determine the optimal order of the Markov Chain. We will use the data for the Prague Stock Exchange Index:PX, which is the overall index including all 50 companies that are currently traded on the Prague Stock Exchange. The data will be analyzed for the time period 2000-2017.

Outline

1. Introduction
2. Methodology of theory of Markov Chains and Bayesian Information Criterion
3. Random walk and efficient markets
4. Markovian Model
5. Analyzing Data
6. Discussion of results and comparison with other conclusions from previous research
7. Conclusion and suggestions for further research

Bibliography

1. Bartoszyanski, R. and Niewiadomska-Bugaj, M. (2008).Probability and statistical inference. 1st ed. Hoboken, New Jersey: Wiley-Interscience.
2. Fama, E. (1965).Random walks in stock-market prices. 1st ed. [Chicago]: Graduate School of Business, University of Chicago.
3. Fama, E. (1970).Efficient capital markets: a review of theory and empirical work. 1st ed. R. Lowbridge (Module Leader).
4. Grimsved, E. and Pavia, J. (2015).Evaluating the Efficiency of the Swedish Stock Market: a Markovian Approach. Undergraduate. University of Gothenburg.
5. Meyn, S. (2012).Markov chains and stochastic stability. 1st ed. Springer.

Contents

Notation

$\mathbb{E}$ expected value

Γ gamma function

$\Lambda_{i,j}$ likelihood ratio criterion of Markov chain order i and j

$\mathcal{L}(p)$ likelihood function

μ_i initial distribution of probabilities

ω vector of expected return times

ω_i expected return time to state i

π stationary distribution of Markov chain

π_i steady state probability of state i

σ standard deviation

f profit function

$l(p)$ log-likelihood function

m number of states of a state space S

n_{ijk} counts of how many times Markov chain has follower a certain path $i \to j \to k$

P^{n-k} k-step transition probability matrix

p_{ijk} transition probability of a certain path $i \to j \to k$

p_{ij} transition probability from state i to state j

$p_{ij}(n-u, n)$ probability of transition from state i to state j in u steps

r_t return of stock index

S state space

S^u	space of all possible sequences of states from S with length u
X_n	random process
P	time homogeneous transition probability matrix

1 Introduction

The predictability of financial markets has been a widely discussed topic ever since the Amsterdam Stock Exchange was established in 1602. There has not yet been a consensus on whether the yields earned by private investors, mutual fund managers and other market participants are credited to some superior knowledge and skill or are just an outcome of pure chance. To address this question, Fama [13] defined a market to be *weak form efficient* if all past publicly available information about traded assets are incorporated in the present prices of these assets. This proposition triggered a great amount of attention as the consequence of the weak form efficient market is the impossibility of its prediction from past prices. Therefore, if a certain market is indeed weak form efficient, any means of technical analysis are found to be purposeless for its predictions. To this day, there has been an extensive amount of literature written on the aforementioned topic with results supporting both its truthfulness and falsehood when tested on various market.

Our thesis aims to contribute to the discussion by focusing on the question whether the returns of Prague stock exchange index PX can be considered weak form efficient or not. We will examine *daily, weekly and monthly* returns of the PX index from the time period 1994-2017. Many methods have been proposed to test the weak form efficiency of a market. Among the most popular are the *auto-correlation, variance ratio and run test* models. In our thesis we will test the weak form efficiency with the usage of the *Markov chain theory.* This approach is perhaps slightly unconventional in the area of financial economics, however, its insensitivity to outlier values and convenient graphical representation make it a profound model from both the interpretation and analytical point of view.

We will develop a Markov chain model consisting of two discrete states "High" and "Low" on which the time series of index returns will be mapped. The *optimal* order of this Markov chain model will be established by the *Bayesian Information Criterion* and *Likelihood ration criterion.* The Markov

chain model of order 0 denotes that the time series follows a random walk, which is the implication of the weak form efficiency. Otherwise, if the order is strictly greater than 0 than a certain dependence structure exists within the time series and thus the market can not be considered weak form efficient. The transition matrix of the Markov chain will be estimated from data by *Maximum likelihood estimator*, which is a consistent estimator of the transition probabilities. We will verify the time homogeneity assumption of the transition matrix by χ^2 test proposed by Anderson and Goodman [2].
The main motivation for analyzing the weak form efficiency of Prague Stock Exchange is the lack of empirical work on this topic. Moreover, most research already conducted, dates back at least 10 years and the results for more recent data are close to non-existent. We consider the usage of *Markov chains* methodology to be another contribution of our thesis. To our best knowledge, only one publication on the weak form efficiency of Prague Stock Exchange with use of Markov chain theory has been published by Hackl and Pošta [21]. Their approach is, however, significantly different from ours[1]. Our empirical results will therefore serve as a valid alternative to already existing ones, which were obtained with the use of more conventional methods of analysis.

Below, the questions of the main interest of our thesis are summarized.

1. Can any of *daily, weekly and monthly* returns of PX index be considered weak form efficient for time period 1994-2017?

2. Is the assumption of time homogeneity valid for any of these return types for the time period 1994-2017

3. In case the market is not weak form efficient, can we exploit this fact for its prediction?

4. Do our results comply with previous findings?

[1]Hackl and Pošta [21] in their paper use the Brownian motion to model the stock returns. Brownian motion is an example of continuous time Markov chain, whereas our approach considers a discrete time Markov chain.

Our thesis is structured as follows: We will review the literature and past results in Chapter 2. In Chapter 3 we will present the efficient market hypothesis and random walk hypothesis. In Chapter 4 we introduce the core concepts of first and higher order Markov chain theory. In Chapter 5 we will derive the Maximum likelihood estimator for first and higher order Markov chains. Chapter 6 is devoted to the Markov chain model and outlines its application for the weak form efficiency test. In Chapter 7 we will describe our three data-sets used for the empirical part and present the results of the weak form efficiency test in Chapter 8. Chapter 9 is devoted to discussion of the results and suggestions for further research. Finally, we conclude our thesis in Chapter 10.

2 Literature review

In this section we will review the empirical evidence of the weak form efficiency. Firstly, we will focus on the results obtained for the Prague Stock Exchange as its analysis will be the central point of our thesis. In the second sub-section we will review the results obtained on foreign financial markets as well. In this second subsection we will preliminary focus on the research conducted by Markov chain methodology.

2.1 Prague Stock Exchange

Diviš and Teplý [8] evaluated the weak form efficiency of Prague stock exchange on monthly and weekly return data from 1993-2004. Their results confirm that the Prague Stock Exchange was indeed weak form efficient in the observed time period. Their other contribution was the comparison of the Prague Stock Exchange index to the US Stock Exchange Dow Jones Industrial Average (DJIA) index, which was considered as a benchmark for efficiency. The conclusion of their paper was that the Czech financial market is still less efficient than the one in the US, however, the dissimilarities are only marginal and are constantly diminishing over time.

Similar results were also obtained by Horská [23]. She used a model of random walk with drift for the means of her analysis. Šíbl [54] considered daily data from the time period 1999-2004. He concludes that when using the run test, daily returns can be considered uncorrelated on 5% confidence level. Čámský [53] employed the run test and auto-correlation models for daily data between 1999-2003. Based on the results, he rejected the weak form efficiency when considering each year separately (short time-horizon). He, however, does not reject the weak form efficiency when testing the whole time period of four years (long time-horizon). These results also confirm the proposition that Prague Stock Exchange has become more efficient in the first years of new millennium.

Different conclusion was obtained by Quang [40], who analyzed daily data from 1996-2006. He argues that there exists a significant linear dependence between consecutive daily returns and thus the Prague Stock Exchange can not be considered weak form efficient. Same results were obtained by Hájek [24], who analyzed daily returns from 1995-2005. Pošta [39] used the Kalman filters technique to further confirm the weak form inefficiency for daily data from time period 1995-2007, which is the same result as Lazar and Ureche [31], who analyzed data from the same time period.

We see that the overall conclusion about the weak form efficiency based on these results is *somehow* inconclusive. The general consensus of authors is that even though the weak form efficiency of Prague Stock Exchange has been rejected in many cases, the market efficiency has been slowly improving with years. This conclusion is in line with the general trend of developing markets. In the next sub-section we will review the results obtained for foreign stock exchanges and primarily focus on the research conducted by Markov chain methodology.

2.2 Foreign stock exchanges

Okonta et al. [38] applied Markov chain methodology to test the weak form efficiency of Nigerian Stock Market. They analyzed weekly returns from time period 2014-2016. The results show that the Nigerian Stock Market can not be considered weak form efficient for the observed time period. Pavia and Grimsved [20] analyzed the weak form efficiency of Stockholm Stock Exchange market. Data they use were daily, weekly and monthly returns from 2000-2015. They find that the Stockholm Stock Exchange is weak form efficient for all types of returns they have used. Kilic [28] evaluated daily data from the Istanbul Stock Exchange from 1988-2012 and concluded the weak form efficiency. However, when he further analyzed smaller time intervals he found that a certain pattern in data exists and proposed a technical analysis for its exploitation.

The general conclusion of authors is that the developed markets are considered weak form efficient for all monthly, weekly and daily data. On the other hand, majority of developing markets is still considered weak form inefficient, especially for high frequency data (daily, hourly, minute). This fact is mostly caused by slow speed of adjustment of prices to new information. Also the social and political conditions are a crucial indicator of whether the market might be considered weak form efficient or not (Rodriguez et al. [42]).

3 Random walk and Efficient markets

In the first part of this section we will explain the definition of the Market efficiency, its various forms and present a brief overview of its historical development. Further, we introduce the Random Walk hypothesis and provide a link between these two topics.

3.1 The Efficient Market Hypothesis

The *Efficient Market Hypothesis (EMH)*, of which the weak form efficiency is a special case, has lead to many discussions between economists throughout the last century. According Sewell [45], the EMH can be regarded as one of the strongest hypothesis in the whole of the social sciences.

The theory of efficient markets was first informally introduced by Hayek [22] in his famous paper *The use of knowledge in society*, where he argues that the financial markets reflect all available information in a given place and time. This somehow *complete* information would otherwise be unknown to an individual as it is impossible to comprehend by a single mind. Hayek concludes that the rational consumers should always rely on information reflected by market behavior instead of relying on their own intuition about economic outcomes.

The first one to give the EMH a proper economic foundations was Samuelson [43] in his article *Proof that properly anticipated prices fluctuate randomly.* In this article he introduces the idea of EMH and argues that the market prices are impossible to forecast if they are fully anticipated, in other words if they fully incorporate the expectations and information of all market participants (Lima, Ohashi [32]). Samuelson also stated the theoretical assumptions that are necessary to hold in order for a market to be considered efficient. These assumptions include perfect competition, zero transaction costs and free access to information.

One of the prominent economic figures, who is credited for much of the theoretical work on efficient markets is Eugene Fama. Fama [13],[10] describes the efficient market as market in which security prices at any given time “fully reflect” all available information that is relevant to the price of this security. Perhaps a more clearer rephrase of this definition is given by Malkiel [34]. He states that market is said to be efficient if the price of an asset would be unaffected by revealing the complete information set about this asset to all market participants. It is of no surprise that many financial economists and in particular investors strongly oppose this hypothesis. Its basic implication declares that any yields obtained on financial markets are solely due to luck and not due to any superior knowledge or experience as Burton Malkiel [34] explains in his book *A Random Walk Down Wall Street*.

Fama first introduced the topic of EMH in his dissertation thesis in 1964 and has been continuously contributing to the area of market efficiency ever since. He provided the link between the EMH and Random Walk hypothesis (RWH) and subdivided the EMH into its three main forms. These three forms are: A weak form efficiency (all past information are available and incorporated in the price of a security), semi-strong form efficiency (all public information are available and incorporated in the security price at any time) and strong form efficiency (all public and private information about the security are available and incorporated in the security price). Below is the hierarchical diagram[2] showing all three aforementioned definitions.

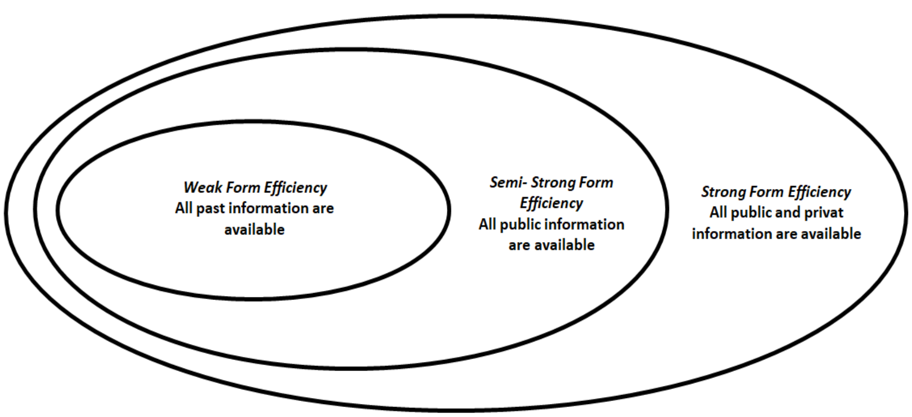


Figure 1: Three forms of market efficiency.

[2]Figure inspired by Veselá [50]

The difference between Fama's interpretation of EMH as opposed to Samuelson's axiomatic approach is that Fama saw efficiency as an actual outcome produced by sophisticated traders (Alajberg et al. [1]), rather than a set of assumptions that need to be satisfied. For the means of our analysis we will focus on the weak form efficiency proposed by Fama, which as described in the next section is closely related to the random walk in a stock returns. Therefore from now on, whenever we refer to EMH we will refer to its weak form, if not stated explicitly otherwise.

In order for the weak form efficiency to be testable we will adapt the following definition proposed by Jensen [25]: *A market is efficient with respect to information set* Θ_t *if it is impossible to make economic profits by trading on the basis of information set* Θ_t. In the case of weak form efficiency Θ_t is taken to be the information contained in the past price history of the market up to time t. This means that investors can not consistently earn profit if their decisions are based solely on the past market prices, which is in line with the definition of weak form efficiency. The figure below taken from J.Wang [51] illustrates the idea of efficient market.

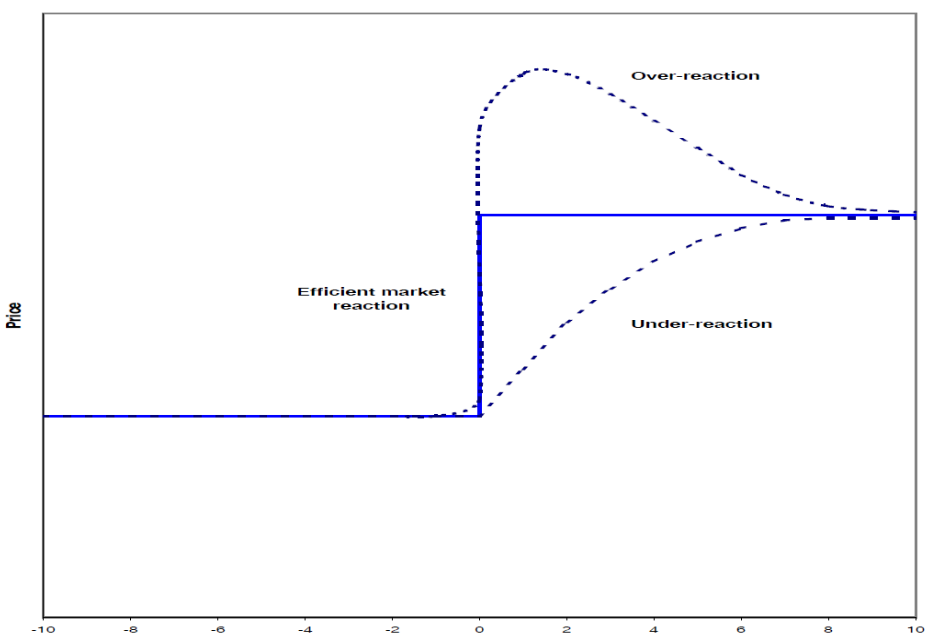


Figure 2: Stock market price adjustment after the announcement of positive publicly available information. Source: J.Wang [51]

Three possible adjustments are illustrated after the information announcement. In case of *over and under reaction* the market does not fully reflect the new available information and incorporates it with time delay. This is the cause of its inefficiency as opposed to efficient market reaction, which reflects the new information immediately after its announcement.

3.2 Joint hypothesis problem

Fama [10] notes that we can only test whether information is properly reflected in prices in the context of a pricing model that defines the meaning of "properly". Fama [11] states that it is not possible to test the EMH on its own and that it needs to be tested jointly with some model that reflects market expected equilibrium returns or prices. This fact will be referred to as the *Joint hypothesis problem*. We will therefore first construct a theoretical model that will represent an efficient market behaviour. This model will be compared to our model estimated from data. The result of this comparison will determine whether we reject the EMH or not.

Fama [14] states that test of market efficiency relates observed prices to equilibrium prices in the sense that under EMH the observed prices should exhibit the properties of the equilibrium prices. Therefore the correct choice of our theoretical asset pricing equilibrium model and correct estimation of our empirical model will be crucial parts of our analysis. We will discuss the choice of appropriate equilibrium model in Section 6 in more depth.

3.3 Efficient Markets and Random Walk Hypothesis

Fama [12] defined a market to behave randomly if successive price changes are independent of each other and therefore the best prediction of a price of a security tomorrow is its price today and not any other price from the past. The random walk hypothesis (RWH) states that past prices cannot be used to find a more accurate forecast of a price of a security next day than today's price. There is then no correlation between the price changes on different days and no information in past prices useful for forecasting future prices (Lima, Ohashi [32]). The above definition in context of a stock market means that it is impossible for investors to consistently outperform the market if their models are based solely on past security returns or prices. The precise relation between RWH and EMH has been highly discussed topic in literature (see Lo, MacKinlay [33] and Jensen [25]). The cause of a random walk behaviour of a market might be due to the fact that investors choose their assets consistently at random (Grimsved and Pavia [20]). We consider such behaviour to be highly unlikely and therefore will take the RWH and EMH to be equivalent. The Random Walk of a time series is in line with the definition of weak form market efficiency and therefore will its rejection imply rejection of the weak form market efficiency as well. In our paper we will test the random walk behavior specified in this sub-section and based on the empirical outcome of our model we will decide, whether the Prague stock exchange is indeed weak form efficient or not.

4 Markov Chains theory

In this section we will lay down the basic theoretical foundations of the Markov Chain theory. Markov chains have firstly been introduced by a Russian mathematician Andrei Markov in 1906. Markov's primary motivation for development of this theory was to extend the Bernoulli's theorem of *weak law of large numbers*, which assumes independence of random events. In 1913 Markov in his famous paper proved that under a certain conditions weak law of large numbers holds even for dependent random variables. We will later see that these are the exact conditions necessary for a Markov chain to converge to its stationary distribution. For more historical context of Markov chains refer to Gagniuc [18].

Note that our text provides only a brief introduction to this extensive theory and reader is encouraged to see other more detailed monographs on this topic[3]. For the purposes of our thesis we will focus solely on the theory of discrete states and time Markov chains, however, further extensions to continuous time and states have been widely studied and found many practical applications[4]. We will predominantly use the texts of Konstantopoulos [29] and Grimmet and Stirzaker [19] as our main reference sources for purposes of this introduction to Markov chain theory.

4.1 First order Markov Chains

Consider a finite set of states $S = \{1, \dots, m\}$, referred to as a state space from now on. Consider a discrete time random process $\{X_n : n \in Z_+\}$ that is mapped onto this state space and transitions from one state of a state space to another. Each transition between states in S will be called a *step.* The key assumption of the first order Markov Chain is called the *Markov Property* (sometimes also called *memory-less property*). The interpretation of Markov Property is that the future action of the process is not dependent upon the steps that led up to the present state, given the present state. The

[3] For example *Markov Chains* by Norris [37] or *Markov Chains: Gibbs fields, Monte Carlo simulation, and queues* by Bremaud [6].

[4] For more information see Yin, Zhang [52]

process that satisfies this property is called a first order Markov chain. The formal definition of a first order Markov chain (or simply Markov chain from henceforward) is given below.

Definition 1. *Let S be a state space. The process $\{X_n : n \in Z_+\}$ is a Markov chain of order 1 if it satisfies the Markov property*

$$\Pr(X_n = j \mid X_0 = i_0, \ldots, X_{n-2} = i_{n-2}, X_{n-1} = i) = \Pr(X_n = j \mid X_{n-1} = i),$$

$\forall i_0, \ldots, i_{n-2}, i, j \in S$ *and* $\forall n \in Z_+$.

When testing the RWH we will also use the *zero order Markov chain* defined below.

Definition 2. *Let S be a state space. The process $\{X_n : n \in Z_+\}$ is a Markov chain of order 0 if it satisfies the independence on previous states*

$$\Pr(X_n = j \mid X_0 = i_0, \ldots, X_{n-2} = i_{n-2}, X_{n-1} = i) = \Pr(X_n = j),$$

$\forall i_0, \ldots, i_{n-2}, i, j \in S$ *and* $\forall n \in Z_+$.

Example 1. *We will demonstrate the definition of Markov chain by a simple problem. Let us consider the classical example of drawing balls from the box. If we allow sampling with replacement that is lagged one time period, we have a process that satisfies the Markov property as any draw depends only on the previous one. In this case each draw ball is put aside and returned back to the box only after the next one is taken out. This way we always have information about exactly one ball, which is in line with the definition of first order Markov chain*[5]*. If we allow the replacement immediately after the ball is drawn we have a Markov chain of order 0.*

[5]Note that the number of time periods the replacement of a ball is lagged will represent the order of a Markov chain. This intuition will be relevant when we consider higher order Markov chains in section 4.3.

From the definition of a Markov chain we can derive the joint probability distribution of this chain that can be expressed as:

$$\Pr(X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i, X_n = j) = \Pr(X_0 = i_0) \times \Pr(X_1 = i_1 \mid X_0 = i_0)$$
$$\times \ldots \times \Pr(X_n = j \mid X_{n-1} = i), \forall i_0, \ldots, i_{n-2}, i, j \in S, \forall n \in Z_+.$$

We will refer to the functions

$$\begin{aligned} &\mu(i_0) = \Pr(X_0 = i_0), i_0 \in S, \\ &p_{ij}(n-1, n) = \Pr(X_n = j \mid X_{n-1} = i), \forall i, j \in S, n \geq 0, \end{aligned} \tag{1}$$

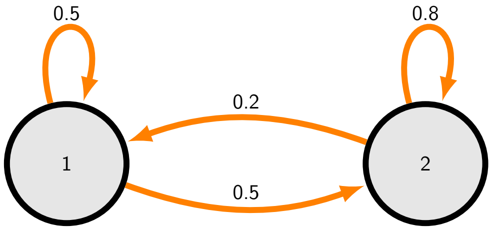
as *initial distribution* and *transition probability* from state i at time n-1 to state j at time n, respectively. Next, we will define the core concept of transition probability matrix (TPM) of the Markov chain X.

Definition 3. *Let S be a state space and X a first order Markov chain. The transition probability matrix*

$$P(n-1, n) = [p_{ij}(n-1, n)]_{i,j \in S} \tag{2}$$

is a $m \times m$ square matrix, where m denotes the number of states in S.

Transition matrix therefore contains all underlying probabilities that connect the states of a state space together. The visualization of a Markov chain is illustrated in the figure below, where we consider state space $S = \{1, 2\}$ consisting of two states. The corresponding transition matrix is also displayed to demonstrate the link between the graphical and analytical representation of the Markov chain.
The transition diagram is constructed from the TPM in a straightforward way. Each transition probability corresponds to one arrow drawn from a particular state. The diagonal entries of the TPM are the probabilities of Markov chain to return back to the same state as it was one step before and are graphically represented by the loops above the states. Note that there is one to one correspondence between the transition matrix and the transition diagram.



(a) Transition diagram

$$\begin{pmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{pmatrix}$$

(b) Transition matrix

Figure 3: Two state Markov Chain

One of the most important properties of a Markov chain is *time homogeneity*.

Definition 4. *Let X be a Markov chain defined on state space S. This chain is said to be time homogeneous if:*

$$\forall n \in Z_+, \forall i, j \in S : p_{ij}(n-1, n) = p_{ij}.$$

This property states that the probability of any state transition is independent of time. From now on we will refer to p_{ij} as one step transition probability from state i to state j (or just transition probability), while

$$P := P(n-1, n) = [p_{ij}]_{i,j \in S} \tag{3}$$

will be referred to as one step transition probability matrix (or simply transition matrix). Markov chain for which the above definition does not hold is said to be *non-homogeneous*. We will later see that the property of time homogeneity is a crucial assumption that needs to be verified if we want to test the EMH by Markov chain methodology. Formal test of time homogeneity of our model will be presented in section 6. From henceforward we will consider only time homogeneous Markov chains if not explicitly stated otherwise.

We can use the transition matrices to calculate transition probabilities k steps ahead in time by taking their products[6].

$$P(n-k,n) = \underbrace{P \times P \times \ldots \times P}_{\text{k times}} = P^k \tag{4}$$

We will refer to the above transition matrix as *k-step transition matrix.*

Example 2. *We will demonstrate the two step transition matrix P of a Markov chain. Consider a Markov chain X defined on a state space $S = \{1, 2\}$ with the one step transition matrix P displayed below:*

$$P = \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{5}{6} \end{bmatrix}.$$

Now consider the two step transition probabilities. Let's say that the Markov chain is in a state 1 and we would like to know the probability that it will be in state 1 again after two steps. There are two possibilities how such event can occur. First, the Markov chain transitions to state 1 in first step and than again in step 2 (the probability of this event is $5/6 * 5/6$) or it transitions to state 2 in step one and than back to state 1 in step 2 (the probability of this event is $1/6 * 1/6$). The sum of these probabilities is $13/18$. The remaining two step transition probabilities are calculated in a straightforward way as shown below:

$$p_{11}(n-2,n) = 5/6 * 5/6 + 1/6 * 1/6 = 13/18$$
$$p_{12}(n-2,n) = 5/6 * 1/6 + 1/6 * 5/6 = 5/18$$
$$p_{21}(n-2,n) = 5/6 * 1/6 + 1/6 * 5/6 = 5/18$$
$$p_{22}(n-2,n) = 5/6 * 5/6 + 1/6 * 1/6 = 13/18.$$

We obtain the same results by multiplying the transition matrix P by itself.

$$P^2 = \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{5}{6} \end{bmatrix} \times \begin{bmatrix} \frac{5}{6} & \frac{1}{6} \\ \frac{1}{6} & \frac{5}{6} \end{bmatrix} = \begin{bmatrix} \frac{13}{18} & \frac{5}{18} \\ \frac{5}{18} & \frac{13}{18} \end{bmatrix}$$

[6]This property of a transition matrices comes from the Chapman-Kolmogorov equation not discussed in our text. For more detail see Konstantopoulos [29]

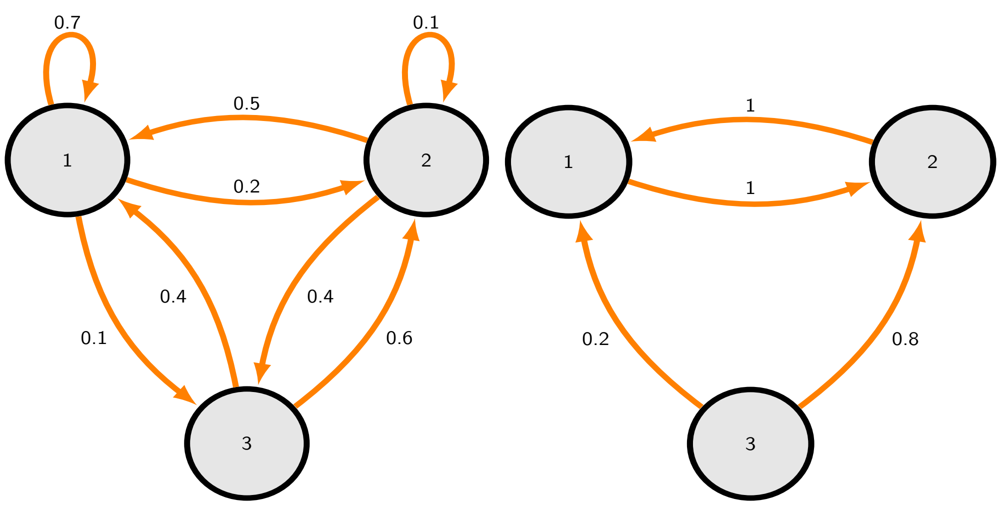
We see that the property of time homogeneity enables us to use the same transition matrix P for each time transition. This fact is very convenient as we do not have to calculate each transition probability separately when we want the probabilities of more than one time periods ahead.

The interesting question that naturally arises is whether these probabilities converges to some limiting values as we increase the number of transition steps. This question will be answered in the subsequent section and later when testing the stability of our Markov model demonstrated in practise.

4.2 Properties of Markov Chains

In this subsection we will introduce some key definitions and concepts regarding Markov chains that we will use henceforward in our thesis.

Definition 5. *Any state $i \in S$ is said to be recurrent if when the Markov chain starts at state i there is always a non-zero probability that the it will return back, otherwise it is called transient.*



(a) Three recurrent states

(b) Two recurrent and one transient state

Figure 4: Recurrent and transient states

Definition 6. *A state $i \in S$ is said to be absorbing if $p_{ii} = 1$.*

Therefore an absorbing state is one from which it is impossible to escape with with probability 1.

Definition 7. *We say that states $i, j \in S$ communicate with each other, if the Markov chain can transition between states i and j with non-zero probability.*

Another important property of a Markov chain that is closely related to *communicating* states is *irreducibility.*

Definition 8. *Let X be a Markov chain defined on a state space S. The chain X is said to be irreducible if all states in S communicate with each other.*

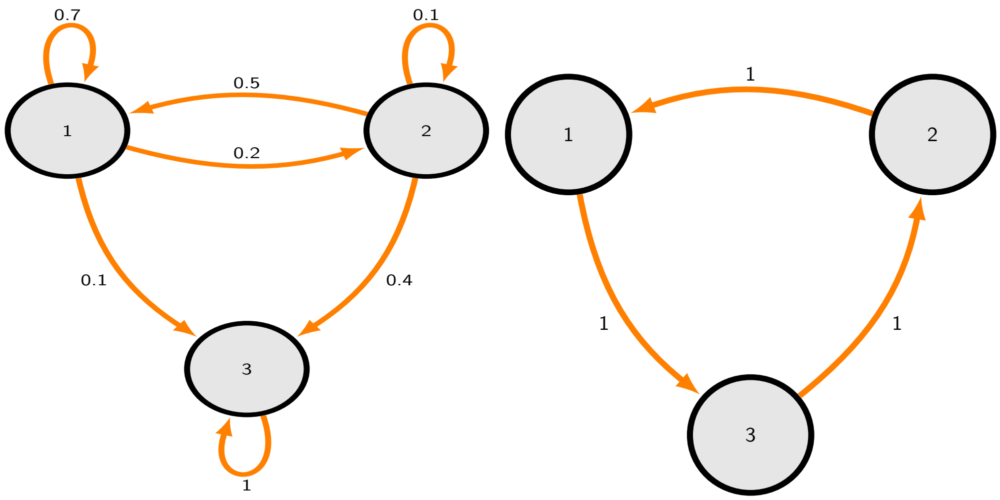
In simple words the definition of irreducibility means that it is possible to get from any state of a chain to any state within a chain with non-zero probability. Therefore the chain in the Figure 4 a) is irreducible, whereas chain in Figure 4 b) is not. It is clear that all states in irreducible chain must be recurrent, assuming the finite state space (note that the equivalence does not hold).

Definition 9. *Let X be a Markov chain defined on state space S. The state $i \in S$ is said to have period d_i, where d_i is:*

$$d_i = gcd\{m : P(X_m = i \mid X_0 = i) > 0\}.$$

Here gdc stands for a greatest common divisor. A chain is said to be aperiodic if $\forall i \in S : d_i = 1$

The following figure summarizes the above mentioned definitions.



(a) Absorbing state

(b) Markov chain with periodic states

Figure 5: Periodic Markov chain and absorbing state

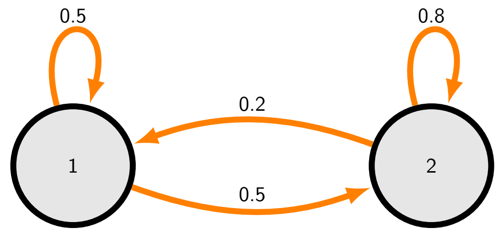
The following proposition will later enable us to validate the assumptions of our Markov chain model conveniently without the need for explicit validation of irreducibility and aperiodicity.

Proposition 1. *Markov chain X defined on a state space S is aperiodic and irreducible if and only if its corresponding transition probability matrix is strictly positive (all its entries are greater than zero).*

4.3 Stationary distribution

Having already the appropriate theoretical background we will study the limiting probabilistic behaviour of the Markov chain. It is possible to show that under certain assumptions we can say that a Markov chain will converge to some stationary distribution (sometimes also refer to as limiting or equilibrium distribution). We will first illustrate the idea of stationary distribution on example before stating the formal definition.

Example 3. *Let X be a Markov chain defined on a state space $S = \{1, 2\}$. The transition diagram of our chain is visualized in the figure below.*



(a) Transition diagram

	n=1	*n=2*	...	*n=100*	*n=101*
$p_{11}(n)$	*0.5*	$0.5 \times 0.5 + 0.5 \times 0.2 = 0.35$	...	$\approx 2/7$	$\approx 2/7$
$p_{12}(n)$	*0.5*	$0.5 \times 0.5 + 0.5 \times 0.8 = 0.65$	...	$\approx 5/7$	$\approx 5/7$
$p_{21}(n)$	*0.2*	$0.8 \times 0.2 + 0.2 \times 0.5 = 0.26$	...	$\approx 2/7$	$\approx 2/7$
$p_{22}(n)$	*0.8*	$0.2 \times 0.5 + 0.8 \times 0.8 = 0.74$	...	$\approx 5/7$	$\approx 5/7$

At time n=0 If the Markov chain starts at state 1 it is certain to be in this state with probability 1, same logic holds for state 2. For time n=1 we will use the one step transitions represented by one step transition matrix P.

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.2 & 0.8 \end{pmatrix} \tag{5}$$

To calculate the probability that when Markov chain starts at state $i \in S$ at time n=0 it will be in state $j \in S$ in two time periods we multiply the transition matrix P by itself.

$$P^2 = \begin{pmatrix} 0.35 & 0.65 \\ 0.26 & 0.74 \end{pmatrix} \tag{6}$$

If we consider the transitions for arbitrary many time periods into the future we notice an interesting phenomenon. We see that the probabilities that the Markov chain is in a state 1 or 2 after the process has run for a long time settle to some steady values.

$$P^{100} \approx \begin{pmatrix} 2/7 & 5/7 \\ 2/7 & 5/7 \end{pmatrix} \tag{7}$$

We also see that $p_{11}(n) = p_{21}(n)$ and $p_{12}(n) = p_{22}(n)$ for some large n. This fact means that the long run probabilities are independent of the initial *starting state of the Markov chain. Notice that once the transition probability converges to its limiting value it will not change. For example consider the transition* $p_{11}(101) \approx 2/7 \times 1/2 + 5/7 \times 1/5 = 2/7$. *We see that the probability that the Markov chain will be in state 2 in any given time is larger than that it will be in state 1. This is not surprising as we see that the Markov chain will on average tend to spend in state 2 more time (the probability of exiting state 2 is 0.2 as opposed to staying there with probability 0.8).*

We will refer to these long run probabilities described in the previous example as *steady state probabilities*. Generally the notation that we will use is π_j for a steady state probability of state j. We will also use the vector notation $\pi = (\pi_1, \ldots, \pi_n)$ to describe *stationary distribution* of a Markov chain. So the stationary distribution of the Markov chain from the previous example is:

$$\pi = \left(\frac{2}{7}, \frac{5}{7}\right).$$

The following proposition will give us a necessary condition for a probability vector π to be stationary distribution of a given Markov chain.

Proposition 2. *A probability vector* π *is a stationary distribution of a Markov chain with transition probability matrix* P *if:*

$$\pi = \pi \times P.$$

We say that π *is invariant by the transition matrix* P.

The steady state probabilities will help us answer questions such as: what is the probability that the Markov chain is in a certain state at any given time or what is the expected return time of a Markov chain to certain state. The expected return time denotes the expected number of steps between consecutive visits of a Markov chain to a certain state. The stationary distribution and time required for the convergence of the Markov chain to such stationary distribution can reveal many important information about a given Markov chain, especially about its stability and limiting behaviour.

The following proposition will state the necessary conditions under which the Markov chain does converge to its stationary distribution.

Proposition 3. *Let X be irreducible Markov chain defined on state space S, then:*

1. *stationary distribution π exists and is unique*
2. *Ratio $\frac{n_j(n)}{n} \to \pi_j$, where $n_j(n)$ is the number of times X visits a certain state $j \in S$ in n transitions and n denotes the total number of transitions of X.*
3. *The expected return time to state $j \in S$ will be denoted ω_j and is equal to $\frac{1}{\pi_j}$*
4. *if the Markov chain is aperiodic, then:* $\Pr(X_n = j) \to \pi_j$ $j \in S$ *as* $n \to \infty, n \in Z_+$.

Note that the expected return times to states 1 and 2 from Example 3 are:

$$\omega_1 = \frac{1}{\pi_1} = 7/2$$
$$\omega_2 = \frac{1}{\pi_2} = 7/5.$$

The Proposition 3 will be a crucial reference point for determining the stationary distribution of our Markov chain model.

4.4 Higher order Markov Chains

In this section we will generalize the definition of a first order Markov chain to higher order Markov chain and allow for time dependency of our random process more than one time period into the past. The definition of a Markov chain of order u is therefore a natural extension of the definition given for the first order Markov chain.

Definition 10. *Let S be a state space. The process $\{X_n : n \in Z_+\}$ is a Markov chain of order u if it satisfies*

$$\begin{aligned} &\Pr(X_n = j \mid X_0 = i_0, \ldots, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}) \\ &= \Pr(X_n = j \mid X_{n-u} = i_{n-u}, \ldots, X_{n-1} = i_{n-1}), \forall i_0, \ldots, i_{n-1}, j \in S, \forall n \in Z_+. \end{aligned}$$

Before formally defining the transition probabilities for higher order Markov chain, we will again illustrate the idea on example.

Example 4. *Consider two Markov chains V and W defined on a state space $S = \{1, 2\}$, where first is of order 1 and the second of order 2. Below are their corresponding transition probability matrices*

$$P_V = \begin{array}{c} \\ 1 \\ 2 \end{array} \!\!\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \end{array}$$

$$P_W = \begin{array}{c} \\ 11 \\ 12 \\ 21 \\ 22 \end{array} \!\!\begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{pmatrix} p_{111} & p_{112} \\ p_{121} & p_{122} \\ p_{211} & p_{212} \\ p_{221} & p_{222} \end{pmatrix} \end{array}.$$

Here the notation $p_{ijk}, i, j, k \in S$ denotes the probability that Markov chain W has followed a certain trajectory $i \to j \to k$ of transitions between states of a state space S. We see that Markov chain W no longer has a square matrix of transition probabilities as in the case of V. This is due to the fact that we have to consider all possible sequences of states two time periods into the past. This remark can further be generalized for transition probability matrices of u:th order Markov chain with m states, that will be defined in the formal definition.

Definition 11. *Let* $S = \{1, \ldots, m\}$ *be a state space and* $S^u = \{i_0 \ldots i_{u-1} : \forall i_k \in S\}$ *be a space containing all possible sequences of length* u *consisting of states* $i \in S$. *Consider Markov chain* X *of order* u. *The transition probability* $p_{ij}(n-u, n)$ *to end up in state* $j \in S$ *at time* n *after having followed the path described by sequence* $i \in S^u$ *is defined as:*

$$p_{ij}(n-u, n) = \Pr(X_n = j \mid X_{n-1} = i_{u-1}, \ldots, X_{n-u} = i_0),$$

where $i = i_0 \ldots i_{u-1} \in S^u, j \in S$. *The transition probability matrix* $P(n-u, n) = [p_{ij}(n-u, n)]_{i \in S^u, j \in S}$ *is then* $m^u \times m$ *matrix of transition probabilities* $p_{ij}(n-u, n)$.

Note also that the number of degrees of freedom needed to uniquely describe the TPM of higher order Markov chain grows exponentially. For example the number of degrees of freedom required to describe the TPM P_V from Example 4 is 2, whereas the TPM P_W requires 4 degrees of freedom. This comes from the fact that each row of a TPM needs to sum up to 1. This fact will play an important role when estimating the optimal order of our Markov chain model. We will define the time homogeneity of a transition matrix for higher order Markov chains.

Definition 12. *Let* X *be a Markov chain of order* u *defined on a state space* S. *Then* X *is called time homogeneous if*

$$p_{ij}(n-u, n) = p_{ij}(0, u),$$

$\forall n \geq u, i \in S^u, j \in S$. *Here* p_{ij} *denotes the probability of each transition following the sequence* i *to* j.

The definitions of *irreducibility and periodicity* of a Markov chain defined for first order can be easily extended for higher orders as well as the definitions of *recurrent, transient, communicating and absorbing* states. Note that these definitions are not related to the order of a Markov chain and therefore will not be restated in this section again. We will not consider the stationary distribution of higher order Markov chains, for more information on this topic see e.g. *Markov Chains and Stochastic Stability* by Meyn and Tweedie [35]. Additional resources used to conduct this introductory section to Markov chain theory include Blitzstein and Hwang [5] and Bertsekas et al. [3].

5 Maximum likelihood estimator of TPM

In the previous section we have seen that the TPM gives us a powerfull tool to analyze various properties and characteristics of the Markov chain model. Unfortunately the *true* TPM is in reality rarely known and we have to estimate it from obtained data. In this section we will derive the Maximum likelihood estimator (MLE) for the finite, discrete Markov chain that will be used to estimate the TPM. For its derivation we will use the optimization method of *Lagrange multipliers*. For the purposes of clarity of the derivation we will first consider a first order Markov chain $X = \{X_n : n \in Z_+\}$ defined on a state space $S = \{1, \dots, m\}$ and later generalize the MLE for higher orders Markov chain.

5.1 First order MLE

From data we will get a particular realization of X:

$$X = \{X_0 = i_0, X_1 = i_1, \dots, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}, X_n = i_n\},$$

$\forall n \in Z_+, \forall i_0, , i_n \in S$. This realization of the Markov chain will specify the joint probability distribution of X.

$$\Pr(X_0 = i_0, X_1 = i_1, \dots, X_n = j) = \Pr(X_0 = i_0) \times \Pr(X_1 = i_1 \mid X_0 = i_0)$$
$$\times \dots \times \Pr(X_n = i_n \mid X_{n-1} = i_{n-1})$$

We will rewrite the joint probability distribution in terms of the transition probabilities and initial distribution. This function is the likelihood function of a given transition matrix denoted $\mathcal{L}(p)$.

$$\mathcal{L}(p) = \mu(i_0) p_{i_0 i_1} \dots p_{i_{n-1} i_n} = \mu(i_0) \prod_{t=1}^{n} p_{i_{t-1} i_t}$$

We will further define n_{ij}, which will denote the number of times state $i \in S$ is followed by state $j \in S$ in our realization of Markov chain X in n steps i.e. $\sum_{i,j \in S} n_{ij} = n$. We will rewrite the likelihood function in terms of n_{ij} and the transition probabilities p_{ij} as:

$$\mathcal{L}(p) = \mu(k) \prod_{i=1}^{m} \prod_{j=1}^{m} p_{ij}^{n_{ij}}.$$

For the convenience of computation we will take the log-likelihood function denoted $\ell(p)$.

$$\ell(p) = log(\mu(k)) + \sum_{i=1}^{m}\sum_{j=1}^{m} n_{ij} log(p_{ij})$$

In order for TPM to be valid the probabilities of making transition from a given state have to add up to 1 and be non-negative. That is, for each i:

$$\sum_{j=1}^{m} p_{ij} = 1, \forall i \in S$$

$$p_{ij} \geq 0, \forall i, j \in S.$$

These will be the constraints of our optimization problem. We have m equality constraints, one for each state of S. Therefore we need m Lagrange multipliers: $\lambda_1, \ldots, \lambda_m$ to formulate the Lagrange function *L(p)* that we will maximize:

$$L(p, \lambda) = \ell(p) - \sum_{i=1}^{m} \lambda_i \Big(\sum_{j=1}^{m} p_{ij} - 1 \Big). \tag{8}$$

Taking into account the non-negativity constraints $p_{ij} \geq 0, \forall i, j \in S$, the resulting first order optimality conditions read

$$p_{ij}^{\star} \frac{\partial L(p^{\star}, \lambda^{\star})}{\partial p_{ij}} = n_{ij} - p_{ij}^{\star}\lambda_i^{\star} = 0, \forall i, j \tag{9}$$

$$\frac{\partial L(p^{\star}, \lambda^{\star})}{\partial \lambda_i} = \sum_{j=1}^{m} p_{ij}^{\star} - 1 = 0, \forall i. \tag{10}$$

Now if we substitute from (9) to (10) and solve for $p_{ij}^{\star}$ we get the desired result of the maximum likelihood estimate of the TPM for first order Markov chain:

$$\widehat{p_{ij}} := p_{ij}^{\star} = \frac{n_{ij}}{\sum_{j=1}^{m} n_{ij}}. \tag{11}$$

The nominator in (13) as already defined denotes the number of times state $i \in S$ is followed by a particular state $j \in S$ in n steps. The denominator counts the total number of transitions from this state i to any state $j \in S$ in these n steps. From the final result we see that the MLE depends solely on this ratio. We will verify that $\widehat{p_{ij}}$ is indeed a point of maximum by taking the second partial derivatives of the Lagrange function with respect to p_{ij}

and λ_i.

$$\frac{\partial^2 \ell(p,\lambda)}{\partial^2 p_{ij}} = -\frac{n_{ij}}{p_{ij}^2}, \forall i,j \tag{12}$$

$$\frac{\partial^2 \ell(p,\lambda)}{\partial^2 \lambda_i} = 0, \forall i \tag{13}$$

We see that the term $-\frac{n_{ij}}{p_{ij}^2}$ is strictly negative (for $n > 0$). This holds for all $i, j \in S$, note that the resulting *Hessian matrix* of second derivatives will have a negative values on diagonal and therefore by definition will be negatively definite. This fact implies that the point $\widehat{p_{ij}}$ is indeed maximum as we have a concave function on convex set. Moreover, from the strict concavity we know that it is the point of *global* maximum. Douc et al. [9] showed that the MLE of TPM is asymptotically normally distributed and that it is a consistent estimator of transition probabilities. We will therefore use the MLE method when estimating our Markov chain model.

5.2 Higher order MLE

Let us now consider a Markov chain of order 2. Consider again a particular realization of the Markov chain of order two.

$$X = \{X_0 = i_0, X_1 = i_1, X_2 = i_2, \dots, X_{n-2} = i_{n-2}, X_{n-1} = i_{n-1}, X_n = i_n\}$$

$i_0, \dots, i_n \in S$. This realization of the Markov chain will specify the joint probability distribution of X.

$$\begin{aligned}
&\Pr(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = \Pr(X_0 = i_0) \times \Pr(X_1 = i_1 \mid X_0 = i_0) \\
&\times \Pr(X_2 = i_2 \mid X_0 = i_0, X_1 = i_1) \times \Pr(X_3 = i_3 \mid X_1 = i_1, X_2 = i_2) \times \dots \\
&\times \Pr(X_n = i_n \mid X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2})
\end{aligned}$$

Same as in the case of first order case we can define n_{kij} to represent the number of times Markov chain has followed a certain path $k \to i \to j$ in n steps. We can than write the likelihood function in terms of the transition probabilities p_{kij} and counts n_{kij}.

$$\mathcal{L}(p) = \mu(i_0) p_{kl} \prod_{k=1}^{m} \prod_{i=1}^{m} \prod_{j=1}^{m} p_{kij}^{n_{kij}}$$

Maximizing the log-likelihood function with respect to $\sum_j p_{kij} = 1, \forall k, i$ and $p_{kij} \geq 0, \forall k, i, j$ we get:

$$\widehat{p_{kij}} = \frac{n_{kij}}{\sum_{j=1}^{m} n_{kij}} \tag{14}$$

We can further generalize the MLE for higher orders Markov chain, however, we will omit the full deduction and only state the MLE formula. The full derivation can be found in Billingsley [4]. In MLE formula, $\{k \dots ij\}$ denotes an arbitrarily long sequence of states. The number of states in the sequence before the state j will represent the order of the chain.

$$\widehat{p_{k \dots ij}} = \frac{n_{k \dots ij}}{\sum_{j=1}^{m} n_{k \dots ij}}. \tag{15}$$

6 Markov chain model

This section will outline the theoretical approach to testing the EMH using the Markov chain methodology. In section 3.2 about the *Joint hypothesis problem*, we have already discussed the importance of choosing the correct theoretical model representing the efficient market to which will our empirical model be compared. For the means of our analysis, this theoretical model will be represented by a Markov chain of order 0. This model can be viewed as a sequence of mutually independent random variables and thus represents a random walk in a stock price returns.
We start constructing our empirical Markov chain model by mapping the time series onto a predefined states of a state space. Next, we test the random walk hypothesis by comparing the Markov chain of order 0 to Markov chain of the *optimal* order that will be established by Bayesian Information criterion[7]. Finally, we will test the time homogeneity of our model. This three step approach has been to large extent inspired by Tan and Yilmaz[8] [49]. The assumptions for our Markov chain model will be *irreducibility, aperiodicity and time homogeneity*. We will validate these assumptions in the section 8. We will highlight the outline of our work in Figure 6 and then address each section individually.

[7]The Bayesian Information Criterion will be described in full detail in section 6.2.3

[8]The method for establishing the *optimal* order of a Markov chain by Goodness of fit test as proposed by Tan and Yilmaz (2002) has been found unreliable. This will be further discussed in section 6, where more precise method for assessing the *optimal* order will be presented.

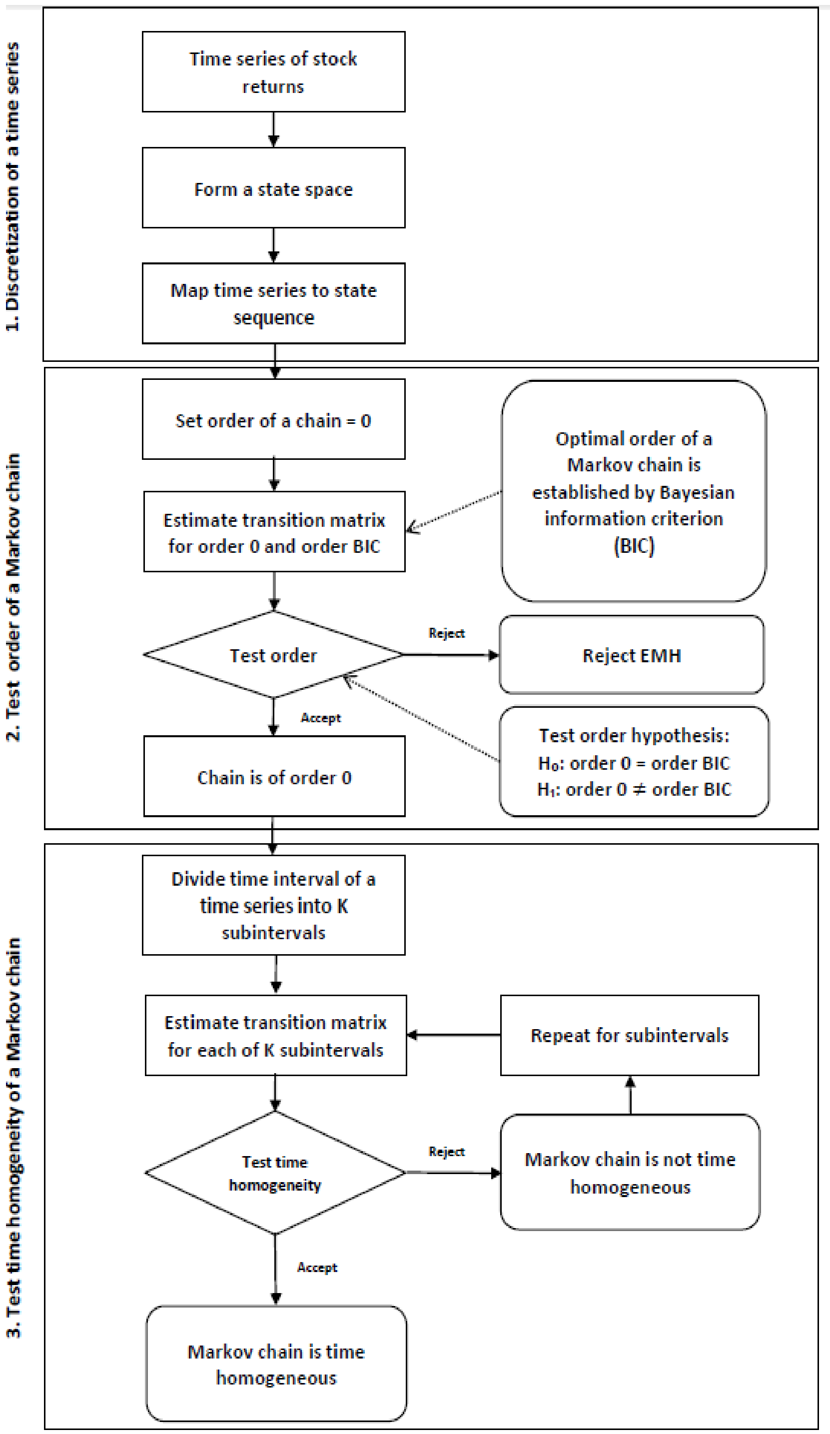


Figure 6: Schematic outline

6.1 Discretization of a time series

First, we need to establish a number of states of our state space, and then define a mapping function that will map our time series onto these states. Many various state space discretizations have been suggested in the literature on Markov chains and stock markets. For example Lukáš and Svoboda [47] used 8 states Markov chain to model Prague PX index. However, as Tan and Yilmaz [49] point out, for purposes of testing the EHM, the most popular approach is to use two state space discretization. These two suggested states represent the above and below average returns of a time series.

We will adapt this proposition in our thesis as well and decide whether to reject the EMH or not based on the results obtained by this model. We will work with *daily, weekly and monthly* stock returns of the Prague Stock Exchange PX index from year 1994 to 2017. They represent the aggregated closing prices for each of the frequencies [9]. Closer analysis of our data will be provided in Section 7. In the following subsections we will first introduce how the stock returns are calculated and then present the two state Markov chain model.

6.1.1 Price returns and Two state Markov chain model

Based on the work of Grimsved and Pavia [20] we will consider the stock returns as follows. Let P_i be a price of a PX index at time $i \in \{0, \ldots, T\}$. The return change between consecutive time periods, denoted $r_i, i \geq 1$, during the time period $(i-1, i]$ is calculated as:

$$r_i = \frac{P_i - P_{i-1}}{P_{i-1}}.$$

The Markov chain model that will be used to test the EMH is constructed below. In this model we consider a state space S of two states "Increase" (I) and "Decrease" (D), therefore our model will be discrete in both time and space. We will further denote $r_t, t \in \{0, \ldots, T\}$ to be a time series of all PX returns r_i. The term $\mathbb{E}(r_t)$ represents the expected return of the time series

[9] Data used in our thesis can be obtained at: https://www.quandl.com/data/PRAGUESE/PX-Prague-Stock-Index-PX.

r_t for the whole time period.

$$X(t) = \begin{cases} I & \text{if } r_t \geq \mathbb{E}(r_t) \\ D & \text{if } r_t < \mathbb{E}(r_t) \end{cases} \tag{16}$$

6.2 Determining the order of a chain

Our primary goal is to test the hypothesis whether the Markov model estimated from data is of zero order or of some higher order k. The former would imply that the stock returns follow a random walk and that we fail to reject the EMH. To test this hypothesis properly we will proceed in two steps. First the *optimal* order of a Markov chain will be established using the Bayesian Information Criterion (BIC). Secondly, the comparison between the zero order Markov chain and the chain of the established *optimal* order will be made using the *Likelihood ration criterion*.

If we reject the null hypothesis that the orders are equal, the hypothesis of weak form market efficiency will be rejected as well and the order of a chain will be taken to be the one established by BIC.

6.2.1 Testing methods

There exists no conventional approach to testing the *optimal* order of a Markov chain. One of the most frequent methods of tackling such problem has been the *Pearson's* χ^2 test of goodness of fit (also called the χ^2 test), proposed by Karl Pearson in his paper in 1900.

The χ^2 test uses the sequential hypotheses testing in order to derive the *optimal* order of a Markov chain. This approach has, however, lead to various forms of criticism throughout the 20^{th} century. Rajarshi [41] argues that the convergence to the *optimal* order of a Markov chain might not be possible when using this method and proposes the usage of the Akaike information criterion (AIC) or Bayesian information criterion (BIC) for obtaining better theoretical results instead. Both of these methods are based on penalizing the increasing order of the Markov chain as will be shown in more detail in the following sub-section.

The order established by BIC is obtained by consecutive tests of the highest allowed order M against all lower orders $0, 1, \ldots, M-1$ as Katz [26] shows. However, if BIC establishes the *optimal* order to be 0, than BIC says only that this order represents the data best when penalty for higher orders is taken into account. It does not reveal if there is any dependence structure in the data as explained in Grimsved and Pavia [20].

For the means of our empirical analysis we will therefore proceed as illustrated in the Figure 6 and first determine the optimal order $i \in \{1, \ldots, M-1\}$ by BIC. This will be done by sequentially testing orders $\{1, \ldots, M-1\}$ against highest established order M and then selecting the most *optimal* among these. The proper meaning of the *optimal* order will be addressed in the upcoming sections. Then, *Likelihood ration criterion* will be used to test whether the Markov chain of order i is statistically different from order 0 or not. We will first derive the Likelihood ratio criterion and then BIC.

6.2.2 Likelihood ratio criterion

The key part of testing whether our empirical Markov chain model is of order 0 or of some higher order k will be conducted with the use of *Likelihood ratio criterion*. For the purposes of demonstration of the *Likelihood ratio criterion* we will consider a Markov chain of order 1 and Markov chain of order 2, defined on a state space $S = \{1, \ldots, m\}$. We will state the hypothesis as follows:

$$H_0 : \text{Markov chain is of order 1}$$

$$H_1 : \text{Markov chain is of order 2 or higher}$$

Under the null hypothesis, $p_{1ij} = p_{2ij} = \ldots = p_{mij} = \widehat{p_{ij}}$. We will define the *Likelihood ratio criterion* to be:

$$\Lambda_{1,2} = \prod_{k=1}^{m} \prod_{i=1}^{m} \prod_{j=1}^{m} \left(\frac{\widehat{p_{ij}}}{\widehat{p_{kij}}} \right)^{n_{kij}} \tag{17}$$

The MLE for p_{kij} is the same as in (14), however the MLE for p_{ij} differs from the one derived in (13):

$$\widehat{p_{ij}} = \frac{\sum_{k=1}^{m} n_{kij}}{\sum_{k=1}^{m} \sum_{j=1}^{m} n_{kij}}. \tag{18}$$

The intuition for the MLE described in (18) is that the state i serves as a intermediate for all transitions coming to this state i and then to state j. In the nominator we count the number of transitions Markov chain has followed a certain trajectory "any state" $\rightarrow i \rightarrow j$ and the denominator counts the number of transitions of a type "any state" $\rightarrow i \rightarrow$ "any state". Here, of course, we consider all states to be from our state space S. Anderson and Goodman [2] show that when the logarithmic transformation of the *Likelihood ratio criterion* is taken, then, under the null hypothesis, the statistics

$$-2log(\Lambda_{1,2}) = 2\sum_{k=1}^{m}\sum_{i=1}^{m}\sum_{j=1}^{m} n_{kij} log\left(\frac{\widehat{p_{kij}}}{\widehat{p_{ij}}}\right)$$

becomes asymptotically χ^2_{df} distributed with $df = (m^2 - m)(m-1)$. Here, m denotes the number of states of a state space S. It is at this point where we consider appropriate to remind the reader of the formula for the χ^2 distribution as this distribution will be used extensively throughout our thesis. The probability density function of χ^2 distribution is equal to:

$$f(x;k) = \begin{cases} \frac{x^{\frac{k}{2}-1}e^{-\frac{x}{2}}}{2^{\frac{k}{2}}\Gamma\left(\frac{k}{2}\right)} & \text{if } x > 0, \\ 0 & \text{if } x \leq 0, \end{cases} \tag{19}$$

where Γ denotes the *Gamma function* and k the number of degrees of freedom. The generalization of the *Likelihood ratio criterion* is derived in a straightforward way when testing whether the Markov chain defined on a state space S is of order u against order v, where $u < v$.

$$H_0 : \text{Markov chain is of order u}$$

$$H_1 : \text{Markov chain is of order v or higher}$$

Under the null hypothesis $p_{1\ldots k\ldots ij} = p_{2\ldots k\ldots ij} = \ldots = p_{m\ldots k\ldots ij} = \widehat{p_{k\ldots ij}}$, where the sequence $\{k\ldots i\} \in S^u$, sequence $\{1\ldots k\ldots i\} \in S^v$ and $j \in S$. The *Likelihood ratio criterion* is then defined as:

$$\Lambda_{u,v} = \prod_{l=1}^{m}\cdots\prod_{i=1}^{m}\prod_{j=1}^{m}\left(\frac{\widehat{p_{k\ldots ij}}}{\widehat{p_{l\ldots k\ldots ij}}}\right)^{n_{l\ldots k\ldots ij}} \tag{20}$$

The MLEs are obtained in the similar fashion as in (14) and (17). The statistics $-2log(\Lambda_{u,v})$ has asymptotic χ^2_{df} distribution has with

$df = (m^v - m^u)(m-1)$ degrees of freedom. In the next subsection we will see that the *Likelihood ratio criterion* is a crucial part of the BIC.

6.2.3 Bayesian Information Criterion

The Bayesian Information Criterion was first proposed by Schwartz [44]. We will define a set of distinct Markov chain models $Y = \{Y_1, \ldots, Y_M\}$, where each $Y_u, u = 1, \ldots, M$ are Markov chains order u defined on a state space S. Here M denotes the highest allowed order of our Markov chain. For our purposes, the order M will be taken to be one for which all entries of the corresponding TPM are strictly positive. This is by Proposition 1 equivalent to taking the maximal order for which the TPM is irreducible and aperiodic. Note that with increasing order of the Markov chain the dimensionality of its TPM increases exponentially (see Example 5 for illustration). As a consequence of this fact the number of non-zero entries of TPM will decrease relatively fast with each additional order as some sequences of state transitions will eventually have probability zero.

Our goal is to find a Markov chain $Y_u \in Y$ with order that best describes our data. Increasing the order of Markov chain increases its complexity that we would like to penalize. Katz [26] shows the BIC introduces a penalty that gives a consistent estimation of the *optimal* order of a Markov chain. In our thesis we will therefore focus solely on the BIC.

To derive the BIC for a Markov chain we will consider two prior probabilities on models from set Y. First one, $\Pr(Y_i), Y_i \in Y$ represents our prior information about a certain Markov chain model and the second one $\Pr(\theta_i \mid Y_i)$ gives us prior information about the parameters of this Markov chain. Our goal is to maximize the posterior probability of a certain Markov chain from Y given the data-set D with n observations. That is to maximize:

$$\Pr(Y_i \mid D) = \frac{\Pr(Y_i)\Pr(D \mid Y_i)}{\Pr(D)} \tag{21}$$

Here we use the Bayes formula. Let us at this moment explain each of the terms in the above equation. The probability that we are most interested in is the term $\Pr(Y_i \mid D)$. That is given data, how well will model Y_i describe them. Term $\Pr(Y_i)$ is the prior probability on our models. This means how well we think model Y_i is suited for our task based on our past experience and accumulated knowledge. Term $Pr(D \mid Y_i)$ denotes how likely are our data to be generated by model Y_i. Lastly term $\Pr(D)$ is the probability of our data. As the probability $\Pr(D)$ is a normalization constant, we can further proceed with just a proportion of equation (21):

$$\Pr(Y_i \mid D) \propto \Pr(Y_i) \Pr(D \mid Y_i). \tag{22}$$

The equation (22) gives us a convenient way to compare two models from set Y. Assume that we want to compare model $Y_j \in Y$ with our initial model Y_i. We will take the ratio of their posterior probabilities and decide which one performs better on a given data-set D based on the magnitude of the ratio:

$$\frac{\Pr(Y_i \mid D)}{\Pr(Y_j \mid D)} = \frac{\Pr(Y_i) \Pr(D \mid Y_i)}{\Pr(Y_j) \Pr(D \mid Y_j)}. \tag{23}$$

If the ratio is greater than 1 than Markov chain Y_i performs better on a given data than Markov chain Y_j and vice versa. If we further assume that the prior probability on a our Markov chains from Y is uniformly distributed, that is: $\Pr(Y_i) = \Pr(Y_j) = \frac{1}{m}, \forall i, j \in \{1, \ldots, M\}$, the model selection problem will depend solely on the so called *Bayes factor* defined below:

$$\text{Bayes factor} = \frac{\Pr(D \mid Y_i)}{\Pr(D \mid Y_j)}. \tag{24}$$

We would like to find a Markov chain from model set Y that has the highest posterior probability among all models from Y given the data D.

This is where the crucial link between the Bayes factor and BIC comes in play. Katz [26] shows that the Markov chain with the highest posterior probability among all models from Y is the one for which the value obtained by BIC is minimal among all models from Y.

As already stated at the beginning of this section we first need to establish the highest allowed order M. We will than sequentially compare the Markov chain of order M to all orders $i \in \{1, \ldots, M-1\}$. The optimal order of a Markov chain will be established based on the results obtained by the BIC estimator defined below.

Definition 13. *Let $Y_i \in Y$ be a Markov chain of order $i < M$. Let the Likelihood ration criterion Λ for testing the Markov chain of order i against order M denote $\Lambda_{i,M}$, than the BIC estimator for the optimal order of the Markov chain is:*

$$f(u_{optimal}) = min_{0<u<M} f(u), \tag{25}$$

where

$$f(u) = -2log(\Lambda_{i,M}) - (m^M - m^i)(m-1)log(n). \tag{26}$$

We will now provide the intuition behind the above definition. Consider first the term:

$$-2log(\Lambda_{i,M}) = 2(\ell(p)^\star_M - \ell(p)^\star_i).$$

Here $\ell(p)^\star_M$ and $\ell(p)^\star_i$ denote the logarithmic maximum likelihood functions of Markov chains of orders M and i respectively. Note that increasing the order of Markov chain never decreases its interpretation power. This comes from the fact that if we allow the time dependency more steps into the past our conditional probability will only increase or remain the same (additional information will never decrease the probability). This is the same intuition as in the linear regression setting, where adding additional parameters to linear regression never decreases the R^2 (Grimsved, Pavia [20]).
With this in mind we see that the term $-2log(\Lambda_{i,M})$ will always be non-negative. We also see that if $\ell(p)^\star_i$ is close to $\ell(p)^\star_M$ than $-2log(\Lambda_{i,M})$ will approach 0 from above. Let us now consider the term $(m^M - m^i)(m-1)log(n)$, which represents the complexity penalty. We will not focus on the term $log(n)$ as it will be the same for every model and further restrict our attention to the term $(m^M - m^i)(m-1)$. This represents the difference between number of degrees of freedoms needed to describe transition matrices of Markov chain Y_i and Y_M. We will illustrate this on an example.

Example 5. *Consider a state space* $S = \{1, 2, 3\}$ *and Markov chains* Y_1 *and* Y_2 *defined on* S *of orders 1 and 2 respectively. We will also consider their transition probability matrices* P_{Y_1} *and* P_{Y_2}*:*

$$P_{Y_1} = \begin{array}{c} \\ 1 \\ 2 \\ 3 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} p_{11} & p_{12} & p_{13} \\ p_{21} & p_{22} & p_{23} \\ p_{31} & p_{32} & p_{33} \end{pmatrix} \end{array}$$

$$P_{Y_2} = \begin{array}{c} \\ 11 \\ 12 \\ 13 \\ 21 \\ 22 \\ 23 \\ 31 \\ 32 \\ 33 \end{array}\begin{array}{c} \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{pmatrix} p_{111} & p_{112} & p_{113} \\ p_{121} & p_{122} & p_{123} \\ p_{131} & p_{132} & p_{133} \\ p_{211} & p_{212} & p_{213} \\ p_{221} & p_{222} & p_{223} \\ p_{231} & p_{232} & p_{233} \\ p_{311} & p_{312} & p_{313} \\ p_{321} & p_{322} & p_{323} \\ p_{331} & p_{332} & p_{333} \end{pmatrix} \end{array}.$$

We need 6 degrees of freedom to describe the transition matrix P_{Y_1} *and 18 degrees of freedom do describe* P_{Y_2}*. Therefore the difference in number of degrees of freedom is* $18-6 = 12$*. Same result is obtained when we calculate:* $(3^2 - 3^1)(3 - 1) = 12$*. By induction it is possible to show that this formula will hold for arbitrary orders.*

Note also that when we increase the gap between orders the difference in degrees of freedom will grow exponentially. We will now put the two discussed terms together. As the $-2log(\Lambda_{i,M})$ is always non-negative we want this term to be as small as possible (remind that the optimal Markov chain has the minimal BIC value). At the same time we want the difference in number of degrees of freedom to be the largest (this will be accomplished when comparing the order 1 to order M) as it allows for the biggest penalization. The lower orders will tend to have relatively large value of $-2log(\Lambda_{i,M})$, however, also high difference in degrees of freedom and consequently high penalization. With increasing order both terms will become smaller and smaller. The Markov chain with *optimal* order established in this way will have the largest posterior probability out of all models in Y as described in Katz [26].

6.3 Time homogeneity

The question of time homogeneity has been widely neglected by many authors[10] when using the Markov chain methodology, even though it is a crucial assumption for validity of the results. As already described in section 4, time homogeneous Markov chain denotes that the transition probabilities do not change with time. In order to test time homogeneity the time interval is divided into $\frac{1}{2^k}$ sub-intervals of the same length and transition matrix estimated for each of these sub-intervals separately using the technique described in section 5.1. In this approach we will therefore first divide the time interval into halves and if we reject the time homogeneity we than recursively split the two time intervals each into halves and proceed in this way until we fail to reject the time homogeneity for some sub-intervals. For each split of the time interval we will test whether the transition probabilities estimated for each sub-intervals are statistically different from the transition probabilities estimated for the whole time intevral. Using the terminology described in Pavia and Grimsved [20] we denote sub-interval k by $I_k, k = 1.\dots, N$. In general we consider a Markov chain X of order u defined on a state space $S = \{1, \dots, m\}$. The transition probability that X has followed a certain path $i \in S^u$ on I_k to $j \in S$ will be denoted:

$$p_{ij}^k = \Pr(X_n = j \mid X_{n-1} = i, \dots, X_{n-u} = i_{n-u}), n \in I_k, i \in S^u, i, j, i_{n-u} \in S.$$

[10] See Tan and Yilmaz [49] for some examples.

The hypotheses of the formal test as described in Anderson and Goodman [2] is

$$\begin{aligned} H_0 &: p_{ij}(t) = p_{ij} \\ H_1 &: p_{ij}(t) = \frac{n_{ij}(t)}{\sum_{j=1}^{n} n_{ij}(t)}. \end{aligned} \tag{27}$$

Using the theorem of Cramér [7] and the *Likelihood ration criterion* Anderson and Goodman [2] derive that the statistics:

$$-2log(\Lambda) = 2\sum_{k=1}^{N}\sum_{i=1}^{m}\sum_{j=1}^{m} n_{ij}^{k}\left(log(p_{ij}^{k} - log(p_{ij})\right) \tag{28}$$

has asymptotic χ^2 distribution with $(N-1)m^u(m-1)$ degrees of freedom.

7 Data description

In this section we will describe data used in our thesis. We will work with *daily, weekly and monthly* stock returns of the Prague Stock Exchange official Index PX[11]. Daily data were collected between 1994-04-03 and 2017-11-26, weekly data between 1994-04-03 and 2017-11-26 and monthly data between 1994-04-30 and 2017-11-30. We will first examine the descriptive statistics of our data. The test results of our empirical model will be presented in next section. Finally the conclusion about the weak form efficiency of Prague stock exchange and its time homogeneity will be made for each of three data-sets considered.

7.1 Descriptive statistics

Descriptive statistics of the PX daily, weekly and monthly returns are presented below. Here *Min.* refers to the minimal value, *Max.* to the maximal value, *Sd.* to standard deviation and *Data* to number of observations.

Table 1: Descriptive statistics of stock returns

	Min.	**Median**	**Mean**	**Max.**	**Sd.**	**Data**
Daily	-14.94	0.03	0.009546	13.16	1.334	5858
Weekly	-14.94	0.06	0.003667	11.73	1.355	1230
Monthly	-5.45	0.14	0.1468	4.21	1.243	283

Figure 7 visualize the box-plot of five number summary for each data-set.

To get a better understanding of our data we will test whether the distributions of daily, weekly and monthly stock returns are approximately normally distributed. Note that the Markov chain methodology does not require normality assumption[12], therefore its testing is merely a matter of exploratory analysis than necessity. We will use the two sample Kolmogorov-Smirnov test[13] to formally test the normality of our data. In this test we first compare the empirical cumulative distribution function of our daily

[11] This index was called PX 50 until March 2006

[12] As opposed to regression based methods that are widely used for testing the EMH.

[13] For more details on this test see Stephens [46]

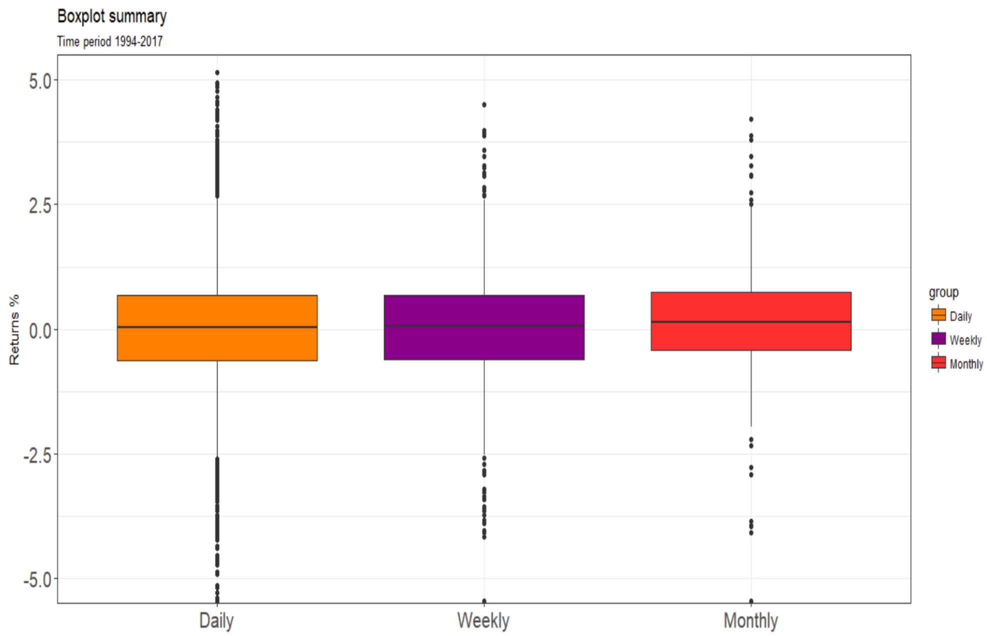


Figure 7: Box-plot visualization

data with 5858 randomly sampled data points from normal distribution with mean 0.009546 and standard deviation 1.334609. We repeat the test for weekly data which are compared to 1230 randomly sampled data points from normal distribution with mean 0.003667 and standard deviation 1.355904 and for monthly returns, which we compare to 283 random observations from normal distribution with mean 0.1468 and standard deviation 1.24298. The null hypothesis for this two sample Kolmogorov-Smirnov test is that the empirical and normal distribution are not statistically similar. Result are shown in the table below.

Table 2: Two sample Kolmogorov-Smirnov test of data normality

	Daily returns	**Weekly returns**	**Monthly returns**
p-value	4.2188e-15	6.0824e-06	0.05365

Based on the results from this test we reject the null hypothesis for daily and weekly returns on 5% significance level. Moreover if we set the significance level to be 10% we reject the null hypothesis for monthly returns as well and conclude that all three return frequencies are approximately normally distributed. This conclusion about normality of our data is in line with the empirical evidence on distribution of long run stock returns proposed by Takaki [48]. The visualization of the density functions for daily, weekly and monthly stock returns is provided below.

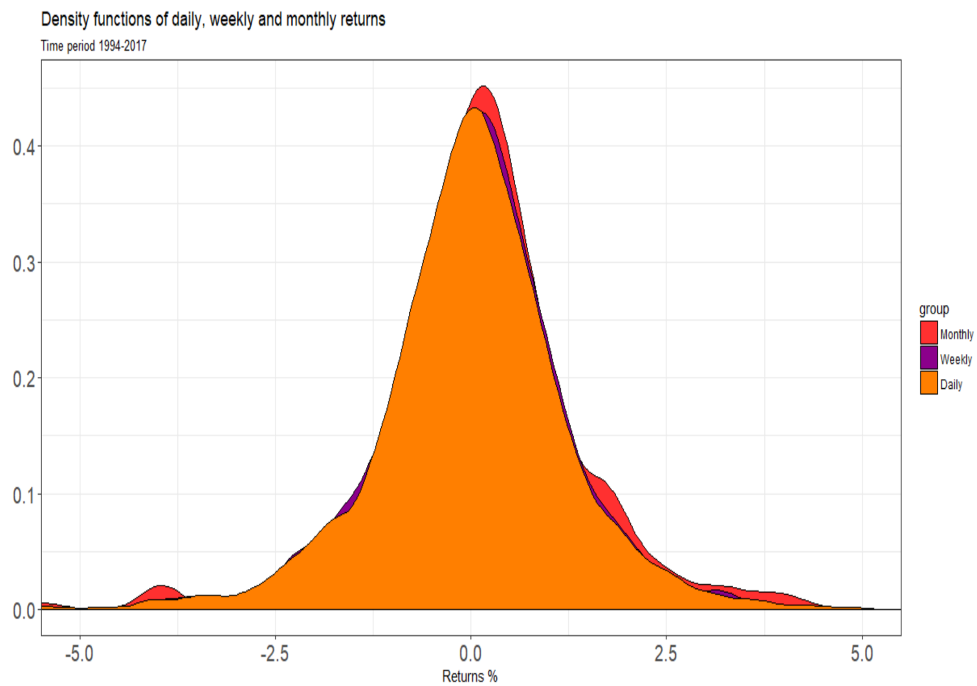


Figure 8: Visualization of the distribution of PX index returns

8 Results

In this section we will present empirical results of our two-state Markov chain model evaluated on all three data-sets that we have used for our analysis. We will present the results in the same order as was outlined in the Figure 6.

8.1 BIC, Likelihood ratio criterion and Time homogeneity results

Therefore, first the results obtained by BIC are presented. Following the methodology described in section 6.2.3 we set the maximal order of a Markov chain M to 4. Order 4 in our case represents the highest order for which all entries of the transition probability matrices estimated on daily, weekly and monthly data sets are strictly positive. Below, the results of the BIC estimation are presented. We compare sequentially the orders 1,2 and 3 to order 4 using the same method described in section 6.2.3.

Table 3: Testing Markov chain orders 1,2 and 3 against order 4 using the BIC

Order	**Daily**	**Weekly**	**Monthly**
1	-92.8599	-90.8817	-60.9031
2	-79.8456	-77.4281	-55.2649
3	-52.4762	-50.1554	-37.3592

We see that the obtained value of the BIC estimation is lowest for order 1 for all three data sets. Using the Definition 13 we take the order 1 to be the *optimal* order in a sense discussed in 6.2.3.

We will now test whether the Markov chain models of order 1 are statistically different from Markov chain order 0, which represents random walk in stock returns. For this comparison we will use the *Likelihood ration criterion* and test the similarity for all three data-sets. The null hypothesis is thus that the Markov chain of order 1 and order 0 are not statistically different. Here the abbreviation *df* denotes the number of *degrees of freedom* used for testing.

Table 4: Testing Markov chains of order 1 against order 0 by *Likelihood ration criterion*

	Daily	**Weekly**	**Monthly**
p-value	8.024e-09	0.1637	0.7018
χ^2 statistic	37.2816	3.6186	0.7081
df	2	2	2

We see that for daily returns we reject the hypothesis that the Markov chain of order 1 is equal to order 0 on all significance levels 1%,5% and 10%. This fact leads to rejection of the weak form market efficiency for daily stock returns of PX index. We fail to reject the hypothesis for weekly and monthly returns when testing on the same significant levels. Before we make any conclusions about whether or not the stock market is efficient for weekly and monthly returns, we have to test the time homogeneity of their corresponding transition probability matrices. We first estimate the TPMs for the whole time period 1994-2017 and than divide this time interval into two sub-periods. The time *period 1* is taken to be between years 1994-2006 and the time *period 2* between years 2006-2017. We estimate the corresponding transition probability matrices for both of these sub-intervals and compare them to the TPM for the whole time period. The results of this test are displayed below, where the null hypothesis is that the TPM for the whole time period is not statistically different from the TPMs estimated for each of two sub-periods.

Table 5: Time homogeneity test for weekly and monthly returns

	Weekly period 2	**Weekly period 1**	**Monthly period 2**	**Monthly period 1**
p-value	0.9796	0.9796	0.8164	0.8702
χ^2 statistic	0.0411	0.0411	0.4054	0.2779
df	2	2	2	2

We fail to reject the time homogeneity for both weekly and monthly data, when testing on 1%, 5% and 10% significance levels. With this test result we conclude that the PX index of Prague stock exchange is weak form efficient for weekly and monthly stock returns.

8.2 Validation of the assumptions

As stated in section 6.2.3 we need our Markov chain models to be irreducible, aperiodic and time homogeneous. We have already tested the time homogeneity assumption for weekly and monthly returns. To test the other two assumptions we will use the Proposition 1. Therefore, if the transition matrices are strictly positive, our assumptions will be automatically validated. We will display all three TPM's for time period 1994-2017 below. Here P_M, P_W and P_D denote TPM's of monthly, weekly and daily data of our two state Markov model, respectively.

$$P_M = \begin{pmatrix} 0.528 & 0.472 \\ 0.479 & 0.521 \end{pmatrix}, P_W = \begin{pmatrix} 0.497 & 0.503 \\ 0.463 & 0.537 \end{pmatrix}, P_D = \begin{pmatrix} 0.528 & 0.472 \\ 0.452 & 0.548 \end{pmatrix}.$$

We see that all three transition probability matrices are indeed strictly positive, therefore the assumptions of our Markov chain model are validated. We will test the time homogeneity of daily return data in the next subsection.

8.3 Four state Markov chain

In the previous section we have concluded that daily returns of the PX index for time period 1994-2017 are best represented by the Markov chain of order 1. Based on this fact we have rejected the EHM for daily returns. We can take advantage of this fact and use the transition probabilities to model the daily stock return behaviour.

The major drawback of two state approach used to test the EMH in terms of interpretation is that it does not reveal much information about our data. In this model only the directions of returns are considered and not the magnitudes. To gain more insight into daily return data, we propose a four state Markov chain model Y, which will more accurately represent the magnitudes of returns. The time series of daily returns r_t will be mapped onto: "Very high" (VH), "High" (H), "Low" (L) and "Very low" (VL) return states described below. Here $\mathbb{E}(r_t)$ denotes the expected value of r_t and $\sigma(r_t)$ its standard deviation.

$$Y = \begin{cases} VH & \text{if } r_t \geq \mathbb{E}(r_t) + \sigma(r_t) \\ H & \text{if } r_t \in (\mathbb{E}(r_t), \mathbb{E}(r_t) + \sigma(r_t)) \\ L & \text{if } r_t \in (\mathbb{E}(r_t) - \sigma(r_t), \mathbb{E}(r_t)] \\ VL & \text{if } r_t \leq \mathbb{E}(r_t) - \sigma(r_t) \end{cases} \quad (29)$$

The state space of this Markov chain will be $S_Y = \{VH, H, L, VL\}$. Before modeling the data, we will first test the time homogeneity of daily returns. The testing procedure and the definitions of time periods are the same as described in the Chapter 8.

Table 6: Testing time homogeneity of daily returns for 1994-2017

	Daily period 2	**Daily period 1**
p-value	0.00014	0.00013
χ^2 statistic	17.7069	17.8416
df	2	2

Based on the results we reject the time homogeneity of the TPM for the whole time period 1994-2017 (when testing on all conventional significant levels 1%, 5% and 10%). We will further proceed recursively by dividing the time period 2006-2017 into another two time sub-periods. Period 3 is taken to be between years 2006-2012 and period 4 to be between years 2012-2017. We than test whether these two TPM's are statistically different from the TPM estimated for the whole time period 2.

Table 7: Testing time homogeneity of daily returns for 2006-2017

	Daily period 4	**Daily period 3**
p-value	0.7330	0.7307
χ^2 statistic	0.6210	0.6274
df	2	2

We fail to reject the time homogeneity for time period 2006-2017 (using the same significance levels as before). We will therefore use daily stock returns data from this time period only for our further analysis. We will not use the daily returns from time period 1994-2006 in our thesis as we consider the analysis of the more recent data from 2006-2017 to be of greater value. The *EMH* test was conducted in the same fashion as in the previous section. The results again indicate that the optimal order of the Markov chain for modeling daily returns is 1. This result is also true for the subset of data from 2006-2017 and therefore the usage of Markov chain of order 1 is justified. We provide the results obtained by MLE method described in Chapter 5 of the transition counts and probabilities of the Markov chain model Y in the table below.

Table 8: Transition counts of Markov chain model Y

	VL	**L**	**H**	**VH**	**Row total**
VL	68	112	77	28	285
L	79	436	534	92	1141
H	93	494	552	98	1237
VH	45	98	75	47	265

The corresponding TPM of a Markov chain Y is:

$$P_Y^1 = \begin{array}{c} \\ VL \\ L \\ H \\ VH \end{array} \begin{array}{c} \begin{array}{cccc} VL & L & H & VH \end{array} \\ \begin{pmatrix} 0.239 & 0.393 & 0.270 & 0.098 \\ 0.069 & 0.382 & 0.468 & 0.081 \\ 0.075 & 0.399 & 0.447 & 0.079 \\ 0.170 & 0.370 & 0.283 & 0.177 \end{pmatrix} \end{array} \tag{30}$$

From the Row total column of Table 8 we see that the count distribution Markov chain Y visited a certain state is very symmetric. This fact corresponds with our conclusion about the normality of our data. It is clear that the transition matrix P_Y is strictly positive and by Proposition 1 the Markov chain Y is *irreducible* and *aperiodic*. The assumptions of Proposition 2 are satisfied and we can analyze the long run behaviour of the stock returns:

$$P_Y^2 = \begin{pmatrix} 0.121 & 0.388 & 0.397 & 0.094 \\ 0.092 & 0.390 & 0.429 & 0.089 \\ 0.092 & 0.390 & 0.429 & 0.089 \\ 0.118 & 0.387 & 0.395 & 0.100 \end{pmatrix} P_Y^3 = \begin{pmatrix} 0.102 & 0.389 & 0.418 & 0.091 \\ 0.097 & 0.389 & 0.424 & 0.090 \\ 0.096 & 0.389 & 0.425 & 0.090 \\ 0.102 & 0.389 & 0.417 & 0.092 \end{pmatrix}$$

$$P_Y^4 = \begin{pmatrix} 0.098 & 0.389 & 0.422 & 0.091 \\ 0.097 & 0.389 & 0.423 & 0.091 \\ 0.097 & 0.389 & 0.423 & 0.091 \\ 0.098 & 0.389 & 0.422 & 0.091 \end{pmatrix} P_Y^5 = \begin{pmatrix} 0.097 & 0.389 & 0.423 & 0.091 \\ 0.097 & 0.389 & 0.423 & 0.091 \\ 0.097 & 0.389 & 0.423 & 0.091 \\ 0.097 & 0.389 & 0.423 & 0.091 \end{pmatrix}.$$

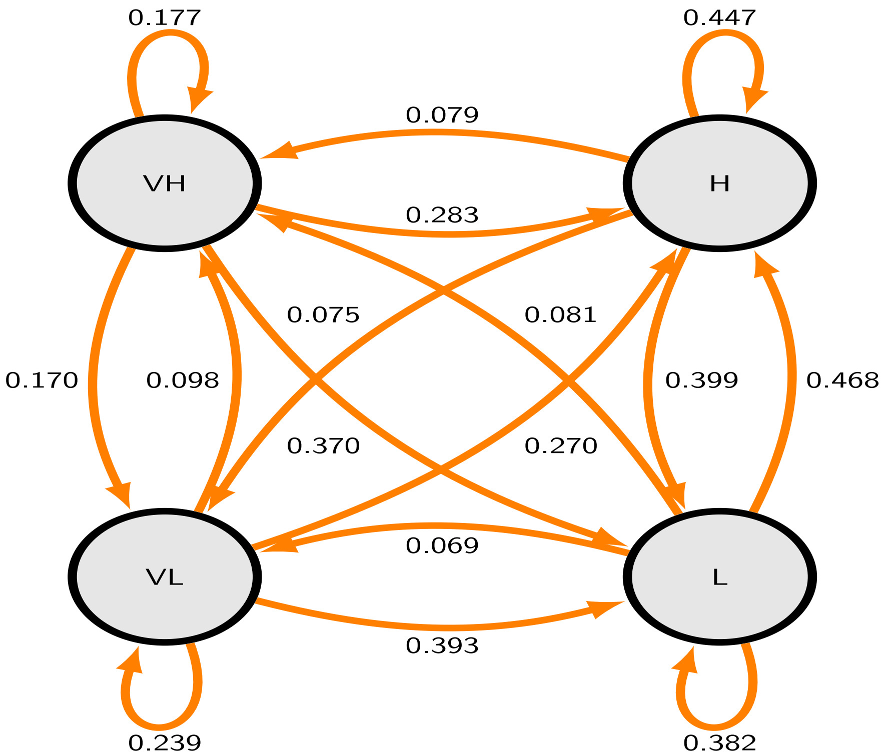
The transition matrix is sufficiently close to stationary distribution π_Y in five steps. The stationary distribution of the Markov chain is then

$$\pi_Y = \begin{array}{c} \begin{array}{cccc} \pi_{VL} & \pi_L & \pi_H & \pi_{VH} \end{array} \\ \begin{pmatrix} 9.7\% & 38.9\% & 42.3\% & 9.1\% \end{pmatrix} \end{array}.$$

We can also calculate the average return times to states

$$\omega_Y = \begin{array}{c} \begin{array}{cccc} \omega_{VL} & \omega_L & \omega_H & \omega_{VH} \end{array} \\ \begin{pmatrix} 10.30 & 2.57 & 2.36 & 10.99 \end{pmatrix} \end{array}.$$

This means that on average the PX index will be in a "Very low" state every 10.3 days, which is almost the same time interval in which it will visit "Very high" state. The transition to states "Low" and "High" occurs on average approximately every 2.5 days. Below we will show the transition graph describing the Markov chain Y.



(a) Transition graph of daily returns between 2006 and 2017

Figure 9: Four states Markov chain transition graph

8.4 Expected returns

In this sub-section we will calculate the expected returns for each of the four states. To do this we will use the definition presented in Feldman and Valdez-Flores [15].

Definition 14. *Let Y be a Markov chain defined on a state space S with transition matrix P and profit function f (i.e. each time a chain visits state i, a profit of f(i) is obtained). The expected profit at* n^{th} *step is given by*

$$\mathbb{E}[f(Y_n) \mid Y_0 = k] = P^n f(k) \tag{31}$$

Therefore first the TPM P_Y^n is multiplied by the vector f and than the k^{th} component of the resulting matrix is taken. For our purposes we will take the profit function to be a vector of average returns of each of the four states of Markov chain Y. The results are presented in the table below, where time 0 is taken to be the initial average return for each state. Note that the long run average of daily returns for time period 2006-2017 is -0.0022.

Table 9: Expected returns for initial starting state in %

Time	**0**	**1**	**2**	**3**	**4**	**5**	**6**
VL	-2.6300	-0.4439	-0.0704	-0.0139	-0.0043	-0.0023	-0.0022
L	-0.5561	0.0768	0.0117	0.0005	-0.0017	-0.0022	-0.0022
H	0.5739	0.0356	0.0097	0.0002	-0.0018	-0.0022	-0.0022
VH	2.5150	-0.0438	-0.0447	-0.0128	-0.0044	-0.0026	-0.0022

From Table 9 we observe interesting information. First of all if investor buys a stock in a "Very high" state its expected return in the next time period decreases from 2.515% to -0.0438% which is below average return. The opposite movement is seen for "Very low" state, where the return is expected to rise from -2.63% to -0.4439%.

The same pattern, however, of a smaller magnitude is seen for "Low" state. Here the expected increase of the stock return is from -0.5561% to 0.0768%. After the first time period we see decrease to 0.0117%. We see that given the initial state all expected returns eventually converge to the long-run expected return of -0.0022% after 6 trading days. This fact is in line with our finding about the stationary distribution. Based on these results we can propose that the optimal trading strategy is to focus on times when the PX index is either in a state "Low" or "Very low" and than sell the stock after 1 or 6 trading days respectively. Once the index reaches states "High" or "Very high" it is advisable to sell the stock as soon as possible as the expected return is most likely going to decrease in the consecutive time periods.

9 Discussion of results and further research

In this section we will first comment on the results obtained in the empirical part of our thesis and then suggest a possible extensions and suggestions for further research based on our work.

9.1 Discussion of results

Our results of weak form market efficiency might perhaps come as a surprise. In section 2 we have argued that the general empirical conclusion of research conducted on data from time period 1993-2010 was that the Prague Stock Exchange was becoming more efficient throughout the years. Yet we have rejected the weak form efficiency for daily returns for the time period 1994-2017. Our explanation of weak form inefficiency for daily returns is the slow adjustment of the stock returns to new information.

With a rapid speed of globalization the amount of information that is affecting the stock return development is the largest ever. We have explicitly derived that the lag with which the stock returns react to all publicly available information is on average one day. This market inability to react to information immediately might perhaps be caused by insufficiency in terms of technology used. For example investors on the New York or London stock markets, that are widely considered to be weak form efficient even for high frequency data, are known to use extensively methods of machine learning and artificial intelligence for return predictions.

These methods are employed to gather available information from as many different sources as possible and evaluate it automatically in a real time. Once these techniques become fully implemented and will become standard tool, rather than exceptions, the convergence towards the weak form efficiency for high frequency data of Prague Stock Exchange might be precipitated. Our results also show that in weekly and monthly returns all publicly available information are already incorporated, which is in line with the results of "*average information lag*" of one day.

9.2 Further research

Our thesis primarily focuses on testing the weak form market efficiency. The modeling part described in section 8.3 allows for much modification. Future studies can exploit the fact that daily stock returns can be modeled by first order Markov chain and suggest more sophisticated models of return behaviour. For example Kilic [27] proposed Markov chain model that represents expected returns. This model accounts for the initial capital investment by investor, buying and selling costs as well as capital gain tax.
Another option is to use the Markov regime-switching model, which captures the structural breaks in the data (for detail description of this model see Fišerová [17]). Four states proposed in the modeling section 8.3 can further be divided into additional states and create a larger and more detailed state space that way. For the inspiration on various forms of state space discretization refer to Lukáš and Svoboda [47]. Also hourly or even minute data could be used to more accurately predict the movement of the return index. Another suggestion for further research is to focus on individual companies traded on Prague Stock Exchange.
Fielitz and Bhargava [16] proposed a method how to aggregate data of individual companies to form a vector Markov chain model. By following this method the efficiency or inefficiency of a market can be explained in terms of individual companies rather than in terms of the aggregated index. Last but not least we suggest to examine the correlation between transitions to certain state and days of the week on which these transitions have occurred. For example the proposition if the state of Markov chain associated with high returns is more likely to appear in the first days of the week or if it is the other way around might be worth examining.
In our thesis we consider only the weak form efficiency of Prague Stock Exchange. Other forms of efficiency, which were described in section 3.1 can be tested. For example Nezdara [36] and Krištoufek [30] tested the semi-strong form efficiency of Prague Stock Exchange in time period 1997-2002 and 2000-2006 respectively.

10 Conclusion

We regard the contribution of our thesis to be in three main forms. First, in our analysis we consider data from time period 1994-2017, which is all together 23 years. To our best knowledge this time period is the longest ever considered when analyzing the efficiency of Prague Stock Exchange. This fact enables us to capture the long-time horizon behaviour of the Prague Stock Exchange in its full richness. The empirical results discussed in section 9 can serve as a valid alternative to already existing ones.

Secondly, we regard the introduction of the Markov chains methodology as a valuable contribution. This methodology is not very frequent in the area of financial economics even though it serves as a powerfull modeling tool. The fact that this methodology does not require the assumption of normality of data means that we can capture a non-linear structural dependencies in our data as opposed to regression based techniques for which the normality assumption is much more vital. Another advantage of Markov chain method is the insensitivity to outliers.

Last but not least our aim was to present all derivation and theoretical aspects of our thesis in the most accessible way. We have also provided many reference links to more advance literature on various topics discussed, where more detailed explanation was not possible. We believe that our thesis may serve as a stepping stone to further research and encourage interested reader to refer back to Chapter 9 for more inspiration on possible extensions of our work.

11 Bibliography

References

[1] D. Alajbeg, Z. Bubaš, and V. Šonje. "The efficient market hypothesis: problems with interpretations of empirical tests" (2012).

[2] T.W. Anderson and L.A. Goodman. "Statistical Inference About Markov Chains". *The Annals of Mathematical Statistics* 28 (1957), pp. 89–110.

[3] D. Bertsekas and J. Tsitsiklis. *Introduction to Probability*. Vol. 2. Athena Scientific Belmont, MA, 2008.

[4] P. Billingsley. "Statistical Methods in Markov Chains". *The Annals of Mathematical Statistics* 32.1 (1961), pp. 12–40.

[5] J.K. Blitzstein and J. Hwang. *Introduction to Probability*. CRC Press, 2014.

[6] P. Bremaud. *Markov Chains - Gibbs Fields, Monte Carlo Simulation and Queues*. Vol. 31. Springer-Verlag New York, 1999.

[7] H. Cramér. *Mathematical methods for statistics*. Vol. 9. Princeton University Press, 2016.

[8] K. Diviš and P. Teplý. "Informační efektivnost burzovních trhů ve střední Evropě". *Finance a úvěr* (2004), pp. 9–10.

[9] R. Douc et al. "Consistency of the maximum likelihood estimator for general hidden Markov models". *the Annals of Statistics* 39.1 (2011), pp. 474–513.

[10] E. F. Fama. "Efficient Capital Markets: A review of Theory and Empirical Work". *Journal of Finance* 25.2 (1970), pp. 383–417.

[11] E. F. Fama. "Efficient Capital Markets: II". *The Journal of Finance* 46.5 (1991), pp. 1575–1617.

[12] E. F. Fama. "The Behaviour of Stock-Market Prices". *Journal of Business* 38.1 (1965), pp. 34–105.

[13] E.F. Fama. "Random Walks in Stock-Market Prices". *Financial Analysts Journal* 21 (1965), pp. 55–59.

[14] E.F. Fama. "Two Pillars of Asset Pricing". *American Economic Review* 104.6 (2014), pp. 1467–1485.

[15] R.M. Feldman and C. Valdez-Flores. *Applied Probability and Stochastic Processes*. Vol. 2. Springer Science & Business Media, 2010.

[16] B.D. Fielitz and T.N. Bhargava. "The Behavior of Stock-Price Relatives A Markovian Analysis". *Operations Research* 21.6 (1972), pp. 1183–1199.

[17] T. Fišerová. "Regime-switching models and their application in the financial markets (Bachelor thesis, Institute of Economic Studies, Prague)" (2011).

[18] P.A. Gagniuc. *Markov chains, from theory to implementation and experimentation*. Vol. 1. Wiley, 2017.

[19] G. Grimmett and D.R. Stirzaker. *Probability and random processes*. Vol. 3. Oxford: Oxford University Press, 2001.

[20] E. Grimsved and J. Pavia. "Evaluating the Efficiency of the Swedish Stock Market: a Markovian Approach (Bachelor Thesis in Financial Economics, University of Gothenburg, school of business, economics and law)" (2015).

[21] Z. Hackl and V. Pošta. "Information Efficiency of the Capital Market: a Stochastic Calculus Approach Evidence from the Czech Republic". *Czech Journal of Economics and Finance* 57.5 (2007).

[22] F.A. Hayek. "The Use of Knowledge in Society". *The American Economic Review* 35.4 (1945), pp. 519–530.

[23] H. Horská. "Český akciový trh - jeho efektivnost a makroekonomické souvislosti". *VŠE v Praze Working Paper* 7 (2005), pp. 283–301.

[24] J. Hájek. "Test slabé formy efektivnosti středoevropských akciových trhů". *Politická ekonomie* 6 (2007), pp. 773–791.

[25] Michael C. Jensen. "Some Anomalous Evidence Regarding Market Efficiency". *Journal of Financial Economics* 6 (1978), pp. 95–101.

[26] R.W. Katz. "On Some Criteria for Estimating the Order of a Markov Chain". *Technometrics* 23.3 (1981), pp. 243–249.

[27] S.B. Kilic. "Estimating Probability of Session Returns for Istanbul Stock Exchange 100 Index as Markov Chain Process". *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 1 (2013), pp. 501–512.

[28] S.B. Kilic. "Estimating Probability of Session Returns for Istanbul Stock Exchange 100 Index as Markov Chain Process". *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi* 22.1 (2013).

[29] T. Konstantopoulos. "Introductory lecture notes on Markov Chains and Random Walks". Vol. 200. 9. Uppsala University, 2009.

[30] L. Krištoufek. "Testování středně silné formy efektivních kapitálových trhů na akciovém trhu v ČR v letech 2000-2006 (Dissertation thesis, Institute of Eonomic Studies, Prague)" (2007).

[31] D. Lazar and S. Ureche. "Testing efficiency of stock market in emerging economies". *The Annuals of the University of Oradea* 62 (2007), pp. 832–836.

[32] R. Ch. Lima and A. Ohashi. "The Efficient Market Hypothesis and the Dynamic Behavior of Sugar Future Prices". *Departamento de Economia, PIMES-Universidade Federal de Pernambuco* 30 (1999), pp. 484–493.

[33] A. W. Lo and A. C. MacKinlay. *A Non-Random Walk Down Wall Street*. Princeton University Press, 1999.

[34] B. G. Malkiel. *A random walk down Wall Street*. Vol. 1. Norton New York, 1973.

[35] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Vol. 1. Springer-Verleg, 1993.

[36] O. Nezdara. "Market efficiency in the Prague stock exchange (Dissertation thesis, Institute of Economic Studies, Prague)" (2003).

[37] J. Norris. *Markov Chains*. Vol. 2. Cambridge University Press, 1998.

[38] C. Okonta, O. Elem-Uche, and C. Madu. "A Markovian Analysis of the Nigerian Stock Market Weekly Index". *International Journal of Statistics and Applications* 7.2 (2017), pp. 152–157.

[39] V. Pošta. "Estimating the Dynamics of Weak Efficiency on the Prague Stock Exchange Using the Kalman Filter". *Czech Journal of Economics and Finance* 58.6 (2008).

[40] T. Quang. "Testování slabé formy hypotézy efektivního trhu na českém akciovém trhu lineárními a nelineárními metodami (Dissertation thesis, VŠE in Prague)" (2007).

[41] M. B. Rajarshi. "Statistical Inference for Discrete Time Stochastic Processes". 31 (2012).

[42] E. Rodriguez et al. "US stock market efficiency over weekly, monthly, quarterly and yearly time scales". *Physica A: Statistical Mechanics and its Applications* 413 (2014), pp. 554–564.

[43] P.A. Samuelson. "Proof That Properly Anticipated Prices Fluctuate Randomly". World Scientific Book Chapters (2015), pp. 25–38.

[44] G.E. Schwartz. "Estimating the Dimension of a Model". *The Annals of Statistics* 6.2 (1978), pp. 461–464.

[45] M. Sewell. "History of the Efficient Market Hypothesis". *RN* 11.4 (2011).

[46] M. A. Stephens. "EDF Statistics for Goodness of Fit and Some Comparisons". *Journal of the American statistical Association* 69.347 (1974), pp. 730–737.

[47] M. Svoboda and L. Lukáš. "Application of Markov chain analysis to trend prediction of stock indices". *Proceedings of the 30th International Conference Mathematical Methods in Economics.* 2012.

[48] H. Takaki. "A discrete-time model of high-frequency stock returns". *Quantitative Finance* 4.2 (2004), pp. 140–150.

[49] B. Tan and K. Yilmaz. "Markov Chain Test for Time Dependence and Homogeneity: An Analytical and Empirical Evaluation". *European Journal of Operational Research* 137.3 (2002), pp. 524–543.

[50] J. Veselá. *Analýzy trhu cenných papírů.* Vol. 1. Oeconomica, 1999.

[51] J. Wang. "Efficient market hypothesis - Chapter 9 (Lecture notes)" (2006).

[52] G.G. Yin and Q. Zhang. *Continuous-Time Markov Chains and Applications.* Vol. 37. Springer-Verlag New York, 1998.

[53] F. Čámský. "Testování hypotéz o efektivnosti. Finančné trhy: Odborný mesačník pre teóriu a praxu finančních trhov". *Finančné trhy: Odborný mesačník pre teóriu a praxi finančních trhov* 1 (2005).

[54] R. Šíbl. "Efektivnost Kapitálových Trhů (Dissertation thesis, Masaryk University, Brno)" (2005).

List of Figures

List of Tables

12 Appendix: Figures

In this appendix section we will present the graphs of both the stock returns of the PX index and the PX index for all three data frequencies considered in our thesis.

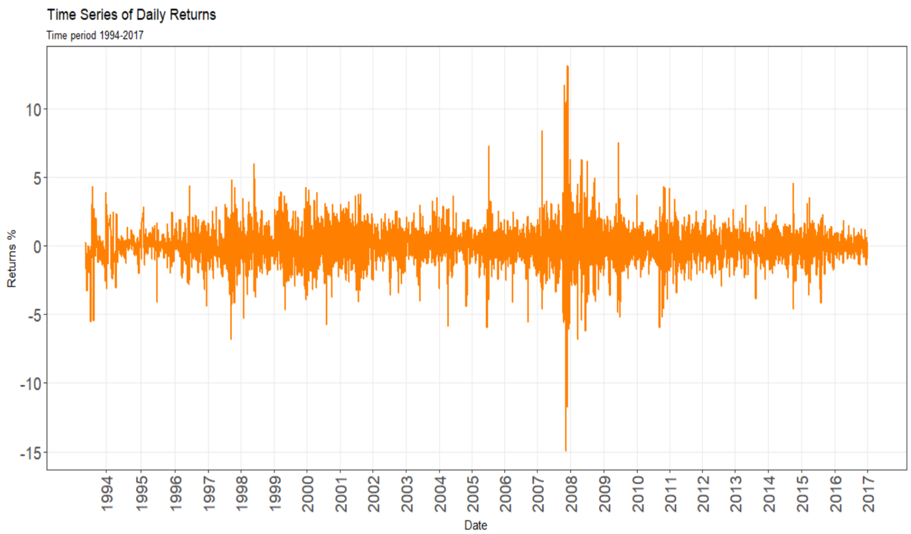


Figure 10: Daily stock returns

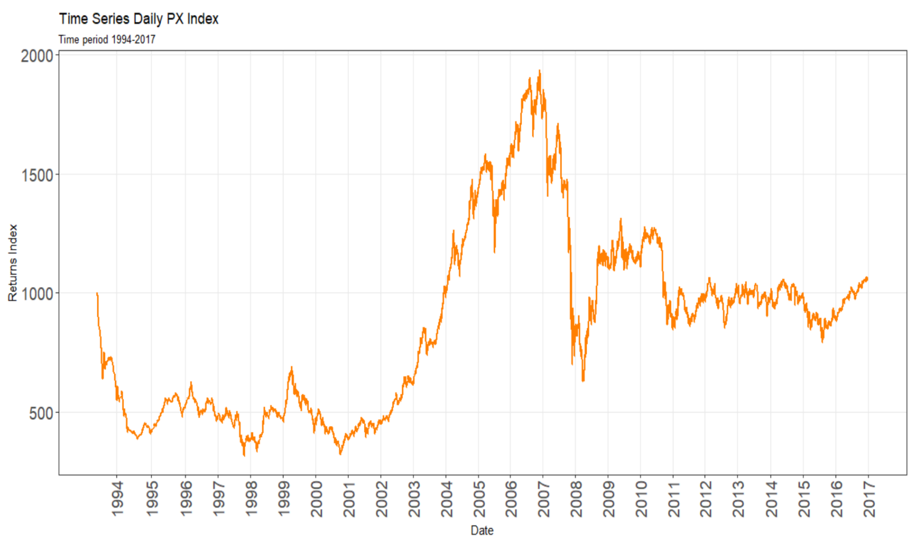


Figure 11: Daily PX index returns

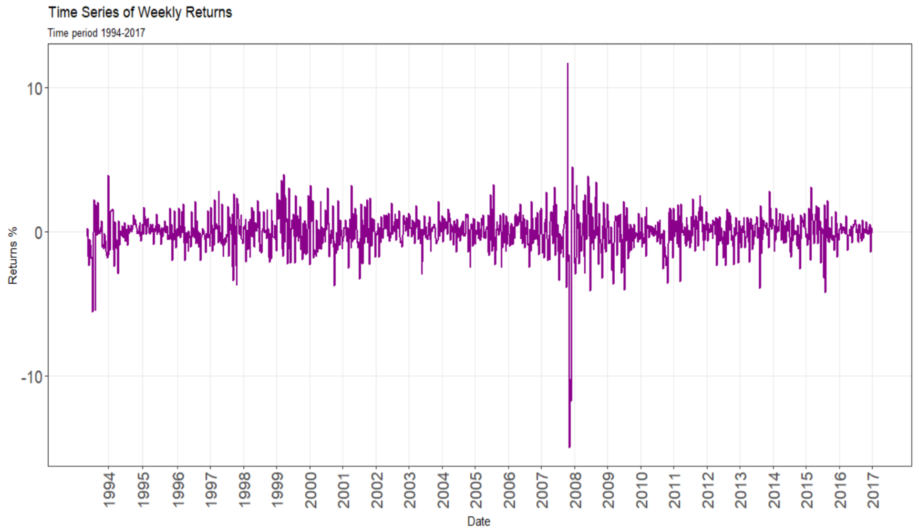


Figure 12: Weekly stock returns

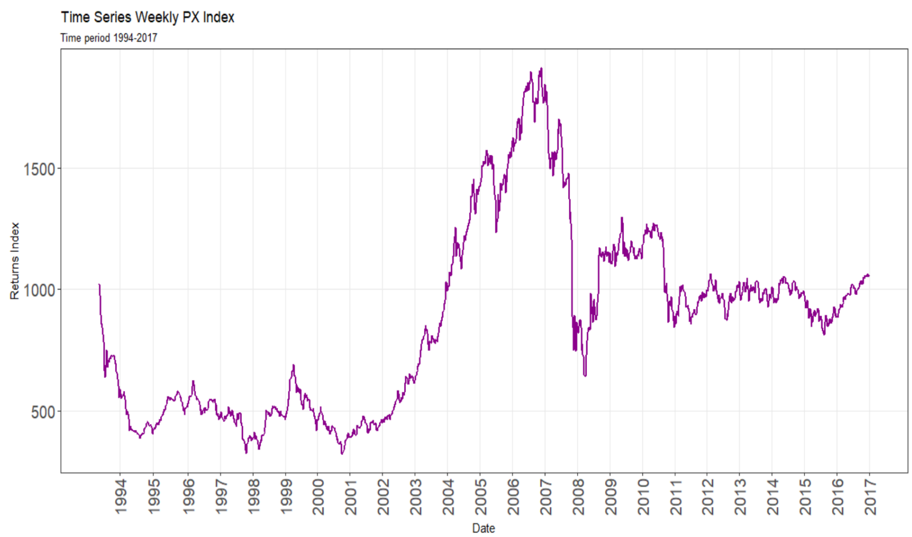


Figure 13: Weekly PX index returns

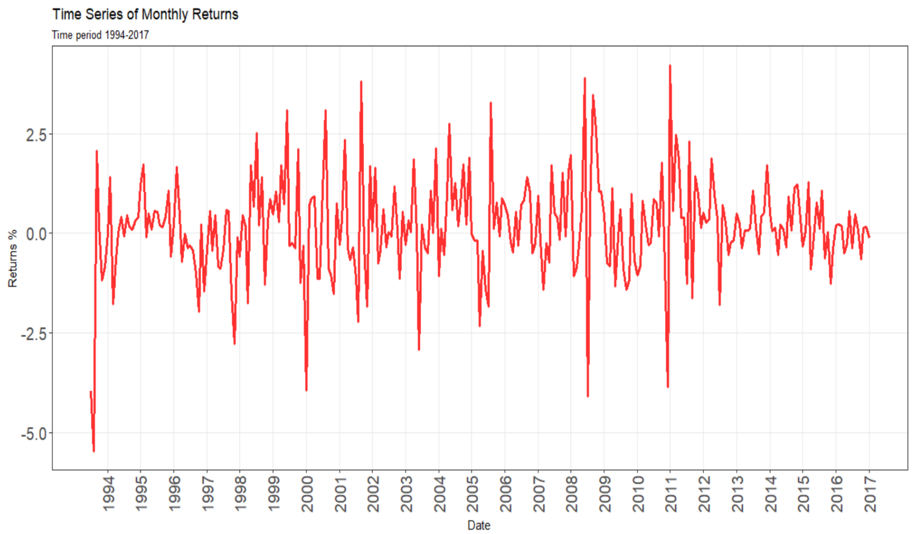


Figure 14: Monthly stock returns

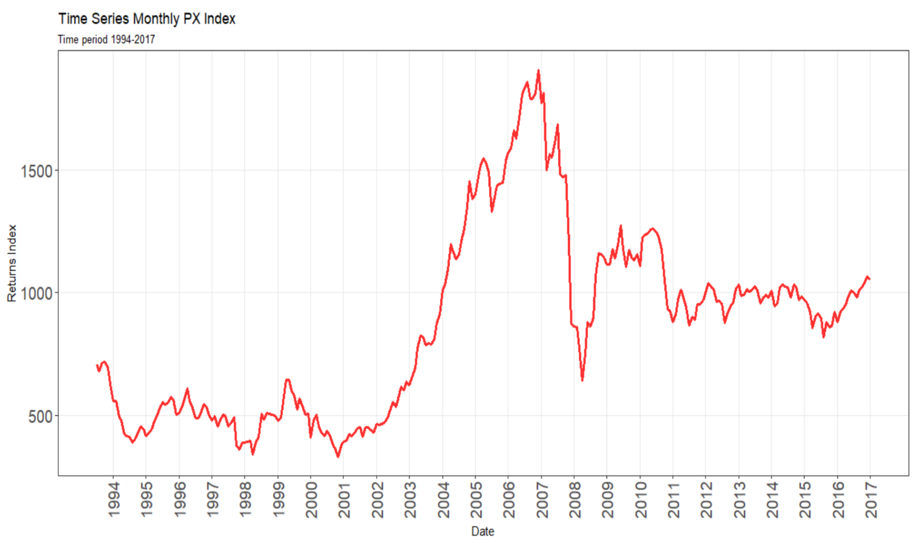


Figure 15: Monthly PX index returns